The Primate Mindreading Controversy: A Case Study in Simplicity and Methodology in Animal Psychology

Simon Fitzpatrick

Department of Philosophy

University of Sheffield

s.fitzpatrick@sheffield.ac.uk

1.   Introduction

What, if anything, do non-human primates (henceforth, "primates") understand about the minds of other agents? Can they mentally represent and reason about the mental states of others? If they can, what kinds of mental states can they represent (e.g. perceptions, goals, intentions, beliefs) and what kinds of reasoning about mental states are they capable of? These are the central questions in the field of primate "mindreading," or "theory of mind." However, some 30 years after Premack and Woodruff (1978) posed such questions, there remains very little consensus on how we should answer them.

Much of the recent debate has centred on an ongoing controversy over whether primates are capable of reasoning about basic aspects of the visual perspective and perceptual awareness of others. Several researchers claim that recent behavioral experiments provide strong evidence for such a mindreading capacity in several primate species. Other prominent researchers, however, vigorously deny that these studies provide any evidence at all for mindreading.

One issue that has played a prominent role in this controversy and throughout the history of the debate over primate mindreading concerns the relative "simplicity" or "parsimony" (these terms are typically used interchangeably) of mindreading and non-mindreading explanations of behavior. In interpreting the available data, both proponents and skeptics about primate mindreading have argued that their chosen explanation is "simpler" or "more parsimonious" than the alternatives, and hence should be preferred. In both instances, considerations of simplicity have been invoked to bolster the case for a particular hypothesis in the face of seemingly equivocal data.

My aim in this chapter is to look at the role that such appeals to simplicity have played in this controversy as a case study for thinking about the proper place of simplicity considerations in inferring cognitive processes in non-human animals, and

in science more generally (see also Sober, this volume). After providing some background to the controversy and describing the appeals to simplicity that have been made, I will pose some problems for such appeals, which call into question the appropriateness of bringing in simplicity considerations when evaluating behavioral data. I will then outline a general philosophical account of simplicity, which, I will argue, makes the best sense of what is going on in the recent controversy over primate mindreading. In doing this, I hope to shed new light on the nature of the evidence for primate mindreading, and on how future work might be able to resolve this controversy.

2.   Perspective-taking in primates?

2.1   Some background

Mindreading is a form of *higher-order* cognition, involving reasoning processes and mental states whose contents concern other agent's mental states. In humans at least, such reasoning plays a key role in our ability to predict and explain the behavior of other agents. A central area of research for comparative psychologists looking for possible mindreading capacities in primates and other non-human animals concerns *visual perspective-taking*, in particular, reasoning about what others can and cannot see and how others are likely to behave in a given situation based on what they perceive. Primates often behave in ways that seem to suggest such reasoning: for example, primates will often hide contested food items from others' view (Byrne and Whiten, 1988) and follow the gaze of others to locate the object of attention (Tomasello and Call, 1997).

However, the problem is that such observations also seem to be consistent with an alternative hypothesis: instead of being *mind*readers primates could just be *behavior*-readers. In contrast to mindreading, behavior-reading involves purely first-order reasoning about non-mental phenomena, such as the regularities in others' behavior. For example, from past experience a chimpanzee might learn that a competitor will be less likely to take a food item if she places it behind certain kinds of objects. Similarly, she might learn that certain movements of a conspecific's eyes toward a location are correlated with the presence of an interesting object at that location. The key aim of experimentalists has thus been to develop behavioral tests for mindreading that can rule out purely first-order explanations for success in the

relevant task. This project has, however, proved to be extremely difficult (Heyes, 1998).

Moreover, in the late 1990s there emerged experimental results that seemed to provide evidence *against* perspective-taking abilities in primates. Famously, Povinelli and Eddy (1996) presented chimpanzees with a choice of begging for food from one of two experimenters: one who could see them and another who could not (e.g. because she was blindfolded or had a bucket covering her head). Povinelli and Eddy found that even after repeated trails subjects begged just as much from unseeing as from seeing experimenters, suggesting that chimpanzees do not reason about what others can and cannot see.

However, in recent years these results have been challenged on methodological grounds. Since the social life of chimpanzees is dominated by *competition* for food rather than *communicative-co-operative* interactions of the sort involved in Povinelli and Eddy's study, it has been argued that subjects may have failed to perform well in this and other similar tasks, not because they lack the ability to reason about others' perceptions, but because they did not properly understand the task (Hare et al., 2000). This concern provided the impetus for Hare, Call and Tomasello to develop a more naturalistic approach to investigating chimpanzees' understanding of visual perspective.

2.2   Food competition

The starting assumption for Hare, Call and Tomasello was that if primates possess mindreading capacities, these would be more likely to show themselves in competitive situations that closely resemble their natural social interactions. In the first of a series of studies, Hare et al. (2000) placed a subordinate and dominant chimpanzee in cages on opposite sides of a middle room that contained two pieces of food: one visible to both chimpanzees, the other placed behind an opaque barrier and only visible to the subordinate. They were initially prevented from entering the middle room by guillotine doors, which, when slightly raised, allowed them to see each other and the layout of the room. They were then released, with the subordinate given a slight head start. Dominant chimpanzees tend to take all the food that is available to them and punish subordinates who challenge them. So the prediction was that if subordinates are capable of reasoning about the visual perspective of others,

they should preferentially target the piece of food behind the barrier, which the dominant cannot see. This is exactly what happened.

