©2015

Frank Richard Batiste

ALL RIGHTS RESERVED

Theory of Mind and the Role of Target Individuals' Group Affiliation

By

Frank Richard Batiste

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Anthropology

Written under the direction of

Lee Cronk

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2015

ABSTRACT OF THE DISSERTATION

Theory of Mind and the Role of Target Individuals' Group Affiliation

by FRANK RICHARD BATISTE

Dissertation Director:

Lee Cronk

Theory of Mind (ToM) is the ability to interpret the behavior of others in terms of underlying mental states such as beliefs, wants, desires (Premack and Woodruff, 1978). The simulation theory of ToM claims that an individual replicates, or mirrors, the assumed mental states of a target individual and processes them using his/her own mental architecture--the same architecture that is used to make decisions based on one's own beliefs, desires, or thoughts. Thus, ToM may be considered as a form of empathy, a process where the perception of a target's state generates a state in the observer that is more applicable to the target's situation than to the subject's own prior situation (Preston and de Waal, 2002). The experience of emotional empathy is influenced by coalitional cues such as familiarity (Liew, Han, and Aziz-Zadeh, 2011), similarity (Xu, Zuo, Wang, and Han, 2009), and shared group membership (Avenanti, Sirigu, and Aglioti, 2010), as well as immediate situational cues such as the color of a target's tee shirt (Kurzban, Tooby, and Cosmides, 2001), or simply referring to a counterpart in a task as a partner or opponent (Burnham, McCabe, and Smith, 2000). To date, the effect of such immediate coalitional cues has not been tested for ToM. In the present study, a ToM task was

designed to test subjects' perspective taking ability in response to one of three different conditions, a neutral frame, a cooperative frame, or a competitive frame. Two types of perspective-taking errors were recorded: incorrect responses and response hesitations. It was predicted that subjects would 1) make significantly fewer errors on the task in the cooperative frame relative to the other two conditions, and 2) make significantly more errors in the competitive condition. Partial support of these predictions was found. ToM was sensitive to cues of coalition, but only for one type of error, hesitations. While cooperative and competitive conditions were marginally significantly different from each other in the expected direction (subjects in the cooperative frame made fewer perspective taking errors than subjects in the competitive frame), neither differed significantly from the control condition.

Acknowledgements

This project was made possible through the assistance and support of many people.

Thank you to my advisor and committee chair, Lee Cronk, and to the other committee members: Susan Cachel, Ryne Palombit, and Stephen Stich. Your guidance along the way was invaluable.

Thank you also to Michelle Night Pipe and everyone in the “EPC” Lab Discussion Group, both past & present, for your feedback and encouragement.

This project would not have been possible without assistants to act as experimental confederates. Robert Huseby & Marley Doring were better than I could ask for.

And last, thank you to the Center for Human Evolutionary Studies for the financial support necessary to carry this project out to completion.

Dedication

To Cindy and Lauren,

For your patience, love, and support.

Table of Contents

List of Tables

List of Illustrations

Chapter 1: Introduction

The Importance of Theory of Mind

Humans possess an impressive array of complex cognitive skills and abilities, including tool use, language, mathematics, and cumulative culture. Among these myriad abilities is one that appears so simple in comparison that it hardly seems to be a skill worth mentioning at all. Yet it is a vital one. This skill, Theory of Mind (ToM, also referred to as mindreading, mentalizing, or folkpsychology in the literature), is the ability to think about things that from an everyday perspective are not only invisible, but do not appear to exist in the physical world at all: minds, thoughts, beliefs, desires[1]. It is the ability to impute these mental states to oneself and to others, to interpret observed behavior in terms of underlying beliefs and desires, and to understand that others can have such mental states that differ from one's own (Premack and Woodruff, 1978). Without it we might not be the social species that we are, we might lack religion as we know it, we might even lack the capacity for cumulative culture. Not only can we think about these things, but we regularly invoke them to explain our behavior, and we also understand that others do the same.

While the term itself implies that ToM is something one *has*, it is clear from the above definition that it should be thought of as an action, something one *does*: one *imputes* mental states and *interprets* others' behavior in terms of underlying motives, beliefs or desires. For this reason, terms such as mindreading or mentalizing, as verbs, better capture the essence of the definition (mentalizing more so, due to unfortunate connotations of mindreading). However, Theory of Mind remains the most commonly

[1] This should not be taken as an endorsement of mind-brain dualism. From a neuroanatomical perspective, of course, these things are all products or processes of brain activity, and are indeed real, physical phenomena.

used, and so I use it herein.

From an evolutionary perspective, the ability to make mental state attributions about others' beliefs is certainly a key factor in humans' ability to form complex and extensive social relationships. It is easy to imagine the selective advantages ToM could confer on members of a social species in cooperative or competitive interactions, and surely it must have been an important factor in human evolution. According to Atran and Norenzayan (2004) ToM is a necessary skill for navigating the social world providing an obvious adaptive advantage to our ancestors in that it allowed them to easily distinguish friends from enemies. The ease with which we infer mental states is suggestive of this importance. The sensitivity of out ToM system is dependent on a so-called hyperactive agency detection system (HADD) (Barrett, 2000; Barrett and Lanman, 2008). The HADD makes it is easy, even automatic, to assume that causal agents that we do not necessarily see are behind occurrences and that those agents are acting with purpose—they are not acting randomly, but have motivations guiding their actions. We readily detect motion indicative of animacy from minimal cues (Tremoulet and Feldman, 2000; Pinto and Shiffrar, 1999) and interpret that motion in terms of underlying emotions (Dittrich, 1993; Dittrich, Troscianko, and Lea, 1996) and other mental states, even when the observed targets are animations of shapes that clearly do not possess a mind to read (Heider and Simmel, 1944; Abel, Happe, and Frith, 2000). Autistic individuals, for whom impaired ToM is a diagnostic symptom, also show a concomitant impairment in this ability as well (Castelli, Frith, Happe, and Frith, 2002).

One major way in which ToM has provided an adaptive advantage is through its key role in religion (Atran and Norenzayan, 2004), which in turn, has been an important

factor in the formation of large-scale societies. The Byproduct Theory of religion (Atran and Norenzayan, 2004) seeks to explain religion by focusing on the widespread features of religions that might be best explained by looking to other, established features of human cognitive processes. ToM has been invoked to explain two of these features: 1) many religions are dualistic, positing some type of immaterial soul or spirit separate from the body, and 2) gods are seen as having human-like minds and like us, they have beliefs, wants, and desires. In addition, Barrett argues (Barrett, 2000; Barrett and Lanman, 2008) the ToM system's default setting is the attribution of true beliefs to others. Bloom (2007), in turn, argues that religion naturally arises from this system. We cannot see nor feel other minds, nor can we see or feel others' mental states, but the ToM system, however, allows us to think about these things. At the same time, we also possess a separate mental system which we use to think about real, physical objects. Given these two separate systems, it is natural to conclude that the mind exists independently from the body/physical world and that some immaterial aspect of ourselves continues to exist after bodily death. Without ToM, this type of religious belief would not be possible. A recent paper lends support to this connection between ToM and religion. Across a series of studies, Norenzayan, Gervais, and Trzesniewski (2012) tested the relationship between ToM deficits and belief in a personal God. They found that autism spectrum is correlated with reduced belief in God, and that ToM deficits (not other symptoms or associated personality traits) are responsible for this relationship.

So then, ToM is a crucial component in religion and belief in gods. Religion, in turn, may be a crucial component in our transition from small hunter-gatherer bands to large, complex societies (Shariff, Norenzayan, and Henrich, 2010). More specifically,

religions with “high gods,” omniscient, morally concerned policing agents, are highly correlated with group size across cultures, and belief in such gods may lead to increased prosocial behavior (Shariff and Norenzayan, 2008), allowing for increased group size. Without religion, prosociality, and the capacity to form large, complex groups, our species would look quite different than it does today.

ToM may also be responsible for the emergence of large-scale societies in another way. Any given religion is a collection of beliefs and practices that must be learned from others in one's society. It is not the product of individual learning, rather it is culturally learned—it is socially transmitted information (Boyd and Richerson, 1985). While social learning is not limited to humans (Laland and Hoppit, 2003; Galef and Laland, 2005; Whiten, Goodall, McGrew et al., 1999; van de Waal, Borgeaud, and Whiten, 2013; Laland and Williams, 1997; Laland and Plotkin, 1990; Baptista and Petrinovich, 1984), we routinely engage in a sophisticated degree of cultural learning that far exceeds that seen in any other species—we acquire a vast amount of beliefs, desires, practices, preferences, and skills from others. All this information is transmitted from their brains into our own (Henrich, 2014). And given the complexity of information and knowledge required to survive day-to-day life in any society, we must rely on cultural learning over individual learning for the vast majority of it (Henrich, 2014). Our capacity for cultural learning may be due, in part, to ToM (e.g., Herrmann, et al., 2007; Tomasello, 2011): beliefs, desires and other mental states are culturally learned—need to have a way of correctly inferring those beliefs and desires are. Likewise, in learning a skill from another person, having an understanding of their goal or intent in performing the component actions can lead to better learning of that skill. And just as we are prone to over-

attribution of mental states, a related result (or side effect) of ToM's involvement in skill learning is that we are also prone to over-imitate, in that we tend to faithfully copy all component actions or steps in an observed process whether they directly relate to the end goal (Neilsen and Tomaselli, 2010). This is a skill our non-human primate relatives lack (Horner and Whiten, 2005), which may further explain the great degree of cumulative cultural learning seen in humans.

A great deal of the research on ToM has been primarily developmental or comparative in nature and the questions asked in those studies fall within the realm of psychology and cognitive science: When do children begin to make mental state attributions about others? What other cognitive abilities assist in our ability to do so? Why are autistic individuals impaired in this area? Do non-human primate species have ToM? A common theme underlying these questions is that they view an individual's ToM as lying on a continuum, from lacking (or severely impaired) to possession of full-fledged ToM. These are certainly important research questions, yet there is another question that may (and should) be asked: Regardless of where one might be on this continuum, is ToM is expressed in the same way regardless of situation, context or target individual? In answering this question, we can move away from viewing ToM in a more binary fashion (as something one does or does not have) and increase our understanding of this cognitive ability and its role in the evolution of human (and non-human?) sociality. This is the broader question I address in this study. More specifically, I examine whether our ability to accurately take another person's perspective is susceptible to cues of group membership, that is, whether a target individual and the observer are members of the same or different groups.

Overview of the Paper

To accomplish this, I review the literature and concepts related to ToM and empathy in order to argue that ToM is best conceptualized as a type of empathy. Chapter 2 begins by expanding on the definition and theories of ToM and explaining its relation to empathy via the Simulation Theory of ToM (Goldman, 2006). Next, the importance of false belief tasks in understanding this skill is discussed. These tasks are designed to assess one's understanding of others' potentially false beliefs, i.e., that they can have beliefs that differ from one's own and from reality (Dennett, 1978, though see Bloom and German, 2000). Such tasks including the Displaced Object (Wimmer and Perner, 1983), Unexpected Contents (Hogrefe, Wimmer, and Perner, 1986), and Surprising Object (Perner, Leekam, and Wimmer, 1987) tasks all measure whether subjects can 1) view them from another individual's perspective and 2) understand that that individual holds a belief contrary to what the subject knows to be reality.

Some false belief tasks, like those described above, may rely on elicited responses, while others remove the need to consciously demonstrate understanding and instead measure spontaneous responses (Baillargeon, Scott, and He, 2010). While explicit tasks reveal that false belief understanding emerges in children between four and five years of age (Wellman, Cross, and Watson, 2001) implicit methods have pushed the age of understanding back. Three-year-olds (Clements and Perner, 1994), 15-month-olds (Onishi and Baillargeon, 2005), and even 13-month-old infants (Surian, Caldi, and Sperber, 2007) demonstrate false belief understanding as well.

Next, Chapter 2 discusses some of the biological/neurological bases underlying ToM that lend support to the claim that ToM is a type of empathy. In particular, the

temporoparietal junction (TPJ) plays an important role in belief attribution, and appears to be active when thinking about one's own as well as others' beliefs (Samson, Apperly, Chiavarino, and Humphreys, 2004; Saxe and Kanwisher 2003; Saxe, Carey, and Kanwisher, 2004). This is followed by a discussion of the role language plays in ToM as well as the extent to which ToM may be considered a core element of cognition independent of language, along with a brief review of the literature testing the existence of ToM in non-human primates. And last, Chapter 2 ends with an argument for the inclusion of another group to study, in addition to children, autistic individuals, and non-human primates: Adults (Apperly, Samson, and Humphreys, 2009). It is important to understand the role ToM plays, not only as a developmentally emerging ability in children, but also in cooperative interactions between adults. Adult subjects can provide us with informative data that cannot be acquired from the study of children.

Chapter 3 reviews empathy and focuses on its “mirroring” aspect: Empathy is any process where observing a target’s state generates a similar state in an observer that is more applicable to the target’s situation than to the observer's own prior state or situation (Preston and de Waal, 2002): when an individual observes another, some aspect of the target’s internal state is replicated, or mirrored, within the observer. This is the case for both motor level and emotional level empathy. For example, viewing images of another in physical pain activates brain regions in the observer that are involved in processing one's own pain (Jackson, Meltzoff, and Decety, 2005). And, if we accept the theory of ToM as a simulation, this is also the case for ToM.

