

Forthcoming in *Mind & Language*

The evolution and development of visual perspective taking

Ben Phillips

Correspondence

School of Historical, Philosophical, and Religious Studies, Arizona State University
975 S. Myrtle Ave, P.O. Box 874302, Tempe, AZ 85287-4302, USA

Email: [ben.s.phillips\[at\]gmail.com](mailto:ben.s.phillips@gmail.com)

Abstract

I outline three conceptions of seeing that a creature might possess: ‘the headlamp conception,’ which involves an understanding of the causal connections between gazing at an object, certain mental states, and behavior; ‘the stage lights conception,’ which involves an understanding of the selective nature of visual attention; and seeing-as. I argue that infants and various nonhumans possess the headlamp conception. There is also evidence that chimpanzees and 3-year-old children have some grasp of seeing-as. However, due to a dearth of studies, there is no evidence that infants or nonhumans possess the stage lights conception of seeing. I outline the kinds of experiments that are needed, and what we stand to learn about the evolution and development of perspective taking.

1 INTRODUCTION

The ability to adopt someone else’s visual perspective confers obvious advantages. If I know what you see, I might gain valuable information about the location of food or predators. This knowledge might also enable me to anticipate your next action, thereby generating opportunities for cooperation and competition. Clearly, visual perspective taking has fitness benefits. But what exactly is it?

There are several relations that an individual might represent in tracking someone else's gaze. In tracking *B*'s gaze, *A* might represent the direction in which *B*'s eyes point; the geometric relation that obtains between *B*'s eyes and the object that her gaze terminates on; or one of the more discriminating relations that are often construed as *seeing* relations. Given the options, we need to be careful in specifying what we mean when we invoke the notion of a 'visual perspective taker.'

Unfortunately, theorists are not always sensitive to these foundational issues. To be sure, most recognize Flavell's (1977) distinction between Level 1 and Level 2 perspective taking. If I'm a Level 1 visual perspective taker then I can know *which* objects you see. If I'm operating at Level 2 then I can also know *how* you see those objects (e.g. the fact that you see the banana *as yellow*).

But despite widespread recognition of the Level 1/Level 2 distinction, details concerning what it takes to qualify for each level are thin on the ground. One pressing issue is that there are likely several agent-object relations one could represent and thereby qualify as a Level 1 perspective taker. A related problem is that theorists are rarely explicit on the lower bounds of Level 1 perspective taking. If we are to make progress on the question of how and why visual perspective taking evolved, getting clear on which species are Level 1 perspective takers, and which species fall just short, is of paramount importance.

It is also rare for researchers to clarify which approach to mental representation they are assuming when they assert or deny that some creature is a visual perspective taker. Engaging with this very issue, Buckner (2014) has recently argued that if we are to move vexed debates about animal mindreading forward, participants should either refrain from invoking representational idioms, or, explicitly commit to an account of mental representation that can be scrutinized.

In what follows, I tackle these foundational issues head-on. I argue that in addition to various geometric notions, such as line-of-gaze, there are (at least) three conceptions of seeing that a gaze-follower might possess. I call the first one the ‘headlamp conception’—possessing it requires an understanding of the fact that there are causal connections between having a direct-line-of-gaze to an object, certain mental states, and behavior. I call the second, the ‘stage lights conception’—possessing it requires an understanding of the selective nature of visual attention. The third notion is seeing-as—possessing it requires an understanding of the fact that we do not just see objects; we see them as possessing certain features.

I then argue that if we want to determine which seeing relation, if any, a subject represents, we need not commit to a specific theory of mental content. Rather, we can remain neutral by focusing on the subject’s discriminative abilities: if the subject cannot discriminate seeing relation, R , from the other relations in question, there is no reason to construe her as representing R ; however, if she does possess this discriminative capacity, that constitutes evidence that she represents R —the more robust and flexible the capacity, the stronger the evidence.

I then review the experimental literature. I argue that there is evidence that nonhuman primates, corvids, and infants possess a conception of seeing that is at least as sophisticated as the headlamp conception; and some evidence that chimpanzees possess the notion of seeing-as. However, I argue that there is no evidence that nonhuman animals or infants possess the stage lights conception of seeing. Importantly, this is not because subjects have failed appropriate tests: it is because appropriate tests have not been carried out. I then outline some experimental strategies that are designed to fill this gap. I conclude by discussing ramifications for competing hypotheses about the evolution of perspective taking.

2 VISUAL PERSPECTIVE TAKING AND THE LOGICAL PROBLEM

If we want to determine whether individual, *A*, is a ‘visual perspective taker,’ two key questions must be addressed: First, which seeing relation is at stake? Second, what conditions must be in place for *A* to qualify as *representing* this relation?

2.1 Which seeing relation?

There are various ways to tackle the first question. On the one hand, we could focus on a notion of seeing that figures in folk psychology. On the other hand, we could focus on a notion that figures in some area of cognitive science. Regardless of which avenue we take, difficult questions lie ahead. For instance, it cannot be assumed, without argument, that there is such a thing as *the* folk notion of seeing, or *the* notion of seeing that figures in cognitive science. In both cases, pluralism deserves serious consideration. For instance, if Godfrey-Smith (2005) is right, the folk deploy behavioristic notions of the mental in some contexts, and representational notions in others. Similarly, it may be that different seeing relations are invoked in different areas of cognitive science, or even within vision science itself (see Phillips, 2016, forthcoming). This is difficult and contentious terrain, but until we approach it head-on, questions of the sort, ‘Does *A* attribute states of seeing to others?’ remain hopelessly unconstrained.

2.2 Which theory of representation?

In negotiating the issues outlined above, suppose we home in on a relation that is aptly construed as a ‘seeing’ relation. The next question is, ‘What conditions must be in place for *A* to qualify as *representing* this relation?’

For instance, if a creature tracks the given seeing relation to a reasonably reliable degree is that sufficient for representing it? What about a primate who only tracks it in competitive interactions with conspecifics? Is that a sufficiently robust form of tracking? In interpreting the results of visual-perspective-taking experiments, these kinds of

questions should be front and center. If we continue to neglect them, debates about which individuals are visual perspective takers will continue to flounder.

2.3 The logical problem

By way of illustration, consider the study by Hare and colleagues (2000). A subordinate and a dominant chimpanzee were housed on either side of a middle room (Figure 1).

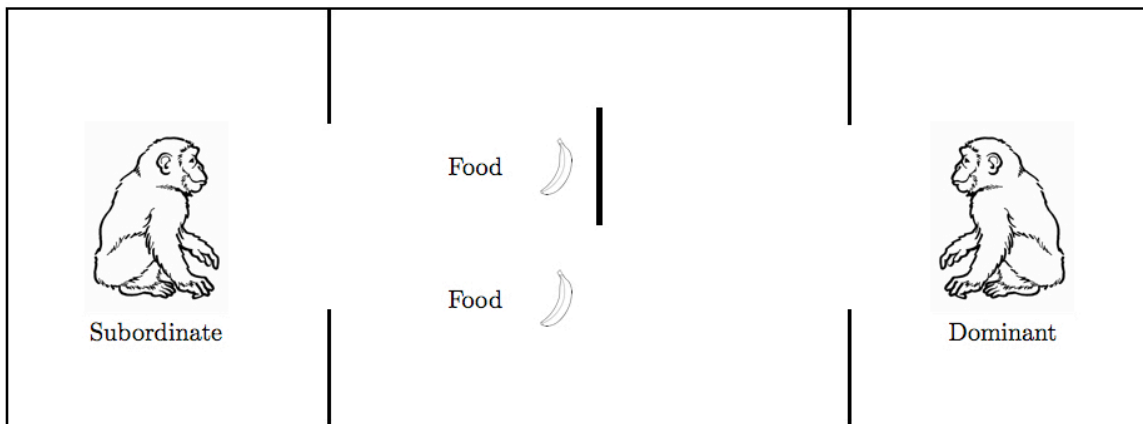


FIGURE 1 Hare and colleague's experimental setup.

One piece of food was placed out in the open so that both could see it when the doors to their rooms were open. In the *occluder test*, a second piece of food was placed on the subordinate's side of an opaque barrier. In the *transparent barrier test*, the second piece was placed on the subordinate's side of a transparent barrier. In both cases, the subordinate was given access to the middle room slightly before the dominant.

Hare and colleagues found that the subordinate had a strong tendency to only retrieve food from behind the barrier in the opaque test. According to the visual-perspective-taking hypothesis, the subordinate exhibited this tendency because he knew which food items the dominant could see.

But not all theorists accept this interpretation of the results. According to what Lurz (2011) calls the *complementary behavior-reading hypothesis*, instead of predicting the

dominant's behavior by representing him as *seeing* some pieces of food and not others, the subordinate did so by representing him as having a *direct-line-of-gaze* to some pieces and not others, where *direct-line-of-gaze* is a purely geometric relation that obtains between an individual's eyes and a target object (more on this relation below).

The systematic availability of this complementary behavior-reading hypothesis has become known as *the logical problem*. According to some theorists (Povinelli & Vonk, 2004; Lurz, 2011), the logical problem is pervasive, for the observable grounds for attributing a state of seeing *O* to some individual will normally include the fact that the individual has a direct-line-of-gaze to *O*. Thus, whenever *A* predicts *B*'s behavior in a way that depends on what *B* can see, there will be two types of hypotheses on offer: a visual-perspective-taking hypothesis, which construes *A* as representing what *B* can see; and a 'behavior-reading hypothesis,' which construes *A* as only representing which objects *B* has a direct-line-of-gaze to.

This is where the need to address questions about mental representation come in, for unless we take a stance on the criteria for *representing*, not just tracking, a distal relation, how can we adjudicate between the following possibilities regarding any individual study?

- (1) *Visual-perspective-taking hypothesis*: *A* represents *B* as seeing certain objects by representing *B* as having a direct-line-of-gaze to them.
- (2) *Behavior-reading hypothesis*: *A* merely tracks what *B* sees by representing *B* as having a direct-line-of-gaze to certain objects.
- (3) *Agnosticism*: the evidence is neutral between (1) and (2).

2.4 Promiscuous theories of mental representation

In raising these very issues, Buckner (2014) points out that teleosemantic theories of content provide promiscuous answers to the question of when tracking a seeing relation is sufficient for representing it. According to Millikan's (1984, 1993, 2000) influential account, the content of a representation is determined by the *biological functions* of the systems that consume it. By way of illustration, consider magnetosomes: organelles within magnetotactic bacteria which are attracted to magnetic north when in the Northern Hemisphere, and magnetic south when in the Southern Hemisphere. According to Millikan (1989, p. 291), magnetosomes represent *the direction of lesser oxygen*, not magnetic north or south, because the mechanisms that consume their outputs have the biological function of moving magnetotactic bacteria toward oxygen-free water (oxygen-rich water kills them).¹

Given that teleosemantic theories assign a content-determining role to biological functions, they are liberal when it comes to the question of when tracking a seeing relation is sufficient for *representing* it. As Buckner (2014) explains:

Under these approaches, an interpretation of current data as evidence for seeing-attributions is plausible, for surely line-of-gaze is only significant in a chimpanzee's cognitive economy as evidence for seeing ... Teleosemantic theories institutionalize this intuition in different ways; for example, even a state that inflexibly responded only to direct-line-of-gaze could count as representing seeing if that state's ability to track seeing explained its fitness benefits (Papineau, 1993), allowed the systems that consume that representation to perform their proper functions (Millikan, 1984), or

¹ Different teleosemanticists provide different accounts of *biological function*. See Godfrey-Smith (1994) for a discussion.

caused the representation to be recruited to a position of behavioural control in the organism's learning history (Dretske, 1988). (2014, p. 579)

Of course, not everyone accepts the teleosemantic approach to content-determination (see Shea 2013). As Buckner argues, empirical researchers therefore face a choice:

... either discard the representational idiom and characterize disagreements about social cognition in terms of more precise models, or be explicit about differences in underlying psychosemantics, exposing them to critical scrutiny. (2014, p. 579)

Notice that our question about which notion of seeing is at stake engenders a similar choice: discard all talk of 'seeing'-attributions and characterize the relevant disagreements in terms of more precise models, or be explicit about the account of seeing one is presupposing, leaving it open to critical scrutiny.

In tackling these important challenges, I start by describing various relations a creature might represent when tracking the gaze of another individual. In section 4, I then address the question of what it would take for a subject to represent one of these relations.

3 THREE CONCEPTIONS OF SEEING

Let's say that *A* has a 'direct-line-of-gaze' to *O* just in case there is an unobstructed line going from *A*'s eyes to *O*. For example, wearing an opaque blindfold would preclude *A* from having a direct-line-of-gaze to *O*, but wearing sunglasses would not.²

Even though direct-line-of-gaze is a purely geometric relation, representing it is a sophisticated ability, for it requires one to group a wide array of behavioral cues into

² If *A* can discriminate between opaque and transparent objects, this does not entail that *A* possesses a concept of seeing. As Lurz (2011, p. 36) points out, 'a chimpanzee's concept of opacity might simply be the concept C* such that if it sees (or seems to see) an object O behind/within a barrier/medium Y, then, ceteris paribus, it is disposed to believe that Y is not C*.'

abstract equivalence classes—think of all the possible combinations of head-direction, eye-direction, and differently-located objects. Computing direct-line-of-gaze is certainly a big step-up from both gaze aversion (detecting and avoiding the gaze of someone else) and co-orientation (adjusting one’s gaze so that it points in the same direction as someone else’s).³

3.1 The headlamp conception of seeing

Direct-line-of-gaze is clearly not a seeing relation: a blind individual could have a direct-line-of-gaze to the very same objects as someone with 20-20 vision. What is missing then? What are the minimal conditions that a relation must meet to qualify as a seeing relation?

A natural starting point is the observation that sophisticated gaze-followers do not just represent geometric relations between objects and eyeballs. What they represent are relations between *intentional agents* and objects. As Andrews (2012, 2017) points out, the ability to discriminate agents from non-agents is what unites all forms of mindreading, from the most basic to the most sophisticated. We can therefore mark an explanatorily fruitful distinction between those creatures who represent agents as standing in the relevant geometric relations to objects, and those who represent these relations in a way that does not respect the agent/non-agent distinction.

What emerges is a functional conception of seeing according to which *A* sees *O* just in case *A*’s state of having a direct-line-of-gaze to *O* occupies certain causal roles. This notion of seeing certainly appears in the work of various empirical researchers (see Emery, 2000; Flavell, 2004; Goossens, 2008; Rosati & Hare, 2009; Butterfill & Apperly,

³ The ability to represent direct-line-of-gaze is often referred to as ‘geometric gaze following.’ There is evidence for geometric gaze following in great apes, Old World monkeys, and New World monkeys (Rosati & Hare, 2009). There is also evidence for geometric gaze following in corvids (Bugnyar et al., 2016). It is controversial whether lemurs engage in geometric gaze following (see Ruiz et al., 2009; Sandel et al., 2011; Bray et al., 2014).

2013; Wellman, 2014). For instance, Wellman argues that infants' understanding of seeing goes beyond the grasp of an 'externalized line of directedness from eyeballs to objects,' for they understand that 'seeing often produces inner, subjective experiences such as emotions and desires as well' (2014, p. 86).

Which causal roles must be occupied for the relation in question to qualify as a *seeing* relation? Is it sufficient that having a direct-line-of-gaze to *O* interacts with the agent's desire for *O*, causing an attempt to retrieve it? Do there need to be causal connections to other states, such as memories and beliefs? It is hard to think of a principled way of answering these questions. Perhaps the best we can do is delineate a continuum of seeing relations, constituted by increasingly complex causal connections among various types of mental states. Nonetheless, if the goal is to home in on the *lower bounds* of visual perspective taking, we should recognize a *basic* functional conception according to which having-a-direct-line-of-gaze to an object occupies basic causal roles, such as those involving intentions (e.g. the intention to obtain a piece of seen food) and emotions (e.g. fear of a seen predator). Which causal roles might a more sophisticated visual perspective taker represent? Two candidates, which I elaborate on below, concern those roles constitutive of selectively attending to an object, and those constitutive of seeing an object *as* having certain features.

For reasons that will become clear below, let's call this basic functional conception of seeing, 'the headlamp conception.' According to this conception, seeing an object is like directing a headlamp towards it. Different agents have their own headlamps, which can be directed at different objects, but so long as an agent is awake, she sees every object at which her headlamp is directed.

3.2 The stage lights conception of seeing

The lower bounds of visual perspective taking are thus populated by those creatures who possess the headlamp conception of seeing. Other than seeing-as, is there any conception of seeing more sophisticated than the headlamp conception? In searching for a notion of interest, consider the following question: Can a sighted agent have a direct-line-of-gaze to an object, but fail to see it? Suppose I'm looking for Waldo, who is standing among a large crowd of people. I have a direct-line-of-gaze to him, and yet I haven't noticed him. Do I *see* him?

One answer is that I see Waldo, but not as well as the objects to which I'm attending. Another answer is that I fail to see Waldo, precisely because I'm not attending to him. I won't attempt to adjudicate between these positions. For our purposes, we can remain neutral and focus on what they have in common: a commitment to the fact that one can have a direct-line-of-gaze to multiple objects, and yet one's attention is divided between them unevenly.

Let's call the conception of seeing that incorporates the distinction between having a direct-line-of-gaze to an object and visually attending to it, 'the stage lights conception.' Just as stage lights illuminate different objects on the stage to different degrees, one can selectively attend to various objects to different degrees, even if one has a direct-line-of-gaze to them all.

It is common for vision researchers to distinguish between *overt* and *covert* visual attention. A shift in overt attention involves a shift in eye direction, while a shift in covert attention does not. In either case, it is possible for *A* to see two objects, but devote more attentional resources to one at the expense of the other (Wright & Ward, 2008). Relatedly, *A*'s attention can shift from one object to the other, even though she still sees both. It is this distinction between seeing multiple objects and selectively

attending to some more than others that is understood by those in possession of the stage lights conception.

The stage lights conception is thus more sophisticated than the headlamp conception. According to the headlamp conception, *A* sees an object just in case *A*'s having a direct-line-of-gaze to it occupies certain causal roles involving intentions or emotions. Importantly, none of these roles encompass the distinction between seeing an object and selectively attending to it.⁴

3.2.1 The stage lights conception as a genus

It is important to leave room for the possibility that some creatures have a more sophisticated understanding of selective attention than others. In doing so, we can think of the stage lights conception as a genus. For instance, some creatures' understanding of selective attention may be limited to the fact that it can be driven by stimulus salience (e.g. a bright yellow banana surrounded by dark green leaves). More sophisticated creatures may understand that selective attention can be allocated to objects in a top-down fashion, based on the agent's goals and expectations.

There will also be differences in creatures' beliefs about which kinds of entities can be selectively attended to. Researchers commonly distinguish between object-based attention, spatial attention, and feature-based attention (see Kravitz & Behrmann, 2011). For example, a creature who understands object-based attention may not understand that *features*, such as colors, can be selectively attended to as well.

⁴ It is common for researchers to liken selective visual attention to a 'spotlight' or 'beam' (see Posner, 1978). In using the stage lights metaphor, I'm leaving room for the view that visual attention can be split among multiple objects, as it appears to be in multiple object tracking.

3.3 Seeing-as and the stage lights conception

The third conception is *seeing-as*. Possessing it requires one to understand that agents don't just see objects: they see them as possessing certain properties.

What exactly is the difference between the stage lights conception of seeing and the notion of seeing-as? One might think that grasping the former requires one to grasp the latter. For instance, if I know that you are not attending to Waldo, even though you have a direct-line-of-gaze to him, isn't that tantamount to knowing that you don't see him *as such*? Not necessarily.

A useful way of characterizing the stage lights conception is to think of an agent with a direct-line-of-gaze to various objects, where each line-of-gaze is weighted. For example, Waldo receives a weight of zero (the agent is not attending to him at all); the mountain in the background receives a weight of 3 (the agent attends to it a little); and the ice cream truck in the foreground receives a weight of 7 (the agent attends to it much more than either Waldo or the mountain). We could characterize these weights in dispositional terms: if the agent's desire to meet up with Waldo is of the same strength as her desire for ice cream, then her allocation of attention means that she will probably approach the truck, not Waldo. What is important about these weightings is that they bracket *how* each object is seen: all they specify is the degree to which each object is attended to by the agent.⁵

In principle, then, a creature could possess the stage lights conception without having any grasp of seeing-as. The converse is also true. For example, *A* could represent *B* as seeing one object as yellow and another object as green, without having any

⁵ *A* could determine that *B* is selectively attending to *X* without understanding what *B* sees *X* *as*. How? By grasping rules that connect distributions of distal features to allocations of attention. For example, *A* might understand that bright objects are attention-grabbing; that when objects are surrounded by like-colored objects they are less attention-grabbing; and so on.

understanding of the fact that *B* is attending to the yellow object (a desirable banana) more than the green object (an unappetizing leaf).

Some obvious questions arise at this point: (1) How do we tell which of the two conceptions, if any, a given subject possesses? (2) Which conception develops first? (3) Which conception evolved first in our lineage? To answer these questions, researchers will need to run comparative studies of the sort outlined in sections 7 and 8.

Comparative studies will also tell us how the stage lights conception of seeing maps onto the Level 1/Level 2 distinction. Possession of the headlamp conception qualifies one as a Level 1 perspective taker: an understanding of seeing-as qualifies one as a Level 2 perspective taker. If it turns out that an understanding of selective visual attention develops after an understanding of seeing-as, then it will make sense to construe the former as ‘Level 3 perspective taking.’ If, however, an understanding of selective visual attention develops before an understanding of seeing-as, it will make more sense to construe the former as ‘Level 1.5 perspective taking.’

4 REPRESENTING SEEING RELATIONS

We have just uncovered three kinds of seeing relations that an individual might represent. How might we determine which one a given individual represents without getting into the weeds and defending a specific theory of mental content? As a first step, consider the relationship between an individual’s discriminative and representational capacities.

4.1 A discrimination constraint

Suppose I train my cat, Oliver, to raise his paw when presented with a cell phone. He learns to discriminate my cell phone from various other objects around the house. However, his discriminative abilities are not systematic enough for him to qualify as representing the abstract category *cell phone*—if I show him anything with the same

three-dimensional shape as my phone, he raises his paw in the requisite fashion.

Presumably, what Oliver *can* represent is the three-dimensional property shared by cell phones and other like-shaped objects. In other words, what Oliver represents is a feature that is not co-extensive with *being a cell phone*, but serves as a reliable indicator of this category.

What would it take for Oliver to qualify as representing the category *cell phone*? At the very least, he would need to harbor an ability to discriminate some cell phones from some cell-phone-shaped objects that are not in fact cell phones. Given that he cannot do this, there is no motivation for construing him as representing the category *cell phone*—attributing this representation to him certainly wouldn't give us any novel predictive or explanatory power.

If we apply the same reasoning to an individual's ability to discriminate between the various relations uncovered above, we might gain traction on the question of which ones she represents without committing to a specific theory of content. For instance, suppose *A* tracks direct-line-of-gaze across a wide variety of circumstances, and is relatively successful in using it to predict the actions of sighted agents. However, suppose *A* cannot discriminate between direct-line-of-gaze and the more abstract relation picked out by the stage lights conception of seeing. We would therefore have good reasons for thinking that *A* represents the relation picked out by the headlamp conception of seeing, but not the one picked out by the stage lights conception—asserting that she represents the latter would not provide us with any novel predictive or explanatory power.

4.2 Discrimination and biological function

Don't some teleosemantic theories flout our discrimination constraint? Consider Fodor's (1990) famous example of frogs snapping at flies. Suppose a frog does not discriminate

between flies and moving black dots. If we adopt the line of reasoning given above, this will lead us to the conclusion that the frog does not represent flies.

However, some teleosemanticists have argued that frogs determinately represent flies. For instance, Sterelney (1990, p. 127) enjoins us to check the relevant counterfactuals: had flies been white, ancestral frogs would have snapped at moving white dots, not moving black ones. According to Sterelney, the truth of counterfactuals like this one motivates the thesis that frogs' visual systems have the biological function of tracking flies.

One might argue that Millikan's consumer-based theory (1984, 1993) gives us the same content-assignment. Arguably, a frog's motor system—which controls the snapping action of the tongue—performs its proper function when it snaps at flies. The state that this snapping mechanism consumes therefore represents flies, not moving black dots.

In a similar vein, consider Pietroski's (1992) useful example of the kimu. The kimu are creatures who were colorblind until a mutation enabled them to detect red things. Each morning, they climb a hill to watch the (partly red) sunrise. As it happens, this behavior is adaptive, for the only creatures who hunt kimu are snorfs, and snorfs are poor climbers: what's more, they only hunt at dawn.

Call the state that a kimu tokens when confronted with red, *M*. What does *M* represent? A natural suggestion is *red*. However, notice that detecting red is only advantageous for kimu when it takes them away from snorfs. Thus, according to a consumer-based teleosemantics, the content of *M* is something like *fewer-snorfs-here*—a conclusion Millikan (2000, p. 149) has explicitly endorsed. For our purposes, the key point is that Millikan is committed to this content-assignment even if we stipulate that the kimu cannot discriminate snorfs from non-snorfs. Perhaps their green color renders snorfs invisible to the kimu (e.g. put a strawberry right next to a snorf, and a kimu will

readily approach it). On Millikan's view, the representational abilities of the kimu outstrip their discriminative abilities.

4.3 Discriminative capacities as evidence of biological function

These examples demonstrate that a teleosemanticist might flout the discrimination constraint. What are we to do then? We could forego neutrality and disregard any theory of content that falls afoul of the constraint. But doing so would only pull us into controversial debates about the explanatory role of content.

Another tack would be to drop all talk of representational capacities, and limit ourselves to questions about which seeing and gazing relations members of a given species can *discriminate* from one another. Having found answers to these sorts of questions—for a wide enough variety of species—we could then address questions about how and why these discriminative abilities evolved. For instance, we could ask: How and why did the ability to discriminate direct-line-of-gaze from the seeing relation picked out by the stage lights conception evolve in species *X* and *Y*? Why is that members of *X* can discriminate between these two relations in circumstances *C*, but members of *Y* cannot? Why didn't the ability to discriminate between these two relations evolve in outgroup *Z*? By helping ourselves to the tools of comparative psychology, we could answer these questions without weighing in on the representational issue at all.

It is certainly tempting to leave representational notions behind, and forge ahead with questions about discriminative capacities. I think we can do better, though. For suppose *S* tracks direct-line-of-gaze across a wide variety of circumstances. If *S* thereby represents one of the seeing relations identified above, how do we determine *which* one it is? Each relation is such that tracking it could have been adaptive for *S*'s ancestors. For instance, tracking the headlamp relation would have enabled them to anticipate the behavior of conspecifics when competing for food; tracking the stage lights relation

would have enabled them to do this in a more nuanced fashion, when some food items were harder to see than others; and tracking seeing-as would have enabled them to anticipate the behavior of conspecifics in illusory settings. Given these distinct advantages, how do we determine which relation, if any, S 's state-type has the *function* of tracking? It is in answering this question that an appeal to discriminative capacities becomes helpful, even for the promiscuous teleosemanticist.

To illustrate, suppose members of a given species consistently pass a battery of tests, each one requiring the subject to discern which objects her competitor has a direct-line-of-gaze to. S is one of the individuals who consistently passes these tests. Call the relevant state-type, T . Which seeing relation does T have the function of tracking? Let's suppose that, despite S 's success in food competition tasks of the sort carried out by Hare and colleagues (2000), she consistently fails any task that requires her to discriminate between a competitor who sees the target object *as F* and a competitor who sees it *as G*, where *being F* and *being G* are any pair of distinct properties (e.g. *being green* versus *being blue*). In that case, we have no reason to construe S as representing seeing-as.

How might a promiscuous teleosemanticist accommodate this conclusion? Presumably, the answer is that S and her conspecifics cannot discriminate between seeing-as and the less fine-grained relations catalogued above. We therefore have no evidence that it was *adaptive* for S 's ancestors to track anything as abstract as seeing-as. If, somehow, we had independent reasons for thinking that S 's ancestors were kinuesque in acquiring an adaptive ability to track seeing-as, their discriminative shortcomings would not matter. But clearly, we do not have independent evidence of this sort.⁶ We

⁶ What would independent evidence look like? I'm not aware of actual examples, but imaginary cases are not hard to construct. For example, suppose snorfs can in fact climb the hill that the kimu climb at sunrise. The light at sunrise makes the kimu look red, which happens to be the color of skimu—like-shaped creatures who are poisonous, and therefore avoided by snorfs. The state-type that the kimu token

therefore have no reason to construe *S*'s state as having the function of tracking seeing-as. Construing *S*'s state as such certainly wouldn't provide us with novel explanatory power. Attributing the headlamp conception to *S* suffices to explain her successful performance in food competition tasks of the sort carried out by Hare and colleagues (2000).⁷

What conclusions would the promiscuous teleosemanticist be entitled to draw if *S* and her conspecifics were in fact able to discriminate seeing-as from the other relations in question? Presumably, it would provide evidence that they possess states with the function of tracking, and therefore representing, seeing-as—the more robust and flexible the ability, the stronger the evidence. Importantly, this converges with a claim that proponents of non-promiscuous theories of content will readily endorse: namely, if *S* discriminates seeing relation, *R*, from the less abstract relations that that are often co-instantiated with it, that provides some evidence that *S* represents *R*—the more robust and flexible the ability, the stronger the evidence.⁸

We therefore have enough agreement on the evidential status of discriminative behavior to guide experimental work on the different varieties of visual perspective taking. For even though not all will agree that an ability to discriminate seeing relation, *R*, from the other relations in question is *necessary* for representing *R*, everyone—including the promiscuous teleosemanticist—should agree on two points: (1) if individual, *S*, fails to exhibit this ability then, unless there is independent evidence for kimuesque tracking abilities, there is no reason to construe *S* as representing *R*, and (2)

when climbing the hill therefore has the function of tracking *how* snorfs see them: namely, *as* poisonous skimu. However, the kimu cannot discriminate between being seen as skimu, and being seen as kimu.

⁷ Rather than remaining agnostic, many will draw the stronger conclusion that *S*'s state does not have the function of tracking seeing-as. The agnostic conclusion will suffice for our purposes.

⁸ See Allen (1999) and Buckner (2014) for further discussion concerning just *how* robust and flexible the subject's discriminative capacities must be.

if S does exhibit this ability then that provides us with some evidence that she represents R .⁹

The upshot is that regardless of one's view on mental content, unless we test subjects on their ability to discriminate the various relations catalogued above, we will be saddled with what we might call 'logical problems'—experimental results that fail to tell us which kind of visual perspective taker we have on our hands. Nonetheless, it is worth reiterating that even if we were to set the issue of mental representation aside, the task of determining how and why various species' discriminative capacities evolved would remain. The comparative studies reviewed below, which test the discriminative capacities of various species, would therefore retain their importance.

5 INFANTS, NONHUMANS, AND THE HEADLAMP CONCEPTION OF SEEING

With these caveats in mind, I'll now present evidence that infants and various nonhumans possess a conception of seeing that is at least as sophisticated as the headlamp conception. I'll also describe the kinds of gaze-followers who fall just short.

5.1 Nonhumans and the headlamp conception

In addition to chimpanzees, various other nonhumans have passed food competition tasks like the one designed by Hare et al. (2000). They include rhesus macaques (Flombaum & Santos, 2005); ringtailed lemurs (Sandel et al., 2011; Bray et al., 2014), and certain corvids (Dally et al., 2006; Bugnyar, 2011; Bugnyar et al., 2016).

For some of these studies, it is debatable whether subjects passed by deploying the headlamp conception of seeing, or, by engaging in mere gaze-following.¹⁰ However, in the

⁹ Evidence that S represents R and constitutive conditions for S 's representing R are not the same thing. We can therefore take a stance on what would provide good reasons for construing S as representing R , while remaining neutral on what *constitutes* S 's representing R . See Allen (1999) on this point.

¹⁰ For example, see Ruiz et al. (2009) for a discussion of ringtailed lemurs and whether they represent direct-line-of-gaze, or, whether they just look reflexively in the same direction as conspecifics.

version run by Hare and colleagues (2000), it is hard to explain chimpanzees' success without attributing to them a conception of seeing that is at least as sophisticated as the headlamp conception. Recall that to pass the test, subordinates had to (i) represent the dominant as only having a direct-line-of-gaze to the food out in the open, and (ii) use this knowledge to predict that the dominant will only retrieve the food out in the open. This requires the subordinate to understand at least one of the causal roles that having-a-direct-line-of-gaze to something plays in generating intentional action.¹¹

It is also hard to explain the performance of ravens without attributing a conception of seeing to them that is at least as sophisticated as the headlamp conception. In a recent study by Bugnyar and colleagues (2016), ravens cached food in a room that was visible from an adjacent room through a peephole. When the sounds of other ravens were played, those subjects who had already looked through the peephole themselves tended to guard their caches. This is compelling evidence that ravens understand the causal connection between seeing a desirable object, and performing actions on it (in this case, intentional pilfering).

There is also evidence that nonhuman primates grasp causal connections between direct-line-of-gaze and certain affective states. For instance, Mineka et al. (1984) found that laboratory-raised rhesus macaques can acquire a fear of snakes via observational learning. Subjects acquired the phobia by watching their wild-reared parents behaving

¹¹ Couldn't the subordinate represent direct-line-of-gaze in a way that falls short of grasping this causal role? For example, perhaps the subordinate deploys the following rule: *Avoid food to which a dominant has a direct-line-of-gaze*. This is sometimes referred to as 'the evil-eye hypothesis' (Povinelli & Vonk, 2004). Notice that, unlike the complementary behavior-reading hypothesis, the evil-eye hypothesis does not constitute a *systematic* alternative to the visual-perspective-taking hypothesis. By utilizing appropriate controls, it can therefore be ruled out. In fact, Schmelz and Call (2016) recently did just this. See also Call and Tomasello (2008) for a review of converging evidence that chimpanzees understand at least some of the causal roles that having-a-direct-line-of-gaze to something plays in generating intentional action.

fearfully while looking at real, toy, or model snakes. Similarly, Goossens et al. (2008) found that longtailed macaques are more likely to follow a human's gaze when the human looks with a negative expression (e.g. a bared-teeth display, which is indicative of submission).

5.1.1 Revisiting the logical problem

As was explained above, Lurz (2011) claims that individuals may succeed on Level 1 tasks by representing facts about direct-line-of-gaze, not facts about 'seeing.' If it is the headlamp conception of seeing we are interested in, though, Lurz's argument has no bite. Recall that according to the headlamp conception, *A* sees *O* just in case *A*'s having a direct-line-of-gaze to *O* occupies the requisite causal roles. Given that the individuals Lurz is referring to succeed on Level 1 tasks by grasping some of the causal roles that having a direct-line-of-gaze to an object plays in generating intentional action, they qualify as possessing the headlamp conception.

In fairness to Lurz, he may be concerned with a notion of seeing more sophisticated than the headlamp conception. It appears unlikely that he has something like the stage lights conception of seeing in mind: none of the experiments he considers were designed to test subjects' grasp of selective visual attention (I suggest some strategies below in section 7). The onus is thus on Lurz to explicate the notion of seeing he is tacitly employing, and how it differs to both the headlamp conception and the stage lights conception.

5.2 Infants and the headlamp conception

Much of the evidence that infants possess a conception of seeing at least as sophisticated as the headlamp conception comes from studies of joint attention. For instance, Liszkowski and colleagues (2004) elicited declarative points from 12-month-olds, while varying an adult's reaction to them. In one condition, the adult alternated gaze between

object and infant, commenting about the object in an interested fashion (e.g. ‘Wow, it’s a dog!’). In another condition, the adult gazed with positive emotion at the infant, but not the object. In a third condition, the adult looked at the object, but not the infant. Liszkowski and colleagues found that infants were only satisfied—e.g. stopped pointing—in the first condition, when the adult alternated gaze between object and infant in an excited fashion.

Similar results have been obtained by Moses et al. (2001); Phillips et al. (2002); Liszkowski et al. (2007); Liebal et al. (2010); and Moll et al. (2006, 2008). In the study by Moll and colleagues (2008), an adult and an infant played with three objects together: they played with one of them in an excited fashion, and the other two in a calm fashion. In the testing phase, the adult reacted excitedly when presented with all three objects on a tray, before requesting one of them in an ambiguous fashion (e.g. ‘Will you hand it to me?’). 14-month-olds gave the adult the object that they had played with together in an excited fashion.

Together, these studies provide compelling evidence that infants understand some of the causal connections between having-a-direct-line-of-gaze to something and mental states, such as intentions and emotions.

5.3 Evidence for mere gaze-following

Which species fall just short of possessing the headlamp conception? Homing in on this side of the border between the mere gaze-followers and the visual perspective takers will shed light on how and why the most basic forms of visual perspective taking evolved.

As an initial step, consider gaze aversion: the act of detecting and avoiding the gaze of another individual (e.g. a predator). Suppose *S* engages in gaze aversion, but no other form of gaze-tracking: for example, *S* cannot follow the gaze of another individual to a specific object. Let’s also suppose that *S* reacts the same way, regardless of whether the

eyes she detects are attached to an agent. For example, if she sees a pair of forward-facing eyes on a rock, she still engages in avoidance behavior.

Does *S* possess the headlamp conception of seeing? Does she understand that there is a causal connection between a predator's having a line-of-gaze to her, and that predator's decision to pursue her? Perhaps she does: it is just that she is prone to false positives. On the other hand, it may be that *S*'s system for gaze aversion is not integrated with her beliefs about the intentions of others: she may not even have the capacity to represent the intentions of others. Either way, let's suppose that the causal pathway going from *S*'s eye-detection system to the motor system that controls her avoidance behavior is direct, bypassing any intention-detection system. *S* thereby fails to discriminate between agents with forward-facing eyes, and non-agents with forward-facing eyes, without this being a matter of false positives. This means that *S* fails to discriminate between line-of-gaze and the seeing relation picked out by the headlamp conception: she is a mere eye-direction-detector.¹²

Are there any species that fit this description? Wild house mice (Topal & Csanyi, 1994) and African jewelfish (Coss, 1979) are good candidates. Members of both species exhibit gaze aversion when presented with a pair of horizontally-aligned black spots. However, to date, there is no evidence that members of either species attribute intentions to others, let alone represent causal connections between having a line-of-gaze to an object and intentional action. Of course, the lack of positive evidence may be due to a lack of studies. The important point is that if we want to locate the border between those species who are mere gaze-detectors, and those who possess the headlamp

¹² Baron-Cohen (1995) posits a specialized *eye-direction detector* (EDD) or 'gaze module.' He also posits an *intentionality detector* (ID); a *shared attention mechanism* (SAM); and a *theory of mind module* (ToMM). If we adopt this framework, then a creature is a *mere eye-direction-detector* just in case the outputs of its EDD are inaccessible to those systems involved in the representation of intentional behavior.

conception of seeing, these are the kinds of creatures that comparative researchers should start with.¹³

6 INFANTS, NONHUMANS, AND THE STAGE LIGHTS CONCEPTION OF SEEING

I have presented evidence that 12-month-old infants and various nonhumans can discriminate direct-line-of-gaze from the more abstract relation picked out by the headlamp conception of seeing. Moreover, the fact that this discriminative capacity is relatively robust in some species constitutes evidence that some of them *represent* the seeing relation picked out by the headlamp conception. Is there any evidence of infants or nonhumans discriminating between direct-line-of-gaze and the relation picked out by the stage lights conception of seeing?

6.1 Do infants possess the stage lights conception?

Miller and Bigi (1977) presented children with 47 shapes of different colors and sizes. In one test, children were asked to construct “easy” and “hard” versions of a game in which the player must find two identical red triangles. The children constructed the game by surrounding the red triangles with other colored shapes. In another trial, children were asked to rank, from “easiest” to “hardest,” gameboards that had already been constructed. First-graders (with an average age of 7) understood that placing more objects around the targets made them harder to find. However, only third- and fifth-graders understood that placing *red triangles* around the targets would produce the hardest game.

¹³ Similar remarks apply to reflexive co-orientation. It is possible that some creatures align their gaze with conspecifics in a way that is divorced from predictions of intentional behavior. See Ruiz et al. (2009) for a highly relevant discussion of gaze co-orientation in lemurs.

Similar studies have been carried out by Miller and Bigi (1979); Pillow (1988, 1989); Flavell et al. (1995); Fabricius et al. (1997); and Parault et al. (2000). These studies provide evidence that by 5 years of age, children understand that visual attention can vary, while line-of-gaze is held constant. However, it would be far too hasty to conclude that children do not acquire the stage lights conception of seeing *until* 5 years of age.

First and foremost, these studies did not use infants as subjects. Second, they all deployed tasks with relatively high demands. For instance, Fabricius and colleagues (1997) simply asked subjects, ‘Can somebody look at something, but not see it?’ Third-graders tended to say ‘no.’ However, this may have been due to the abstract nature of the question. Moreover, third-graders may think that it is *harder* to see some objects than others, even though every object looked at is seen: a belief that would qualify them as possessing the stage lights conception of seeing.

To test infants’ grasp of the stage lights conception, we clearly need experiments with lower task demands. In section 7, I outline an appropriate experiment. But first, I want to discuss a rare study that was designed to test infants’ grasp of the fact that one’s allocation of visual attention can vary, even when one’s gaze remains fixed.

In the study in question, Moll and colleagues (2006) got 14-, 18-, and 24-month-olds to watch as an adult gazed excitedly towards a sticker, which was placed on one side of an object. In one condition, the object was new for the adult—she had not been present when the infant and another adult played with it together. In the other condition, the adult was familiar with the object—she and the infant had played with it together. Importantly, in both conditions, the adult’s direction-of-gaze was identical.

Moll and colleagues found that when the object was new for the adult, the infant responded to the adult’s excited utterance (e.g. ‘Oh, great, look!’) by attending to the object as a whole. In contrast, when the object was old for the adult, the infant

attempted to either localize something specific on it or find another object in the adult's visual field. Moll and colleagues drew the following conclusion:

These children thus assumed different attentional foci between the conditions even though the person was behaving in the same way and looking at the exact same spot in both conditions ... Infants not only know something about what others see, or have visual access to; they also know something about what people attend to selectively within their visual fields. (2006, p. 426)

If Moll and colleagues are right, this is a striking result. It would mean that at 14 months of age, infants have acquired the stage lights conception of seeing. But do the results really show this? In fact, I think that the results, while important, are neutral between the hypothesis that infants understand that an individual's attention can be divided unevenly between the objects in her visual field, and the hypothesis that infants only grasp the selective nature of linguistic reference.

Just as one can have a direct-line-of-gaze to various objects, even though one selectively attends to just one of them, one can have a direct-line-of-gaze to various objects, even though one selectively *refers* to just one of them. What is the developmental relationship between an understanding of selective attention, and an understanding of the selective nature of reference? It is certainly plausible that an increasingly robust grasp of selective attention fosters an understanding of the selective nature of reference: if I know that you are visually attending to the banana, not the apple, even though you see both, that will help me in determining the referent of your utterance, 'I want that!' It is also plausible that the influence flows in the other direction: if I'm good at reference disambiguation, that will aid me in determining which objects, within your visual field, you are selectively attending to. For example, if I know that by 'Wow, that's a nice sticker!' you are referring to sticker on the side of the toy,

not the toy itself, that will help me to understand that you are visually attending to the sticker, not the object to which it is attached.

Of course, it may be that an understanding of selective visual attention develops at the same time as an understanding of linguistic reference, each one reinforcing the other. On the other hand, it may be that one develops before the other. There is no a priori guarantee that understanding the selective nature of one type of state brings with it an understanding of the selective nature of another. For instance, it is widely held that chimpanzees understand that an individual's intention can be about a single object, even though the individual has a direct-line-of-gaze to numerous objects (see Call & Tomasello, 2008). However, no one, to my knowledge, has argued that chimpanzees thereby understand the selective nature of visual attention.

Thus, in concluding that the infants in their study exhibited an understanding of selective visual attention, Moll and colleagues (2006) overreach. They need to rule out the alternative hypothesis that in responding to the adult's excited utterance of 'Oh, great, look!' infants were merely exhibiting an understanding of the selective nature of linguistic reference. More specifically, in gazing towards the sticker in the old-object condition, infants may have believed that the adult saw every object in her visual field: it is just that *the sticker* was the one that she was referring to in uttering 'Oh, great, look!',^{14,15}

6.2 Nonhumans and the stage lights conception

At present, there is thus little evidence that infants grasp the selective nature of visual attention: the right sorts of tests have just not been carried out. Similar remarks apply

¹⁴ 'Oh, great, look!' is elliptical for 'Oh, great, look at that!' The speaker's intended referent therefore varies across the old- and new-object conditions, even though she doesn't utter a term that refers to either the sticker or the toy.

¹⁵ Developmental studies of *fast mapping*—the ability to learn the referent of a lexical item after minimal exposure—lend independent plausibility to this hypothesis (see Carey, 2010).

to nonhuman animals. When it comes to joint attention, nonhuman primates are outperformed by human infants (Tomasello, 2014). Moreover, given their relative lack of linguistic competence, no one has tried to perform the kind of study carried out by Moll and colleagues (2006) on nonhumans. As things stand, we therefore have very little to go on. In what follows, I make steps towards filling this gap.

7 HOW TO TEST FOR THE STAGE LIGHTS CONCEPTION OF SEEING

Before suggesting some experimental strategies, it is worth considering why it might be adaptive to possess the stage lights conception of seeing. As was pointed out above, an understanding of selective attention may serve as a catalyst for the infant's growing facility with speaker reference. Alternatively, the infant's developmental trajectory might start with an understanding of speaker reference, which then gives rise to an understanding of selective attention. And, of course, it is possible that these abilities emerge at the same time. Unless we test for an understanding of selective attention in the absence of the need for reference disambiguation, any evidence will be neutral between these hypotheses.

Why might an understanding of selective visual attention have evolved in nonhumans? Perhaps it gave creatures an advantage in competitive foraging contexts. For example, if a subordinate knows that a dominant has a direct-line-of-gaze to multiple food items, but one of these items is much harder to see than the others, the subordinate might anticipate that the dominant will retrieve one of the easier-to-see items first, leaving the harder-to-see item open for the taking.

7.1 Testing infants

Using the insights gleaned above, here is one way to test children's understanding of selective visual attention. A child will sit on a chair facing two arrays of colored balls (see Figure 2). In the pre-testing phase, the child will be asked to locate two green balls:

one contained in each array. One of the green balls will be highly salient (e.g. it will be immediately surrounded by yellow balls). The other green ball will be far less salient. The order in which the child locates the two balls will be observed. The purpose of this pre-test is to ensure that, from the child’s perspective, one of the green balls is more salient than the other one.

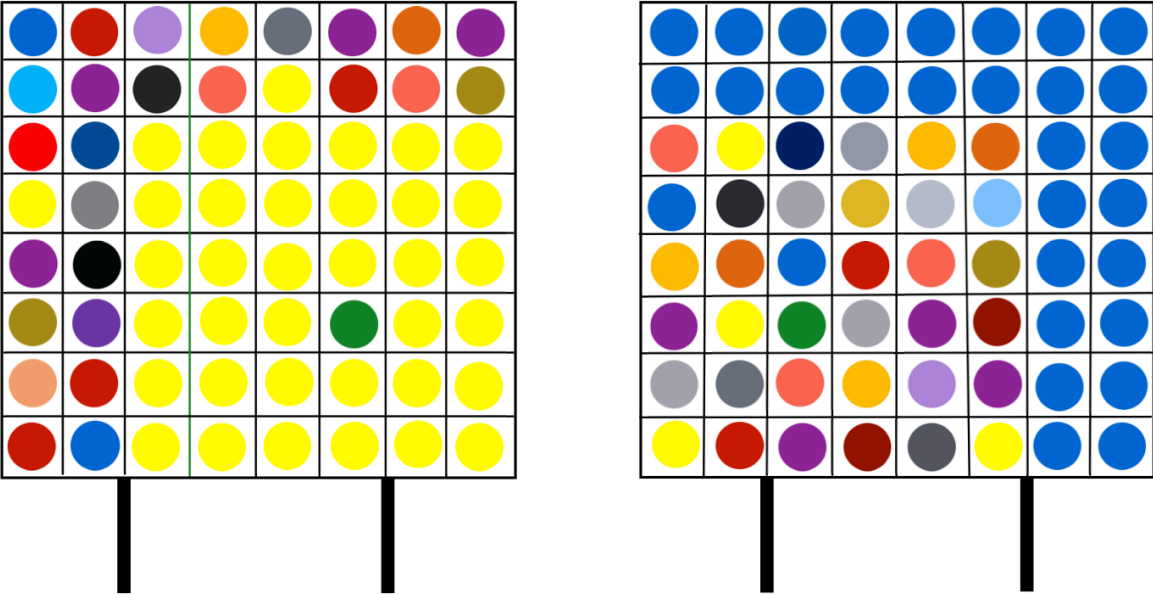


FIGURE 2 Experimental set-up for study on infants’ grasp of selective visual attention. A color image is available in the online version of this paper.

A researcher will then place both balls in new positions, before touching both and saying, ‘Look! Here’s one, and here’s the other one.’ The child will then be asked to locate both balls again. The purpose here is to ensure that the child *now* knows where both balls are located, even though, initially, one was less noticeable. Once a camouflaged object has been made salient, it is often hard to ‘unsee’ it.

In the testing phase, an experimenter, *A*, will face the child while holding both green balls. Another experimenter, *B*, will then take each ball and place them in the arrays (*A* will continue to face the child). *A* will then turn around and retrieve one of the green balls. In the *low-salience condition*, *A* will retrieve the green ball that is least salient. In the *high-salience condition*, *A* will retrieve the green ball that is most salient.

The purpose is to determine whether the child is surprised when *A* retrieves the ball that is least salient, and unsurprised when *A* retrieves the ball that is most salient. Gazing behavior could be used to determine whether the child is surprised by *A*'s choice. If the child expects the adult to retrieve a given ball, she should look at it in anticipation of the adult's action. Moreover, if the adult's choice violates the child's expectation, the child should look towards the adult more than she does when her expectation is met.¹⁶ Verbal tests could also be run. For example, the experimenter could ask the child, 'Which ball do you think she'll find first?'¹⁷

One worry is that the child may predict *A*'s choice by deploying a behavioral rule of the form: *People tend not to retrieve objects that are immediately surrounded by like-colored and like-shaped objects*. To control for this possibility, a condition could be included in which experimenter *A* sees the placement of both balls before retrieving one. If the child is using a rule of this sort, she should be surprised when *A* retrieves the harder-to-notice ball, even though *A* saw its initial placement.

7.2 Testing nonhumans

A similar study could be run on nonhuman primates. A subordinate and a dominant could compete for grapes, housed inside blue boxes. There could be two arrays of

¹⁶ See Baillargeon et al. (2010) who have used both anticipatory-looking and violation-of-expectation tasks to test the mindreading abilities of pre-verbal infants.

¹⁷ Variants of this experiment could be used to determine which varieties of selective attention children understand (e.g. bottom-up versus top-down; object-based versus feature-based; and so on).

colored boxes, each one containing a blue box, surrounded by other colored boxes (all non-blue and empty).

In the training phase, subjects could learn that each blue box contains grapes, while non-blue boxes are empty. Given that chimpanzees have demonstrated an ability to learn which colored boxes contain which rewards in past studies (e.g. see Krachun et al., 2010), it is likely that they will reach criterion on these tests.

In the testing phase, the degree to which each target pops out could easily be manipulated by varying the colors of the distractors. Importantly, before each trial starts, both targets could be made salient to the subordinate while the dominant's view is occluded. For example, an experimenter could pick them up and indicate their location to the subordinate.

If the subordinate understands the selective nature of visual attention, the box he tends to retrieve should be the one that is less salient to the dominant. Moreover, his tendency to retrieve the less salient target should increase as its relative saliency decreases.

In a variant of this test, looking times could be measured. Instead of setting up a scenario in which subordinate and dominant compete with one another, the subject could just watch as a competitor retrieves one of the boxes. The direction and duration of the subject's gaze could be used to gauge which box he expected the competitor to retrieve first.

Variants of this strategy could be used with corvids. As was pointed out above, corvids guard their caches against potential pilferers if their initial attempts at caching were seen or heard. To test whether they possess the stage lights conception, subjects could be presented with multiple caching options: trays with contents that make the location of cached food easy to see, and trays with contents that make it much harder. For example, trays could be housed inside semi-transparent colored boxes, which make

the landmarks in certain trays harder to see than the landmarks in others. If subjects understand that some objects are harder to see than others, this may be reflected in which tray they prefer to cache in.¹⁸

8 INFANTS, NONHUMANS, AND SEEING-AS

When does the ability to attribute seeing-as develop? Which nonhumans, if any, represent seeing-as? Developmental research suggests that children begin to understand seeing-as from about 3 years of age, while research on nonhumans has been very limited.

8.1 Children's understanding of seeing-as

In Moll and Meltzoff's (2013) study, two blue pictures (e.g. of a dog) were placed behind transparent screens. One screen was fully transparent, while the other served as a yellow filter. The position of each screen was such that from the child's perspective, both pictures looked blue, whereas, from the experimenter's perspective, the one behind the yellow filter looked green. In the *perspective-taking task*, the experimenter requested the picture that looked green. Most children 3 years of age and older correctly selected the picture behind the yellow filter.

In the *perspective-confronting task*, the experimenter asked, 'How do I see the dog from my side over here? Do I see it like this [pointing at a blue color sample] or like this [pointing at a green color sample]?' The experimenter also asked, 'How do you see the dog from your side over there? Do you see it like this [pointing at a blue color sample] or like this [pointing at a green color sample]?' Only the children who were at least 4.5 years old correctly responded above chance levels.

¹⁸ Corvids prefer to cache their food far away from onlookers. Interestingly, though, they tend not to use distance as a protective strategy when they are not visible (Dally et al., 2006). Corvids may well exploit distance in this nuanced fashion because they understand that objects are *harder to see* when they are farther away. Alternatively, they may have just learned a simple rule: *Cache far away from onlookers*. Only further experiments of the sort outlined above will enable us to determine whether corvids really do understand that some objects are harder to see than others.

The evidence therefore suggests that 3 year olds possess a notion of seeing-as: they just can't deploy it in those contexts that require them to contrast their own perspective with that of others (see also Masangkay et al., 1974, Karg et al., 2014, and Krachun & Lurz, 2016).

8.2 Chimpanzees' understanding of seeing-as

A handful of studies have tested chimpanzees' understanding of the distinction between visual appearance and reality (as far as I'm aware, no other species have been tested). Krachun et al. (2016) used distorting lenses, mirrors, and tinted filters to alter the apparent size, number, and color of food items. They found that chimpanzees could maximize food rewards by making a choice based on the real, as opposed to apparent, properties of stimuli. For instance, in the distorting lens experiment, subjects correctly chose the grape that *looked* smaller, even though it was in fact larger than the other grapes on offer (see also Karg et al., 2014).

There is thus some evidence that chimpanzees understand the difference between the way things look and the way they are. But is there any evidence that chimpanzees can use this knowledge to attribute false perceptions *to others*? In the only study to date, Karg and colleagues (2016) obtained negative results. This is not enough to conclude that chimpanzees cannot attribute false perceptions to others—obviously, more studies are required. I won't elaborate on which experimental strategies are the most promising (for an extended discussion see Lurz, 2011). Instead, I want to address an important issue concerning the evolution of Level 2 perspective taking.

8.3 The evolution of Level 2 perspective taking

Suppose a dominant chimpanzee mistakenly sees a ripe banana as green and leaf-shaped (the same color and shape as surrounding leaves). If a nearby subordinate understands this fact about the dominant's visual experience, he could use this to predict that he will

leave the banana alone. But notice that if the subordinate understands that the camouflaged banana is much *harder* to see than other bananas in the vicinity, this would also be enough for him to predict the dominant's behavior. In other words, to predict that a competitor will not retrieve a camouflaged object, it suffices to grasp the stage lights conception of seeing.

This point has important ramifications. First, it means that in testing for an understanding of seeing-as, we need to avoid scenarios involving camouflage: otherwise, positive results will be neutral between the hypothesis that the subject grasps seeing-as, and the hypothesis that she merely grasps the stage lights conception of seeing. Second, it means that we need to be careful in assessing hypotheses as to how and why Level 2 perspective taking evolved.

Take Lurz's (2011) suggestion that Level 2 perspective taking evolved because it enabled creatures to anticipate the effects that *camouflage* has on the actions of others. There are good reasons for focusing on camouflage: in contrast to distorting lenses, mirrors, and tinted filters, the habitats of ancestral primates would have been littered with camouflaged objects (e.g. camouflaged insects, fruits, and predators). Focusing on camouflaged objects therefore gives Lurz's hypothesis ecological validity. Importantly, though, it also brings our alternative hypothesis into play: namely, that it was the need to read minds in the presence of hard-to-see-objects that gave rise to the stage lights conception of seeing.

What this shows is that in constructing and testing hypotheses about the evolution of Level 2 perspective taking, we need to address the following questions: (1) Which types of illusions were most prevalent in the ancestral habitats of a given species: were they camouflage-based illusions or some other variety? (2) Which types of illusions would it have been adaptive to know about for the purposes of predicting the behavior of others? (3) Which creatures can predict the behavior of others in illusory settings involving

camouflage, and which creatures can also predict the actions of others in illusory settings that do not involve camouflage? Unless we address these sorts of questions, we are in danger of running two distinct forms of perspective taking together.¹⁹

9 CONCLUSION

I have argued that there are at least three conceptions of seeing that a creature might possess. When we examine their discriminative capacities, we find evidence that infants and various nonhumans possess the headlamp conception of seeing; and that 3-year-old children, and possibly chimpanzees, understand seeing-as. However, there is no evidence that either infants or nonhumans possess the stage lights conception of seeing.

Importantly, this is not because they have failed relevant tests: it is because appropriate tests have not been carried out. I then suggested some experimental strategies.

There is much to gain from these sorts of experiments. Applying them to infants will provide important insights into the relationship between word learning and a developing grasp of selective attention. These experiments will also shed light on the transition from Level 1 to Level 2 perspective taking. It may be that an understanding of selective attention is an important stepping stone in the child's ability to grasp the constructive nature of the mind. For example, understanding that someone can see two objects, but see one of them more easily than the other, might be an important step towards understanding that different people can see the same object under different aspects.

Determining which species possess which conception of seeing also promises to shed more light on how and why perspective taking evolved. An understanding of selective visual attention may have been an important stepping stone in the evolution of Level 2

¹⁹ Level 2 perspective taking is not limited to an understanding of illusions. If I know that, given our respective vantage points, you see the turtle as upside down, whereas I see it as upright, that counts as Level 2 perspective taking (Masangkay et al., 1974). This means that, initially, Level 2 perspective taking could have been adaptive for reasons other than the prediction of behavior in illusory settings.

perspective taking. Alternatively, an understanding of selective attention may have evolved as a byproduct of our ability to determine speaker reference, well after the branch containing modern humans split off from the branch containing chimpanzees and bonobos. Only comparative research of the sort outlined above will give us traction on these issues.

ACKNOWLEDGEMENTS

I would like to thank Cameron Buckner, Carla Krachun, and Jessica Grady for very helpful comments on the paper. I am also very grateful for the extensive comments I received from four anonymous reviewers for *Mind & Language*. Thanks to Laura Larocca for helping with the illustrations.

REFERENCES

- Allen, C. (1999). Animal concepts revisited: The use of self-monitoring as an empirical Approach. *Erkenntnis*, 51(1), 33–40.
- Andrews, K. (2012). *Do apes read minds? Toward a new folk psychology*. MIT Press.
- Andrews, K. (2017). Chimpanzee mindreading: Don't stop believing. *Philosophy Compass*, 12(1).
- Baillargeon, R., Scott, R. & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Science*, 14(3), 110–118.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA.: MIT Press.
- Bray, J., Krupenye, C. & Hare, B. (2014). Ring-tailed lemurs (*Lemur catta*) exploit information about what others can see but not what they can hear. *Animal Cognition*, 17(3), 735–744.
- Buckner, C. (2014). The semantic problem(s) with research on animal mind-reading. *Mind & Language*, 29(5), 566–589.
- Bugnyar, T. (2011). Knower-guesser differentiation in ravens: others' viewpoints matter. *Pros. R. Soc. B.*, 278, 634–640.
- Bugnyar, T., Reber, S. & Buckner, C. (2016). Ravens attribute visual access to unseen competitors. *Nature Communications*, 7(10506), DOI: 10.1038/ncomms10506
- Butterfill, S. & Apperly, I. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28(5), 606–637.
- Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Science*, 12(5), 187–192.
- Carey, S. (2010). Beyond fast mapping. *Lang Learn Dev*, 6(3), 184–205.
- Coss, R. G. (1978). Perceptual determinants of gaze aversion by the lesser mouse lemur (*Microcerbus murinus*): The role of two facing eyes. *Behaviour*, 64, 248–267.
- Dally, J.M., Emery, N.J. & Clayton, N.S. (2006). Food-caching western scrub-jays keep track of who was watching when. *Science*, 312, 1662–1665.
- Emery, N. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, 24, 581–604.

- Fabricius, W., Schick, K., Prost, J. & Schwanenflugel, P. (1997). We don't see eye to eye: Development of a constructivist theory of mind. Poster presented at the biennial meeting of the Society for Research in Child Development, Washington, D.C., April.
- Flavell, J. H. (1977). The development of knowledge about visual perception. *The Nebraska Symposium on Motivation* (Vol. 25, pp. 43–76). Lincoln, NE.: University of Nebraska Press.
- Flavell, J. H. (2004). Development of knowledge about vision. In D.T. Levin (ed.), *Thinking and Seeing: Visual Metacognition in Adults and Children* (pp. 13–36). Cambridge, MA.: MIT Press.
- Flavell, J. H., Green, F. L. & Flavell, E. R. (1995). The development of children's knowledge about attentional focus. *Developmental Psychology*, 31, 706–712.
- Flombaum, J. I. & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology*, 15(5), 447–452.
- Fodor, J. (1990). *A Theory of content and other essays*, Cambridge, MA: MIT Press, Bradford Book.
- Godfrey-Smith, P. (1994). A modern history theory of functions. *Noûs*, 28, 344–362.
- Godfrey-Smith, P. (2005). Folk psychology as a model. *Philosophers' Imprint* 5(6), 1–16.
- Goossens, B. M. A., Dekleva, M., Reader, S. M., Sterck, E. H. M. & Bolhuis, J. J. (2008). Gaze following in monkeys is modulated by observed facial expressions. *Anim. Behav.*, 75, 1673–1681.
- Hare, B., Call, J., Agnetta, B. & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59, 771–785.
- Karg, K., Schmelz, M., Call, J. & Tomasello, M. (2014). All great ape species (*Gorilla gorilla*, *Pan paniscus*, *Pan Troglodytes*, *Pongo abelii*) and two-and-a-half-year-old children (*Homo sapiens*) discriminate appearance from reality. *Journal of Comparative Psychology*, 128(4), 431–439.
- Karg, K., Schmelz, M., Call, J. & Tomasello, M. (2016). Differing views: Can chimpanzees do level 2 perspective-taking? *Animal Cognition*, 19, 555–564.
- Krachun, C., Carpenter, M., Call, J. & Tomasello, M. (2010). A new change-of-contents false belief test: Children and chimpanzees compared. *International Journal of Comparative Psychology*, 23(2), 145–165.
- Krachun, C. & Lurz, R. (2016). I know you see it wrong! Children use others' false perceptions to predict their behaviors. *Journal of Experimental Child Psychology*, 150, 380–395.
- Krachun, C., Lurz, R., Russell, J. & Hopkins, W. (2016). Smoke and mirrors: Testing the scope of chimpanzees' appearance-reality understanding. *Cognition*, 150, 53–67.
- Kravitz, D. J. & Behrmann, M. (2011). Space-, object-, and feature-based attention interact to organize visual scenes. *Atten Percept Psychophys*, 73(8), 2434–2447.
- Liebal, K., Carpenter, M. & Tomasello, M. (2010). Infants' use of shared experience in declarative pointing. *Infancy*, 15(5), 545–556.
- Liszkowski, U., Carpenter, M., Henning, A., Striano, T. & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science*, 7, 297–307.
- Liszkowski, U., Carpenter, M. & Tomasello, M. (2007). Reference and attitude in infant pointing. *J. Child. Lang.*, 34, 1–20.
- Lurz, R. (2011). *Mindreading animals: The debate over what animals know about other minds*. Cambridge, MA.: MIT Press.
- Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E. & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child Development*, 45, 357–366.
- Miller, P. H. & Bigi, L. (1977). Children's understanding of how stimulus dimensions affect performance. *Child Development*, 48, 1712–1715.

- Miller, P. H. & Bigi, L. (1979). Development of children's understanding of attention. *Merrill-Palmer Quarterly*, 25, 235–250.
- Millikan, R. (1984). *Language, thought and other biological categories*. Cambridge, MA.: MIT Press.
- Millikan, R. (1993). *White queen psychology and other essays for Alice*. Cambridge, MA.: MIT Press.
- Millikan, R. (2000). *On clear and confused ideas: An essay about substance concepts*. Cambridge: Cambridge University Press.
- Mineka, S., Davidson, M., Cook, M. & Keir, R. (1984). Observational conditioning of snake fear in rhesus monkeys. *Journal of Abnormal Psychology*, 93(4), 355–372.
- Moll, H., Koring, C., Carpenter, M. & Tomasello, M. (2006). Infants determine others' focus of attention by pragmatics and exclusion. *Journal of Cognition and Development*, 7(3), 411–430.
- Moll, H. & Meltzoff, A. (2013). Taking versus confronting visual perspectives in preschool children. *Developmental Psychology*, 49(4), 646–654.
- Moll, H., Richter, N., Carpenter, M. & Tomasello, M. (2008). Fourteen-month-olds know what “we” have shared in a special way. *Infancy*, 13, 90–101.
- Moses, L. J., Baldwin, D. A., Rosicky, J. G. & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, 72, 718–735.
- Parault, S. J. & Schwanenflugel, P. J. (2000). The development of conceptual categories of attention during the elementary school years. *Journal of Experimental Child Psychology*, 75, 245–262.
- Phillips, A., Wellman, H. & Spelke, E. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, 85, 53–78.
- Phillips, B. (2016). Contextualism about object-seeing. *Philosophical Studies*, 173(9), 2377–2396.
- Phillips, B. (forthcoming). The shifting border between perception and cognition. *Noûs*.
<https://doi.org/10.1111/nous.12218>
- Pietroski, P. (1992). Intentional and teleological error. *Pacific Philosophical Quarterly*, 73, 267–281.
- Pillow, B. H. (1988). Young children's understanding of attentional limits. *Child Development*, 59, 38–46.
- Pillow, B. H. (1989). The development of beliefs about selective attention. *Merrill-Palmer Quarterly*, 35, 421–443.
- Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Earlbaum.
- Povinelli, D. J. & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind & Language*, 19, 1–28.
- Rosati, A. G. & Hare, B. (2009). Looking past the model species: diversity in gaze-following skills across primates. *Curr. Opin. Neurobiol.*, 19, 45–51.
- Ruiz, A., Gómez, J. C., Roeder, J. J. & Byrne, R. (2009). Gaze following and gaze priming in lemurs. *Animal Cognition*, 12, 427–434.
- Sandel, A., Maclean, E. & Hare, B. (2011). Evidence from four lemur species that ringtailed lemur social cognition converges with that of haplorrhine primates. *Animal Behavior*, 81, 925–931.
- Schmelz, M. & Call, J. (2016). The psychology of primate cooperation and competition: a call for realigning research agendas. *Phil. Trans. R. Soc. B*, 371, 20150067.
- Shea, N. (2013). Naturalising representational content. *Philosophy Compass*, 8(5), 496–509.
- Sterelney, K. (1990). *The representational theory of mind: An introduction*. Cambridge, MA.: Blackwell.
- Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA.: Harvard University Press.
- Topal J. & Csanyi V. (1994). The effect of eye-like schema on shuttling activity of wild house mice (*Mus musculus domesticus*): Context-dependent threatening aspects of the eyespot patterns. *Anim. Learn. Behav.*, 22, 96–102.
- Wellman, H. (2014). *Making minds: How theory of mind develops*. Oxford: Oxford University Press.
- Wright, R. D. & L. M. Ward. (2008). *Orientating of attention*. Oxford: Oxford University Press.