Controls were introduced for various behavior-reading hypotheses. Subordinates were given a head start to help rule out the hypothesis that they merely chose the food that the dominant *did not* go for. Also in some conditions the dominant's door was completely closed until the subordinate made a choice to go for one of the pieces. Subordinates still preferred the hidden food, so could not have merely reacted to the dominant's behavior upon being released. In another condition the opaque barrier was replaced with a *transparent* one. Here subordinates did not show a preference, thus seeming to be aware that the dominant had visual access to both pieces of food and ruling out the hypothesis that subordinates merely prefer food behind barriers.

In a follow-up study (Hare et al., 2001) the middle room contained two opaque barriers. A single piece of food was placed on the subordinate's side of one of the barriers, with the experimenters varying whether the subordinate and the dominant or just the subordinate saw this. In a control condition a dominant that witnessed the food being placed was switched with a dominant that had not, just before the subordinate was released. The result was that subordinates tended to go for the food more when the competitor had *not* witnessed the placement of the food.

Hare et al. have thus concluded that chimpanzees *are* in fact sensitive to what others can and cannot see and also to at least some aspects of what others *know* about a situation based on what they have and have not seen in the recent past.

Subsequent work has expanded on these initial studies in various ways. Flombaum and Santos (2005) have produced similar results with rhesus monkeys, suggesting that it is not just apes that have such capacities. Melis et al. (2006) gave chimpanzees the opportunity to steal food from an experimenter who was inside a booth covered by opaque glass on three sides, apart from a slit at the front. On the left and right hand sides of the booth were holes connected to tunnels through which subjects could reach into the booth. One tunnel was transparent, the other opaque. During the test condition the experimenter placed food next to the ends of both tunnels and then stared directly ahead. In a control condition the experimenter placed food next to the tunnels and walked away. From the first trial subjects had a preference for grabbing food through the opaque tunnel in the test condition but *not* in the control condition. Melis et al. conclude that this shows that chimpanzees engage

in quite subtle mentalistic reasoning: even though the experimenter has visual access to both pieces of food, chimpanzees use the opaque tunnel to conceal their *reach* to the food, giving the experimenter less chance to take it away.

2.3   Povinelli and Vonk's critique

While Hare et al. and colleagues suspect that primates lack more sophisticated human mindreading capacities, such as the ability to reason about false beliefs (Call and Tomasello, 2008), these new studies are claimed to show that visual perspective-taking is at least one important mindreading capacity that is *not* uniquely human. A number of theorists, however, are extremely skeptical about these claims. Most strikingly, Povinelli and Vonk (2006) have argued that these studies are incapable *in principle* of providing any evidence at all for mindreading.

The problem, according to Povinelli and Vonk, is that since mental states are not directly observable, subordinates can only infer particular mental states in dominants based on features of the situation that they *can* observe. For example, in the Hare et al. (2000) study the subordinate can only infer that the dominant can or cannot see the food based on whether or not an opaque barrier stands in between the dominant's eyes and the food. Povinelli and Vonk claim that there is therefore a purely first-order explanation for the results: subjects make a direct inference from these observable cues to a behavioral prediction *without* any reasoning about others' mental states. Subordinates could, for instance, reason according to a behavioral rule (acquired by learning or known innately) which says that dominants are less likely to go for a particular food item if an opaque barrier stands in between their eyes and the food. Similarly, in the Hare et al. (2001) study subordinates may reason that dominants are less likely to go for food behind an opaque barrier if they were not present when the food was placed at that location.

Since these experiments do not control for this possibility they are, according to Povinelli and Vonk, completely *irrelevant* to the question of whether primates understand anything about the minds of others. If we are to have a genuine behavioral test for perspective-taking a very different kind of experimental paradigm will have to

be developed—one that does not give subjects any observable cues that could be used to predict behavior.[1]

2.4   Which is simpler?

For the most part, Hare et al. have been prepared to concede that Povinelli and Vonk's alternative explanation for their results is not directly ruled out by their control conditions. Nonetheless, they insist that they still provide strong evidence for visual perspective-taking and that Povinelli and Vonk's alternative explanation is extremely *ad hoc*. But why do they claim this? It is here that considerations of simplicity/parsimony have been seen to enter the fray. For example, Tomasello and Call argue that:

> The results of each experiment may be explained by postulating some behavioral rule that individuals have learned that does not involve an understanding of seeing. But the postulated rule must be different in each case, and most of these do not explain more than one experiment. This patchiness of coverage gives this kind of explanation a very *ad hoc* feeling, especially since there is rarely any concrete evidence that animals have had the requisite experiences to learn the behavioral rule—there is just a theoretical possibility. It is thus more plausible to hypothesize that apes really do know what others do and do not see in many circumstances. (Tomasello and Call, 2006, p. 371)

Tomasello and Call do not claim to be proposing a simplicity argument here, but this is primarily how their argument has been interpreted.[2] Their first claim is that though explanations of the sort proposed by Povinelli and Vonk can in principle be offered for each of their results, since the relevant observable cues and behavioral predictions are different in each experiment, skeptics will need to posit a huge number of different rules governing subordinates' reasoning across all the cases. For instance, different rules will be required in the condition where the food is already behind an opaque barrier, the condition where the dominant observes food being placed behind an opaque barrier, and the condition where the food is behind a transparent barrier. An entirely different kind of rule will also be needed in the Melis et al. (2006) experiment, since here the competitor has a direct line of sight towards both pieces of

---

[1] This conclusion is meant to apply to *any* study where the relevant differences in others' mental states and behavior correlate with observable differences between conditions. Povinelli and Vonk do suggest alternative paradigms, which they claim avoid this problem, though it far from clear why these are not open to the same criticisms levelled at the Hare et al. studies.
[2] According to Povinelli and Vonk (2006, p. 394) Hare et al. have claimed that "although it is possible that chimpanzees form concepts solely about behavior, the case for this is unproven, and they seem to imply that parsimony should push us toward assuming that they do, in fact, represent mental states."

food; the key thing for the subject is to conceal its *reach* from the experimenter. Skeptics will also need to explain how subjects have learnt each of these specific rules (or somehow possess them innately). The claim then seems to be that we can offer a *simpler*, more unified, explanation for the results by just granting chimpanzees a psychological understanding of visual perspective that underlies their reasoning in all the various conditions.

This type of simplicity argument has had a long history in the debate over primate mindreading (see e.g. Premack and Woodruff, 1978; Whiten, 1996). The idea is that attributing higher-level mentalistic concepts (such as a concept of *visual awareness*) allows us to avoid attributing to subjects a cluster of rules that specify a one-to-one mapping between particular observable cues and predictions of how others are going to behave in specific circumstances.

Povinelli and Vonk, however, are not impressed by this kind of argument:

> [R]easoning about mental states must entail observing and reasoning about behavior (in all its subtleties) and, on the basis of such observed features, generating and reasoning about representations of unobserved mental states. Thus the capacity to reason about mental states does not somehow relieve the burden of representing the massive nuances of behavior or the statistical invariances that sort them into more and less related groups. In either event, these behavioral abstractions must be represented… [Thus] there is no sense in which a system that makes inferences about behavioral concepts alone provides a *less* parsimonious account of behavior than a system that must make all of those same inferences *plus* generate inferences about mental states. (Povinelli and Vonk, 2006, p. 393-394 [emphasis in original])

Since mindreading must involve reasoning about behavior and other observable features of the situation—due to the fact that mental states must be inferred indirectly—the only difference between mindreading and behavior-reading is that the latter process skips the middle step of inferring and representing others' mental states before generating a behavioral prediction. Consequently, a mindreading explanation must in fact be *less* parsimonious than a behavior-reading explanation.

Povinelli and Vonk do not explicitly claim that this greater parsimony gives us a reason to prefer a behavior-reading explanation of Hare et al.'s results. However, they are clearly motivated by parsimony considerations since they repeatedly claim that it is "sufficient" to explain the current data by attributing chimpanzees a capacity for behavior-reading alone. Other skeptics are more explicit: after replicating the results of Hare et al. (2000) with marmosets (a species of new world monkey),

Burkart and Heschl (2007, p. 468) claim that "at least at the present state of evidence, the more parsimonious explanation seems to be that they deal with a directed gaze without understanding visual perspective." This is in spite of the fact that their results are perfectly consistent with perspective-taking.

How should this dispute be adjudicated? Should we side with proponents like Tomasello and Call, with skeptics like Povinelli and Vonk, or with neither of them? My aim in the rest of the chapter is to focus on the logic of these conflicting arguments and go some way towards answering these questions. My conclusion will be that Tomasello and Call's argument against Povinelli and Vonk does have a high degree of plausibility to it and that Povinelli and Vonk's criticisms are misguided. First, however, I want to pose some general problems for appeals to simplicity in this kind of context. Then in section 4 I will outline a philosophical account of simplicity that provides us with a productive framework for interpreting and evaluating these arguments.

3.   Substance or procrastination?

As we have seen, the controversy over the new food competition experiments has in part taken the form of a controversy over what is the "simplest"/"most parsimonious" explanation of subjects' behavior. In the absence of further data that might prove more decisive, both sides, it seems, explicitly or implicitly take simplicity considerations to provide an additional criterion that can be used to choose between competing hypotheses. However, they seem to have very different ideas of how such a criterion should be applied in this context.

Though the appropriateness of these kinds of simplicity arguments is normally just taken for granted, there are a number of serious worries that can be raised about the appeals to simplicity that have been made in this debate. To begin with, the notions of simplicity/parsimony involved here seem to be extremely slippery. If we look closely, the dispute between Tomasello and Call and Povinelli and Vonk highlights the fact that there are several *different* and *conflicting* respects in which hypotheses about the cognitive basis of primate social behavior could be said to be "simple" or "complex."

In their response to Tomasello and Call, Povinelli and Vonk view themselves as rebutting the claim that a mindreading explanation is more parsimonious than a

behavior-reading explanation: a mindreading explanation cannot be more parsimonious because mindreading *entails* representing and reasoning about observables, but behavior-reading does not entail representing and reasoning about mental states. However, Tomasello and Call's claim is not concerned with whether or not subjects represent and reason about observable features, but rather the psychological *unity* of the explanation being offered. On their view we can unify the subjects' behavior in the different conditions by attributing to them a general concept of *visual awareness* that underlies their responses in all the conditions. Even though, as Povinelli and Vonk point out, applying this concept must involve representing and reasoning about observable features this does not undermine the claim that a behavior-reading account is less psychologically unified than a mindreading account, in the sense that no comparably general concept is activated in all the different conditions; instead there has to be a cluster of specific rules linking particular observations with predictions of others' behavior.

If this is right, it seems that we have two *different* kinds of simplicity here—*simplicity as psychological unity*, which seems to favor a mindreading account, and *simplicity as parsimony of mental representation*, which seems to favor a behavior-reading account. In fact, I suggest that a number of other different kinds of simplicity could equally be taken into consideration here; these also come to contrasting verdicts. For example:

- *Simplicity as (less) cognitive sophistication.* A long tradition in comparative psychology holds that the simplest explanation of animal behavior is the one that attributes to the animal the "lowest" or least sophisticated cognitive processes consistent with the available data.[3] Behavior-reading is usually regarded as a less sophisticated (and hence simpler) cognitive process than mindreading since it involves purely first-order order reasoning and representation, while mindreading involves higher-order reasoning and representation. This is another, slightly different, kind of simplicity that seems to favor a behavior-reading account.
- *Simplicity as analogy.* Another common notion of simplicity in the literature is that the simplest explanation for similar behavior in different species is one that cites the operation of similar cognitive mechanisms. Thus given that we would attribute perspective-taking abilities to a human that behaved in the

---

[3] A common symptom of this view is enthusiasm for a methodological principle known as *Morgan's Canon*, which has been widely seen as a simplicity principle: "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale" (Morgan, 1894, p53). See Fitzpatrick (2008), Sober (this volume).

same way as the Hare et al. subjects, it would seem that the simplest explanation of the similarity is that chimpanzees also have such capacities.

- *Simplicity as evolutionary parsimony.* de Waal (1991; see also Sober, 2000) has argued that when closely related species such as chimpanzees and humans are found to behave in similar ways it is simpler, evolutionarily speaking, to posit *homologous* cognitive mechanisms in order to account for this similarity—shared mechanisms inherited from a common ancestor. The alternative is to posit different cognitive mechanisms that evolved independently of each other but nonetheless produce similar behavior. In this case de Waal would claim that the simplest explanation for the results is that a mechanism for visual perspective-taking was present in the common ancestor of humans and chimpanzees and is retained by both species.

Here we have several different and conflicting respects in which mindreading and behavior-reading explanations can each be seen as "simple" or "complex." This list could be extended indefinitely. For example there are many different ways in which we could assess the "amount of stuff" that psychological explanations have to posit: we could count the number of cognitive processes involved, the number of iterations of each cognitive process involved, the number of rules or principles involved in any given instance of reasoning, the number of steps involved in any reasoning process, etc. etc.

This poses a very significant problem for an attempt to make sense of which direction simplicity considerations should take us in this context. It is not enough just to say that we should prefer "simpler" explanations of behavior—*which* aspects of the respective explanations are we to be concerned with when attempting to assess their relative simplicity? From discussions in the literature it is not at all clear why we should measure simplicity in one of these ways rather than any of the others. Nor is it clear how different kinds of simplicity should be balanced against each other—in principle there are many different ways of trading-off a given amount of simplicity of kind X against a given amount of simplicity of kind Y with respect to the "overall" simplicity of the rival explanations. What we need, it seems, is a principled account of why we should value *any* of these putative simplicity properties and of what weight, if any, they should carry relative to each other.

This naturally focuses attention on another worry about these arguments. Both sides seem to assume that simplicity carries some degree of *epistemic* weight here: if one explanation is simpler than another this makes it more plausible, or somehow better supported by the data. But what *justification* is there for this? After all, there

doesn't seem to be any *a priori* reason to think that simpler explanations are more likely to be true.

Similar sorts of worries about appeals to simplicity have been raised earlier in the literature. Heyes (1998) describes the various simplicity arguments that have been proposed in the primate mindreading debate as empty "procrastination." Here she echoes a general philosophical skepticism about simplicity: though many scientists and philosophers have advocated principles of simplicity in theory choice, we do not have adequate answers to general versions of the worries just raised: how should the simplicity of theories be measured and what justification is there for preferring simple theories to less simple ones? Heyes takes the view that if the experimental data is open to both mindreading and behavior-reading interpretations, we should just be agnostic as to which hypothesis is correct until decisive evidence is found: "To answer Premack and Woodruff's question, we need more strong experiments, not more weak arguments" (Heyes, 1998, p. 112).

Prima facie Heyes' skepticism with respect to simplicity does seem to have much to recommend it. It seems difficult to avoid the concern that all this talk of simplicity is just arcane bean counting and of dubious relevance to the fundamental questions that these researchers are interested in—are primates capable of visual perspective-taking and do the food competition studies really provide evidence for such a capacity?

However, I want to resist dismissing all of the simplicity arguments that have featured in the primate literature as mere "procrastination." Tomasello and Call's argument in favor of a mindreading interpretation of their results *does* seem to have an important ring of plausibility to it, though the arguments of Povinelli and Vonk and Burkart and Heschl seem much less plausible. The problem then is to explain how there can be plausibility in some appeals to simplicity and not others, in a way that addresses, or somehow avoids, the worries about simplicity discussed above, and to supply a principled assessment of what bearing these considerations have on the broader debate about primate mindreading.

What I will do now is to describe my own view about the role of simplicity considerations in science, which, I will argue, can help us to resolve these issues.

4.   The deflationary account of simplicity

The general philosophical account of simplicity that I endorse is what I call the "deflationary account" (Fitzpatrick, 2006). This account represents a significant departure from the standard view that philosophers have taken on the role of simplicity in science. The standard view (among those who are not skeptics about simplicity) is that simplicity constitutes one of the central criteria that scientists do, and should, use for evaluating and choosing between rival theories: other things being equal, simple theories should be chosen over less simple ones. Simplicity is thus seen as a general *theoretical virtue*, alongside such things as empirical adequacy, consistency with established background theories, and so on.

The deflationary account departs from the standard view in that it denies that simplicity should be seen as a general theoretical virtue and criterion for theory choice *in its own right*. There is no adequate general justification for favoring simple theories over less simple ones. However, the claim is not that we are *never* justified in preferring "simple" theories to less "simple" ones. Rather the claim is that in cases where we do seem to be justified in preferring theories that are "simpler" in some particular respect, some *other* consideration is doing the real epistemic work. Typically, what is doing the real work are various background theoretical considerations, often specific to the scientific context at hand. So for instance the justification for preferring theory T1 to T2 comes not from the fact that T1 is more "parsimonious," but because the specific additional entities that T2 posits (and T1 does not) are theoretically *implausible*—e.g. they contradict, or pose problems for accepted background theories in that domain. Thus on the deflationary account, we may be warranted in preferring theories that are "simpler" (in some particular respect) in some contexts but *not* in others. Moreover, the reason why we are warranted in one context may be very different from why we are warranted in another context. This is because different kinds of background consideration do the epistemic work in the different cases.

Broadly deflationary views of simplicity have been defended by philosophers such as Richard Boyd (1990), John Norton (2003) and, most notably, Elliott Sober (1990; this volume). Such a view, I think, has very significant advantages over more traditional accounts of simplicity. In particular it makes the problem of justification tractable, reducing it to the problem of explaining what are the other considerations that do the real work in specific cases where certain kinds of simplicity considerations seem to be epistemically motivating. This is a much less onerous problem than

explaining how a completely general preference for "simplicity" can be justified across all scientific contexts. Moreover, the deflationary account provides a more accurate understanding of scientific practice. When reflecting on their methodology scientists often *misdescribe* highly context-specific preferences for theories with very specific kinds of properties as a general preference for "simple" theories. Thus the deflationary account provides us with a better understanding of the actual underlying reasons for preferring putatively "simpler" theories in real cases in science (see Sober, 1990; Fitzpatrick, 2006 for examples).

I will now argue that the deflationary account provides substantial illumination on the debate about simplicity and primate mindreading.[4] In so doing I will also illustrate the attractions of the deflationary account.

5.   Tomasello and Call's argument revisited

I noted earlier that Tomasello and Call do not claim that their response to Povinelli and Vonk is a simplicity argument—indeed they actually express similar skepticism to Heyes about such arguments (2006, p. 380-381). However, it is no surprise that their argument has been understood in this way. They seem to be claiming that behavior-reading explanations just become too complicated when applied to a diverse range of situations in which subjects correctly anticipate others' behavior. In this respect the argument has been compared with a common simplicity argument against behaviorism: when patterns of behavior become very elaborate it is often simpler to ascribe sophisticated cognitive processes to organisms rather than having to postulate an enormous web of learnt associations between individual stimuli and responses.

Now, I think it is easy to make sense of this anti-behaviorist argument from a deflationary point of view. It is really a sort of *poverty of the stimulus* argument: as patterns of behavior become more elaborate—i.e. subjects respond adaptively to a wider range of situations, many of which are novel for them—associative learning explanations become increasingly implausible, not because the web of stimulus-response associations becomes too "complex", but because it is difficult to explain

---

[4] Sober (1998, 2000) has also discussed issues surrounding simplicity and primate mindreading. My analysis of Tomasello and Call's argument in section 5 is different, but complimentary to Sober's (1998) discussion of a similar argument from Whiten (1996). See also Sober (this volume) for application of a model selection framework to this debate.

how subjects could have *acquired* the requisite associations.[5] The more subtle and specific these associations have to be and the more of them that have to be posited, the less tenable it becomes to claim that subjects will have reliably encountered all the environmental stimuli a behaviorist account would require in order for them to be acquired. Thus more sophisticated cognitive machinery is required to adequately explain subjects' behavior.

Tomasello and Call's argument is best understood in a similar fashion:[6] we cannot just *assume* that the Hare et al. subjects will have had experiences sufficient to acquire generalizations about others' behavior that would enable them to respond appropriately to the situations they are confronted with. Consider the subordinate chimpanzee that has to decide whether or not to go for food behind a transparent barrier. Suppose, as Hare et al. (2000, p. 783) claim, that this situation is highly novel for subjects, so they "could not have had many opportunities to learn specific contingencies between [transparent] objects and the behavior of their groupmates." As I see it, Tomasello and Call's claim is that it is difficult to see how a behavior-reader who has no understanding of visual perception could predict that the relevant observable cue in this situation—the transparency of barrier—is going to make a difference to the dominant's behavior without generalizing from past experience in similar situations. A mindreader, however, who possesses a concept of *visual awareness* might plausibly do without such prior experience. All it needs to do is reason about whether or not the dominant can *see* the food, and the transparency or opacity of the barrier is likely going to be a salient consideration for an agent that possesses such a higher-level concept. Plausibly then, a mindreader could potentially respond appropriately in this situation even if it has never previously interacted with other agents in situations involving transparent barriers, or perhaps never seen a transparent barrier. But there is no reason to think that a behavior-reader could do this since it needs to have some basis for associating this observable cue with a behavioral outcome.

---

[5] Poverty of the stimulus arguments claim that the environmental input to a given set of cognitive mechanisms possessed by an organism underdetermines the demonstrated behavioral capacities of the organism. Thus *additional* cognitive structure—additional knowledge, concepts or cognitive mechanisms—must be posited to adequately explain these capacities. Such arguments are most well known from Chomsky's work on human language acquisition (see Laurence and Margolis, 2001), but they have also played a significant role in arguments for sophisticated cognitive mechanisms in animals.

[6] This is not to suggest that Povinelli and Vonk are behaviorists (they are not), only that there are commonalities in the structure of the arguments.

Mindreading is thus assumed to be more of a *generative* capacity than behavior-reading—allowing subjects to adapt to novel situations in a way that goes beyond the contingencies of their previous experience. Povinelli and Vonk object to this generative view of mindreading on the grounds that mental states can only be attributed to others based on a *prior* understanding of the correlations between particular observable features and others' behavior (recall the passage quoted in section 2.4). Discussing Hare et al. (2000) they claim that:

> [T]he subordinate's reaction… can be explained either by the subject's possession of a concept about the statistical invariants that exist in head/eye/body orientation toward food, on the one hand, and future behavior, on the other, or all of that *plus* a representation of an unobservable mental state. (2006, p. 396 [emphasis in original)

Here mindreading is characterized as a process of inference from observed statistical associations between observable features and others' behavior, where it is the recording of these associations that leads the mindreader to infer particular mental states in others. On this view it is easy to see how reasoning about mental states can in principle be dropped from an account of how individuals make predictions about others' behavior: the only difference between mindreading and behavior-reading is the intervening (and superfluous) attribution of mental states in between recording a set of token observations and computing a prediction of behavior based on statistical associations between previous observations of the same type and subsequent behaviors of a given type.

However, this is an extremely impoverished view of how mindreading must work. On any serious theoretical model of mindreading, mindreaders do not make inferences to mental states and behavioral predictions based purely on statistical associations between observables. The key function that higher-level mentalistic concepts play in behavior prediction is that they endow subjects with *inferential abilities* that are necessarily lacking in subjects possessed of only low-level concepts about correlations between observables. These are crucial for picking up on the *significance* of novel observable cues for predicting others' behavior in situations where subjects have had *no opportunity* to observe correlations between the relevant cues and others' behavior.

Different theoretical models of mindreading provide rather different accounts of the nature of these abilities (see Nichols and Stich, 2003; Goldman, 2006). For

instance, on a "theory-theory" model, reasoning about visual awareness relies upon a tacit psychological theory about the basic mechanics of visual perception. This psychological knowledge drives subjects' mental state inferences from what they observe. It is assumed to be rich and general enough that, combined with other reasoning processes, it can enable subjects to work out the significance of features of novel social situations in the absence of previous experience of similar situations (indeed explaining such flexibility is a key a motivation for theory-theory accounts). For instance, it could include an abstract theory of occlusion or visibility which specifies in general terms the sorts of conditions in which objects are occluded or visible to other agents, and which would enable a mindreader to *work out* that the dominant can *see* the food behind the transparent barrier, even if she has never previously interacted with other agents in situations involving transparent barriers. She will then reason that the dominant will likely compete for it, since she knows that dominants normally compete for food that they can *see*.

On a "simulationist" model the story would be rather different. Mindreaders don't base their inferences on a tacit theory about visual perception; rather, they imagine themselves in the situation that the other agent is in and then *mentally simulate* their own behavioral responses "off-line" to generate a prediction of what the other agent will do. Here the subordinate would utilize her own perception of being able to *see* through the transparent barrier and her own practical reasoning about what *she* would do if she were in the dominant's position and could see the food.

There is no need here to stake a position on which theoretical model is more plausible for primate perspective-taking.[7] On both of these sorts of models, by conceptualizing the situation in mentalistic terms mindreaders possess powerful *inferential abilities*—underlied by tacit theory, or simulation—that allow them to go beyond their previous experience. (Note the importance on both models of attributing states of *seeing* to the dominant for making an inference from the novel observation— the transparency of the barrier—to a prediction of her behavior). The presence of such abilities in mindreaders, but not in behavior-readers, is what allows us to pose a poverty of the stimulus problem for a behavior-reading account: if we cannot plausibly assume appropriate experiences—e.g. interaction with others involving

---

[7] Recent models of human mindreading combine elements of tacit theory *and* simulation, plus mechanisms not easily categorized on either view (e.g. Nichols and Stich, 2003). It is likely that if primates have any mindreading capacities theirs will also be made up of a variety of mechanisms.

transparent barriers—in order for subjects to be able to correctly associate the relevant observable cue with a prediction of how the other agent will behave, then subjects' success in this kind of task would constitute evidence for some form of mindreading mechanism that supplies this extra information.

An obvious response is to claim that knowledge of the relevant correlations between specific cues and behavior could be innate rather than learned. But this just pushes the explanatory burden back a level: we need to explain how evolution could have endowed subjects with specific rules that would correctly pick up on the significance of the transparent barrier. If transparent barriers were not a typical feature of the habitats of ancestral primates then we merely replace an ontogenetic poverty of the stimulus problem with a phylogenetic one.[8]

Now, I want to emphasize that it is a further question how compelling this kind of poverty of the stimulus argument should be in support of visual perspective-taking, given the data that is currently available for chimpanzees and other primates. But before discussing this, I want to make some remarks about the worries about simplicity from section 3. This deflationary interpretation of Tomasello and Call's argument provides a much better understanding of the main simplicity argument that has been proposed in support of primate mindreading. Whiten (1996) for example puts the argument in terms of "economy of representation": reasoning about others' mental states allows agents to dispense with a body of rules that encode a one-to-one mapping between sets of observable cues and behavioral predictions. But it is not really the *amount* of mental representation that matters here. It is the fact that having higher-level mental concepts allows agents to recognize the significance of features of novel situations that cannot be picked up on by mere behavior-readers. In both the anti-behaviorist argument and the reading of Tomasello and Call's argument I am proposing, it is clear that what is doing the real epistemic work is not the "simplicity" of behaviorist versus cognitive, or behavior-reading versus mindreading explanations

---

[8] Could subjects have a purely behavioral analogue to a concept of *visual awareness* that could do the same work? I am skeptical. The problem is spelling out such a concept so that it is: i) not just a mentalistic concept in disguise; ii) sufficiently broad that it can account for the flexibility of subjects' responses to different types of stimuli (it can't be tied to a small set of very specific observable cues); iii) not so broad that it becomes massively implausible that subjects could have such a concept, as a product of learning or innately (it can't specify the behavioral significance of every perceptual subtlety that a subject might be confronted with). With respect to i) it should be remembered that when behaviorists tried to offer behavioral analyses of mental concepts they often turned out to be just as mentalistic as the concepts they were trying to replace. Indeed I suspect that there are serious problems in even making sense of apparently low-level concepts like *barrier, transparent,* or *opaque,* without implicitly relying on psychological concepts like *goal, agent* or *visibility.*

per se—e.g. the sheer number of stimulus-response connections/behavioral rules required. Rather, it is a set of background assumptions concerning the previous experiences (or innate endowment) of subjects, that cast doubt on the plausibility of the kind of story that behaviorist/behavior-reading accounts have to tell in order to account for the data. Seen in this light, the worries about simplicity raised in section 3 can be addressed reasonably straightforwardly. The justification for these arguments, depends not on a general justification for preferring simple theories to less simple ones, but rather on the justification for the relevant background assumptions. If these are independently motivated, and they do in fact discriminate between the rival explanations in the right way, then it is easy to see how these arguments could have epistemic force. Heyes' general skepticism about simplicity arguments in this context can thus be resisted.

In addition, on deflationary readings of these arguments we can avoid worries about the slipperiness of notions of simplicity/parsimony, and the possibility that there are conflicting respects in which behaviorist/behavior-reading accounts can be seen as "simpler" than cognitive/mindreading accounts. As it stands, Povinelli and Vonk's response to Tomasello and Call—that behavior-reading explanations must be more "parsimonious" because representing and reasoning about mental states entails representing and reasoning about observables, but not *vice versa*—is completely beside the point, since it is not some general notion of "simplicity"/"parsimony" that is doing the real work here. Of course it is open to skeptics to elucidate other background considerations that might ground simplicity arguments in the opposite direction. A key advantage of the deflationary account is that it reduces the problem of adjudicating between conflicting claims about "simplicity" to a problem of weighing these kinds of background considerations, where they exist. Though this can be difficult, weighing different kinds of background evidence is a standard sort of problem that scientists face. Hence on the deflationary view, we do at least seem to stand a chance of getting at the real epistemic force of these kinds of arguments, and being able to distinguish what seem to be genuinely plausible appeals to simplicity from less plausible ones.[9]

---

[9] An under-recognized aspect of the role of simplicity in science is that it is often employed as methodological cudgel to beat the opposition. For instance skeptics about animal minds have often appealed to "parsimony" merely in order to pour scorn on attributions of sophisticated cognitive capacities to animals (Fitzpatrick, 2008). An attraction of the deflationary account is that it allows us to tease apart *ad hoc* rhetorical appeals to simplicity from those that do have some epistemic force.

Now for a more difficult question: should Tomasello and Call's argument, combined with the current data, persuade us that chimpanzees (and other primates, such as rhesus monkeys) *do* in fact reason about visual perspective? A problem here is that it does seem difficult to evaluate claims about the previous experiences of experimental subjects. For instance, according to Melis et al. (2006, p. 161) "it seems highly unlikely that our subjects had had experience stealing food from a competitor whom they could only partially see and whose behavior could not be monitored during the ongoing transgression." But is there really good evidence for this and can we plausibly rule out the hypothesis that behavior-readers could have made appropriate generalizations from other experiences of approaching food in concealed ways?

I don't think that these issues are by any means settled as yet, and much further empirical work needs to be done to see if proponents of primate mindreading are right.[10] However, my analysis in this chapter shows how an argument of the sort proposed by Tomasello and Call and others, properly understood, may contribute to the resolution of this and other similar debates. We can now diagnose what is wrong with Povinelli and Vonk's blanket rejection of any study in which the relevant difference in the other agent's mental states and behavior is correlated with observable differences between the conditions. The mere fact that there *is* such a correlation does not suffice to show that a behavior-reading explanation must be at least as plausible as a mindreading explanation for subjects' success in predicting behavior. Studies like Hare et al.'s *can* provide evidence for mindreading, given appropriate background assumptions about subjects' previous experience (or innate endowment), which question the plausibility of a behavior-reader (but not a mindreader) being able to *discover* the relevant correlations and use the observable cue(s) in the right way.

We can also see how future empirical work might bolster a case for primate mindreading that utilizes this kind of argument. For instance, in so far as a behavior-reading account does depend on some kind of learning hypothesis, we can indirectly test this hypothesis by actually probing the learning abilities of primates. Povinelli and Vonk's claims assume that primates are very good at noticing statistical regularities between observable states of affairs. So we can actually look to see how

---

[10] Here we should not forget recent work on primates' understanding of other kinds of mental states such as goals and intentions (Call and Tomasello, 2008).

good they are at doing this (see Call and Tomasello, 2008, for some suggestive remarks). Also, if primates have no insight into others' mental states and reason purely according to observed regularities then we can expect that they will sometimes form *over-* or *under-generalized* expectations about how others will behave in various situations. Hunting for errors of over- or under-generalization has been a successful approach in developmental psychology. For instance, Baillargeon (2004) has found striking errors of under-generalization in young infants' reasoning about the physics of hidden objects, suggesting that infants may not have particularly rich understanding of some of the underlying principles until later in development. Thus one thing to do here is to continue to extend the range of perspective-taking scenarios that primates are tested in. Keeping in mind issues of ecological validity, we can see if transporting tasks to perceptually different but logically similar contexts, or perceptually similar but logically different contexts, alters subjects' performance. For example, does adding a frame, changing the color, size, or some other arbitrary feature of a transparent barrier that has no influence on its occluding properties, affect subjects' performance? We might also try to *manipulate* the regularities that subjects actually observe. In one of their conditions Melis et al. (2006) rewarded subjects for responding to an entirely arbitrary cue that actually had nothing to do with the competitor's mental states, but through experimental conceit did in fact influence the competitor's behavior. They found that subjects did not learn to respond differently according to the presence or absence of this cue even though they were rewarded for doing so. Though this sort of finding is by no means conclusive evidence against learnt behavioral rules, proponents of such an account will need to explain how subjects can easily acquire *some* behavioral rules—presumably on what will have to be quite meager evidence—but not others.

There are therefore a number of avenues for research to pursue that could shed light on whether primates actually reason in the way that Povinelli and Vonk suggest. If we fail to find evidence for this type of reasoning this would not by itself be a knockdown argument for mindreading, but it would substantially strengthen Tomasello and Call's argument—particularly if we find further evidence for primates' flexibility in predicting behavior across different situations.

Acknowledgements

References

Baillargeon, R. (2004). Infants' reasoning about hidden objects: evidence for event-general and event-specific expectations. *Developmental Science*, 7, 391-424.

Boyd, R. (1990). Observations, explanatory power and simplicity. In R. Boyd, J. Glasper and J.D. Trout (eds.), *The Philosophy of Science*. Cambridge, MA: MIT Press.

Burkart, J. and Heschl, A. (2007). Understanding visual access in common marmosets, *Callithrix jacchus*: perspective taking or behavior reading? *Animal Behaviour*, 73, 457-469.

Byrne and Whiten, (1988). *Machiavellian Intelligence*. Oxford: Clarendon Press.

Call, J. and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12, 187-192.

Fitzpatrick, S. (2006). *Simplicity, Science and Mind*. Doctoral dissertation, Department of Philosophy, University of Sheffield.

Fitzpatrick, S. (2008). Doing away with Morgan's Canon. *Mind & Language*, 23, 224–246.

Flombaum, J. and Santos, L. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology*, 15, 447-452.

Goldman, A. (2006). *Simulating Minds*. Oxford: Oxford University Press.

Hare, B., Call, J., Agnetta, B. and Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59, 771-785.

Hare, B., Call, J. and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, 61, 139-151.

Heyes, C.M. (1998). Theory of Mind in non-human primates. *Behavioral and Brain Sciences*, 21, 101-148.

Laurence, S. and Margolis, E. (2001). The Poverty of the Stimulus argument. *British Journal for the Philosophy of Science*, 52, 217-276.

Melis, A., Call, J., and Tomasello, M. (2006). Chimpanzees (*Pan troglodytes*) conceal visual and auditory information from others. *Journal of Comparative Psychology*, 120, 154–162.

Morgan, C. Lloyd (1894). *An Introduction to Comparative Psychology*. London: Walter Scott.

Nichols, S. and Stich, S. (2003). *Mindreading*. Oxford: Oxford University Press.

Norton, J. (2003). A material theory of induction. *Philosophy of Science*, 70, 647-70.

Povinelli, D. and Eddy, T. J. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development*, 61, 1–152.

Povinelli, D. and Vonk, J. (2006). We don't need a microscope to explore the chimpanzee's mind. In S. Hurley and M. Nudds (eds.), *Rational Animals?* Oxford: Oxford University Press.

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515–526.

Sober, E. (1990). Let's razor Ockham's Razor. In D. Knowles (ed.), *Explanation and its Limits*. Cambridge: Cambridge University Press.

Sober, E. (1998). Black box inference: when should intervening variables be posited? *British Journal for the Philosophy of Science,* 49, 469-498.

Sober, E. (2000). Evolution and the problem of other minds. *Journal of Philosophy*, 97, 365–386.

Tomasello, M. and Call, J. (1997). *Primate Cognition*. Oxford: Oxford University Press.

Tomasello, M. and Call, J. (2006). Do chimpanzees know what others see—or only what they are looking at? In S. Hurley and M. Nudds (eds.), *Rational Animals*. Oxford: Oxford University Press.

de Waal, F. (1991). Complimentary methods and convergent evidence in the study of primate social cognition. *Behavior*, 118, 297-320

Whiten, A. (1996). When does smart behavior-reading become mind-reading? In P. Carruthers and P.K. Smith (eds.) *Theories of Theories of Mind*. Cambridge: Cambridge University Press.