The mirror neuron (Gallese, Fadiga, Fogassi, and Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, and Fogassi, 1996) and mirror systems (Rizzolatti and Craighero, 2004)

in the brain appear to be the underlying mechanisms allowing this representation of others' mental states in our own brains to be possible. There are individual neurons as well as brain regions that activate when an individual performs an action or experiences an emotion or other mental state and also when observing another individual performing the same action or expressing the same mental state. In particular, the proposed function of action perception (Wilson and Knoblich, 2005) is that mirror neurons/systems are for the perception of the behavior of conspecifics and allow individuals to draw inferences about the motives or purposes of others' actions. Next, variation in empathic responses and the sources of that variation are discussed, including factors such as motivation of the empathizer, race of target, familiarity with target, target's past behavior. To aid in the argument that ToM is empathy—one of three, along with motor mimicry and emotional empathy (Blair, 2005; Preston and de Waal, 2002) similar sources of variation are reviewed (e.g., Baron-Cohen, Wheelwright, Skinner, Martin, and Clubley, 2001).

Again, research with adults sheds light on this aspect of ToM. Rather than remaining a fully developed skill throughout life, ToM tends to decline into old age, more so than other age-related cognitive losses (Cavallini, Lecce, Bottiroli, Palladino, and Pagnin., 2013). And there is evidence that engaging in cooperative tasks activates brain regions implicated in ToM (Elliott, Völlm, Drury, McKie, Richardson, and Deakin, 2006), that ToM is positively correlated with cooperative traits, but negatively correlated with "Machiavellian" traits (Paal and Bereczkei, 2007).

Next, in Chapter 4 I review the literature arguing that our coalitional psychology evolved to be flexible. Empathy is sensitive to variations across and within individuals and contexts: Previous research has shown differential effects for emotional and motor

empathy (e.g., Liew, Han, and Aziz-Zadeh, 2011; Xu, Zuo, Wang, and Han, 2009) in which subjects favor in-group members over out-group members. The degree to which humans experience empathy towards others is influenced by many factors, including familiarity with the target individual and group membership (e.g., Cikara and Fiske, 2011; Gutsell and Inzlicht, 2010). Happiness and sadness elicit more empathy compared to anger and shame (Duan, 2000). Singer et al. (2006) found observers' gender and targets' previous behavior (playing an economic game fairly or unfairly) have a differential effect on empathizing.

Race is another factor that affects empathy. Xu, Zuo, Wang, and Han (2009) found that subjects showed increased activations in pain processing- and empathy-related brain areas when viewing members of their own race in pain, but not other races. But this discrimination is also modulated by attitudes towards the out-group. Avenanti, Sirigu, and Aglioti (2010) found similar results, but also that the response was tempered by subjects' level of implicit racial bias as well as preexisting cultural biases. painful situations. This rightly suggests that group boundaries are not necessarily fixed. In fact, coalitions form and dissolve as need arises. Feelings of group membership can be experimentally induced by merely informing subjects that their responses on a questionnaire place them in one of two categories (Tajfel, Billig, Bundy, and Flament, 1971); subjects then show a clear pattern of in-group vs. out-group favoritism in subsequent tasks, even without ever meeting their fellow members. Levine, Prosser, Evans, and Reicher (2005) found when subjects were primed to think about their group affiliation (in terms of favorite soccer team), they were more likely to aid an apparently injured person if they wore the same team's shirt, but not a rival team's. Yet when primed

to think of themselves more broadly as soccer fans, subjects aided rivals, but ignored others wearing generic sports apparel. Even divisions based on seemingly permanent traits such as race are not fixed (Kurzban et al., 2001). While race may serve as quick heuristic for group membership, Kurzban and his colleagues found that divisions along such lines can easily be overridden by other cues that are reliably associated with coalition membership, such as colored tee-shirts worn by others. They argue that there is no evolved racism module, since it was highly unlikely one would encounter anyone that differed that drastically in physical appearance in our environment of evolutionary adaptation (EEA). But there were different groups, some of which would be enemies, and so there was (and is) a need to identify others' group membership.

Finally, Chapters 5 and 6 present the study, the results and discussion of this project. To review the foundations upon which my hypothesis rests: Theory of Mind (ToM) is the ability to interpret others' behavior in terms of underlying mental states (Premack and Woodruff, 1978). The simulation theory of ToM proposes that we simulate ToM may be thought of as a type of empathy in that we take others' minds in our own in order to understand or predict their actions (Goldman, 2006). The degree to which humans experience emotional and motor empathy towards others is influenced by many factors, including familiarity and group membership (e.g., Cikara and Fiske, 2011; Gutsell and Inzlicht, 2010) and is particularly sensitive to immediate cues of coalition (Kurzban et al., 2001, Levine, Prosser, Evans, and Reicher, 2005). In this study, I investigate the effect that our perceptions of a target individual's group membership has on one's subsequent ability to make accurate Theory of Mind (ToM) attributions about that individual: Does ToM show a pattern of variation similar to motor and emotional

empathy? Though firmly established by five years of age (Wellman, Cross, and Watson 2001), ToM is prone to variation between individuals (e.g., Baron-Cohen et al., 2001), and within individuals: Adults can make errors similar to those made by young children (Keysar, Lin, and Barr, 2003), and ToM declines as adults age (Cavallini et al., 2013). But are these variations influenced by our flexible coalitional psychology? Can immediate coalitional cues influence ToM attributions?

If ToM is rightly considered as a form of empathy, then the answer the these questions is yes. ToM should show a pattern of variation similar to other forms of empathy in terms of its sensitivity to cues of coalition. Given the fluid nature of group membership and that we've evolved a flexible coalitional psychology, then it should be possible to manipulate a subject's ability to make correct ToM and false belief attributions by manipulating variables related to the nature of the relationship between them and the target individual. That is, subjects will be more accurate in assessing the mental states of target individuals they perceive as in-group members relative to those they perceive as out-group members. I am testing the hypothesis that ToM is not a fixed trait within individuals, but rather a (flexible) skill that is context/target dependent much in the way motor and emotional empathy are. Specifically, I am testing the hypothesis that since we are more empathetic towards in-group members vs. out-group members, we will also better at inferring the mental states of others when we are primed to think of them as in-group members rather than out-group members. To do so, I have combined a perspective taking methodology (after Keysar, Lin, and Barr, 2003) with a priming method (Burnham, McCabe, Smith, 2000) designed to establish in-group versus out-group relationships between subjects and confederates.

Next, I discuss the results of the study and address potential sources of error arising from methodological issues that may call my findings into question. I identify steps to correct these issues and turn to an overview of future work necessary to improve upon and strengthen the present methods and findings.

Last, I consider two main implications of this work. First, it will lead to a better understanding of how making accurate inferences of others' mental states depends on perception of target individuals. While a large body of work explores how emotional empathy varies in response to such perceptions, ToM research has instead focused on developmental emergence in infants as well as deficits exhibited by autistic individuals. If inference of others' mental states is dependent on whether we perceive them, for example, as friend or foe, it is important to learn to what extent those coalitional cues can affect ToM. Second, this work also has the potential to suggest ways to increase accurate ToM among members of cultural groups who might otherwise be antagonistic towards each other. Group boundaries and coalitions are not permanent; they form and dissolve as need arises. Likewise, our underlying evolved coalitional psychology--tuned to detect and act upon cues of group membership--is flexible also. Engaging in cooperative tasks or being presented with subtle cues of shared group membership may not only affect how we respond emotionally to others, but also how we come to understand the inner cognitive worlds of others, whether they are friends or members of opposing groups in conflict. This could lead to more effective peace-making or reconciliation techniques for disputes between neighbors to long-standing animosities between cultural, political or racial groups.

Chapter 2: Theory of Mind

Theory of mind (ToM) is the ability to impute mental states to oneself and to others, to interpret observed behavior in terms of underlying beliefs and desires, and to understand that others can have mental states that differ from one's own (Premack and Woodruff, 1978). While it is clear from Premack and Woodruff's definition that ToM is a cognitive ability, something one *does*, they consider ToM to be a *theory* because first, "such states are not directly observable, and second, because the system can be used to make predictions, specifically about the behavior of other organisms" (p515). In contrast, Leslie (2000) defines it as a representational system that captures the cognitive properties underlying observed behavior—he has humorously noted that ToM is neither a theory nor a theory of mind. According to Leslie, there is an innate, preverbal *mechanism* (the Theory of Mind Mechanism, or ToMM) that allows one to interpret the actions of an individual in terms of its underlying beliefs, desires, or other mental states. Considering both definitions, the key terms are Premack and Woodruff's "imputes" and Leslie's "interpret". Theory of mind is not something one has (as Leslie says, it is not a theory), but something one does: we interpret others' behavior in terms of underlying mental states.

Theories of Theory of Mind

Several major functional models have been proposed to explain how ToM happens: The Theory-Theory or Child Scientist model (Gopnik, 1993; 1996) posits that the ordinary person constructs a naive folk-psychological theory that guides assignment of the mental states of others. As such, it views ToM as a literal theory, contrary to

Leslie's aphorism. In contrast, Simulation Theory (Gallese and Goldman, 1998; Goldman, 2006), views ToM as an action—an individual fixes the target's mental states by trying to replicate or emulate them. Jern and Kemp (2015) argue for incorporating decision networks into our understanding of ToM. These networks, based on Bayesian networks, include the assumption that individuals make choices in order to achieve their goals. We observe others make choices and use this information to infer their goals, knowledge, or beliefs. In this regard, it can be viewed as a formal version of the child-scientist theory, though they do not rule out simulation theory. Third, Rationality Theory (Dennett 1987), argues that the ordinary person a) functions according to rational rules, and b) assumes that others do, too. Thus, the individual seeks to map one's own thoughts and choices to others by means of this rationalizing: "I would do *X* in this situation, therefore, this other individual will do *X* also." A fourth functional model of ToM is the Associationist account (De Bruin and Newen, 2012), This model seeks to provide an alternate explanation for preverbal infants' performance on implicit measures of ToM and false belief (see below) without the need to actually understand false belief. De Bruin and Newen posit an interaction between two modules in the brain: an "association module" that allows infants to recognize congruent associations between agents and objects, and an "operating system" that relies on inhibition and selection to process incongruent associations.

Schematically, both the Child Scientist (CS) model and Simulation Theory (ST) assume the same general decision-making mechanism: an individual has a desire g (e.g., "I want pizza.") and a belief $m \rightarrow g$ ("Calling Lou Malnati's will get me a pizza") which are processed in a 'decision-making' mechanism that outputs a decision d based on those

inputs (“I will call Lou Malnati's”) (see Figure 2.1). However, they differ in their explanation of how the individual extends this system to make a mental attribution of another individual’s mental state. CS postulates that an observer relies on a set of beliefs about the target, *T* desires *g* (“Joe wants pizza”), *T* believes $m \rightarrow g$ (“Joe believes using the phone will get him a pizza”), as well as a belief in some general psychological decision-making law in order to determine the target’s mental state (“People who want pizza call Lou Malnati's”). These beliefs are processed in a formal reasoning mechanism, which outputs a belief about the target’s decision (Joe will call Lou Malnati's to get a pizza) (see Figure 2.2). In contrast, ST argues that the observer uses the same decision making mechanism that is used for one’s self in order to infer the mental state of the target. One’s own belief *~g* (“I think Lou Malnati's does not sell pizza”) and desire *h* (“I prefer tacos to pizza”) are set aside, or “quarantined” and replaced with pretend states, assumed to be those of the target, substituted for one’s own (Figure 2.3). Note that in the figure the quarantine is indicated by a dashed line, which symbolizes that the quarantine is not perfect—some of the observers beliefs can still potentially contaminate the simulation. Rationality Theory is not as much a theory of the contents of others’ minds as it is a theory of the contents—rational rules—of one’s own mind (“When I want pizza, I call Lou Malnati's”), and an assumption that others also follow the same rules (“If Joe wants pizza, he will call Lou Malnati's”).

Simulation Theory is of particular interest here as it suggests a mechanism similar to those moderating motor and emotional empathy: in all three systems, the actions, emotions, or other cognitive states of a target individual are mirrored in the observer. Simulation also fits the broad definition of empathy given by Preston and de Waal

(2002)—it is any process where the attended perception of a target's state generates a state in the observer that is more applicable to the target's state or situation than to the subject's own prior state or situation. Again, in all three cases (motor, emotion and cognition), attending to a target generates a similar state in the observer.

A critical component of ST is that it views simulation as an *attempt* to put one's self in another person's shoes; it does not focus on the accuracy of the simulation (Goldman, 2006). In this regard, ToM operates in a similar fashion to other forms of empathy. Various factors, such as one's familiarity with target individuals, the nature of the social relationship between observer and target (Kozak, Marsh, and Wegner, 2006), the perceived fairness of their actions (Singer, Seymour, O'Doherty, Stephan, Dolan, and Frith, 2006), or testosterone levels in the observer (Hermans, Putnam, and van Honk, 2006) are known to influence the accuracy of a simulation. Figure 2.3 illustrates the presence and potential influence of one's own beliefs and desires; they are set aside, or "quarantined" during the simulation. However, the quarantine is indicated by a dashed line, acknowledging that the observer's beliefs may leak into the simulation and be projected (falsely) onto the target (e.g., Nickerson, 1999). In addition, even adults can succumb to the "curse of knowledge" (Birch and Bloom, 2003; 2004) in making belief attributions about others: their own knowledge of reality can interfere with the attribution process and they mistakenly attribute this knowledge to target individuals.

Theory of Mind and False Beliefs

How do we determine whether someone has ToM, whether they impute mental states to another individual or understand actions in terms of underlying mental states?

Something far simpler could underlie our actions and responses to others—if-then rules: If aggressor has facial expression X, then perform action Y. If a potential mate exhibits behavior B, respond in kind, else do nothing. Underlying mental states driving action are unnecessary in this case. In addition, Gopnik's Child Scientist and Dennett's Rationality Theory are limited as well, in that they rely on more general assumptions about how other people behave (a naive folk-psychological theory or the assumption that others follow the same rational rules as the observer). In contrast, a simulation account of ToM takes the additional step of attempting to put one's self in the perspective of another, in order to see the world as they do and draw a conclusion about the underlying mental state driving their behavior.

This is an important distinction to make, because our “default setting” is to attribute true beliefs to others (Leslie, German, and Polizzi, 2005). In the case where a target individual holds all the same beliefs and knowledge as the observer, how can we know if the observer is making an accurate assessment of the target's mental state or if he is simply drawing a conclusion based on his own mental state, as one might do from a CS or RT perspective? Or, for that matter, how do we know that an observer is not simply relying on behavioral observations (“After Joe paces around the room, he will call Lou Malnati's, order a pizza and eat it”). Dennett (1978) has suggested that we rely on false beliefs to determine whether an individual is actually imputing a mental state to a target individual in the interpretation of their behavior. Leslie argues that, in addition to the automatic ToMM, a slow to develop “executive” Selection Processor allows us to inhibit the tendency to make default true belief attributions in cases where observed target individuals hold false beliefs about reality (Leslie, German, and Polizzi, 2005). Why false

belief? As Wellman, Cross, and Watson (2001) summarize, “Mental-state understanding requires realizing that such states may reflect reality and may be manifest in overt behavior, but are nonetheless internal and mental, and thus distinct from real-world events, situations or behaviors. A child's understanding that a person has a false belief—one whose content contradicts reality—provides compelling evidence for appreciating this distinction between mind and world” (p655).

False Belief Tasks

Based on this idea, Wimmer and Perner (1983) developed the Displaced Object or Location Change task (also referred to as the Sally Anne task (Baron-Cohen, Leslie, and Frith, 1985). For over 30 years now, this has been the primary method for testing whether a subject understands false belief: A subject watches as “Maxi” (a puppet used in the original experiment) hides a favorite object such as a toy in one of two containers. Maxi then temporarily leaves the room. While Maxi is gone, a second person, the experimenter or an assistant, moves the object to the other container in full view of the subject. Maxi then returns to retrieve her object. At this point, the subject is asked two control questions to check their understanding and attention: where did Maxi put the object, and where is the object now. Then the key question is asked, Where does Maxi think the object is? In order to answer correctly, subjects must ignore their own knowledge and recognize that Maxi now has a false belief about the object’s location.

In addition to this task, there are two others commonly used, the Unexpected Contents (Hogrefe, Wimmer, and Perner, 1986) and the Surprising Object (Perner, Leekam and Wimmer, 1987) tasks. Unexpected Contents presents subjects and a

companion (an experimenter or puppet) with a container such as a box of crayons. Subjects are asked what they think is in the box, and after they answer their companion leaves. The box is opened to reveal something else, such as candles. Control questions are also asked (for example, “What is really in the box?”) and here the key question asks the subject what their companion (who has not seen the revelation) thinks is in the box. The Surprising Object task is similar, presenting subjects with what appears to be a rock, but turns out to be a sponge only painted to look like a rock. Though somewhat different in set-up, the Displaced Object, Surprising Object, and Unexpected Contents tasks are similar in that to answer correctly, subjects need to understand that their companion holds a belief contrary to reality. It is not uncommon to see all three tasks presented as a single false-belief test battery.

There are actually numerous variations in the way false-belief tasks are presented. In their meta-analysis, Wellman, Cross and Watson (2001) found that the key question in the Location Change task can be asked in terms of action (“Where will Sally *look* for her toy?”), thoughts (“Where does Sally *think* her toy is?”), or speech (Where will Sally *say* her toy is?”). The target object may be moved as an intentional deception (“Let's play a trick on Maxi.”), or it may be moved inadvertently. The protagonist may be a puppet, a doll, a real person presented live, or any one of these portrayed in a video. Similar variants accompany the Unexpected Contents and the Surprising Object tasks as well.

ToM vs. False-belief

Before proceeding, there is an important distinction between ToM and false-belief understanding that needs to be addressed. Because false belief is so important in

demonstrating ToM, and false belief tasks are so widely used to test for ToM, it is easy to forget that ToM and false-belief understanding are not one and the same. False belief provides a convenient way to test whether someone can interpret the behavior of others in terms of their underlying beliefs. It is only one special case, one component of the larger ability that is ToM. Therefore, it would be erroneous to conclude that merely because a child fails the false belief task that he or she lacks ToM. For example, autistic children and adults have difficulty passing false belief tasks (Baron-Cohen, Leslie, and Frith 1985), yet Senju, Southgate, White, and Frith (2009) found that subjects with Asperger syndrome can correctly make false belief attributions, but they apparently do so by different means—it is not the automatic process it is in normal individuals. Bloom and German (2000) have criticized the use of false belief task as a test for ToM for these reasons: Passing the false belief task requires more than ToM (for example, inhibition of one's own knowledge of reality), and ToM does not necessarily entail false belief reasoning. These are important criticisms, yet the false-belief task in all its variations does remain the most widely used test of ToM; the logic behind its use is sound. And as discussed below, some more recent methodological developments in false belief task presentation do take some of these issues into account.

Biological Bases of ToM

Brain imaging studies have consistently revealed four areas that are active during ToM tasks, studied to date: the medial prefrontal cortex (mPFC), temporal poles, posterior superior temporal sulcus (STS), and the temporoparietal junction (TPJ) (Frith and Frith, 2006; Saxe et al. 2004; Saxe and Powell, 2006). The temporal poles are

thought to be ‘convergence zones’ where simple features from various modalities are brought together, so that our understanding of objects can be modified by context. The mPFC is the most simulation-like of these areas; it is activated when people think about their own as well as when they think of others’ mental states. It has also been considered to be the primary location of our ToM ability, showing the most consistent activation in ToM tasks (Frith and Frith, 2001). In addition, it is located in the prefrontal cortex, which is broadly involved in planning for the future, and with anticipation of what others will think or feel (Frith and Frith, 2006). The posterior STS and TPJ are involved in eye movement observation, provide information about where others are looking, and for representing the world from different visual perspectives. In addition, the TPJ has been thought to be involved in preliminary stages of social cognition that work in service of ToM (Saxe and Kanwisher, 2003). However, the TPJ may be more important for mentalizing than previously recognized, particularly with regards to belief attribution (Samson et al., 2004; Saxe and Kanwisher, 2003; Saxe et al., 2004). Patients with lesions in the left TPJ have been found to respond to both verbal and nonverbal false belief tasks at chance levels (Samson et al., 2004).

The TPJ shows a greater response to images of people compared to objects and to nonverbal versus verbal stimuli (Saxe and Kanwisher, 2003). It shows an increased response to descriptions of mental states versus other social information, especially when there is an incongruence between the two (Saxe and Wexler, 2005). The temporoparietal junction is involved in many tasks. In addition, to these mentalizing functions, there are at least four other functions that we might assign to the right temporoparietal junction: It is a perceptual area involved in detection of global versus local (whole versus part)

aspects of visual (Robertson, 1996; Robertson, Lamb, and Knight, 1988) and auditory stimuli (Justus and List 2005), monitoring incongruent input across sensory modalities (visual versus proprioceptive) (Balslev, Nielsen, Paulson, and Law, 2005), processing speed and visual short-term memory (Peers, Ludwig, ROrden, Cusack, Bonfiglioli, Bundesen et al., 2005), and possibly identification of congruent versus incongruent rhythmic stimuli in musically untrained subjects (Vuust, Pallesen, Bailey, van Zuijen, Gjedde, Roepstorff, and Østergaard, 2005).

ToM may simply be an aspect of this global-specific stimulus processing. If this is so, then the different results of Samson et al. (2004) pointing to the left hemisphere and Saxe et al.(Saxe and Kanwisher, 2003; Saxe and Wexler, 2005) pointing to the right can be explained. Both hemispheres' functions are necessary in attributing beliefs to others. The right hemisphere interprets when there are no violations of the whole percept, but reacts to incongruencies, which may then be processed by the left hemisphere. The Vuust study (Vuust et al., 2005) may suggest a learning component to this, as well, given the differences between trained musicians, who processed rhythmic incongruencies as musical information and non-musicians, who heard them only as violations of the whole.

While the false belief task is the primary test of mentalizing, false beliefs are the exception to the rule; everyday reasoning about other minds involves mostly attributions of true beliefs (Dennett, 1996, cited in Saxe and Kanwisher, 2003). If this is the case, then it follows that we would not need to rely on details, or local-level stimuli, to make these attributions. Turning to a behavioral conditioning model, whole percepts (global-level stimuli) could come to trigger particular mental state attributions in a stimulus-response classical conditioning sense.

In an experiment where subjects were presented with stories of protagonists from familiar versus unfamiliar social backgrounds who held either normal or norm-violating beliefs, Saxe and Wexler's (2005) data suggested that subjects attempted to form integrated impressions of the protagonist and resolve incongruent situations between social backgrounds and stated beliefs. Here, the right TPJ showed a lower response to background information. This response was not modulated according to the familiarity or unfamiliarity of the background described. However, an increased response in this hemisphere was seen when mental states were described, with an additional response when the protagonist's background and mental state were incongruent. In contrast, the left TPJ showed a strong response to the social background information (which was not significantly different from its response to mental states). Additionally, its response to the unfamiliar social backgrounds was significantly higher than its response to descriptions of familiar ones. From this, they speculate that the left TPJ might have a broader role in the attribution of enduring socially relevant traits, while the right TPJ is restricted to making attributes of relatively more transient mental states.

Emergence of ToM

The False Belief Task Findings

One of the more robust findings in ToM research has been the pattern of emergence of the ability in children to pass false belief tasks between three and five years of age (Wellman, Cross, and Watson, 2001). Children do not begin to correctly respond until their fourth year, and by the time they are five, they are able to do so consistently. This has been taken as evidence that ToM comes about as a developmentally emerging

conceptual change; children younger than four years fail false belief tasks because they do not yet understand others' behavior in terms of underlying mental states.

An oft-repeated critique of psychological research is that it tends to rely on a limited subject pool (college students in the U.S.) and subsequently makes pronouncements about human universals (see Henrich et al., 2010 for a recent in-depth review and discussion). Clearly, any claim about human universals would be strengthened if similar results are obtained across many different sample populations across many different settings. This concern has been addressed with regards to ToM, and cross-cultural research appears to support the standard view of false-belief understanding as a developmentally emergent phenomenon and suggest that it is a human universal.

In an early study of cross-cultural false belief understanding, Avis and Harris (1991) presented a location change task to Baka children in Cameroon and found that by 4-5 years of age, the children “are good at predicting a person's action and emotion in terms of his or her beliefs and desires about a situation rather than in terms of the objective situation itself” (p464-5), and that this belief-desire reasoning competence is not as developed in 3-year-olds, results that are consistent with findings using Western children as subjects.

Vinden (1996) conducted a surprising objects task and deceptive container (unexpected contents) task to Junín Quechua children. Her results are more difficult to compare to the standard findings as she divided her subjects into two groups—those under 6 years of age and those 6 and over, an age where false belief understanding is firmly established. She found that in both age groups the majority of children who

responded correctly to appearance and reality questions did not provide consistently correct responses to false-belief questions. While her results do indicate an expected increase in performance with age, these children did not perform as well on false belief tasks as children in other studies.

Wellman, Cross and Watson (2001) undertook a large meta-analysis of false-belief studies examining 77 articles and reports, which comprised 178 separate studies and 591 different conditions. While they included the country of the participants in each study, it is not clear what to conclude regarding cross-cultural differences, or lack thereof. First, it is difficult to determine the extent of cultures sampled; a comprehensive list of countries is not included in the analysis. Only seven countries—those in which six or more total conditions were run—are explicitly discussed: the United States, United Kingdom, Korea, Australia, Canada, Austria, and Japan. Elsewhere, though, Wellman, Cross, and Watson (2001) add that their data include children from two nonliterate, more traditional communities, the two discussed above: hunter-gatherer Baka from Africa (Avis and Harris, 1991) and speakers of Quechua from Peru (Vinden, 1996), for a total of nine countries. Second, it is difficult to interpret the findings. Wellman and his colleagues found that country of origin significantly influences children's performance on false-belief tasks, but it does not interact with age. “For these [nine communities], children's false belief performance increases across years in equivalent age trajectories, although at any one age children from different countries and cultures can perform differently” (Wellman, Cross, and Watson, 2001:669). In addition, they conclude that this analysis argues against proposals (such as Lillard, 1998) that the understanding of belief—both true and false—is the result of socialization within literate, individualistic European and

American cultures. Rather, an understanding of others that includes a sense of their internal mental states (beliefs, desires, intentions) is widespread. The trajectory is the same, though there is variation in onset and overall performance, depending on the different cultural communities and language systems in which the children are reared (Wellman, Cross, and Watson, 2001).

In a study not included in Wellman's meta-analysis, Vinden (1999) addressed the heavy focus on Western subjects. She presented two different location change tasks (a "look" and a "think" version) to children of four different cultures: a Western control group (consisting of children from Australia, North America, and Europe), Mofu from Cameroon, and Tainae and Tolai, both from Papua New Guinea. She found that "Even though few 4-year-olds and no 3-year-olds were available for testing in the non-Western cultures, a clear trend toward understanding of false belief is visible, though at a somewhat later age in two of the three cultures" (p40). Tolai and Mofu children both show an understanding of false belief at about the same time: 4- and 5-year-olds performed below chance levels, 6-year-olds at chance, and 7- to 10-year-olds at above chance levels. For the Tainae children, Vinden reported that 4- to 8- year-olds performed well above chance, but it is unclear when false-belief understanding emerges, because it was near impossible to recruit younger children to participate.

And more recently, Callaghan and her colleagues (Callaghan, Rochat, Lillard, Claux, Odden. Itakura et al., 2005) also studied false belief understanding across cultures. They presented a location change task to 3-, 4-, and 5-year-old children from Canada, India, Peru, Samoa, and Thailand. They found, for all groups, a significant number of 3-year-olds failed the task, while in 4-year-olds performance was mixed, and a significant

number of 5-year-olds passed. These results are in line with expectations, both in terms of a developmental emergence, as well as the timing.

In all of these studies, culture is associated with variation in the age of onset of false-belief understanding and false belief task performance. The question remains, why? While many factors have been suggested to affect false belief understanding, including children's socioeconomic status (Holmes, Black, and Miller, 1996; Shatz, Diesendruck, Martinez-Beck, and Akar, 2003), amount of schooling (Vinden, 1996; 1999; 2002), or exposure to different parenting methods (Vinden, 2002), these can all easily explain variation within a given culture as well as across cultures.

The Role of Language

Language appears to be an important factor in the pattern of emergence of ToM (Astington and Jenkins, 2001), not necessarily language in general, but particular grammatical elements (De Villiers and Pyers, 2002). Differences across languages can and do affect thought in a variety of ways. Not surprisingly, research in number sense and other areas of core knowledge (see below) has revived interest in this idea of linguistic influence of thought (Whorf, 1956/2001, see Gleitman and Papafragou (2004) for a review). It is important to remember that the literature does not argue that language determines thought, but rather that language and thought (and culture) mutually influence each other (Ahearn, 2011; Hill and Mannheim, 1992). Categorizing colors (Winawer, Witthoft, Frank, Wu, Wade, and Boroditsky, 2007), the shape versus materiality of objects (Lucy, 1992), the positioning of objects in space in relation to one's self (Levinson, Kita, Haun, and Rasch, 2002; but see Li and Gleitman, 2002) and to other

objects (Bowerman and Choi, 2003) all show language-dependent variations, though there remains a shared, underlying core ability.

Just as various vocabularies and languages carve up the external world differently, they can also carve up the internal world differently. For example, how are mental states represented across languages? Basic emotions are often considered universal, yet there is some disagreement over what those basic emotions are. Ekman (1972) named six: happiness, sadness, anger, fear, surprise and disgust. Izard and Buechler (1980) also named those same six, but also include interest, contempt, shame/shyness and guilt. And Panksepp (2000) lists seeking (expectancy), rage (anger), fear (anxiety), lust (sexuality), care (nurturance), panic (separation), play (joy). Clearly there is a good deal of overlap, suggesting that researchers are focusing on the same clusters of facial expressions and physiological data. Wierzbicka (1986), however, points out something that should perhaps be obvious: if lists such as these are "supposed to enumerate universal emotions, how is it that these emotions are all so neatly identified by means of English words?" (p584). Polish does not have a word that corresponds to the English word disgust (Wierzbicka, 1986), the Australian language Gidjingali does not distinguish between fear and shame (Hiatt, 1978, in Wierzbicka, 1986), and the Rarámuri of Mexico have one word for both shame and guilt (Breugelmans and Poortinga, 2006). "English terms of emotion constitute a folk taxonomy, not an objective, culture-free analytical framework, so obviously we cannot assume that English words such as disgust, fear, or shame are clues to universal human concepts, or to basic psychological realities" (Wierzbicka, 1986:584). As we move from basic emotions to higher social emotions (Panksepp, 2000) such as envy, humor, empathy or jealousy this can only become more of an issue. Still,

lacking a word for disgust does not preclude Polish speakers from understanding the concept, just as English speakers understand the concept of the German word *schadenfreude*, though we do not have a single word to denote it.

Returning to some of the cross-cultural work described above sheds some light on this issue, with Vinden's work on the Junín Quechua children (Vinden, 1996) and Mofu, Tainae, and Tolai (Vinden, 1999) providing an insightful starting point. She chose Junín Quechua children as subjects because adults in this culture differ from Western adults in the extent to which they use mental state terms—in this language, as in other Quechua languages, mental concepts such as "thought" or "belief" are not referred to directly. In Junín Quechua, the English question "What do you think?" would translate roughly into "What do you say?" Vinden found that, in the absence of explicit mental state terms, Junín Quechua children's development of mental state understanding was not comparable to that of Western children, lagging behind in terms of age of onset and overall performance.

Vinden's (1999) work with the Mofu, Tainae, and Tolai also points to a linguistic factor affecting children's false belief task performance. Again, it lies along a division between two different conditions, Think questions versus Look questions. She found that "not all children in every culture, or in each age group, responded to the Think false belief question in the same way as they did to the Look question. This is true of the Western sample, as well as the non-Western samples. Further research is therefore warranted to explore how children conceptualize the difference between asking where someone will look for something, and where someone *thinks* that thing is" (p41, emphasis original).

This begins to lead us towards the more serious issue that must be considered in language-based false belief tasks: obligatory grammatical structures—what ***must*** be expressed. Two more recent papers (Matsui, Rakoczy, Miura, and Tomasello, 2009; Shatz, Diesendruck, Martinez-Beck, and Akar, 2003) further clarify this interaction between language and false belief understanding. Shatz and her colleagues (2003) investigated the effect that the explicitness with which a language expresses false belief may have on children's performance on false belief tasks. They contrasted four groups of 3- and 4-year-old preschool children: speakers of Turkish, Puerto Rican Spanish, Brazilian Portuguese, and English. The first two are languages with explicit terms that explicitly indicate false belief, whereas the second two lack such terms languages without such terms. For example, in English, the word *think* is used report a belief whether the speaker is neutral about the truth value of a statement ("Joe thinks it is a good day to exercise") or knows the belief is false ("Joe thinks that New Jersey is west of Illinois"). Puerto Rican Spanish distinguishes between *creer* for the neutral truth-value cases and *creer-se* when the speaker is certain a false belief is held. They compared success on false belief tasks using both "Where does X think the object is?" question and "Where will X look for...?" questions. When explicit false-belief terms were used in Turkish and Puerto Rican Spanish, children in those languages answered significantly more Think questions correctly compared to the other two languages, but there was no difference for the Look questions. This was seen in both 3- and 4-year-olds, although the older group still outperformed the younger children. In other words, the presence of an explicit term to express a false belief improves children's performance on false belief tasks, even for 3-year-olds, an age where children are not expected to pass such tasks.

In similar study, also exploring the relationship between speaker certainty and false belief reasoning, Matsui, Rakoczy, Miura, and Tomasello (2009) focused exclusively on 3-year-olds, comparing Japanese and German preschoolers. Conversational Japanese uses two different sentence-ending terms to indicate one's certainty of belief: “The Japanese certainty particle *yo,* when affixed to an assertion, emphasizes the speaker's strong commitment to the truth of the statement...The uncertainty particle *kana,* on the other hand, expresses strong uncertainty, and hence indicates that the speaker does not commit herself to the truth of the statement” (p604). Matsui notes that these terms are stylistically normal in spoken Japanese and do not appear in formal writing. Japanese and German 3-year-olds were presented with a location change and an unexpected content task. For each task type, children were presented with three variations, a standard version in which the target (a puppet) remained silent, one in which the puppet provided an explicit false belief certainty utterance, and one in which the puppet provided an uncertainty utterance. German lacks analogous terms to *yo* and *kana*, but the subjects in this group were presented with comparable declarative statements (certainty) and “perhaps” (uncertainty) statements.

Matsui's team found that Japanese children are very sensitive to *yo* and *kana* conditions, answering significantly more false belief questions correctly in the *yo* conditions compared to the *kana* conditions. German children did not differentiate between conditions. In a second part to the study, the German children were presented with a stronger statement of certainty vs. uncertainty (must vs. may), while the Japanese children heard simple statements without *yo* or *kana*. The stronger statements of certainty did not affect the German children's performance, but interestingly, the Japanese children

lost their sensitivity. These researchers conclude that “the overall findings of the present studies strongly indicate that certain mental state expressions have the potential to bootstrap the ability of young children who fail to pass the standard false-belief tasks to understand the speaker's false belief in verbal communication” (Matsui et al., 2009:611).

False Belief and the Syntactic Complement

Taken together, the studies reviewed above (Vinden, 1996; 1999; Shatz et al., 2003; Matsui et al., 2009) strongly suggest that specific aspects of a language can influence false belief understanding in children across cultures. However, none of these have been obligatory structures, i.e., structures that must be present in an utterance for it to be considered grammatical or well-formed. Work with deaf children points to one such structure, the syntactic or sentential complement—a specific skill thought to be a necessary component for passing the false-belief task. Deaf children who are otherwise normal, and who are raised by speaking parents do not begin to pass the traditional presentation of the false belief task at the age where other children begin to do so (Figueras-Costa and Harris, 2001; Schick, de Villiers, de Villiers, and Hoffmeister, 2007). While lack of exposure to normal social situations may play a role in this delay (Peterson and Siegal, 2000; Russell, Hosie, Gray, Scott, Hunter, Banks, and Macaulay, 1998), once they do pass the task, it coincides with their mastery of the syntactic complement (Schick et al., 2007), the embedding of tensed propositions under a main verb. Hale and Tager-Flusberg (2003) note that two types of verb are able to take syntactic complements, verbs of communication and—important to this discussion—verbs of mental state. Furthermore, “the embedded clause is an *obligatory* linguistic

argument that may have an independent truth value" (p4, emphasis added). In other words, one does not say, "Joe believes" without also embedding a complement: "Joe believes *X*". The statement can be true regardless of the truth of *X* ("Joe believes that New Jersey is west of Illinois" and "Joe believes that Illinois is the Prairie State" can both be true, though the complement in the first example is false, and true in the second). This structure "invites us to enter a different world...and suspend our usual procedures of checking truth as we know it. In this way, language captures the contents of minds, and the relativity of belief and knowledge states. These sentence forms also invite us to entertain the possible worlds of other minds, by a means that is unavailable without embedded propositions" (de Villiers, 2000:90, quoted in Hale and Tager-Flusberg 2003:4).

Once children—deaf or hearing—can understand and use these embedded forms, they can understand and produce sentences such as "Sally thinks that the toy is in the bucket," answer questions such as "Where does Sally think the toy is?" and so correctly respond to the false belief task. De Villiers and Pyers (2002) suggest that mastery of the complement not only allows children to understand the questions posed to them, but that without it, children do not even understand the concept of false belief. Without the syntactic complement, one cannot even entertain the possibility that someone can hold a false belief.

Implicit vs. Explicit ToM Tasks

Despite the relative consistency of cross cultural emergence of false belief understanding and the apparent importance of language and grammatical structures, there

is also evidence for an earlier competence, suggesting that children younger than five years of age, and even preverbal infants also understand false belief. When the verbal and/or cognitive demands are simplified, (e.g., asking “Where will Sally look first?”) three-year-old children do appear able to implicitly understand false belief (e.g., Clements and Perner, 1994). In completely non-verbal tasks, 15 month-old infants demonstrate understanding of false belief (Onishi and Baillargeon, 2005) and violations of pretense (Onishi, Baillargeon, and Leslie, 2007). Even 13-month old infants (Surian, Caldi, and Sperber, 2007) were able to successfully pass false belief tasks and attribute a false belief about the location of an object. 18-, 12-, and 9-month old infants have been shown to discriminate between adults' intentions, whether unwilling vs. unable to give the infants a toy (Behne, Carpenter, Call, and Tomasello, 2005). The results of these newer studies, with their reduction of other cognitive demands inherent in the false-belief task, are challenging the view that, in humans, false belief understanding follows the developmental trajectory described above.

One possible interpretation of these two sets of findings is that there are two distinct abilities being tapped. On the one hand, there is the ability to implicitly understand false belief, and on the other is the ability to explicitly communicate that understanding. In fact, Baillargeon (2008) has noted that the best way to determine if children and infants understand false belief is to not ask them any questions about it: As discussed above, additional cognitive demands placed on a child (or chimpanzee), such as the need to inhibit their own knowledge of the correct location of a target object, increase the probability of giving an incorrect answer.

Baillargeon, Scott and He (2010) separate false belief tasks into two different

categories based on the type of response required; tasks can rely on either elicited responses or spontaneous responses. Historically, tests of false belief understanding relied on elicited-response tasks, where subjects must answer direct questions about an agent's false belief. The Displaced Object task is an example of this type of test: “Where does Maxi think the toy is hidden?” Such tests do not necessarily have to be verbal. Nonverbal tests of ToM can also require that the subjects explicitly demonstrate their knowledge, through pointing, for example (e.g., Call and Tomasello, 1999).

More recently, spontaneous response tasks have become common; these are the methods used to test false-belief understanding in infants. In these tests, a subject's understanding of an agent's false belief is inferred from behaviors subjects spontaneously produce while observing a typical scene unfold. These spontaneous response tasks can be further divided into two main types, violation of expectation (VOE) tasks and anticipatory looking (AL) tasks.

VOE tasks are based on research findings that shows subjects look longer at unexpected outcomes relative to expected outcomes (e.g., Gergely, Nádasdy, Csibra, and Bíró, 1995). In the context of a false belief task, subjects look longer when an observed agent acts in a manner that is inconsistent with her false belief (Onishi and Baillargeon, 2005): a Sally-Anne-like task unfolds where the subject observes an actor place an object in one of two containers. Next, a blinder is used to block the agent's view. After the blinder is in place, one of two conditions is presented to the subject, either the object stays in the original container or moves to the alternate container. Last, the blinder opens and instead of asking the subject where the actor will look, the actor continues and performs one of two possible actions, she reaches into the original container or the new

container, for a total of four possible outcomes (two possible locations and two possible actions). The time the subject spends looking at each outcome is measured. Subjects spend significantly more time looking when the actor reaches in the container to where the object has moved; she is responding in a manner inconsistent with her (false) belief about where the object is (Onishi and Baillargeon, 2005). This condition represents a violation of subjects' expectations about what was the agent should have done, expectations based on a representation of the actor's beliefs.

AL tasks involve a similar set-up as well, but as the name suggests, measure whether the subjects visually anticipate the outcome before it happens. In other words, the subject look to where they expect the agent to reach. Infants (Southgate, Senju, Csibra, 2007) and adults (Senju, et al., 2009) do this, and their responses suggest an understanding of false belief: subjects look towards the original container in anticipation of the agent's action.

Results from both types of spontaneous response tasks suggest ToM emerges earlier in development than previously thought. Infants as young as 13 months successfully pass VOE tasks (Surian, Caldi, and Sperber, 2007). If preverbal infants understand false belief, why do 3- and 4-year-olds continue to have difficulty with standard tests? Why are elicited-response tasks so much more difficult for children? Scott and Baillargeon (2009, in Baillargeon, Scott, and He, 2010) suggest that elicited response tasks involve at least three different processes: 1) a false-belief representation process (subjects must be able to represent the agent's false belief), 2) a response selection process (given a test question, subjects need to access their false belief representation in order to select the appropriate response), and 3) a response inhibition process (it is not

enough to know the correct response; subjects must also be able to avoid answering based on their own knowledge). Baillargeon, Scott, and He (2010) note the available neuroscience findings suggest that while the actual false belief representation occurs in the TPJ, the response inhibition and selection occurs in the frontal cortex. Connections between the temporal and prefrontal brain areas develop later and more slowly than other connections, making elicited-response tasks (which involve both these areas) much more difficult for young children. In fact, Atance, Bernstein, and Meltzoff (2010) found evidence that 3-year-olds actually take longer to process and respond correctly to false belief questions than to respond incorrectly.

Increased cognitive demands impair performance on various tasks in adult subjects as well, and this effect is not limited to ToM. One example is deception, which requires suppression of a true response in order to intentionally give a false response (Nunez, Casey, Egner, Hare, Hirsch, 2005). One of the most striking examples of response inhibition is the Stroop color naming task (Stroop, 1935), subjects are presented with a list of color words printed in ink that is a different color than the color named. For example, the word “Red” is printed in blue ink, the word “Black” in yellow, and so on. The task involves naming out loud the colors the words are printed in. To do so, subjects need to suppress the more salient response, the reading the words themselves. As a result, response times for the color naming task are much slower than times for simply reading the words aloud. Thus, it should not be surprising that implicit ToM tests, with their reduced extraneous cognitive demands, are better measures of the presence of the ability, independent of the ability to verbally express understanding.

ToM and Core Knowledge

If false belief understanding, and by extension, ToM are not developmentally emerging skills but present from infancy, this significantly changes how we should think of them in terms of their evolutionary emergence and their presence in our non-human primate relatives. Spelke and Kinzler (2007) suggest that humans (and animals) possess a small number of basic, systems of *core knowledge* upon which new skills and belief systems are built: "Each system centers on a set of principles that serves to individuate the entities in its domain and to support inferences about the entities' behavior" (Spelke and Kinzler, 2007:89). In other words, these systems are based on a small number of associated skills that do not overlap with other core systems. There are at least four such systems, for representing 1) the mechanical interactions of inanimate objects; 2) the goal directed actions of agents; 3) the numerical relationships of ordering, addition and subtraction; and 4) the geometric relationships of places in spatial layouts (Spelke, 2003). Or, to put it more succinctly, there are systems for representing objects, action, number, and space.

Core knowledge research is particularly relevant to anthropology because our understanding of core knowledge systems is based on a great deal of cross-cultural (and non-human) studies. To illustrate the concept of core knowledge, and how it can provide useful methods for studying ToM from an anthropological perspective, I briefly review some of the findings in one particular domain, numerical cognition.

Numerical Cognition

Counting systems such as our base-10 are generative; with it we can represent any

quantity exactly, however large or small. In contrast, many societies, including the indigenous Amazonian Pirahã (Frank, Everett, Fedorenko, and Gibson, 2008; Gordon, 2004) and Mundurukú (Pica, Lemer, Izard, and Dehaene, 2004), as well as some Australian Aborigine (Warlpiri and Anindilyakwa) (Butterworth, Reeve, Reynolds, and Lloyd, 2008) and Melanesian groups (Beller and Bender, 2008) possess extremely limited, non-generative counting systems. The Pirahã only have quantity terms for 'one,' 'two,' and 'many.' Mundurukú lack number words beyond five, and in Anindilyakwa there are only generic number words for singular, dual, trial and plural. Yet the studies cited above found that members of these groups, like Western subjects, can exactly represent and recognize small quantities without counting, a skill known as subitizing (Kaufman, Lord, Reese, Volkmann, 1949). In addition, they can approximately represent large quantities (anything greater than 4) (Feigenson, Dehaene, and Spelke, 2004). Shown two bowls of stones, for example, they cannot determine that one has 53 and the other 37. But they *can* accurately determine which has more.

These two nonverbal systems for representing number make up the numerical core knowledge system. These nonverbal systems have been tested for and found in a wide variety of non-human animals, including salamanders (Uller, Jaeger, Guidry, Martin, 2003), rats (Capaldi and Miller, 1988), birds (Pepperberg, 1994), dolphins (Jaakkola, Fellner, Erb, Rodriguez, Guarino, 2005), apes (Beran and Beran, 2004), and monkeys (Cantlon and Brannon, 2006; 2007). Not only are certain types of mathematical operations possible without language or counting, there is a remarkable conservation across cultures and species for this innate number sense. At the same time, having a verbal, generative counting system allows for the development of mathematical abilities

(counting, exact large number representation, multiplication, algebra, etc.) far beyond what is possible without.

A Core Knowledge Approach to ToM

Viewing ToM as a core knowledge system suggests we should expect to find 1) a more basic, language-independent system that is not only universally shared in all humans, but also with non-human primates (chimpanzees at the very least) and possibly with other non-human social species as well, and 2) a language-dependent system that allows a more developed ToM in humans compared to non-humans and shows variation across cultures that are associated with linguistic differences. If ToM has an identifiable language-independent component, this may provide substantive contributions to the debate over whether non-human primates have ToM. At the same time, a better understanding of how language interacts with ToM helps illuminate why ToM appears to be so much more advanced in humans compared to non-human primates, as well as add to our understanding of the relativistic effects of language on cognition.

Leslie (Leslie, 2000; Leslie, Friedman, and T. P. German, 2004) theorizes that human ToM is governed by an innate, preverbal Theory of Mind Mechanism (ToMM), that allows the representation of beliefs and desires. More recently, Spelke and Kinzler (2007) have proposed a fifth core system, one for representing social partners, coalitions, and in- versus out-group members. In addition to the research described earlier in this chapter on infants and implicit measures of ToM, other research also suggests the existence of what can be thought of as underlying "core" elements of ToM: Belief attribution appears to be automatic (Cohen and German, 2009, but see Apperly, Riggs,

Simpson, Chiavarino, and Samson, 2006), as does taking the spatial perspective of another (Tversky and Hard, 2009). Furthermore, human subjects react more quickly when calculating others' beliefs compared to calculating other types of public representations (Cohen and German, 2010). This indicates a domain-specificity for ToM, which we should expect if it were a type of core knowledge.

Leslie's view of ToM as preverbal suggests that ToM could be found beyond humans, at least in chimpanzees and perhaps in other non-human primate species as well. Tomasello and his colleagues (Tomasello, Carpenter, Call, Behne, and Moll, 2005) have suggested that in order to understand the evolution of our ability to understand the intentions of others, we should look for a biological adaptation rooted in primate cognition. ToM research in non-human primates, however, has yet to reach a consensus on the seemingly simple question first posed by Premack and Woodruff (1978): does the chimpanzee have a theory of mind?

ToM and Non-Human Primates

In their seminal study, Premack and Woodruff presented a chimpanzee, Sarah, with videotapes of a human actor in a cage similar to her own. In these videos, the actor attempted to retrieve some bananas that were inaccessible—either attached to the ceiling, outside the cage, or blocked by a box. Sarah was also given photographs that depicted solutions to each of the conditions. In a given trial, she was shown one of the videos and presented with two pictures, one showing the correct solution and the other not. She completed 24 trials and was correct 21 times. Sarah's “consistent choice of the correct photographs can be understood by assuming that the animal recognized the videotape as

representing a problem, understood the actor's purpose, and chose alternatives compatible with that purpose" (abstract). From this study, it appears that perhaps chimpanzees do have ToM. However, there are two concerns. First, we should be cautious generalizing from a study with $n = 1$, and second, it is not a test of false belief.

Subsequent research presents a more complicated picture. Call and Tomasello (1999) presented a non-verbal false belief task to chimpanzees and orangutans, along with 4- and 5-year-old children that required the subjects to interact with two human adults, one acting as a "communicator" and the other as a "hider" who would move the location of a hidden object when the communicator was not looking (see the original paper for a complete description of the methods). In order to retrieve the object, subjects needed to recognize when the communicator held a false belief about the location of the hidden object. While the children all easily passed the task, none of the apes were able to do so. However, there was one potentially serious problem with the methodology: it placed unfamiliar demands on the animals, requiring behavioral responses not natural to them—cooperation in locating food and sharing. Thus, what appears to be a lack of ToM may have been a failure to overcome these additional demands.

Placing animals in a more ecologically valid situation, Hare and his colleagues (Hare, Call, Agnetta, and Tomasello, 2000; Hare, Call, and Tomasello, 2001) ran a series of experiments on social problem solving in subordinate chimpanzees. The animals were pitted against dominant individuals in competition over two food items. In each trial, a subordinate chimpanzee was able to observe both food items and a dominant individual. Conditions varied so that both, one or neither of the pieces of food were visible to the dominant animal. Next, both animals were given access to the area where the food is

located. In order to obtain food, the subordinates needed to modify their behavior based on what the dominant chimpanzee could see. Subjects were successful in these tasks, and appeared able to understand what the dominant individuals see and know, suggesting they do impute mental states to others. It should be noted that this is a competitive task, and chimpanzees appear to perform better on cognitive tasks that have a competitive element (Hare and Tomasello, 2004).

Chimpanzees also appear to understand intentional action (Call, Hare, Carpenter, and Tomasello, 2004) and goals of others (Yamamoto, Humle, and Tanaka, 2012). Call and his colleagues sat chimpanzees across from an experimenter, separated by a clear barrier with slots. The subjects could see the experimenter had a desirable food item, and the experimenter either acted as if they were unable to share the food through the openings (dropping it, or having difficulty transferring it through the slots) or unwilling to share (teasing the subject). The chimpanzees were able to spontaneously distinguish between these two conditions and respond accordingly: in the “unable” condition, the chimpanzees waited longer and remained calmer before leaving the testing situation, while in the “unwilling” condition, they became more agitated and left sooner. Yamamoto, Humle, and Tanaka, (2012) found that their subjects were able to engage in appropriate helping behavior on request, selecting and giving an appropriate tool to a conspecific in need. They were only able to do so if they could observe the conspecific in context—it appears they were able to assess the situation and infer the other animal's goal.

As a whole, these studies suggest that chimpanzees can and do interpret the behavior of others in terms of mental states. However, the results are mixed, some studies

suggesting chimpanzees do have ToM, others not. In a comprehensive review of this and other primate ToM research, Call and Tomasello (2008) conclude that “chimpanzees probably do not understand others in terms of a fully human-like belief-desire psychology in which they appreciate that others have mental representations of the world that drive their actions even when those do not correspond to reality” (p191). In particular, Call and Tomasello focused on chimpanzees' apparent inability to pass the false belief task—understanding that another individual can hold a belief that is contradicted by reality. Still, Tomasello (Call and Tomasello, 2008; Tomasello, Call, and Hare, 2003) is not completely closed to the possibility that non-human primates may share some of these abilities. In contrast, Povinelli (Penn and Povinelli, 2007; Povinelli and Bering, 2000; 2002; Povinelli and Vonk, 2003) is highly critical of the non-human primate ToM research. He has put forth an even stronger position, emphatically stating that chimpanzees do not interpret seeing as a mentalistic event involving internal states, and that “additional experiments will be unhelpful as long as they continue to rely upon determining whether [non-human primates] interpret behavioral invariances in terms of mental states” (Povinelli and Vonk, 2003:abstract). Only humans, Povinelli and his colleagues argue, have ToM.

One possible interpretation of the studies discussed above involves the distinction between explicit and implicit ToM tasks and the effect that additional cognitive demands have on test subjects. Chimpanzees failed tasks with high demands that require explicit responses (e.g., Call and Tomasello, 1999), which lead to the conclusion that they do not have ToM. When tasks were simplified, such that they required spontaneous or implicit responses (e.g., Call, Hare, Carpenter, and Tomasello, 2004; Hare, Call, Agnetta, and

Tomasello, 2000, Hare, Call, and Tomasello, 2001) chimpanzees were more successful, suggesting that they do understand others' behavior in terms of mental states. In fact, the unwilling/unable methodology was also used in a study of human infants (Behne et al., 2005), with similar results. And while there do not appear to be any VOE or AL tests done with chimpanzees or other non-human primates, Krachun, Carpenter, Call, and Tomasello (2009) reported a promising observation. In a competitive nonverbal false belief task they presented to chimpanzees and human children, they noted that while the chimpanzees failed the task, "the apes looked more often at the unchosen container in the false belief trials than in the true belief control trials, possibly indicating some implicit or uncertain understanding" (abstract).

Call and Tomasello (2008) are certainly correct to conclude that chimpanzees probably do not have a fully human-like belief-desire psychology. And indeed, why should they? Chimpanzees are not human. While they may not have a human-like ToM, it certainly seems appropriate to consider the existence of a full-fledged chimpanzee ToM. And if so, human and chimpanzee ToM should share some basic components. In fact, to understand human ToM, Tomasello and his colleagues (Tomasello, Carpenter, Call, Behne, and Moll, 2005) have suggested that we look for adaptations rooted in primate cognition.

These points demonstrate precisely why ToM should be considered as a type of core knowledge. To consider a cognitive ability as core knowledge requires a bottom-up approach (de Waal and Ferrari, 2010): Few researchers take issue with the continuity of anatomy, genetics or development across species—we should also include higher cognitive abilities in this list. We should not make the mistake of defining cognition from

the perspective of humans and then determining to what (limited) extent animals share these abilities. We should ask instead what component abilities we share with other animals that ultimately allow the expression of advanced cognitive abilities in humans. Rather than dividing ToM “into one 'true' form and other forms—which apparently do not deserve the name—the most fruitful approach would be to return to the classical definition and include all forms of imitation in a single framework” (Tomasello et al., 2005:205). ToM is simply the ability to impute mental states to oneself and to others (Premack and Woodruff, 1978). This definition does not make any assumptions about what species can do this, which mental states are involved, how the imputing or understanding works, other key components or abilities, or any threshold below which it is no longer considered ToM.

But in viewing ToM as core knowledge shared with non-humans, however, one assumption we must make is that it has a language-independent component. Tomasello and his colleagues (Tomasello et al., 2005) have noted that while many other theorists suggest language is what makes human cognition unique among animals, their own argument is that language should not be considered a basic ability but a derived one, and so do not attribute ToM to language. Rather, both language and ToM rest “on the same underlying cognitive and social skills that lead infants to point to things and show things to other people declaratively and informatively, in a way that other primates do not do, and that lead them to engage in collaborative and joint attentional activities with others of a kind that are also unique among primates” (Tomasello et al., 2005:690). In humans there are many other cognitive and social abilities—including language—interacting with ToM that allow for a level of expression beyond other animals, as would be expected

from a core knowledge framework.

Theory of Mind and Human Adults

Returning to human subjects, whether using the more traditional Displaced Object and related methods, or the newer, implicit measures of ToM, the focus of research in humans has been on children and infants: when do we first see evidence of its appearance, when is it firmly established? In completely non-verbal spontaneous response tests, 15-month-old (Onishi and Baillargeon, 2005) and 13-month-old infants (Surian, Caldi, and Sperber, 2007) demonstrate implicit false belief understanding. Three-year-olds demonstrate implicit false belief understanding on simplified verbal tasks (Clements and Perner, 1994). Explicit false belief understanding emerges in children between four and five years of age (Wellman, Cross, and Watson, 2001). It is therefore understandable to assume that by five years, ToM is fully realized. However, it can be useful to study the phenomenon in older subjects. In fact, Apperly, Samson, and Humphreys (2009) argue that in order to fully understand ToM, adults must be studied as well.

Consider this analogy. Apperly, Samson, and Humphreys (2009) compare studying the emergence of ToM to observing the construction of a building. During construction, we see scaffolding—temporary structures that are there only to aid in the erecting of the building—along with the permanent structure itself. It would be a mistake to view the building only during construction and assume that all the structures present are necessary parts of the final version. Likewise, if researchers “only ever study children while they are developing the ability to reason about beliefs, then it will be difficult to

find out whether language or executive function are necessary only for development, or whether they are necessary in children's belief reasoning because they are an integral part of the mature system" (p191).

False belief tasks become trivially easy after age five, but under certain conditions, adults also succumb to the "curse of knowledge" (Birch and Bloom, 2007), fail to inhibit their knowledge of reality, and make errors similar to those made by children (Birch and Bloom, 2007; Keysar, Lin, and Barr, 2003; Newton and de Villiers, 2007). Birch and Bloom (2007) found that when sufficiently sensitive measures are used, that adult subjects make the same types of false belief attribution errors that 3- and 4-year-old children make. They presented their subjects with a more complex version of the Location Change task that differed from the original in three ways: first, this version included four containers rather than two; second, they asked the subjects to give the probability that the protagonist would look in each container upon returning; and third, they rearranged the containers in order to manipulate the plausibility the protagonist would look in each one. They found that for adults, "knowledge becomes a more potent curse when it can be combined with a rationale (even if only an implicit one) for inflating one's estimates of what others know" (Birch and Bloom, 2007:385). In other words, when a subject had a potential explanation for why the protagonist might act in accord with the subject's knowledge, rather than her own false belief, this "curse of knowledge" led them to make errors similar to those made by young children on the standard version of the task.

Keysar, Lin, and Barr (2003) found similar results. Adults will either reach for or initially look to an object that has been hidden from a partner's perspective rather than

one that is visible to their partner. Adults can also inform us about individual differences in ToM (Baron-Cohen et al., 2001), and the extent to which other cognitive functions may be necessary only for its development or if they are key parts of adult ToM (Apperly, Samson, and Humphreys, 2009). Older subjects also provide us with informative data about ToM and false belief understanding, for example, whether language is necessary only for the development of ToM, or if it is a key part of the adult ToM system (Apperly, Samson, and Humphreys, 2009).

Apperly and his colleagues (Apperly, Back, Samson, and France, 2008) explored the extent to which one's own knowledge may interfere with adult subjects' making mental state attributions. They presented their subjects with two written statements, one describing the actual color of an object and second describing a man's false belief about its color. After this, subjects judged the accuracy of pictures depicting the previously read sentences. Adult subjects responded more slowly and/or made more errors when they read sentences about a belief which conflicted with reality (the picture) compared to when they read sentences about a belief that did not conflict with reality, as well as when and compared to control conditions, where the man's belief was unrelated to the object in question. Apperly et al.concluded that subjects' difficulty with the task was not related to encoding information about reality along with a conflicting false belief, but rather with keeping this information in mind and using it to base subsequent judgments upon. Other research with adults reveals additional aspects of ToM. For example, others' mental states appear to be automatically encoded by the observer in the absence of explicit instruction to do so (Cohen and German, 2009), and such encoding in turn influences subjects' expectations of outcomes (Kovács, Teglas, and Endress, 2010): targets' beliefs affect

observers' beliefs about the outcome of an observed event. And rather than remaining a fully developed skill throughout life, ToM tends to decline into old age, more so than other age-related cognitive losses (Cavallini et al., 2013).

Last, there is evidence that engaging in cooperative tasks activates brain regions implicated in ToM (Elliott et al., 2006), and ToM is positively correlated with cooperative traits, but negatively correlated with "Machiavellian" traits--using others as tools merely to achieve one's own goals (Paal and Bereczkei, 2007). Elliot and colleagues' brain imaging study presented subjects with the image of a coin, asked them to choose heads or tails in one of four conditions: 1) playing with another player, with financial rewards available (if both players' guesses match the computer, they win money); 2) playing with another player, no financial rewards; 3) playing alone, financial rewards available (if subject's guess matches the computer, he or she wins money); and 4) playing alone, no financial rewards. The “other player” in the cooperation conditions did not exist--it was the computer. The researchers found that playing a game in cooperation with another person was also associated with activation of brain areas involved in ToM (medial prefrontal cortex, temporal pole and temporoparietal junction). Financial rewards for winning trials did not appear to affect these areas--thus, it was the cooperation and not the financial reward that activated the ToM regions.

Paal and Bereczkei (2007) also found a link between cooperation and ToM. The ability of adult subjects to engage in ToM was correlated with the likelihood they would engage in cooperative interactions with others and provide support if needed. In contrast, they found a negative correlation between Machiavellian traits and both social cooperative skills and ToM. The researchers found that these subjects default to

attributing negative intentions to others and as a result, do not expect to receive cooperation. They begin with the assumption that others will exploit them, if they fail to exploit those others themselves (Repacholi, Slaughter, Pritchard, and Gibbs, 2003; Wilson, Near, and Miller, 1998).

Work with adult subjects has also shed light on the automaticity of ToM. When subjects are tested on the contents of target individuals' false belief, they respond more quickly than when responding to questions about reality, but only when the question occurs close in time to the situation that lead to the targets' beliefs (Cohen and German, 2009). This effect is found even when subjects are not given any overt instructions to do so and is as quick as when subjects are given such instruction. Cohen and German take this as evidence that we automatically encode others' beliefs in certain situations, but unless instructions are given to attend to/remember those beliefs, that encoding will not last. Not only do we to automatically encode others' beliefs, but they also appear to affect us similarly to our own beliefs (Kovács, Teglas, and Endress, 2010). When given a visual object detection task, both subjects' own beliefs as well as the beliefs of an agent (that were irrelevant to performing the task) affected adult subjects' reaction times (as well as infant subjects' looking times). Simply including another agent during the task was enough to automatically trigger subjects' belief computation, an effect that persisted even after those agents left the scene. Finally, subjects who are actively engaged in a ToM task process others' beliefs sooner than passive observers (Ferguson, Apperly, Ahmad, Bindemann, and Cane, 2015).

And finally, there is also evidence that in addition to being a developmentally emerging ability, ToM may also degrade over adulthood. Cavallini et al. (2013) presented

young, young-old, and old-old adults (20-30 years of age, 59-70, and 71-82, respectively) with the Strange Stories (Happè, 1994) task and a test of executive brain function. Strange Stories consist of a series of short stories followed by questions that require subjects to infer the characters' thoughts and feelings. To respond correctly, subjects need to understand that a character's underlying intention behind a statement is not literally true, such as in jokes, lies, sarcasm, figures of speech, or pretense (Cavallini et al., 2013). Cavallini and colleagues found that young adults outperformed both of the older adult groups, even when controlling for decline of executive functions (working memory and inhibitory control) that is also seen in old age; ToM appears to decline in old age, independent of other age-related cognitive decline.

Conclusion

To end this overview of ToM, I return to an idea discussed at the beginning of this chapter: the simulation account of ToM (Goldman, 2006) allows us to view ToM as a type of empathy in that it, along with motor and emotional empathy all function similarly—they reproduce in the observer's mind a state that is more appropriate to the situation of an observed individual than their own (Preston and de Waal, 2002). That is, these processes mirror the mental state of another individual in our own minds. This idea of empathy as mirroring is explored in the next chapter.

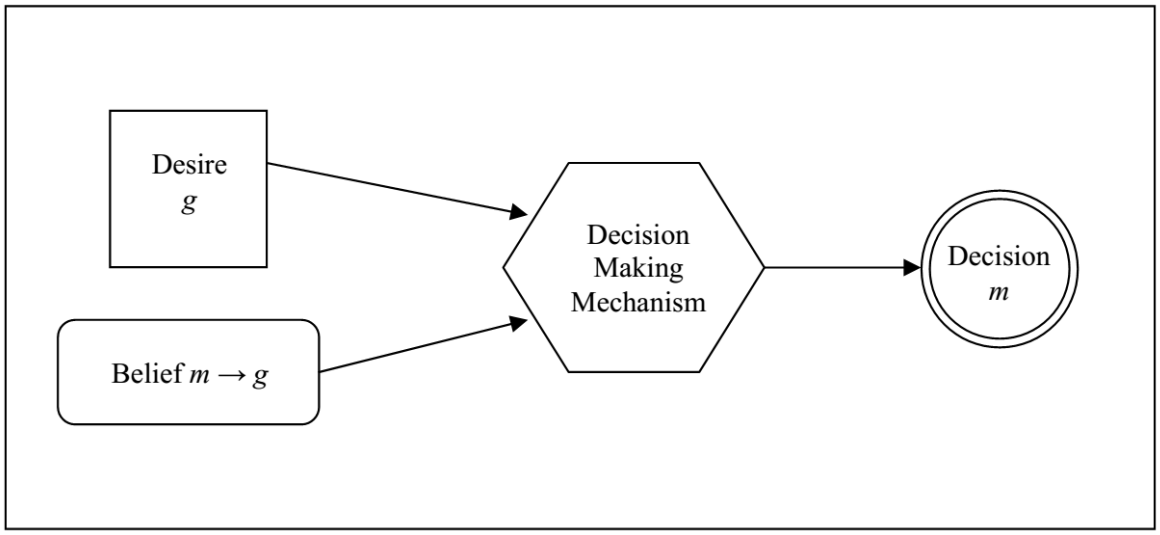


Figure 2.1: Decision by an individual to do *m*. The individual has a desire *g* and believes that *m* results in *g*. The desire and belief are processed in a decision making mechanism, which outputs the decision to do *m* (adapted from Goldman, 2006).

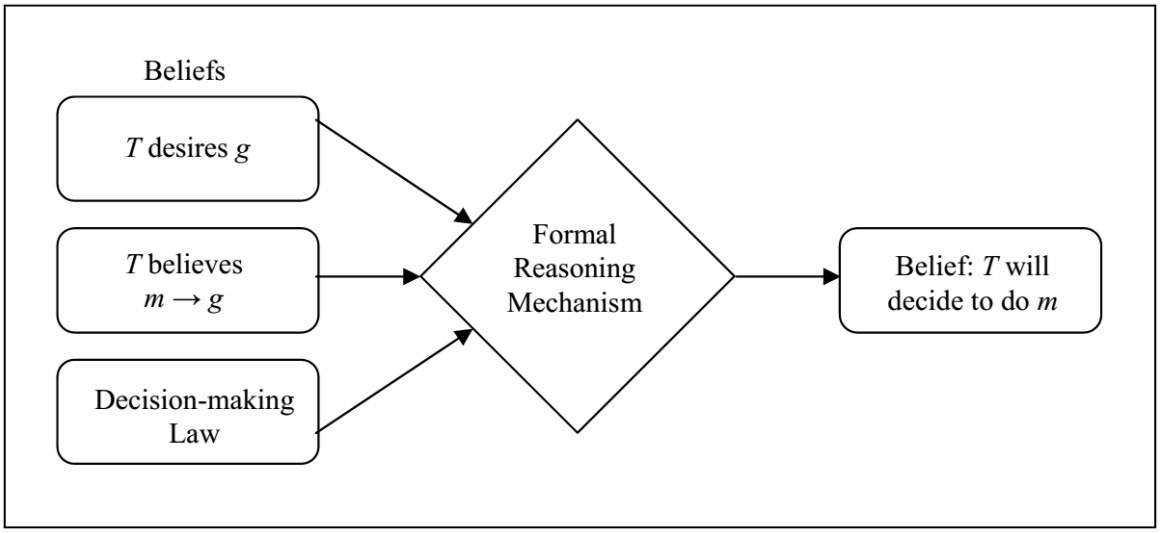


Figure 2.2: Decision attribution reached by theory-based inference. An observer holds beliefs about an individual *T*'s beliefs and about a decision-making law. These beliefs are processed in a formal reasoning mechanism, which outputs the belief that *T* will do *m* (adapted from Goldman, 2006.)

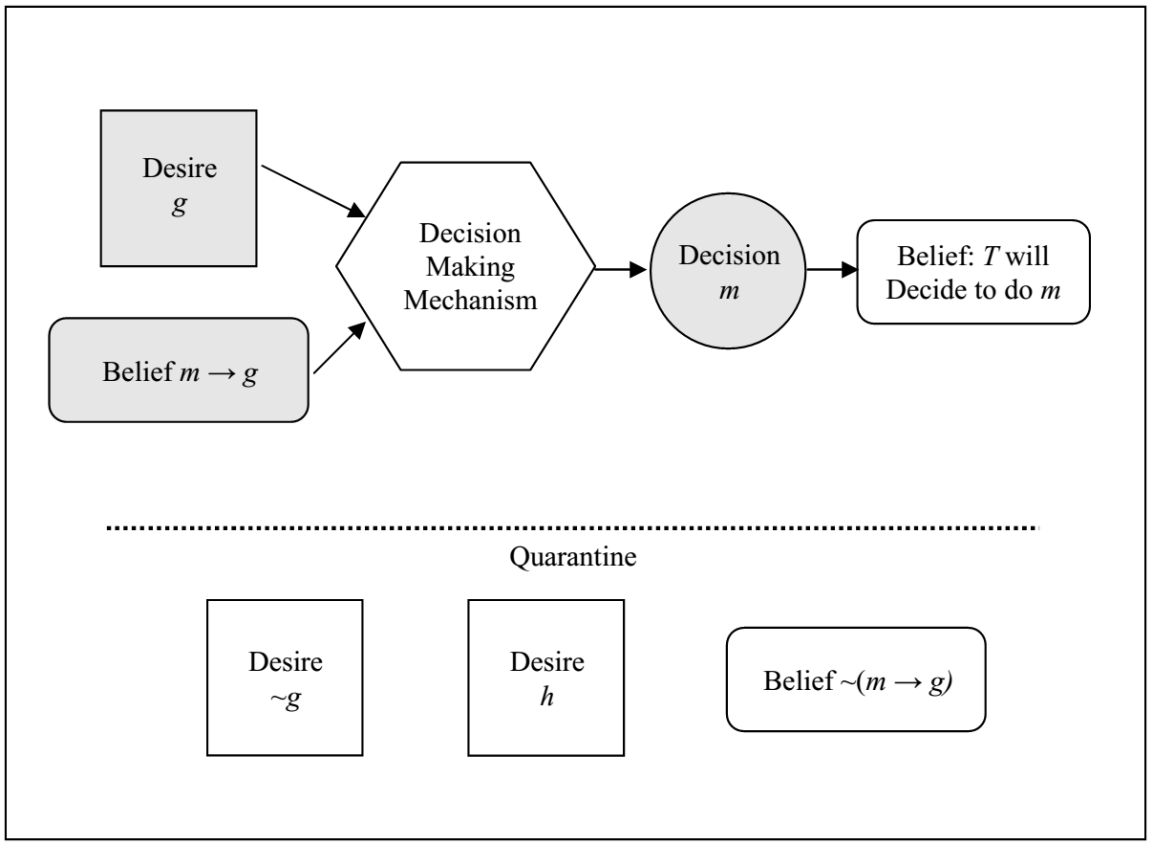


Figure 2.3: Decision attribution reached by simulation. Here, the gray shapes represents 'pretend' states, assumed to be held by *T*. The observer quarantines own beliefs and desires (which may differ from *T*'s) and runs the pretend states through one's own decision making mechanism, notes the result, and forms the belief that the target will also decide to do *m* (adapted from Goldman, 2006).

Chapter 3: Empathy, Mirroring, and Theory of Mind

Empathy is commonly considered to be the ability to put one's self in another person's shoes, to not only see things from their perceptual perspective, but from their emotional perspective as well. More technically, empathy is any process where the attended perception of a target's state generates a state in the observer that is more applicable to the target's state or situation than to the subject's own prior state or situation (Preston and de Waal 2002): when an individual observes another, some aspect of the target's internal state is replicated, or mirrored, within the observer. The observer experiences the same internal states as the other, though not necessarily at the same intensity. It is not strictly a cognitive ability, but also a perceptual one: as is discussed below, the visual stimulus triggers the same brain regions in the observer that are responsible for producing the observed behavior/mental state. For example, viewing images of another in physical pain does not merely evoke a sympathetic understanding of pain and what that pain must feel like. Rather, it activates brain regions in the observer that are involved in processing one's own pain (Jackson, Meltzoff, and Decety, 2005).

ToM is a cognitive ability. But to fully capture its essence, it might be better to consider ToM as a perceptual ability as well. While we can, of course, sit and think about our own and others' mental states, ToM is a skill we actively use on a daily basis interacting with others. Understanding what we perceive and how we perceive is fundamental for understanding the resulting attributions. ToM appears to be based in a system that responds not only when an individual experiences a mental state, but when *observing* another express the same state (Gallese, 2007). Given this definition, ToM (specifically, the simulation account) can be considered as a type of empathy as well—

one of three main types, along with motor empathy and emotional empathy (Blair 2005; Preston and de Waal, 2002), discussed below.

Motor Empathy

Motor empathy/mirroring is "the tendency to automatically mimic and synchronize facial expressions, vocalizations, postures and movements with those of another person" (Blair, 2005:700). Certainly, this is a vital skill to possess, else it would not be possible to imitate others and learn the many skills needed in day to day life and acquire the cultural practices accumulated in one's society. However, to be successful, motor empathy does not require that one explicitly or overtly copies the actions of a target individual or to even move at all. Rather, as with emotional empathy, it is only necessary that the action is recognized for what it is, by activating at a sub-threshold level the same area in the brain of the observer that is involved in the production of the action. In fact, there is evidence that inhibiting one's tendency to imitate may lead to improvement on perspective taking ToM tasks (Santiesteban, White, Cook, Gilbert, Heyes, and Bird, 2011). At the same time, however, the tendency to imitate others is strong, and automatic imitation can occur even when there are incentives to avoid doing so (Belot, Crawford, Hayes, 2013).

Mirror Neurons

The function of mirror neurons provides a promising starting point in explaining how visual stimuli may be processed and understood as social information. These neurons, unknown until almost twenty years ago, were found by accident, during

Rizzolatti and colleagues' (Gallese et al., 1996; Rizzolatti et al., 1996) study of motor neurons (now referred to as canonical motor neurons) located in area F5 of the macaque's premotor cortex (see Figure 3.1). Cannonical motor neurons in this part of the brain are involved in the planning and execution of various actions, such as reaching for a food item. However, while setting up trials for individual monkeys, these researchers discovered that some individual neurons fire not only when a monkey performed these actions, but also when it observed the experimenters performing the same actions.

The term 'mirror neuron' may imply a single type of neuron and homogeneity of function. While this is true in the broadest sense—they fire in response to both motor and perceptual tasks—it is more accurate to consider them as various types constituting a class of neuron. Within this class, neurons can vary across two main dimensions: the types of observed actions they respond to and the types of performed actions to which they respond. In addition, they can be grouped according to the degree of congruence they exhibit in their responses between observed and performed actions.

Goal-Directed Actions: Mirror neurons do not merely fire in response to the presentation of a target object stimulus, nor to an action in and of itself. Rather, they respond to goal-directed actions; they require both the presence of a target object and an entire action sequence, not to any of its isolated or specific components. In other words, an individual mirror neuron may respond (whether observing or performing) to the entire sequence of reaching for and grabbing an item of food, not to any one of these components alone. Typically, mirror neuron respond to only one type of action. Gallese et al. (1996) identified several, including grasping, placing, manipulating, hand interaction, and holding neurons.

Interestingly, while they do not fire (at least in macaques) in response to mimed actions in the absence of a target object, there is a subset of mirror neurons that do respond to an observed goal-directed action when the target object is hidden from view (Umilta, Kohler, Gallese, Fogassi, Fadiga, Keysers, and Rizzolatti, 2001). Macaques in these experiments first observed a goal object being placed behind a blind, then, when they observed a researcher or another monkey reaching behind the blind, certain mirror neurons would fire in response. Simply observing a researcher's hand reaching behind the blind did not activate the neurons; the monkeys first need to observe the target object before it is hidden, and then see it being hidden. This serves to underscore the fact that mirror neurons are sensitive to goal-directed actions, and without awareness of a target object, it seems they cannot recognize an action as such.

Visuo-motor Congruence: Mirror neurons can be further described in terms of the specificity of their firing. Gallese et al. (1996) classify three broad levels of visuo-motor congruence: First, s*trictly congruent* neurons only fire in response to both a specific action (such as grasping) and the way in which the action was executed (such as a precision grip). Second, *broadly congruent* neurons respond more generally to varying combinations of observed and executed actions. Broadly congruent neurons can be further subdivided into three variants: a) some fire to a specific executed action (e.g. a precision grip) but fire more generally in response to any type of observed grasping; b) in others this pattern is reversed; and c) some appear to respond simply to the general goal of an observed or executed action, regardless of how it is achieved. And third, *non-congruent* neurons do not appear to show any specific relationship between their firing for executed versus observed actions. In other words they seem to fire in response to any

produced or observed goal-directed action.

Additional mirror neuron classes: The mirror neurons discussed to this point all respond to arm and hand movements. In addition to these, two further types of mirror neurons have been identified. Kohler, Keysers, Umilta, Fogassi, Gallese, and Rizzolatti (2002, cited in Rizzolatti and Craighero 2004) found *audio-visual* mirror neurons that respond to auditory stimuli, firing when a monkey observed a noisy action (such as crumpling up a piece of paper) and also when they were presented only with the noise that accompanies the action. And subsequent research by Rizzolatti's colleagues (Ferrari, Gallese, Rizzolatti, and Fogassi 2003, cited in Rizzolatti and Craighero 2004) has identified yet other mirror neurons that respond to mouth movements, particularly ingestive (e.g., grasping food with the mouth) and communicative (e.g., lip smacking) actions. This last type of neuron is particularly interesting in light of the present paper; their involvement in a facial expression mediating social behavior suggests a possible connection between mirror neurons and an evolutionary origin for ToM.

Mirror neurons in humans

Of course, the implications of mirror neurons for understanding human ToM and sociality would be pointless if they were unique to macaque brains. Initial studies with human subjects indirectly confirmed their presence in the human brain using methods such as transcranial magnetic stimulation (TMS) to detect motor evoked potentials (MEPs) recorded from observers' muscles (Fadiga, Fogassi, Pavesi, and Rizzolatti 1995; Gangitano, Mottaghy, and Pascual-Leone, 2001). TMS is a method that uses a directed magnetic field to temporarily excite or inhibit specific areas of the brain (see Hallet,

2000), and MEPs are electrical signals that can be detected from within muscles following magnetic stimulation of the corresponding brain area. If a signal is detected from a motor neuron during a subject's observation of an action, it can be inferred that the neuron has mirror properties. These studies show that the human motor system appears to function similarly to the macaque's, responding to both the goal of an observed action and the manner in which it is carried out. There is an important difference, however: the human mirror system can also respond to mimed or meaningless gestures without apparent goals (Buccino, Binkofski, Fink, Fadiga, Fogassi, Gallese et al., 2001; Grezes, Armony, Rowe, and Passingham, 2003).

There is now direct evidence for mirror neurons in humans. Mukamel, Ekstrom, Kaplan, Iacoboni, and Fried (2010) obtained single-neuron recordings in 21 subjects and found a subset of neurons in the supplementary motor area respond to both the observation and execution of facial expressions and grasping motions in a manner similar to the response pattern of mirror neurons seen in macaques.

Mirror Systems

Mirror neurons do not simply fire in isolation. There is a basic cortical mirror neuron circuit in both macaque and human brains (Rizzolatti and Craighero, 2004), which in addition to area F5 of the ventral premotor cortex, also includes area PF of the rostral inferior parietal lobule (in humans, it includes the rostral parietal lobule, the caudal sector of the inferior frontal gyrus (IFG), and the adjacent part of the premotor cortex). In macaques, the rostral inferior parietal lobule projects to area F5. It, too, has both visual and motor properties and in turn, receives inputs from the superior temporal sulcus

(STS). The STS is a visual area that responds to a broader range of movements than PF and F5. Unlike the other two areas, the STS does not exhibit any motor properties and though it provides input, it is not considered part of the mirror neuron circuit (see Figure 3.1). (For a full description of the mirror neuron circuit and related perceptual and motor areas, see Rizzolati and Craighero 2004, pp 171-172, 176-178, and Figure 1).

The Functional Role of Mirror Neurons

What adaptive problem do mirror neurons solve? Since their discovery, many functional roles have been suggested. The original hypothesis was that they mediate action recognition (Buxbaum, Kyle, and Menton, 2005; Gallese et al., 1996; Rizzolatti et al., 1996). Other propositions include action perception (Thornton and Knoblich, 2006; Wilson and Knoblich, 2005), imitation (Iacoboni, 2005), perspective taking (Jackson et al. 2006), language (Fadiga and Craighero, 2006; Rizzolatti and Arbib, 1998; 1999), empathy (Blair, 2005; Duan, 2000; Preston and de Waal, 2002) and theory of mind (Gallese and Goldman, 1998; Singer, 2006). The inclusion of language, perspective taking, and ToM in this list indicates a turn towards attempting to understand the role of mirror neurons in human behavior. Unfortunately, mirror neurons became a fashionable topic in the years following their discovery, and an additional and somewhat odd assortment of other roles has also been proposed. They have been implicated in smoking behavior (Pineda and Oberman, 2006), sexual orientation (Ponseti, Bosinkski, Wolff, Peller, Jansen, Mehdorn et al., 2006) and even contagious yawning (Schurmann, Hesse, Stephan, Saarela, Zilles, Hari, and Fink, 2005)—interesting, but some of these are quite removed from any plausible evolutionary function, let alone any role they might be

expected to play in macaque behavior.

Despite this array of possibilities, enough is known to draw some inferences about what mirror neurons actually do—many of these proposed functions share an underlying commonality. Mirror neurons and systems are thought to "embody" (Gallese, Keysers, and Rizzolatti, 2004; Gallese, 2005, 2007) observed actions, emotions, beliefs, desires or other mental states, and they do so by replicating, or simulating them within the mind of the observer. In this, mirror neurons are a physical basis for empathy, as defined at the beginning of this chapter. Rather than analyzing one's observations of another individual's behavior by mapping it to some type of theory or set of behavioral rules, as in the Theory-Theory account of ToM, one literally experiences the other's actions, emotions and beliefs directly, without a need for translation or higher-level cognitive interpretation.

Even a fully human ability such as language is consistent with this core function of mirror neurons. The macaque brain's Area F5 is homologous to Broca's area in humans—one of the brain regions involved in speech production. Because of this link between F5 and Broca's area, mirror neurons have been implicated in providing a basis for the development of human language from gestural precursors (Rizzolatti and Arbib, 1998). The reasoning here is that mirror neurons fire only for complete goal-directed actions, not for the individual components thereof. Thus, in such actions there is an ordering and fluency, a "syntax" of motion. Just as one could not say in English, "Going am object I grab to the", one cannot grab an object prior to reaching for it. (People with damage to Broca's area are still able to pronounce individual words, but lose their fluency and cannot speak in full sentences.) Rizzolatti and Arbib suggest that verbal

fluency and syntax developed as vocalizations were increasingly paired with gestures. It is an interesting hypothesis, but while it could indeed be the case that the mirror neuron system was a precursor to verbal language, language itself is an unlikely candidate for an answer to the question of why they exist in the first place. Rather, for mirror neurons to embody observed actions, actions must be meaningful. To be meaningful, they must be complete and ordered.

Imitation and action understanding

Rizzolatti and his colleagues (Rizzolatti et al., 1996, Gallese et al., 1996) define action understanding as an "automatically induced, motor representation of the observed action [that] corresponds to that which is spontaneously generated during active action and whose outcome is known to the acting individual" (Rizzolatti and Craighero, 2004:172). In other words, observing another's actions causes a sub-threshold activation of the same motor plan in the observer, as if the observer were to engage in the same action without actually triggering the action. This motor plan allows the observer to understand the goal or intent of the viewed action, which is potentially beneficial in allowing more accurate predictions of others' behavior. There is some evidence for variation in this function from macaques to humans. As discussed above, the presence of an object is necessary to activate mirror neurons of the macaques; they fire only during the observation or production of goal-directed actions, such as reaching for a food item. In humans, however, they respond to a broader range of actions, and the presence of a goal object (or the knowledge of its presence) is no longer a necessary condition in order to mirror another person's actions. This suggests that the evolution of mirror neurons was

due to some selective pressure to recognize goal-directed actions.

This could have interesting implications for human sociality and the spread of cultural behaviors. Mirror neurons provide the basis for imitation in humans (Iacoboni et al.1999). Culture can be simply defined in evolutionary anthropology as socially transmitted information (Alvard, 2003; Barkow, 1989; Cronk 1995; 1999). Considering that our mirror neuron system is not limited to goal-directed actions—we can imitate simply for the sake of imitation—mirror neurons might provide a mechanism for the spread of cultural variation.

This should not be taken to imply that cultural transmission, at least in humans, is only possible through direct imitation of observed actions; recent evidence suggests otherwise (Caldwell and Millen, 2009), nor that imitation is necessary for action recognition. While monkeys do not appear to imitate (Lyons, Santos, and Keil, 2006, but see Dindo, Thierry, de Waal, and Whiten, 2010), they do appear to recognize when they are being imitated (Paukner, Anderson, Borelli, Visalberghi, and Ferrari, 2005). Chimpanzees, however, do exhibit some imitative capacities, but what they do has been characterized as *emulation* (Horner and Whiten, 2005), where the emphasis is on achieving the same goal as the observed actor, while imitation is the production of an exact copy. In contrast to emulation, imitation is thought to be a higher species-level cognitive function (Iacoboni 2005; but see de Waal and Ferrari, 2010). Humans are capable of true imitation, at times to a fault, over-imitating and copying parts of an overall action sequence that are not relevant to or helpful in achieving a particular goal (Horner and Whiten, 2005). However, one limitation of Horner and Whiten's (2005) study was that it required chimpanzees to copy the actions of a human. Studies utilizing

chimpanzee models reveals that these animals are capable of imitation, even of arbitrary actions (Bonnie, Horner, Whiten, and de Waal, 2007; Horner and de Waal, 2009; Whiten, Horner, and de Waal, 2005).

Action perception

The terms “action recognition” and “action perception” may sound similar, but they do refer to different abilities. Action recognition, as described above, is postdictive, "the action I just saw was *X*" whereas action perception is predictive, "based on his movements, I predict the target will do *X*". While it is certainly useful to recognize an observed action, why it was performed, or even be able to reproduce the action, it would be even more useful if the mirror neuron system allowed prediction, to allow an observer to infer what the goal of another individual will be. This predictive aspect is included in models of ToM reasoning (Child Scientist and Simulation Theory, Figures 2.2 and 2.3, Chapter 2), and tests such as the anticipatory looking task, but it is not explicitly included in the action understanding function of mirror neurons. For that, it is necessary to consider another, related, proposed mirror neuron function: action perception (Wilson and Knoblich, 2005).

Wilson and Knoblich begin their discussion of action perception by noting that although humans can imitate, imitation is not the typical response to watching others' actions. In fact, they note that there are structures in the spinal cord specifically for inhibiting undesired imitative action. This leads them to ask why the brain generates a motor plan that goes nowhere. They suggest it is not for action understanding, but for the perception of the behavior of conspecifics. This may seem to be the same thing, but there

is a crucial difference. Action understanding is *postdictive*: It does not posit an impact on ongoing perceptual processing of the external event, but rather draws inferences about the motives or purposes of actions. In contrast, action perception is *predictive*—it projects the probable future course of an ongoing event (although Wilson and Knoblich do not preclude a hybrid theory, incorporating both predictive and postdictive functions).

Despite their recent popularity, many questions about mirror neurons remain unanswered. How evolutionarily ancient or conservative are they? Given that both macaques and humans have mirror neurons, is it reasonable to expect that they will be found in other non-human primates as well, at least in apes and Old World monkeys? Should we expect to find them in New World monkeys or prosimians? What about other mammals? Recently, it has been demonstrated that birds have auditory-vocal mirror neurons (Prather et al.2008). While this could be a case of convergent evolution in birds and primates, it seems more likely that mirror neurons will be found in a wide variety of other animal species animals. Predicting and reacting to the behavior of conspecifics (and prey or predators) is a useful trait.

Besides simply confirming the distribution of mirror neurons across species, it would be informative to determine the full range of actions (beyond hand/arm and mouth gestures) that activate them. In addition, is there a correlation between their presence and brain size, perhaps a concomitant increase in the absolute number of mirror neurons or in the ratio of mirror neurons to canonical motor neurons as relative brain size increases? In other words, does mirroring become more sophisticated or complex as brain size increases across species? Can anything be said about a possible trade-off between mirror and olfactory systems in the primate brain compared to other species (similar to the trade-

off between olfactory and visual systems)? Also, how might the great ape (particularly chimpanzee) mirror neuron system respond in the absence of a goal: will they perform like macaques, like humans or somewhere in between?

Despite these as yet unanswered questions, mirror neurons and mirror systems provide an underlying neurological explanation for empathy as it has been defined herein--a process in which the perception of a target's state (e.g., behavior or emotional expression) generates a similar state in the observer that is more applicable to the target's situation than to the subject's own prior situation (Preston and de Waal, 2002). Mirror neurons/systems do just this.

Emotional Empathy

In humans, the recognition and production of some emotions and emotional expressions also rely on shared neuroanatomical regions. This mirroring does not result in actual mimicry of the emotions and expressions, but occurs at a sub-threshold, preconscious level (Goldman, 2006). The best evidence of this process comes from studies of brain lesions or disorders related to negative emotions such as anger, fear and disgust (Goldman and Sripada, 2005). For example, Adolphs and colleagues (Adolphs, Tranel, Damasio, and Damasio, 1994) studied patient "SM", who suffers from the bilateral destruction of his amygdalae. These researchers showed that SM is abnormal in both his experience of fear and his ability to acquire a conditioned fear response. In addition he was unable to recognize the facial expression of fear in others, an inability that was limited to this single emotion. More recently, Ashwin, Baron-Cohen, Wheelwright, O'Riordan, and Bullmore (2007) also studied fearful face processing, but

looked at the performance in autistic subjects versus controls. These researchers also found that the amygdala was involved, demonstrating its differential activation (along with other brain areas involved in social cognition) between the two groups. The first study directly demonstrating a mirroring mechanism in emotions was done by Wicker, Keysers, Plailly, Royet, Gallese, and Rizzolatti (2003). Using fMRI, they showed there is a common neural basis for seeing and feeling disgust. Whether experiencing disgust themselves, after inhaling disgusting odors or viewing video clips of the faces of others experiencing disgust, subjects showed similar activation in the insula, the brain region involved in processing this emotion. However, it should be noted that no single-neuron tests have been performed on these other areas in non-human or human primates. Thus, we cannot distinguish between single neurons that respond to both production and observation versus a single area that responds to both types of input via separate types of neurons.

Variation in Empathy

As with ToM, neither emotional nor motor empathy are perfect nor uniformly applied to all target individuals. Liew, Han, and Aziz-Zadeh (2011) presented Chinese subjects with short video clips of actors, either Chinese or Caucasian, performing familiar and unfamiliar symbolic gestures which have been shown to activate both ToM and mirror neuron areas of the brain. The subjects' task was to infer the intentions of the actors in the clips. Liew et al. found that familiarity was a key factor in brain activation: 1) viewing clips of actors of the same race was associated with greater activation of mirror neuron regions, and 2) familiar gestures were associated with greater activation of

ToM areas, but unfamiliar gestures were associated with greater activation of the mirror neuron regions. However, when it comes to actual imitation, or rather, overimitation, familiarity of the target individual may not affect behavior (Nielsen and Tomaselli, 2010).

Emotional empathy shows a considerable degree of variation across individuals in response to different targets and situations. To begin, Jackson, Meltzoff, and Decety (2005), showed subjects pictures of others' hands and feet (from a first-person perspective) in either neutral or painful situations (such as opening a door, or caught in a closing door). They found subjects' anterior cingulate cortex (ACC) and anterior insula, brain areas involved in processing one's own pain, were active when subjects viewed the painful situations. As expected from the mirroring account of empathy discussed above, simply viewing others in pain activates some of the same brain areas involved in processing one's own pain. But it is not quite so straight forward. Earlier by work Duan (2000) demonstrated evidence of an interaction between subjects' motivation to empathize and targets' emotions. The study considered two types of empathy: intellectual empathy, when one takes another's perspective; and empathic emotion, the degree to which one feels the same emotions another displays. Duan found happy or sad targets elicit more empathic emotion compared to those expressing anger and shame. Motivation to empathize increased subjects' intellectual empathy when a target was sad, and increased empathic emotion when the target was happy.

The degree to which humans experience emotional empathy towards others is influenced by other factors as well, including familiarity with the target individual and group membership. Strength of empathic responses to in-group members varies directly

with the degree of perceived similarity of the observer and target individual (Stürmer, Snyder, Kropp, and Siem, 2006). Xu, Zuo, