The Evolution of Human Mindreading:

How Non-Human Primates Can Inform Social Cognitive Neuroscience

Laurie R. Santos, Jonathan I. Flombaum, and Webb Phillips

Yale University

1. The Secret of Our Evolutionary Success

For social species like our own, evolutionary success requires more than the basics of finding some food and a mate. In order to survive and reproduce, humans must successfully navigate a rather complicated social world. Each day, we are required to interact with countless other humans, all behaving in ways that we must predict, interpret, and in some cases manipulate. It's a daunting task, even for large-brained primates like us. Luckily for each of our genetic legacies, we're pretty good at it.

We owe this social sophistication in large part to a remarkable cognitive shortcut—the capacity to think about behaviors in terms of mental states like intentions, desires, thoughts and beliefs, an ability commonly referred to as a *theory of mind*. The term "theory of mind," originally coined by Premack and Woodruff (1978), serves as a somewhat literal portrayal of why this cognitive shortcut is considered so remarkable. As Premack and Woodruff (1978) originally noted, our theory of mind is *theory*-like in that it is a set of inferences about things that are not directly observable—namely, states of the *mind*—that are then used to make predictions about an individual's future behaviors. That we, as human adults, have developed a theory about something we cannot directly observe is a fascinating computational feat.

In the decades since Premack and Woodruff's landmark paper, theory of mind (hereafter ToM) has become a hot topic of research in many different fields: cognitive development, social psychology, comparative psychology, abnormal psychology, and philosophy of mind, to name only a few. Not surprisingly, researchers in cognitive neuroscience have followed suit with investigations into the mechanisms underlying human ToM abilities. Such work has revealed that a constellation of neural areas seem to take part in our processing of other minds. Researchers have identified areas as diverse as parts of visual cortex, the amygdala, the anterior

cingulate cortex (ACC) and the superior temporal sulcus (STS). At present, a vast array of experimental work seeks to understand the unique role played by each of these centers in our ability to represent the mental states of others. This work has been elegantly summarized in a number of recent reviews (see Allison, Puce, & McCarthy, 2000; Gallagher & Frith, 2003; Saxe, Carey, & Kanwisher, 2004 for example), and so we will not focus on this work here.

Instead, we will highlight a potential methodological challenge that neuroscientific investigations of ToM have now begun to face. To date, explorations into the neural substrates of ToM have focused primarily on the subjects to whom we can confidently attribute these abilities: ourselves. Consequently, much of the research exploring the neural basis of ToM has utilized non-invasive imaging techniques, such as fMRI. Although non-invasive techniques have already provided some marvelous insights into the global systems involved in human ToM processing, these techniques bear some considerable limitations for studying some aspects of ToM. The first of these is the problem of dissociating the smaller components of ToM processing at the neural level. ToM is made up of a host of elemental cognitive abilities, some of which are likely to activate very close neural populations. Present human functional imaging techniques are not well suited to examine adjacent neural populations in this way, and thus may miss critical differences between the smaller component processes that give rise to ToM (Saxe et al., 2004). Likewise, anatomically separate brain regions may become jointly activated by a number of distinct cognitive processes, making it difficult to decipher the role of each of these areas in different aspects of our ToM abilities. Finally, functional imaging techniques are not well suited for one of the major goals of a neuroscientific inquiry into ToM abilities: the development of neuropharmacological treatments for clinical disorders of ToM reasoning, such as autism (e.g. Baron-Cohen, 1995; Schultz et al., 2000; Siegal and Varley, 2002). The exclusive

use of human subjects ethically precludes the pharmacological experimentation essential for eventually treating such disorders (Amaral, 2002; Machado and Bachevalier, 2003). Faced with these limitations on the use of non-invasive techniques, the field of cognitive neuroscience could greatly benefit from the development of a primate model of human ToM abilities. Using a primate model would afford the opportunity to engage the substrates of ToM with more sensitive measures of neural activity and the potential to develop new neurochemical assays.

The need for a primate model of ToM raises some tough questions for primate researchers: Do any primates other than humans actually possess a ToM? Or is ToM a uniquely human cognitive specialization?   Thankfully for neuroscientists, researchers in comparative cognition have gained considerable insight into these controversial questions in just the past few years. Here, we review recent empirical and philosophical advances in the study of ToM abilities in non-human primates (hereafter, primates), presenting some of our own work investigating these abilities in macaque monkeys.

Before we turn to this review, however, we should note an important caveat. The challenge of investigating representations of unobservable entities such as mental states is only amplified when one hopes to study these representations in an organism that cannot speak.   As a consequence, the comparative study of ToM has enjoyed lively (and often productive) debates as to which (if any) experimental paradigms successfully demonstrate ToM abilities in primates. Although others may well disagree (Povinelli, this volume), we will make the strong case that the available data demonstrate that some primates do in fact theorize about the mental states of others in some of the same ways as human adults do. However, we hope that differences in the specific interpretation of recent experimental findings will not obscure what we believe are the most vital conclusions drawn from these experiments, conclusions that we hope even

methodological skeptics can agree with: (1) Some primates succeed in competitive versions of ToM tasks, even though they fail on nearly identical cooperative versions. (2) Primates in these experiments succeed in interpreting exactly the kinds of stimuli that individuals with clinical deficits, such as autism, appear unable to interpret. And (3) there remains the possibility that the cells in primate cortex known to encode the location of another individual's eye gaze are involved in sophisticated aspects of social reasoning, such as stealing food from a competitor.

We have therefore organized the reminder of this chapter into three separate sections. The first reviews the most recent findings in primate ToM with an emphasis on how primates appear to succeed in social reasoning tasks that are extremely similar to ones that they have famously tended to fail. This section focuses more on historical changes in the study of primate social reasoning, and less on the kinds of ToM representations that these experiments reflect. The second section then revisits this suite of experimental work considering both skeptical and generous views of what these findings mean, and with an eye towards more theoretical aspects of this work (i.e., are primate representations of other minds different from human ones?). Finally, we turn to the implications that recent behavioral work might have for neuroscientists studying the underlying circuitry of human ToM and the etiology of diseases such as autism.

2. A History of Primate Theory of Mind: From Cooperative to Competitive Paradigms

Anyone who has observed primates knows that they spontaneously exhibit a number of natural behaviors that seem consistent with a rich understanding of the mental states of others. Primates demonstrate countless examples of functional deception both in the wild and captivity—they refrain from announcing the presence of food if more dominant individuals are nearby and they conceal objects by hiding them from other individuals' views. (see Byrne and Whiten, 1988; Mitchell, 1999; Whiten and Byrne, 1997). Similarly, a number of primate species

naturally recognize where other individuals are looking (Tomasello and Call, 1997 for review). Chimpanzees (*Pan troglodytes*), for example, spontaneously follow the gaze of both human experimenters (Itakura, 1996; Povinelli and Eddy, 1996b; Povinelli and Eddy, 1997; Tomasello, Hare, and Fogelman, 2001) and conspecifics (Tomasello, Call, and Hare, 1998), and can track an individual's line of regard past barriers and distracters to a target object (Povinelli and Eddy, 1996b; Povinelli and Eddy, 1997; Tomasello et al., 1999).

But despite this impressive range of seemingly insightful social reasoning, primates, until very recently, have typically failed in experimental tasks that attempt to expose whether these animals represent the contents of others' mental states, as opposed to merely responding, albeit adaptively, to others' behaviors (see reviews in Heyes, 1998; Povinelli, 2000; Tomasello and Call, 1997). That is, although animals may respond to cues that reflect the knowledge of others -- for example, they may turn their heads to follow the gaze of others -- they may not actually represent the mental states of others in any meaningful way, or be able to use the information that they represent for solving more sophisticated social problems. Rhesus monkeys (*Macaca mulatta)*, for example, fail to recognize the difference between guessing and knowing in a hidden food task (Povinelli, Parks, and Novak, 1991), and overlook the knowledge state of their offspring when alerting them to the presence of predators (Cheney and Seyfarth, 1990). Primates have even failed in one of the most simplified ToM tasks possible: a game of pick-the-one-the-experimenter-is-looking-at, also known as the object choice task. Here, a human experimenter (or sometimes a trained conspecific) attempts to communicate the location of a hidden food by either looking directly at it or pointing and gesturing towards it. Monkeys and apes in a number of different labs have failed this task, choosing randomly between indicated and ignored locations (Anderson et al., 1995; Anderson et al., 1996; Call et al., 1998; Call et al., 2000;

Itakura et al., 1999; Itakura and Anderson, 1996; Peignot & Anderson, 1999; Povinelli et al., 1999; Povinelli and Eddy, 1996c). Similarly, chimpanzees with extensive experience interacting socially with human caretakers fail to take into account what different experimenters can see when choosing whom to ask for food (Povinelli and Eddy, 1996a). Without training, chimpanzees failed to distinguish between an experimenter with a bucket on her head and one holding a bucket in her arms (though these chimpanzees also had extensive experience playing with buckets and wearing them on their own heads), and an experimenter whose entire head orientation was directed at them and one who was directed away from them. Even after training, the apes in these experiments continued to beg for food from experimenters unable to see them, leading the authors to conclude that chimpanzees possess little understanding of the nature of visual attention—let alone mental states such as beliefs, desires, and intentions.

The apparent discrepancy between naturalistic observations of primate social behavior and negative experimental findings in laboratory tasks recently prompted some researchers to seek a new approach to examining what primates know about the minds of other individuals. This new approach employs the general insight that some kinds of cognitive abilities (particularly those that are domain-specific, see Cosmides and Tooby, 1994) may only become engaged in experimental settings that mimic the ecologically-relevant context for which these abilities evolved. Thus, as a number of researchers have proposed (Byrne and Whiten, 1988; Hare, 2001; Humphrey, 1976), if primates actually understand the mental states of other individuals, then the most likely domain in which they would use this information is when vying with other individuals in ecologically-relevant situations.

In a series of pioneering experiments, Hare, Call, Tomasello and their colleagues (Hare et al., 2000; Hare et al., 2001, see review in Tomasello et al., 2003) noticed that previous empirical

tests on primate ToM differed in at least one important way from the situations in which primates in the wild appear to reason about mental states. Specifically, primates tend to appear the most socially adept when competing with one another (e.g., Hare, 2001), for example, when attempting to conceal the presence of contested food or gain sexual access to contested females. Unfortunately, most previous experiments on primate ToM did not incorporate this feature of competition; most past tests instead required cooperation among individuals. In the object choice paradigm, for example, the importance of where a human is looking can only be understood if the subject first appreciates the experimenter's cooperative intent to communicate the location of the hidden food. Similarly, in the context of Povinelli and Eddy's (1996) begging experiments with chimpanzees, the visual awareness of human caretakers is only relevant if the subject appreciates that a human's response to the begging gesture will be the voluntary decision to give the beggar food. It remains to be seen if this is indeed how chimpanzees and other primates perceive these inherently cooperative interactions.
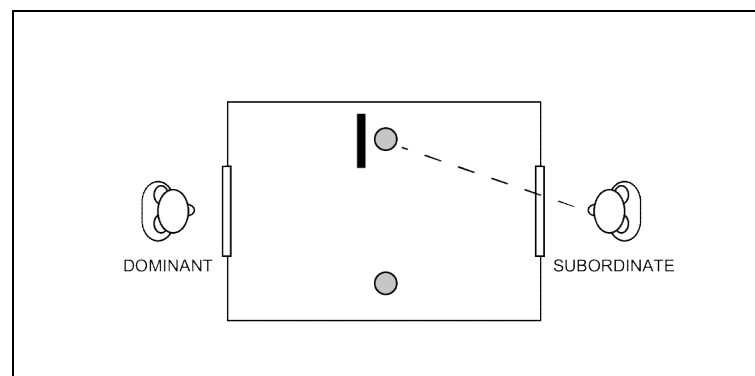
Figure 1. An aerial view of the chimpanzee competition paradigm (adapted from Hare et al., 2000). Subordinate chimpanzees selectively approached the piece of food that the dominant individual could not see.

Hare and colleagues (2000) attempted, therefore, to capitalize on what appears to be the primates' natural inclination to reason about mental states when competing with others for scarce food resources. To do so, they set up a situation in which chimpanzees could compete over access to hidden pieces of food. In this competitive foraging paradigm, subordinate and

dominant chimpanzees were positioned on alternate sides of a middle cage that contained two pieces of food (see Figure 1). In some cases, the food was placed such that the dominant individual could see only one of the two pieces. Hare and colleagues hypothesized that if the subordinate individual was sensitive to what the dominant individual could and could not see, then he should selectively attempt to retrieve only those foods that the dominant individual could not see. Subordinate chimpanzees did just this, successfully retrieving the food that the dominant individual could not see. Similarly, subjects were more likely to approach food that was hidden from the dominant's view by a visual barrier (Hare et al., 2000) and food that the dominant individual did not know was placed in a particular location (Hare et al., 2001). In a more recent series of studies, Hare and colleagues (in press) presented chimpanzees with the opportunity to approach a competitive human experimenter in an attempt to obtain contested food. Again, chimpanzees were more likely to try to approach food that the experimenter was not watching. Taken together, the competitive studies of Hare and colleagues suggest that chimpanzees do have a rich notion of what other individuals can and cannot see, at least in certain competitive situations.

But chimpanzees are not the only species to succeed in ToM experiments involving competition. Recently, our lab has extended the findings of Hare and colleagues to another primate species—the rhesus macaque. To do so, we attempted to develop a competitive paradigm that was similar to classic ToM experiments in the hope that such a methodology would help to illuminate just why competition is such a fertile test-bed for ToM-like reasoning. To this end, we (Flombaum and Santos, 2005) developed a competitive foraging task for a population of free-ranging macaques living at the Cayo Santiago field station in Puerto Rico.

The Cayo Santiago macaques spend much of their day foraging for monkey chow provisioned to them by human experimenters. However, due to the long history of research on this island, the monkeys in this population have also developed a compelling interest in the foods that they see human experimenters eating. (Human foods tend to be sweeter and more exciting than the monkey chow they normally eat.) Because of their curiosity for human foods (see Santos et al., 2001), the macaques often approach experimenters and attempt to steal their food, so much so that we and our colleagues typically use this approach behavior as a dependent measure in empirical studies with this population (see Flombaum et al., 2004; Hauser, 2001; Hauser et al., 2000; Phillips and Santos, in submission; Santos, 2004; Santos et al., 2001; Santos et al., 2002 Sulkowski and Hauser, 2001). In these studies, subjects always appeared somewhat apprehensive of getting close to the experimenters involved in testing, suggesting the macaques may view humans as potentially dangerous competitors. As such, we reasoned that these monkeys should be motivated to take human food only when they can do so without being detected.
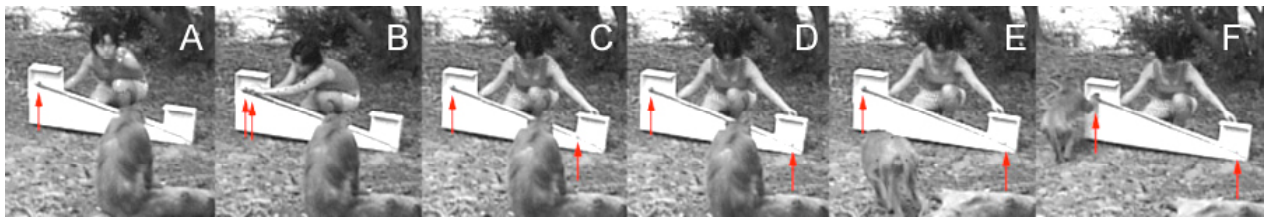

Figure 2. Stimuli used in Flombaum and Santos (2005): A) Experiment 1. Subjects chose the competitor facing backwards, B) Experiment 4. Subjects chose the competitor with his eyes to the side, C) Experiment 6. Subjects chose the competitor with his eyes covered.

Using this logic, we investigated whether the monkeys spontaneously took into account the direction of a human experimenter's gaze when attempting to steal food (Flombaum and Santos, 2005). In each experiment, two male experimenters—aka the "competitors"—approached a lone monkey. Both competitors then placed a platform holding a

grape down on the ground and then turned in a particular way relative to the grape. In the first study, one competitor turned to face the grape, while the other turned his back to the grape (see Figure 2A). Both experimenters then froze for one minute and allowed the subject to attempt to steal the grape. We predicted that subjects should steal the grape from the competitor who could not see the grape, namely the one with his back turned. Our rhesus monkey subjects did just this, choosing to approach the experimenter with his back to the grape. We then explored whether the monkeys could use more subtle cues to what competitors can see. As in the first study, the macaques selectively retrieved the grape from a competitor whose head and eyes were oriented away (Experiment 3), whose eyes alone were oriented away (Experiment 4, see Figure 2B), or whose gaze was blocked by a small barrier (Experiments 5 and 6, see Figure 2C). Taken together, these results demonstrate that macaques, like chimpanzees (Hare et al., in press), spontaneously use information about what a human competitor can see when determining which contested grape to steal. Interestingly, our subjects selectively chose between two humans whose postures were nearly identical to those used in Povinelli and Eddy (1996a)'s famous begging experiments with chimpanzees (see Figure 2); unlike our subjects, the chimpanzees in Povinelli and Eddy's (1996) studies chose randomly between the two experimenters. The only substantive difference between our experiments and those with chimpanzees is that success in the chimpanzee studies required approaching and then cooperating with the experimenter who *could* see, whereas success in our experiments required avoiding and competing with the experimenter who *could not* see.

In a second series of studies (Flombaum & Santos, 2004), we examined whether rhesus monkeys further understand that seeing leads to knowing. Do they know that a competitor who has not *seen* where a piece of food is doesn't *know* where that food is? And do they exploit such

false knowledge when competing for food? To examine these questions, we used a slightly different set-up (see Figure 3): a single human competitor crouched behind a ramp apparatus with two horizontal platforms, one at its top and one its bottom. Above each platform stood a canopy that prevented the experimenter from seeing what was on the platform. The platforms were connected by a slightly inclined horizontal ramp and were positioned such that a food object placed on the upper platform could potentially roll down the ramp and come to rest on the lower platform.

Figure 3. Depiction of the competition event used in Flombaum and Santos (2004): (A-B) The experimenter places two grapes at the top of the platform (grapes indicated by the red arrows), (C-D) The experimenter then sits back, releases a hidden switch and then allows one of the two grapes to roll down the ramp as the subject watches, (E-F) Subjects then have one full minute to approach one of the grapes; Here the subject chooses the top grape

During testing, the apparatus was positioned such that the subject, who was seated in front of the apparatus, could see a piece of food when placed on either of the two platforms, but the competitor, who was seated behind the apparatus, could not. This positioning set up a situation in which the subject knew exactly where the food was at all times, but the competitor, who lacked direct visual access, had to rely on his knowledge of where the food was originally placed. In one condition, the competitor began by placing two grapes at the top of the ramp. One of these grapes then appeared to secretly roll down the apparatus (in fact, its rolling was triggered covertly by the experimenter). Because the experimenter was unable to see the grape roll to its new position, he had no knowledge that it had moved. If subjects recognized that the competitor no longer *knew* where the grape that rolled was, then they should selectively try to

steal that grape rather than the one that was still in its original location, a location that the competitor knew about. When the competitor was behind an opaque barrier, subjects did just this; without training, macaques knew to approach the grape that the experimenter didn't know had rolled. In contrast, when no visual barrier was in place (and thus the experimenter had knowledge of the grape's rolling, see Figure 3 for this set-up), subjects approached the two grapes randomly, no longer showing a preference for the rolled grape. Again, like chimpanzees (Hare et al., 2001), our macaque subjects seem to take into account what competitors do and do not know about a food's location.

Overall, we believe that our recent studies together with the experiments of Hare and colleagues build a strong case for ToM reasoning in apes and monkeys. In these studies, primates behave as though they appreciate the mental states of others and take specific account of how one's eyes and the presence of visual barriers constrain the contents of mental states. We consider whether this is the best way to interpret these results in the next section. As we have emphasized previously, however, what seems to us the most valuable contribution of these experiments is the demonstration that animals who have previously performed abysmally in tasks requiring a certain set of skills—in this case, reasoning about where individuals are looking—suddenly perform excellently when the same skills are required in a different kind of task. As others have highlighted (see Hare, 2001; Tomasello et al., 2003), the most critical difference between past and present ToM tasks seems to involve this aspect of competition over scarce resources. Therefore, it appears that competition may come to play a central role in our understanding of the nature of ToM reasoning more generally, in both human and non-human primates. Additionally, as we discuss more in the final section, understanding why primates succeed in competitive tasks may shed some light on how the networks involved in human ToM

become engaged during our normal social interactions, and the reasons why these networks do not appear to function in clinical disorders such as autism (Baron-Cohen, 1995).

3. *Primate Theory of Mind: An Open-and–Shut Case?*

Thus far, we have tried to describe the state of the art in primate ToM experiments in as neutral representational terms as possible, but it will come as no surprise that we favor a rather generous view of the kinds of representations that these experiments endow to our closest living relatives. In particular, we believe that these competitive experiments establish that chimpanzees and rhesus monkeys are able to attribute perceptions, sometimes known as perspective taking, as well as knowledge to others. Thus, they represent what others *see* as well as what others *know*.

However, over the course of the last several years, a compelling deflationary alternative to this account has been put forward by Daniel Povinelli and his colleagues (Povinelli, 2000; 2004; this volume; Povinelli and Bering, 2002; Povinelli and Vonk, 2003; 2004). According to their view, what they call the "behavioral-abstraction" hypothesis, primates may appear to be insightful social thinkers in competitive experiments, but they do not actually represent *mental states*. Instead, primates represent the *behaviors of others* and the correlations among sets of behaviors, postures, and actions. Take for example rhesus monkeys' performance in one of our studies, successfully stealing food from the competitor facing forward versus one facing backwards (Flombaum and Santos, 2005, Experiment 1). The behavioral-abstraction hypothesis would claim that monkeys succeed by representing "his back is facing me, therefore he probably will not respond to my approach aggressively" as opposed to "his back is facing me, therefore he cannot *see* me." Primates, under this view, do not represent the mental states of others, but rather the likely future behaviors of others given their current behaviors and postures. To be sure, the behavioral-abstraction hypothesis is not behaviorism; Povinelli and colleagues *do* credit

non-human primates with the capacity for mental representations—but such representations only involve others' *behaviors*, not their *mental states*. Under their view non-human primates are not naïve psychologists, but instead naïve behaviorists.

At the core of our disagreement is the fact that, as Povinelli and colleagues correctly point out, almost any experiment can be described with both a mentalisitic account, such as our own, or their behavioral-abstraction account. Essentially, in the context of any experiment with a behavioral dependent measure, one can always replace "therefore, he cannot *see* me" or "therefore, he does not *know*" with a phrase such as "therefore, he probably will not respond aggressively." But before we look more closely at this issue, it should be clear that more is at stake in this debate than merely the semantically most appropriate way to describe some experiments that really appear justly described either way. What is at stake, as we see it, is the extent to which primates represent the events of these experiments *like humans*. As we emphasized in the introduction, the goal of this research enterprise is not only to discover whether primates possess a ToM; our goal is also to learn whether primates possess a ToM so that we can develop a neuroscientific animal model of our own abilities. According to the behavioral-abstraction hypothesis, this project is dead in the water, so to speak. Our own introspections and natural language make it clear that we humans can reason about unobservable mental states, not just about behavioral correlations. Primates, on the other hand, do not reason about mental states, and moreover, if they did, we couldn't really know it in the first place. Comparing our brains to theirs, then, viz. ToM, is like comparing apples and oranges.

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. The problem with the behavioral-abstraction hypothesis, as we see it, is that

the better primates perform in increasingly difficult and controlled experiments, the more the

behavioral-abstraction hypothesis becomes indistinguishable from a mentalistic account, and the

more it becomes implausible. As noted above, one can always substitute a mentalisitic term such

as 'he desires X' with a set of potential behaviors such as 'he will probably approach X.' This is

because the purpose of representing another individual's mental states is, after all, to accurately

predict their future behaviors. Thus, the chain of events in any assessment of another

individual's mental states includes the prediction of behavior: something like, "an occluder

blocks his view, therefore he cannot see, therefore does not know, and, therefore he will not

approach."

This is not to say that an animal could not succeed with a set of representations that

skipped mental states all together, and correlated behaviors directly to environmental cues,

including other behaviors. For example, an animal that only represents "an occluder blocks his

view, therefore he will not approach" is not representing mental states, though that animal would

obviously succeed in some experiments. But it becomes increasingly unlikely that this is *all*

animals are doing when the set of environmental cues that they can accurately correlate with

certain behaviors appears to be highly sophisticated and specific. That is, correlations only work

when one knows what features to use as units in the correlation. Most of the time, when an

animal's eyes point forward, their mouth points forward as well.   Therefore, we might expect

that a primate interested in knowing when someone else is going to approach them will believe

that approach is likely when either their eyes or their mouth point forward. Indeed, in Povinelli

and Eddy's (1996) famous begging experiments, chimpanzees appeared to fail for just this

reason: without direct training, chimpanzees exhibited no begging preference between someone

with a barrier covering her eyes compared to someone with a barrier covering her mouth, nor

between someone whose entire face, including their eyes, was oriented forward and someone whose eyes alone where oriented away. A 'simple' system of correlations is well suited for explaining errors of overgeneralization like the ones demonstrated by chimpanzees' in these earlier experiments.

But the animals tested in competitive paradigms more recently appear *not* to make errors of overgeneralization (see Hare et al., 2000; 2001; in press; Flombaum & Santos, 2005). Instead, they appear finely tuned to exactly the variables, postures, and behaviors in their environments that are good predictors of future behavior (surely because these features are factors that constrain mental states such as seeing and knowing.) This suggests that primates possess a specialized system that identifies and analyzes *only those factors that are relevant to problems of social reasoning*: what in essence boils down to a ToM system. In our experiments (Flombaum and Santos, 2004), monkeys correctly avoided an experimenter with an occluder over his eyes but not his mouth (Experiment 6), an experimenter with his eyes alone were averted to the side (Experiment 4), and even an experimenter whose profile faced the monkey, but whose eyes faced the contested food (compared to an experimenter with his profile to the monkey as well, but with his back to the contested food, see Experiment 2). If the monkey's goal is to predict which of the human competitors will not foil their plan to steal the contested food, then it must do so by correlating all of the human competitors' past patterns of approach with where the competitors eyes are pointing relative to 1) the monkey's current position, 2) the monkey's position at all stages along its approach to the food, and 3) a variety of things in the environment that could cover the competitor's eyes. This alone is an astounding feat, but even more astounding is the fact the monkeys successfully ignore many other correlated but irrelevant variables, like where one's mouth or nose is pointing; these features may have reliably correlated with the

competitor's approach in past encounters, but they are not causally-related to what the experimenter can see, and therefore do not actually *constrain* where he will approach. Identifying and then analyzing exactly the variables that will best predict future behavior (because they constrain mental states) is, functionally, what a system that is specialized for ToM does, and it is a system that, in our opinion, we appear to share with monkeys and apes.

No doubt, monkeys and apes, as well as humans, also possess the ability to compute massive correlations between current behaviors and future behaviors. In fact, we believe that it is the difference between a general system for computing statistical probabilities and a finely tuned ToM system that helps to explain a pattern of data that would otherwise appear inconsistent. Namely, *how can we account for the fact that primates tend to fail cooperative ToM tasks, but succeed in nearly identical competitive ones*? It is insufficient to explain these results merely by saying that the cooperative experiments do not, for whatever reason, engage the system that allows for success in the competitive tasks. Such an explanation neglects to specify the system that accounts for behavior in failed cooperative tasks. Thus, if the behavioral-abstraction hypothesis accounts for the success of primates in competitive experiments, as Povinelli and colleagues would have it, then what accounts for their failure in cooperative ones? Primates may behave incorrectly, even foolishly, in cooperative experiments, but not always randomly: chimpanzees know to beg from a human facing forward and not one facing backwards, they just don't know to distinguish between one with his eyes averted and one with his eyes forward (among other similarly subtle differences). This type of behavior, as we described previously, is precisely the symptom that one would expect of a system that computes nothing more than correlations. Indeed, after further investigation, Povinelli and Eddy (1996a) found that chimpanzees tended to beg for food from the experimenter that showed more 'frontal

aspect' (i.e., more surface area on the front of the body). The chimpanzees had, therefore, tried to correlate posture and behavior, but they failed the begging task simply because they correlated behavior to the wrong kind of posture. It's important to note, however, that if a system for correlating postures and behaviors successfully accounts for the stereotypical failures exhibited in cooperative experiments, then it cannot also account for the systematic successes exhibited in nearly identical competitive experiments. Essentially, the pattern of failures in cooperative experiments and successes in competitive experiments necessitates two sets of explanations, and therefore, *the use of two different cognitive systems.* We agree that behavioral-abstraction (or whatever system computes massive correlations between behaviors) likely accounts for the patterns of failures observed in primate cooperative experiments. It seems to us, however, that a different system—namely a ToM system—must then account for successes in competitive experiments.

Finally, we would like to acknowledge that the kinds of arguments we have put forward here in support of primates possessing a ToM rely upon revisiting a large suite of experiments that include primate failures as well their success. Therefore, they do not by any means constitute and open-and-shut empirical demonstration that animals, under unambiguous terms, possess the kind of theory of mind that we humans do. But these issues aside, we feel there is another important sense in which it can be said that primates possess a ToM. That is, in our opinion, they possess a ToM in the practical sense that they possess abilities that are clearly relevant studying human ToM neurally. This is the major focus of the next section of the paper. Our working hypothesis will be that whenever primates reason about the minds of others—whether by abstracting from their behaviors or with the use of a ToM system—they appear to engage the same brain areas that humans appear to use in similar situations (Allison et al., 2000). In

addition, they appear to reason about others in exactly the ways that autistic children seem

unable (see Baron-Cohen, 1995 for review). Thus, even if we cannot be sure of how monkeys

are representing the actions of others, we can be sure that we are studying systems in monkeys

that are highly relevant to understanding human ToM.

4. *Brains That Think About Minds and Brains That Don't*

We now turn our attention to a somewhat more practical and pressing issue in the study

of primate ToM. As reviewed above, recent competitive studies have provided some of the first

behavioral evidence that primates use the direction of an individual's eye gaze to make

predictions about that individual's knowledge and perceptions. What implications do these

findings have for our understanding of ToM neurally? Do competitive paradigms bring us any

closer to a better primate model of ToM abilities? In short, can these behavioral experiments

somehow help us to understand how brains come to think about minds?

As a first pass, it would seem that these new findings may help clarify data that are

already available on social processing in the primate brain, data that are in some cases now over

a decade old. Very elegant single-unit work by David Perrett and his colleagues (Jellema et al.,

2000; Jellema et al., 2002; Perrett et al., 1985; Perrett et al., 1990; Perrett and Mistlan, 1990), for

example, has shown that that the primate brain possesses some areas that are dedicated for the

processing of head and eye gaze.   This work suggests that cells in the rhesus monkey superior

temporal sulcus (STS) are sensitive to the position of another individual's eyes, often times

independent of that individual's head and body orientation. Perrett and colleagues originally

interpreted the function of these areas as encoding where another individual's eyes were pointed.

Functional imaging studies in humans (reviewed by Saxe et al., 2003), however, have

demonstrated that the human STS is sensitive to more than just eye position; in contrast to the

single unit data, the human STS seems to represent goal directed actions and perceptions—what other individuals *want* and *see*. Using fMRI techniques, Pelphrey and colleagues, for example, have observed that that the STS is modulated both by the context of where another person is looking (Pelphrey et al., 2003) and that person's perceived goal (Pelphrey et al., 2004). Observing such an STS function in humans make sense in light of the macaque single unit findings; because the STS is able to disambiguate head direction and eye orientation (see Perrett et al., 1990), it seems ideally suited for the further process of computing the content of what another individual sees. But could this neural area be encoding something similarly complex in macaques as well? Previously, few researchers would have thought so. Nonetheless, our behavioral experiments with macaques, the very animals studied in this physiology work, make it clear that this species is capable of representing more than just where others are looking—they too seem to represent what others see. For this reason, it is possible that the macaque STS is also capable of encoding what others see, and therefore is necessary to mediate the decisions that macaques successfully make in our competition experiments, decisions about which competitor to avoid. Thus, the macaque STS may do more than simply record where eyes are pointed. This region may generate representations of *seeing* and *not seeing* that can be used in task relevant ways by the organism. As such, though it remains uncertain whether the monkey STS actually 'does ToM,' it is likely that this system provides as an important foundation for sophisticated social reasoning, and therefore, can serve as a site for understanding the neural basis of such reasoning even in a primate model.

Auspiciously, the capacity to detect what others can and cannot see is a capacity ripe for cognitive neuroscientific study—it is one of a suite of ToM abilities lacking in autism, and therefore one for which an animal model would be most beneficial. In addition to their well-

known problems representing the mental states of others (Baron-Cohen et al., 1985; Perner et al., 1989), autistics also have a well-documented deficit in attending to and interpreting eye gaze (see Baron-Cohen, 1995 for review), even at a young age. (Charman et al., 1997). Similarly, autistic children rarely use proto-declarative gestures intended to change another individual's direction of gaze (Baron-Cohen, 1989). Beyond just attending to gaze, however, autistics seem to have problems interpreting the meaning of one's gaze—in particular, that it affects a person's mental states. For example, Baron-Cohen and colleagues (1995) found that autistic children do not use the direction of an individual's gaze to determine what that individual wants or knows, even though they succeed in following gaze, and can accurately determine where an individual is looking. Rhesus macaques, on the other hand, seem to do both. In our experiments, for example, macaques relate the orientation of a competitors gaze to his future dispositions with respect to the grape—more than autistics seem to do. For these reasons, a primate model of STS function may provide important insights into the reasons for cognitive deficits in autism. Such a model may even be necessary. The subtle cognitive impairments observed in autism likely reflect deficits at the neural level that are similarly subtle—some autopsy studies even suggest, perhaps involving impairments in a select population of neurons or even specific varieties of cells (Courchesne, 1997). Neural investigations at such a detailed level require a physiological model.

Rhesus monkeys not only suffice as such a model, but their idiosyncratic patterns of performance in ToM experiments might make them especially well suited for this charge. This is because their failure in cooperative ToM experiments juxtaposed with their success in nearly identical competitive ones affords the chance to understand not only what makes their social reasoning skills work, but also what makes them not work. At the heart of the logic of cognitive

neuroscience is the importance of dissociations, and in rhesus monkeys we seem to have a natural dissociation between reasoning about eye gaze in competitive experiments and tracking it in cooperative ones. Such a dissociation behaviorally may provide some clues to understanding why autistic children, as described above, seem similarly able to track gaze, but not to reason about it.

Moreover, this dissociation in monkeys between reasoning and tracking in different contexts might naturally map onto a dissociation between different neural systems. One outstanding question in the study of human ToM abilities, for instance, concerns the role that the amygdala plays in reasoning about eye gaze (see Baron-Cohen & Ring, 1994). It has been established both in monkeys (Leonard et al., 1985) and humans (Kawashima et. al., 1999) that the amygdala shows sensitivity to eye gaze orientation. Moreover, amygdala damage results both in a generalized impairment to reasoning about mental states and in reasoning about gaze in particular (Fine et. al., 2001). Likewise, removal of the amygdala results in an impairment in the perception of gaze direction (Young et. al., 1995). Nevertheless, several brain areas including the STS are sensitive to eye gaze prior to the amygdala in the processing of visual information (see Brothers, 1990). Because of this, neuroscientists are still unclear about the unique role that the amygdala plays in processing eye gaze information. One possibility is that the amygdala serves to amplify STS responses to eye gaze (Leonard et. al., 1985, Allison, et al. 2000). Such a view fits with some known facts about the amygdala including its function in binding emotional content with sensory information (Adolphs et al., 1999). Thus the amygdala may provide the critical link in reasoning not only about where another individual is looking, but what they *see* as well (see Baron-Cohen and Ring, 1994).

Indeed, primates may fail in cooperative versions of ToM because these kinds of contexts fail to engage the amygdala emotionally, although competitive interactions obviously do. The amygdala is known to become engaged when monkeys participate in competitive situations. Moreover, monkeys with amygdalectomies show little understanding of their own place in the social dominance hierarchy (Kling et al., 1970), suggesting again the importance of this structure in negotiating competitive interactions. For these reasons, a monkey viewing stimuli in a competitive experiment should evidence a highly active amygdala interacting with a highly active STS. A monkey viewing the same stimuli in cooperative experiments might show no such activation in the amygdala. Perhaps this lack of amygdala activity in cooperative studies is at the root of poor performance. In fact, we would hypothesize that monkeys with amygdala lesions should demonstrate an ability to *follow eye gaze*, but should be unable to use this information correctly in competitive paradigms like the one that we have developed. Consequently, they should fail to favor a competitor who cannot see them when selecting which contested food item to approach.

We would like to stress here that dissociating the functions of the amygdala and STS is only one among many neural dissociations to be made if we are to understand how the human ToM network works. Similar experiments could afford a better understanding of how face perception and reasoning about eye gaze may rely on subtly different cognitive and neural resources (Campbell et. al., 1990; Allison et al., 2000), as well as the role of executive functions and regions of frontal cortex in ToM (Fine et. al. 2001; Stuss et. al., 2001). Overall, the availability of a sensitive behavioral paradigm with a laboratory primate allows for investigations into the neural substrates of ToM with a degree of sensitivity unavailable to imaging methods with humans.

5. *Conclusions*

In 1978, Premack and Woodruff posed the question of whether the chimpanzee had a theory of mind. Several decades later, their question has sparked not only rich comparative work on social cognition in primates, but also tremendous interest and empirical progress in other areas of cognitive neuroscience as well.   In this chapter, we have tried to argue that we can now go beyond their original question. The evidence we have reviewed, incorporating both experimental successes and failures, supports the view that we share a number of ToM capacities with at least some of our primate relatives. With this evidence in place, cognitive neuroscientists together with comparative psychologists are poised to begin asking more detailed questions about the mechanisms underlying social cognition, and potentially treating disorders involving these mechanisms. It is clear to us that only a broad interdisciplinary perspective— like the one advocated in this volume— can provide a richer picture of how brains of all kinds succeed in representing the minds of others.

References

Adolphs, R., Russell, J. A., Tranel, D. (1999). A role for the human amygdala in recognizing emotional arousal from unpleasant stimuli. Psychological Science, 10. 167-171.

Allison, T., Puce. A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. Trends in Cognitive Sciences 4, 267–78.

Amaral, D. G. (2002). The primate amygdala and the neurobiology of social behavior: Implications for understanding social anxiety. Biological Psychiatry, 51. 11-17.

Anderson, J. R. Montant, M., & Schmitt, D. (1996). Rhesus monkeys fail to use gaze direction as an experimenter-given cue in an object-choice task. Behavioural Processes, 37. 47-55.

Anderson, J. R., Sallaberry, P. & Barbier, H. (1995). Use of experimenter-given cues during object-choice tasks by capuchin monkeys. Animal Behaviour, 49. 201-208.

Baron-Cohen, S. (1989). Perceptual role-taking and proto-declarative pointing in autism. British Journal of Developmental Psychology, 7. 113-127.

Baron-Cohen, S. (1995). Mindblindness: Essay on Autism and the Theory of Mind. Cambridge: MIT Press.

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., & Grant, J. (1995). Are children with autism blind to the mentalistic significance of the eyes? British Journal of Developmental Psychology, 13. 379-398.

Baron-Cohen, S. Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? Cognition, 21. 37-46.

Baron-Cohen, S. & Ring, H. (1994). A model of the mindreading system: Neuropsychological and neurobiological perspectives. In C. Lewis & P. Mitchell, Peter (Eds). (1994). Children's Early Understanding of Mind: Origins and Development. Hillsdale, NJ: Erlbaum. pp. 183-207.

Brothers, L. (1990). The neural basis of primate social communication. Motivation & Emotion, 14. 81-91.

Byrne, R. W. & Whiten, A. (1988). Machiavellian Intelligence: Social Expertise and The Evolution of Intellect in Monkeys, Apes, and Humans. Oxford: Oxford University Press.

Call, J., Agnetta, B., Tomasello, M. (2000). Cues the chimpanzees do and do not use to find hidden objects. Animal Cognition, 3. 23-34.

Call, J., Hare, B. A., & Tomasello, M. (1998). Chimpanzee gaze following in an object-choice task. Animal Cognition, 1, 89–99.

Campbell, R. Heywood, C. A., Cowey, A. Regard, M. and Landis, T. (1990). Sensitivity to eye gaze in prosopagnosic patients and monkeys with superior temporal sulcus ablation. Neuropsychologia, 28, 1123-1142.

Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., & Drew, A. (1997). Infants with autism: An investigation of empathy, pretend play, joint attention, and imitation. Developmental Psychology, 33. 781-789.

Cheney, D. L. & Seyfarth, R. M. (1990). Attending to behaviour versus attending to knowledge: Examining monkeys' attribution of mental states. Animal Behaviour, 40. 742-753.

Cosmides, C. & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld & S. A. Gelman, (Eds). Mapping the Mind: Domain Specificity in Cognition and Culture. New York: Cambridge University Press. pp. 85-116.

Courchesne, E. (1997). Brainstem, cerebellar and limbic neuroanotomical abnormalities in autism. *Current Opinion in Neurobiology, 7*, 269-278.

Fine, C., Lumsden, J., and Blair, J. R. (2001). Dissociation between 'theory of mind' and executive functions in a patient with early left amygdala damage. Brain, 124, 287-298.

Flombaum, J. I., Kundey, S. M., Santos, L. R., & Scholl, B. J. (2004). Dynamic object individuation in rhesus macaques: A study of the tunnel effect. *Psychological Science, 15,* 795-800.

Flombaum, J. I. & Santos, L. R. (2004). What rhesus monkeys (Macaca mulatta) know about what others can and cannot see. Symposium presented at the 14th Biennial Meeting of the International Society on Infant Studies, Chicago, Illinois.

Flombaum, J. I. & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. Current Biology. In press.

Gallagher, H. L. & Frith, C. D. (2003) Functional imaging of 'theory of mind'. Trends in Cognitive Sciences, 7. 77-83.

Hare, B. (2001). Can competitive paradigms increase the validity of experiments on primate social cognition? Animal Cognition, 4. 269-280.

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.

Hare, B., Call, J., and Tomasello, M. (in press). Chimpanzees deceive a human competitor by hiding. Cognition.

Hauser, M. D. (2001). Searching for food in the wild: a nonhuman primate's expectations about invisible displacement. Developmental Science, 4. 84-93.

Hauser, M. D., Carey, S. and Hauser, L. B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. Proceedings of the Royal Society of London: Biological Sciences, 267. 829-833.

Humphrey, N. K. (1976). The social function of intellect. In P. B. G. Bateson & R. A. Hinde (Eds.) Growing Points in Ethology. Cambridge: Cambridge University Press. 303-317.

Heyes, C. M. (1998). Theory of mind in nonhuman primates. Behavioral & Brain Sciences, 21. 101-134.

Itakura, S. (1996). An exploratory study of gaze-monitoring in nonhuman primates. Japanese Psychological Research, 38. 174-180.

Itakura, S., Agnetta, B., Hare, B., and Tomasello, M. (1999). Chimpanzee use of human and conspecific social cues to locate hidden food. Developmental Science, 2. 448-456.

Itakura, S. and Anderson, J. R. (1996). Learning to use experimenter-given cues during an object-choice task by a capuchin monkey. Cahiers de Psychologie Cognitive, 15. 103-112.

Kawashima, r. Sugiura, M., Kato, T.,, Nakamura, A., Hatano, K., Ito, K., Fukuda, H., Kojima, S., and Nakamura, K. (1999). The human amygdala plays an important role in gaze monitoring: A PET study. Brain, 122, 779-783.

Kling, A., Lancaster, J., & Benitone, J. (1970). Amygdalectomy in the free-ranging vervet. Journal of Psychiatric Research, 7. 191-199.

Leonard, C. M., Rolls, E. T., Wilson, F. A. W., and Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. Behavioural and Brain Research, 15, 159-176.

Machado, C. J. & Bachevalier, J. (2003). Non-human primate models of childhood psychopathology: The promise and the limitations. Journal of Child Psychology & Psychiatry & Allied Disciplines, 44. 64-87.

Mitchell, R. W. (1999). Deception and concealment as strategic script violation in great apes and humans. In S. T. Parker & R. W. Mitchell (Eds.) The Mentalities of Gorillas and Orangutans: Comparative Perspectives. 295-315.

Peignot, P., & Anderson, J. R. (1999). Use of experimenter-given manual and facial cues by gorillas (Gorilla gorilla) in an object-choice task. Journal of Comparative Psychology, 113, 253–260.

Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: The influence of context. Neuropsychologia, 41.156-170.

Pelphrey, K.A., Viola, R. J., & McCarthy, G. (2004). When strangers pass: Processing of mutual and averted social gaze in the superior temporal sulcus. Psychological Science, 15. 598-603.

Perner, J., Frith, U., Leslie, A. M., & Leekham, S. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. Child Development, 60. 689-700.

Phillips, W. and Santos, L. R. (in submission). How rhesus monkeys (Macaca mulatta) reason about the insides of objects. Cognition.

Povinelli, D. J. (2000). Folk Physics for Apes. Oxford: Oxford University Press.

Povinelli, D. J. (2004). Behind the ape's appearance: Escaping anthropomorhism in the study of other minds. Daedalus: Journal of the American Academy of Arts and Sciences , Winter, 29-41

Povinelli, D. J. and Bering, J. M. (2002). The mentality of apes revisited. Current Directions in Psychological Science, 11. 115-119.

Povinelli, D. J., Bierschwale, D. T., and Cech, C. G. (1999). Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. British Journal of Developmental Psychology, 17. 37-60.

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

Povinelli, D. J. and Eddy, T. J. (1996b). Chimpanzees: Joint visual attention. Psychological Science, 7. 129-135.

Povinelli, D. J. and Eddy, T. J. (1996c). Factors influencing young chimpanzees'(Pan troglodytes) recognition of attention. Journal of Comparative Psychology, 110. 336-345.

Povinelli, D. J. and Eddy, T. J. (1997). Specificity of gaze-following in young chimpanzees. British Journal of Developmental Psychology, 15. 213-222.

Povinelli, D. J., Parks, K. A. & Novak, M. A. (1991). Do rhesus monkeys (Macaca mulatta) attribute knowledge and ignorance to others? Journal of Comparative Psychology, 105. 318-325.

Povinelli, D. J. & Vonk, J. (2003). Chimpanzee minds: suspiciously human? Trends in Cognitive Sciences, 7. 157-160.

Povinelli, D. J. & Vonk, J. (2003). We don't need a microscope to explore the chimpanzee mind. Mind and Language, 19. 1-28.

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

Santos, L. R. (2004). Core knowledges: A dissociation between spatiotemporal knowledge and contact mechanics in a non-human primate? Developmental Science. In press.

Santos, L. R., Sulkowski, G. M., Spaepen, G. M., & Hauser, M. D. (2002). Object individuation using property/kind information in rhesus macaques (Macaca mulatta). Cognition, 83. 241-264.

Santos, L. R., Hauser, M. D., & Spelke, E. S. (2001). Representations of food kinds in the rhesus macaques (Macaca mulatta): An unexplored domain of knowledge. Cognition, 82. 127-155.

Sulkowski, G. M & Hauser, M. D. (2001). Can rhesus monkeys spontaneously subtract? Cognition, 79. 239-262.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. Annual Review of Psychology, 55. 87-124.

Schultz, R. T. Romanski, L. M. & Tsatsanis, K. D. (2000). Neurofunctional models of autistic disorder and Asperger syndrome: Clues from neuroimaging. In A. Klin and F. R. Volkmar (Eds.) Asperger Syndrome. New York: Guilford Press. 172-209.

Siegal, M. & Varley, R. (2002). Neural systems involved in theory of mind. Nature Reviews Neuroscience, 3, 462-471.

Stuss, D. T., Gallup Jr, G. G. and Alexander, M. P. (2001). The frontal lobes are necessary for 'theory of mind.' Brain, 124, 279-286.

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

Tomasello, M., Call, J., & Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. Animal Behaviour, 55. 1063-1069.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states—the question is which ones and to what extent. Trends in Cognitive Sciences, 7. 153-156.

Tomasello, M., Hare, B., and Agnetta, B. (1999). Chimpanzees follow gaze geometrically. Animal Behaviour, 58. 769-777.

Tomasello, M., Hare, B., & Fogleman, T. (2001). The ontogeny of gaze following in chimpanzees and rhesus macaques. Animal Behaviour, 61. 335-343.

Whiten, A. & Byrne, R. W. (1997). Machiavellian Intelligence II: Extensions and Evaluations. Cambridge: Cambridge University Press.

Young, A. W., Aggleton, J. P., Hellawell, D. J., Johnson, M., Broks, P. & Hanley, J. R. (1995). Face processing impairments after amygdalotomy. Brain, 118. 15-24,