The Comparative Delusion:  the ‘behavioristic’/’mentalistic’ dichotomy in comparative Theory of Mind research

Derek C. Penn and Daniel J. Povinelli

1 Introduction

It is no secret that research on the Theory of Mind (ToM) abilities of nonhuman animals has been “fraught with controversy” (Shettleworth 1998). ‘Low-level’, ‘behavioristic’ hypotheses purportedly claim that nonhuman animals learn about the statistical regularities in others’ observable behaviors using low-level mechanisms akin to Pavlovian conditioning without any ability to reason about the causal relation between those behaviors in an abstract or inferentially coherent fashion. ‘High-level’, ‘mentalistic’ hypotheses, on the other hand, propose that nonhuman subjects attribute (at least some) mental states to others and reason about the causal role played by those mental states in a fashion roughly analogous to the way that we (the folk) do. The debate between these two dichotomous alternatives dominated comparative ToM research for the first two decades after Premack and Woodruff’s (1978) original paper (see Heyes 1998; Tomasello & Call 1997 for reviews) and continues unabated in press to this day (see, for example, Emery & Clayton in press; Santos et al. in press; Suddendorf & Whiten 2003; Tomasello & Call 2006; Tomasello et al. 2003a, 2003b; Wood et al. 2007).

To be sure, the debate has become an exercise in shadow-boxing. The overwhelming consensus among comparative psychologists today is that a mentalistic explanation of some kind is warranted for at least certain species and certain tasks (e.g., Emery & Clayton in press; Santos et al. in press; Suddendorf & Whiten 2001; Tomasello & Call 2006; Tomasello et al. 2003a; Wood et al. 2007). Few experimentally-minded comparative researchers claim that nonhuman animals have a ToM equivalent to that of a normal adult human. But most would probably now agree with Tomasello et al. (2003a p. 156) when they conclude that nonhuman animals “understand some psychological states in others—the only question is which ones and to what extent.”
Notwithstanding the eminent consensus arrayed against us, we do not believe that a mentalistic explanation for nonhuman social cognition is warranted. But not because we believe that the behavioristic alternative is any more compelling. In our opinion, both alternatives are equally implausible and the entire dichotomy is specious—or, as Heyes and Papineau (2006) aptly put it, “just Descartes dressed up in modern garb.” In the present chapter, we attempt to argue for a more empirically grounded and theoretically cogent middle way.

Admittedly, Povinelli and colleagues have long argued that comparative researchers should explore the possibility that nonhuman animals may be able to reason about the world in an representationally rich and inferentially coherent fashion without necessarily being able to reason about unobservable, hypothetical entities such as mental states and causal forces (Penn & Povinelli 2007a, 2007b; Povinelli 2000; Povinelli et al. 2000; Povinelli & Eddy 1996; Povinelli & Vonk 2003, 2004; Vonk & Povinelli 2006). But unfortunately, Povinelli’s ‘Reinterpretation’ hypothesis has been persistently misinterpreted as a kind of “derived behaviorism” or “sophisticated behavior-reading” and widely criticized as “unfalsifiable” and “unparsimonious” (see, for example, Andrews 2005; Emery & Clayton in press; Santos et al. in press; Suddendorf & Whiten 2003; Tomasello & Call 2006; Tomasello et al. 2003b). So in the present chapter, we try a very different tack.

We start by reviewing why a ‘behavioristic’ account of nonhuman social cognition is, indeed, entirely implausible. We then show why the mentalistic consensus that dominates comparative research today is nevertheless unwarranted. And we finish by arguing that comparative researchers have failed to critique the most obvious limitation in Povinelli’s ‘Reinterpretation’ hypothesis: to wit, Povinelli’s original hypothesis significantly overestimated the cognitive abilities of nonhuman animals. We propose a new version of the Reinterpretation hypothesis that corrects this error.

2 Flogging the behavioristic strawman

The terms, ‘behaviorist’ and ‘behavior reader’, have played a crucial role in the comparative ToM debate: typically as pejorative labels for an implausible null hypothesis
(for recent examples, see Emery & Clayton in press; Hare et al. 2006; Santos et al. in press; Tomasello & Call 2006; Tomasello et al. 2003b). To be sure, it is often not clear that any real-life researcher has ever held the views attributed to this strawman.

Krebs and Dawkins (1984 p. 386), for example, famously hypothesized that the ‘mindreading’ ability of nonhuman animals was due to their ability to keep track of the “statistical rules” that govern “sequences of behavior.” But it is hard to see how even Krebs and Dawkins’ low-level model could be implemented using purely Pavlovian learning or non-representational processes. Heyes (1998) roundly criticized the evidence for ToM-like abilities in nonhuman primates but explicitly acknowledged the importance of “inferences based on nonmental categories” in addition to purely associative learning processes. And Povinelli and colleagues suggested that chimpanzees might not actually cognize the referential or goal-directed nature of others’ looking behavior in mentalistic terms but were always very clear that chimpanzees were fully cognitive creatures with a rich suite of representations at their disposal (e.g., Povinelli 2000; Povinelli et al. 2000; Povinelli & Eddy 1996).

In any case, the evidence that has accumulated over the past ten years has thoroughly eliminated any doubt that a behavioristic explanation of nonhuman cognition is untenable. In particular, we know that nonhuman animals form highly structured representations about the past as well as the occurrent behavior of other agents. They are not limited to learning about statistical contingencies between behavioral cues in an ad hoc or bottom-up fashion. They often do represent the special causal structure among behavioral cues, particularly the goal-directed relation between other agents’ perceptual acts and how those agents are likely to behave. And they are able to generalize this top-down causal knowledge in an inferentially flexible and ecologically rational (i.e., adaptive) fashion. Perhaps most importantly, we now know beyond a shadow of a doubt that all of these sophisticated social-cognitive abilities are not limited to primates or even to mammals.
2.1 Animals form abstract, structured representations about others’ past as well as occurrent behaviors

On one particularly implausible interpretation of the behavior-reading hypothesis, nonhuman subjects are only able to respond to occurrent stimuli in the immediate environment and are incapable of using representations of past events or states of affairs in order to infer how to behave in novel situations. We are not aware of any contemporary comparative researcher who has actually advocated a stimulus-bound version of behavior-reading in print—certainly not Povinelli and colleagues, who have always emphasized the importance of abstract off-line representations in nonhuman social cognition (Povinelli et al. 2000; Povinelli & Vonk 2003, 2004). But just in case there is an advocate of stimulus-bound behavior-reading lurking silently in the bushes, it is worth noting, for the record, that any non-representational account of nonhuman social cognition has been dead in the water for quite some time. Indeed, thanks to the research of Emery, Clayton and colleagues (see Emery & Clayton in press for a review), we now know that sophisticated representations of both past and present states of affairs are not limited to primates or mammals alone.

Dally et al. (2006) for example, had scrub-jays cache food successively in two trays, each in view of a different observer. At recovery, the original storers recached significantly more food caches when they retrieved their caches in the presence of a previous observer than when they retrieved their caches in private. Furthermore, if a previous observer was present, storers tended to recache from the tray that the previous observer had actually observed. A stimulus-bound explanation of these results might postulate that the storers were simply responding to intimidating perceptual cues being emitted by the observers (what Tomasello et al. 2003a call the ‘evil eye’ hypothesis). But Dally et al. (2006) designed a control experiment in which the trays were removed from the original storer and another bird was allowed to cache food in both trays in the presence of an additional ‘control’ observer. If storers based their re-caching decisions solely on the occurrent behavior of the observer bird, they should have re-cached items in the presence of the control bird just as often as they did in the presence of the actual observers because the control bird had witnessed caching in both trays (although by a different storer) and was presumably emitting the same intimidating cues. In fact, the
original storers exhibited very little re-caching in the presence of the control bird (see discussion by Emery & Clayton in press).

Results such as these rule out any explanation in terms of occurrent behavioral cues alone. Clearly, scrub jays are able to keep track of ‘who’ was present for ‘what’ events in the past; and they use this information in an adaptive fashion to protect their cache sites from potential pilferers. Moreover, these birds are keeping track of the relation between numerous cache sites and various potential pilferers in a combinatorial and productive fashion. The ability to represent the combinatorial relation between numerous cache sites and competitors would be literally unthinkable (i.e., computationally infeasible) without some mechanism for encoding the compositional relation between particular constituents of a representation such that when different constituents have the same relation to each other, the fact that it is the same relation in each case is somehow manifest in the structural similarity between the representations. Horgan and Tienson (1996) argue that this is all it should take in order for a representational system to qualify as “syntactically structured”; and we agree (see Penn et al. 2007 for a more extensive discussion of this point).

In our view, it is indisputable that scrub jays form syntactically-structured representations about the ‘what’, ‘when’, ‘where’ and ‘who’ properties associated with concrete events in the past as well as about the abstract statistical regularities that have held across a number of similar situations. Furthermore, they can combine these concrete and abstract representations in order to respond in a rational (i.e., adaptive) fashion to the relation between a particular competitor and a particular cache site. Presumably these representational and inferential abilities are not limited to corvids.

2.2 Animals possess quite a lot of causal knowledge about the perceptual behavior of other intentional agents

It has been well known for quite some time that nonhuman apes spontaneously follow the gaze of other anthropoid subjects and will track another subject’s gaze to areas outside of their own immediate visual field (Povinelli et al. 2000; Povinelli & Eddy 1996; Tomasello et al. 1998; Tomasello et al. 1999). On a behavioristic explanation of this phenomenon, apes follow the gaze of other apes because they have learned that when
they do so, they often see interesting events or objects in the world: the gaze of another anthropoid agent is simply a statistically reliable cue of something interesting to look at and the nonhuman ape has no real understanding of the special epistemic relation between the other agent’s gaze and the state of affairs being observed. A considerable number of experiments have demonstrated that this behavioristic account of gaze-following is untenable.

There is now compelling evidence that chimpanzees and other primates are sensitive to many (though not all) of the observable factors determining other agents’ line of sight and ‘what’ other agents are looking at (Barth et al. 2005; Brauer et al. 2007b; Call et al. 2000; Call et al. 1998; Flombaum & Santos 2005; Hare et al. 2000; Hare et al. 2001; Kaminski et al. 2004; Okamoto-Barth et al. 2007; Povinelli et al. 2003). For example, chimpanzees tend to ‘check back’ if another agent’s gaze does not seem to lead to anything interesting; and they will move around barriers in order to see what other individuals are looking at. Chimpanzees treat transparent barriers differently from opaque ones when approaching food that a competitor might be able to see (Hare et al. 2000; Hare et al. 2006) or when inferring the external target of another agent’s line of sight (Okamoto-Barth et al. 2007). And in certain circumstances, they will even use a human experimenter’s gaze as a cue to the location of food (Barth et al. 2005). Many of these abilities are not limited to apes or even to mammals. Flombaum and Santos (2005), for example, showed that rhesus monkeys prefer to steal food from human competitors who are looking away or whose eyes are covered rather than from human competitors who are looking in their direction. And Dally et al. (2004; 2005) have shown that western scrub-jays prefer to hide food in dimly-lit and/or more distant sites when being observed by potential pilferers and take into account whether a competitor has an unobstructed line of sight to a given location when making their caching and re-caching decisions.

None of this evidence suggests that nonhuman animals cognize “seeing” as an epistemic act or are reasoning in terms of other agents’ mental experience (we will critique these mentalistic claims below). But the animals’ behaviors are clearly too flexible to be explained in terms of innate orienting reflexes alone. And it would be equally implausible to suggest that nonhuman animals acquire all of these sophisticated gaze-relevant behaviors through general purpose, bottom-up statistical learning alone. As
Santos et al. (in press) point out, without some additional top-down structure or knowledge, the statistical regularities causally relevant to tracking another agent’s gaze to its real-world target would be swamped amongst a practically infinite number of equally salient but spurious correlations. Moreover, unstructured statistical models cannot explain how nonhuman animals are able to infer what another agent is looking at even when they are confronted by novel combinations and configurations of cues (e.g., Brauer et al. 2007b; Hare et al. 2000; Hare et al. 2006).

To be sure, the limitations of uninformed, unstructured statistical learning are hardly news (Clark & Thornton 1997). As far as we can see, uninformed, bottom-up, unstructured statistical learning models are all but extinct among computational and cognitive researchers (see Chater et al. 2006 for a recent introduction). It is widely agreed that even basic object perception requires a sophisticated interplay between top-down and bottom-up statistical processes (Kersten et al. 2004). The same is true for computational models of gaze-following (see, for example, Breazeal & Scassellati 2001; Hoffman et al. 2006).

At the very least, then, gaze-following in primates requires quite sophisticated, flexible, embodied, representational mechanisms for inferring the external target of another agent’s looking behavior from unique combinations of contextual and behavioral cues using prior knowledge about the physical factors relevant to determining other agents’ line of sight. Thus, inferring the external target of another agent’s ‘looking’ behavior clearly requires cognitive mechanisms that go far beyond the limitations of uninformed statistical behavior-reading or associative learning. And there is now compelling evidence that nonhuman animals’ understanding of others’ looking behavior goes a good deal farther than just looking where others are looking (see also Povinelli et al. 2002).

2.3 Animals understand some aspects of the goal-directed relation between agents’ perceptual acts and the world

It is increasingly clear that nonhuman animals do, in fact, recognize some aspects of the special, goal-directed relation between intentional agents’ perceptual acts and the world. For example, Hare et al. (2001) put two chimpanzees, one dominant to the other,
on opposite sides of a middle chamber. The experimenters hid a piece of food in one of two containers in the middle of this chamber. In the Informed condition, both the subordinate and the dominant chimpanzees were able to observe where the food was hidden. In the Uninformed condition, the dominant’s door was shut and only the subordinate subject was able to see where the food was placed. When released into the middle chamber, subordinates tended to approach the food reward more frequently in the Uninformed than in the Informed condition even when they were given a slight headstart and were provided no occurrent cues as to the competitor’s intentions.

This seminal experiment does not warrant the high-level mentalistic interpretation it is routinely given (e.g., Hare et al. 2001; Santos et al. in press; Suddendorf & Whiten 2003; Tomasello & Call 2006; Tomasello et al. 2003a). As Povinelli and Vonk (2003; 2004) point out, it is likely that the subordinate chimpanzees simply realized that they should avoid competing for desirable and monopolizable resources with dominants who have had an unobstructed line of sight to the resource in the recent past. It was not necessary for subordinates to reflect on the visual perspective of the dominant as a distinctively mental experience nor was it necessary for subordinates to reason about what the dominant had seen in terms of an unobservable epistemic mental state. Thus, the additional claim that the subordinates behaved the way they did because they knew that <the dominant had seen the food and therefore knows where it is> may satisfy our all-too-human need to posit a conscious (i.e., folk psychological) reason for the subordinates’ behavior, but it is not warranted by the evidence.

This said, even Povinelli and Vonk’s non-mentalistic account of the chimpanzees’ behavior implies that the subordinates understood quite a lot about the causal relation between the competitor’s line of sight, the nature of the object being observed, and how the competitor was likely to behave in the near future. As Tomasello and Call (2006) point out, the subordinates would not have made the same inference if the dominant had been oriented towards a rock and subordinates did, in fact, make the same inference regardless of whether the food was an apple or a banana (see also Brauer et al. 2007b; Hare et al. 2000). Thus, chimpanzees—and perhaps other animals as well—use some of the characteristics of the object located at the endpoint of a competitor’s line of sight in order to infer how that competitor is likely to act in the near future. This relational
inference does not require mentalistic reasoning; but it is certainly no mean cognitive feat.

Picking out causally relevant relations in the world amidst all the salient but spurious correlations presents uninformed statistical learning mechanisms with a computational quagmire. Yet, as Clark and Thornton (1997) point out, life is rife with problems that require relational solutions; and, at least outside the laboratory, animals routinely solve these problems, typically quite well. This implies that both human and nonhuman animals are eminently relational reasoners; not just uninformed statistical learners. Clark and Thornton (1997) suggest that biological cognizers circumvent the limitations of uninformed statistical learning by employing a range of top-down heuristics, ploys and biases to recognize and reason about the relations that matter. And we agree. The comparative evidence strongly suggests that nonhuman animals possess a variety of top-down heuristics, ploys and biases for picking out the causally features of other agents’ occurrent behaviors and for reasoning about other agents’ future behavior in terms of their goal-directed relation to the world.

We hasten to add that the ability to reason in terms of the goal-directed relation between an intentional agent and a particular state of affairs in the world does not necessarily entail, in addition, the ability to cognize others as intentional agents with their own distinct and unobservable psychological states. This latter ability, we will argue below, is of a different kind altogether.

3 On the lack of evidence for anything even remotely resembling a Theory of Mind among nonhuman animals

Faced by the sophistication of nonhuman animals’ social behaviors and the manifest failure of low-level behavioristic models to account for these impressive cognitive abilities, many comparative researchers have come to the intuitively appealing conclusion that nonhuman animals must reason about other minds in largely the same way that we (the folk) do. Santos et al. (in press) put it bluntly: “we would like to propose that, in fact, primates do reason about unobservable mental states, and that they do so with the same basic cognitive systems that we humans use to reason about mental states.”
Generous mentalistic claims such as these are no longer limited to primates or mammals. Emery and Clayton (in press), for example, argue that there is now “good evidence” for theory of mind in scrub-jays as well.

We disagree. As we argue below, not only is there a lack of compelling evidence for anything remotely resembling a ToM among nonhuman animals, there is consistent evidence of an absence (see also Penn & Povinelli 2007b; Povinelli et al. 2000; Povinelli & Vonk 2003, 2004).

3.1 There is no evidence that nonhuman animals represent or reason about other agents’ epistemic mental states

Following Premack and Woodruff’s (1978) seminal paper, many philosophers and psychologists argued that the acid-test of a representational ToM is whether a given subject can reason about the causal implications of another agent’s false beliefs (Bennett 1978; e.g., Dennett 1978; Harman 1978; Wimmer & Perner 1983). Certainly, reasoning about false-beliefs is not the only distinctive component of a ToM or the only valid test of a ToM (see Bloom & German 2000). But it is exceedingly hard to claim that a nonverbal subject understands another agent’s true beliefs as epistemic mental states if the subject is not capable of reasoning about the causal consequences of that agent’s false beliefs as well.

To date, nonhuman primates have failed every well-controlled test of their ability to reason about the epistemic contents of another agent’s counterfactual representations (e.g., Call & Tomasello 1999). As Tomasello et al. (2003a p. 156) frankly acknowledge, “there is no evidence anywhere that chimpanzees understand the beliefs of others” (see also Tomasello et al. 2005). Worse, all of the experiments to date that purport to show that nonhuman animals can reason about ‘false beliefs’ lack the power, even in principle, of showing that subjects are reasoning about the epistemic contents of others’ mental states as distinct from observable behavioral cues (e.g., Bugnyar & Heinrich 2005; Hare et al. 2001, 2006; Povinelli et al. 1990; Santos et al. 2006). The difference is crucial and persistently overlooked.

For example, Hare et al. (2001) tested whether or not subordinate chimpanzees would distinguish between trials in which the food had been moved and re-hidden while
the dominant was not looking (“Misinformed” condition) and trials in which the
dominant competitor had witnessed the correct placement of the food (“Informed”
condition). As it turned out, the subordinates did not approach the food significantly more
frequently in the “Misinformed” condition than in the “Informed” condition. But even if
they had, this would not have warranted a mentalistic explanation. To pass this test, it
suffices to recognize that a potential competitor is less likely to compete for food if the
food was last hidden when that competitor was not present. In other words, the
subordinates only had to reason in terms of the competitor’s propensity to try and retrieve
the food, not about what the competitor (falsely) believed was located beneath each
occluder. Thus, for the same reasons that Dennett (1978) and many others pointed out 30
years ago, Hare et al.’s (2001) protocol lacks the power, even in principle, of
demonstrating that chimpanzees know what other chimpanzees do and do not know (for
effects of the kind of protocols that could, at least in principle, provide such evidence,
see Penn & Povinelli 2007b).

The same deflationary analysis applies, mutatis mutandis, to the more recent (and
more impressive) performance of corvids (cf. Emery & Clayton in press). In every
experiment reported to date, it suffices for the scrub-jays to keep track of ‘who’ was
present for which caching event without, in addition, keeping track of the distinct
counterfactual representations being maintained by each individual competitor. Indeed,
Dally et al. (2006 p. 1665) themselves acknowledge that scrub jays’ ability to keep track
of which competitors have observed which cache sites “need not require a human-like
‘theory of mind’ in terms of unobservable mental states, but […] may result from
behavioral predispositions in combination with specific learning algorithms or from
reasoning about future risk.”

3.2 Nonhuman animals reason do not appear to reason about ‘goals’ as internal
representational states

As we argued above, at least some nonhuman animals recognize the perceptual
relations causally relevant to predicting an intentional agent’s goal-directed behaviors.
Indeed, there is even evidence that nonhuman apes understand some of the perceptual
cues causally relevant to discriminating between ‘intentional’ and ‘accidental’ behaviors
in other anthropoid agents (see Call et al. 2004; Call & Tomasello 1998). But this
evidence does not suggest that nonhuman animals reason about other agents’ goals qua mental representations. As Tomasello et al. (2005) point out, the psychological literature has been plagued by a pervasive ambiguity in the way that the term, “goal”, is used:

“The word goal contains a systematic ambiguity that has contributed to much confusion… When it is said that a person wants a box open, for example, we may distinguish the external goal – a certain state of the environment such as an open box— and the internal goal—an internal entity that guides the person’s behavior (e.g., a mental representation of a desired state such as an open box).”

(Tomasello et al. 2005 p. 676)

When hypothesizing about how cognizers understand the goal-directed behavior of other agents, it is crucial, Tomasello et al. (2005) argue, to distinguish between external goals (i.e., the external state of affairs used as a reference point by the other agent’s cognitive process) and internal goals (i.e., the internal representation of that external goal). We agree. Reasoning about the relation between an external goal and a given agent’s behavior is not the same thing as reasoning about the relation between an internal goal and an agent’s behavior: Only in the latter case do unobservable mental states play a distinctive causal role.

Numerous comparative researchers have claimed that nonhuman animals do, in fact, reason about other agents’ internal goals (Flombaum & Santos 2005; Hare et al. 2000; Hare et al. 2001, 2006; Santos et al. in press; Santos et al. 2006; Tomasello et al. 2003a; Wood et al. 2007). Indeed, Tomasello et al. (2005) themselves claim that nonhuman animals reason about internal goals\(^1\). But all of the evidence to date is consistent with the more modest hypothesis that nonhuman animals reason solely about an agent’s relation to external goals. As Tomasello et al. (2005) point out, the defining feature of being a goal-directed subject is that one is sensitive to the current value of one’s internal goals and the dynamic, causal relation between one’s own actions and the desired outcome. There is extensive evidence that animals as humble as laboratory rats

---

\(^1\) At least, Tomasello et al. (2005 p. 676) stipulate that they will “reserve the term goal for the internal goal, and for the external goal we will use such expressions as ‘the desired result’.” And then go onto claim that “apes understand that others have goals” (p. 685).
are, indeed, goal-directed subjects in this sense (see, for example, Dickinson & Balleine 2000). But there is no evidence that any nonhuman animal recognizes that another agent’s goal-directed behaviors are sensitive to that agent’s representation of the current value of the goal and to that agent’s representation of the instrumental efficacy of a given action as distinct from the subject’s own representations of the goal’s value and the instrumental efficacy of a given action. Indeed, there is not simply an absence of evidence that nonhuman animals reason about other agents’ goals in this mentalistic manner, there is consistent evidence of an absence.

The crux of Tomasello et al.’s (2005) hypothesis is that human children undergo a qualitative change in their social-cognitive abilities at around 12 to 15 months of age: specifically they start to engage in triadic interactions with other individuals involving shared goals and socially coordinated action plans. These collaborative interactions show that children in their second year of life start to recognize that different agents may have different internal goals and different representations of the instrumental efficacy of various possible actions. Tomasello et al. (2005) argue, rightly we believe, that ‘shared intentionality’ is a uniquely human capability. As they point out (pp. 685-686), nonhuman animals do not naturally point, show, or even actively offer things to conspecifics. There is no convincing evidence for instructional teaching among any nonhuman species. And there is no evidence for anything remotely resembling the extensive collaborative interactions between humans. As Tomasello et al. (2005 p. 685) put it, “it is almost unimaginable that two chimpanzees might spontaneously do something as simple as carry something together or help each other make a tool.”

In short, nonhuman primates lack precisely those social-cognitive behaviors that require an ability to reason about others’ goals as internal representational states—susceptible to revaluation and collaborative alignment—and manifest only those behaviors that are possible when reasoning about other agents’ goal-directed behavior in terms of those agents’ relation to external states of affairs. Although nonhuman primates clearly understand and reason about others’ external goals, the ability to reason about internal goals appears to be a uniquely human specialization.
3.3 Nonhuman animals do not appear to possess anything remotely resembling a mentalistic understanding of others’ perceptual acts

The comparative evidence firmly demonstrates that at least some nonhuman animals recognize how particular combinations of contextual and behavioral cues can be used to infer what another agent is ‘looking at’. But contrary to the general consensus, we do not believe that this finely-tuned inferential ability—as sophisticated as it may be—is functionally or representationally equivalent to reasoning about another agent’s “visual perspective” (cf. Brauer et al. 2007b; Emery & Clayton in press; Hare et al. 2000; Hare et al. 2006; Okamoto-Barth et al. 2007). Comparative researchers often seem to believe that there is nothing more to claiming that a nonhuman animal attributes a visual epistemic perspective to others’ perceptual acts than that the subject engages in behaviors that are interpretable as attributing a visual epistemic perspective to others’ perceptual acts. In other words, comparative researchers rarely differentiate between subjects who have an implicit understanding of another agent’s perceptual mental states (i.e., one that is implied by the subject’s behavior and imputed by a human observer but not necessarily tokened or predicated by the subject’s own cognitive system) and subjects who have an explicit understanding of another agent’s perceptual mental state (i.e., one in which the subject’s cognitive system has predicated the causal relation between the other agent’s internal, unobservable mental state and the other agent’s external, observable perceptual behavior). One notable exception is Whiten (1996), who proposed an elegant specification of what it means to say that a nonverbal creature possesses an explicit concept of another agent’s mental state.

To make a credible distinction between implicit and explicit mental state concepts, Whiten (1996) suggested that researchers look for the ability to recognize the relational similarity between perceptually disparate behavioral patterns in terms of the common causal role play by some unobservable mental state. For example, a chimpanzee that encodes the observable patterns, “X saw Y put food in bin A”, “X hid

---

2 If comparative psychologists mean the term “visual perspective” to be taken in a “non-epistemic” sense (e.g., Dretske 1969), this is certainly in need of explication: it is not clear how understanding another agent’s ‘visual perspective’ in a non-epistemic fashion would warrant the kind of mentalistic claims being routinely made (e.g., Hare et al. 2000).
food in bin A”, and “X sees Y glancing at bin A” as members of the same abstract equivalence class with analogous causal consequences could be said, on Whiten’s account, to recognize that “X knows food is in bin A” and thus possess an ‘explicit’ concept of ‘knowing’ as a mental state.

Whiten’s definition may very well qualify as the minimal reasonable criterion for claiming that a nonverbal animal possesses an explicit concept of a mental state. His definition does not require a subject to possess metarepresentational mental states, a theory-like understanding of the mind or the ability to reason about counterfactual mental representations. Indeed, a subject might possess an explicit mental state concept of ‘seeing’ or ‘hearing’ sensu Whiten without any inkling that other animals have mental states ‘in their heads’. Nevertheless, not only is there a striking absence of evidence for anything remotely resembling an explicit concept of perceptual mental states sensu Whiten in any nonhuman species, there is converging and growing evidence of an absence.

As Whiten (1996) points out, demonstrating that a nonverbal subject possesses an explicit mental state concept requires ‘triangulating’ across disparate protocols and showing that the subject cognizes the common causal role play be a given mental state across perceptually disparate task contexts (see also Heyes 1998). But in the last ten years, a growing body of evidence suggests that nonhuman animals do not, in fact, recognize the mental states relations that are common across disparate task contexts. The lack of an over-arching ‘theory of mind’ is particularly striking when comparing the behavior of nonhuman primates on ‘competitive’ and ‘cooperative’ protocols.

With respect to ‘seeing’, for example, nonhuman primates appear to be highly sensitive to the disposition of an anthropoid competitor’s eyes when attempting to steal food from them (e.g., Brauer et al. 2007b; Flombaum & Santos 2005) but are relatively insensitive to the disposition of another agent’s eyes in cooperative-communicative tasks (Barth et al. 2005; Call et al. 2000; Call et al. 1998; Kaminski et al. 2004; Povinelli & Eddy 1996). In other words, at least for nonhuman primates, the disposition of the eyes in the context of stealing food apparently does not belong to the same abstract mental state equivalence class as the disposition of the eyes in the context of begging for food.
With respect to ‘hearing’, Call (2004) showed that nonhuman apes are able to use the noise of food being shaken in a container to determine in which container the food is hidden. And Santos et al. (2006) showed that rhesus monkeys try to avoid making unexpected noises when stealing food from a competitor who is not already looking in their direction. But Brauer et al. (2007a) has shown that, if someone else makes the noise, chimpanzees do not realize that their competitors can use the noise of food being hidden to locate the food. To put these contradictory results in Whiten’s terms, nonhuman primates apparently do not classify the observable patterns “The container makes noise when I try to steal food from A” and ‘X makes noise hiding food in the presence of A” as indicative of the same mental state relation, ‘A hears where the food is located’.

Many comparative researchers have interpreted the discrepancy between nonhuman primates’ behavior on competitive and cooperative tasks as evidence that nonhuman primates are employing two different cognitive systems: a behavioristic system in cooperative situations and a mentalistic system in competitive ones (e.g., Hare 2001; Santos et al. in press; Tomasello et al. 2003a). Certainly, the discrepancy between nonhuman primates’ behavior on competitive and cooperative tasks suggests that they are employing cognitive mechanisms that are tuned to the ecologically relevant features of each task. But possessing cognitive mechanisms finely tuned to particular social tasks is not the “essence” of a ToM system (cf. Santos et al. in press). The essence of an ToM—in both the theory-like sense defined by Premack and Woodruff (1978) as well as the minimalist sense set forth by Whiten (1996)—is the ability to explicitly represent (i.e., predicate) and reason about the causal role played by a given mental state across disparate behavioral contexts. This ability, Whiten (1996) argues convincingly, should allow a mentalistic subject to draw inferences about the causal consequences of a given mental state—e.g., ‘seeing’, ‘hearing’—in novel behavioral contexts. But the evidence to date suggests that this is precisely the ability that nonhuman animals lack—and Whiten’s analysis suggests why.

Whiten’s example of ‘explicit’ mindreading is a textbook example of analogical reasoning: Whiten’s hypothetical chimpanzee must infer a systematic relational correspondence among perceptually disparate behavioral patterns that have nothing in common other than a common but unobservable causal mechanism: i.e., what X
“knows”. If this is an “intervening variable”, it is an intervening variable that requires reasoning about the analogical similarity between disparate causal relations in order to be tokened. But apart from the remarkable and unreplicated feats of a single highly enculturated chimpanzee (i.e., Gillan et al. 1981), there is no evidence that nonhuman animals are able to reason by analogy and considerable evidence that they are not (see Penn et al. 2007). *Ex hypothesi*, the reason why only human subjects possess explicit mental state concepts is because only humans have the representational architecture necessary to reason by analogy.

4 Povinelli’s hypothesis revisited

If a low-level behavioristic hypothesis was the only alternative to a high-level mentalistic hypothesis, then the current mentalistic consensus might be justified on the grounds that it is less implausible than the alternative. But Povinelli and colleagues have long pointed out that there is a vast and largely unexplored middle ground between construing animals as nothing more than operant learners and claiming that they have a mentalistic appreciation of other minds (Povinelli 2000; Povinelli & Bering 2002; Povinelli et al. 2000; Povinelli & Eddy 1996; Povinelli & Giambrone 1999; Povinelli & Prince 1998; Povinelli & Vonk 2003, 2004; Vonk & Povinelli 2006).

4.1 The original Reinterpretation hypothesis

For many years now, Povinelli and colleagues have argued that chimpanzees and other animals are fully “cognitive creatures” endowed with mental representations and inferential abilities similar to those of humans but that nonhuman animals’ representational capabilities might not encompass *all* the same semantic possibilities as human subjects. In particular, given the lack of compelling evidence for mentalistic representations and abstract causal reasoning among chimpanzees, Povinelli hypothesized that chimpanzees are unable to reason about unobservable entities such as mental states and causal mechanisms (see, in particular, Povinelli 2000; Povinelli et al. 2000).

To explain both the profound similarities and dissimilarities between human and nonhuman cognition, Povinelli proposed that chimpanzees and humans share a suite of
representationally sophisticated systems for coping with the social and physical worlds. In addition to these shared mechanisms, humans possess a unique representational system that allows us, and us alone, to interpret the outputs from these ancestral systems in a novel, and quite peculiar, manner—namely, to interpret the behavior of the self and others as being strongly influenced by unobservable entities known as ‘mental states’.

According to the ‘Reinterpretation’ hypothesis, this uniquely human system for representing mental states did not replace the ancestral systems we share with other primates. Rather, the human mind is still composed of both kind of systems and both are inextricably intertwined with each other. Indeed, the original Reinterpretation hypothesis proposed that most of the social-cognitive mechanisms currently employed by normal humans are largely shared with other primates and were in full operation long before humans acquired the means for reinterpreting others’ behaviors in terms of unobservable mental states.

The original Reinterpretation hypothesis has much to recommend it. It was developed primarily in response to the inherent logical weaknesses in the then (and now once again) current claims that chimpanzees and other animals possess a ToM. And the fundamental philosophical and methodological challenge thrown out by the Reinterpretation hypothesis has never been acknowledged or refuted by those advocating a mentalistic explanation of nonhuman cognition: to wit, comparative researchers have consistently failed to specify what unique causal work is being performed by nonhuman subjects’ ToM system that could not have been performed by a sophisticated cognitive system representing and reasoning about observable behaviors alone (see also Penn & Povinelli 2007b; Povinelli & Vonk 2003, 2004).

In hindsight, however, the original Reinterpretation hypothesis has a glaring limitation: Because of its myopic focus on exposing the weakness of existing claims for mentalistic and abstract causal reasoning in nonhuman animals, the Reinterpretation hypothesis seriously overestimated the cognitive abilities of nonhuman animals and the degree of similarity between the mental systems of chimpanzees and humans. Indeed, the original Reinterpretation hypothesis never proposed any limitation on the inferential abilities of nonhuman animals other than an inability to reason about unobservable entities such as causal mechanisms and mental states (see, for example, Vonk & Povinelli
2006). Even though it suggested some general parameters, Povinelli’s original Reinterpretation hypothesis never offered a detailed account of the differences between the representational architectures of human and nonhuman minds. We now believe that the discontinuity between human and nonhuman cognition is much broader and deeper than an inability to reason about unobservable entities alone.

### 4.2 The Relational Reinterpretation hypothesis

Penn, Holyoak and Povinelli (2007) have recently proposed a revision to Povinelli’s original Reinterpretation hypothesis that provides a preliminary specification of the kind of representational-level changes necessary to account for the discontinuity between human and nonhuman cognition. Like the original Reinterpretation hypothesis, the ‘Relational Reinterpretation’ hypothesis proposes that both human and nonhuman animals possess a rich suite of heuristics, biases, top-down knowledge and inferential mechanisms that allow them to pick out the causally relevant relations in the world amidst all the salient but spurious correlations and to form syntactically-structured mental representations about these relations that can be used in a flexible, reliable and ecologically rational (i.e., adaptive) fashion. Unlike the original Reinterpretation hypothesis, however, our new hypothesis argues that only human animals possess a cognitive architecture capable of systematically reinterpreting perceptual, embodied relations in terms of the kind of higher-order, role-governed relational representations found in a physical symbol system (Newell 1980; Newell & Simon 1976)—or, to be more precise, only human subjects possess a cognitive architecture capable of approximating these higher-order features of a physical symbol system, subject to the evolved, content-specific biases and processing capacity limitations of the human brain. Indeed, after reviewing the comparative evidence from numerous domains of research, Penn, Holyoak and Povinelli (2007) argue that almost all of the salient functional discontinuities between human and nonhuman minds – including our species’ unique linguistic, mentalistic, cultural, logical, and causal reasoning abilities – result in large part from the substantial difference in degree to which the human and nonhuman cognitive architecture are able to approximate the higher-order, systematic, relational capabilities of a physical symbol system.
With respect to social interactions, for example, we believe that both human and nonhuman animals possess a variety of mechanisms for recognizing those relations that are causally relevant to predicting the goal-directed behavior of other intentional agents. These heuristics enable both human and nonhuman animals to pick out the causally relevant relation between ‘what’ an agent is ‘looking’ at and how that agent is likely to behave in the near future without computing massive correlations among all possible statistical regularities. However, only humans cognize the higher-order analogical similarities between perceptually disparate behaviors and thus only humans possess the ability to reinterpret other agents’ goal-directed relations in terms of abstract mental state relations disembodied from any particular task context.

Importantly, we are not claiming that our higher-order relational capabilities are sufficient to explain all of our species’ unique ToM abilities. It is clear that human culture (Tomasello et al. 2005), specialized neural systems (Saxe 2006), and language (de Villiers 2000)—to name only the most obvious factors—all play crucial roles in subserving the unique features of human social cognition. Much like Whiten (1996), we are simply suggesting that the ability to recognize and reason about higher-order analogical similarities between perceptually disparate behaviors is a necessary—but not sufficient—condition for enabling our human ToM.

The suggestion that mentalistic explanations are more ‘parsimonious’ than non-mentalistic explanations is a constant refrain in the comparative literature (e.g., Call 2006; Emery & Clayton in press; Tomasello & Call 2006; Whiten 1997; Whiten & Byrne 1991). Our new Relational Reinterpretation hypothesis provides a representational-level explanation for why the putative ‘parsimony’ of mentalistic explanations is illusory. If the last few decades of computational cognitive research have taught us anything, it is that higher-order relational reasoning and analogical inferences are not particularly easy to implement in a biologically plausible neural network: indeed, they require an enormous degree of representational complexity above and beyond the functionality necessary to reason about first-order perceptual relations alone (Gentner et al. 2001; Holyoak & Hummel 2000). Positing that nonhuman animals are limited to reasoning about first-order relations between observable states of affairs is thus both more consistent with the comparative evidence and more parsimonious from a computational
and representational point of view (see Penn et al. 2007 for a more extended version of this argument).

Notwithstanding the monumental impact our uniquely human system for reasoning about higher-order relations and analogical inferences has had on human cognition, we suspect that humans nevertheless overestimate the importance and cognitive efficacy of our symbolic-relational abilities. As Povinelli’s original Reinterpretation hypothesis first suggested, the vast majority of humans’ everyday social interactions do not engage our uniquely human ToM system (see also Bermudez 2003). The role of explicit mentalistic theorizing in human affairs is more post-hoc than we folk would like to admit—and often misguided to boot. Indeed, our species’ cognitive system for reasoning about higher-order symbolic relations does not merely subserve our unique linguistic, logical, causal reasoning and mentalistic abilities. It also subserves our inveterate predilection to reinterpret the behavior of heterospecifics in mentalistic terms… and many other uniquely human delusions.

References


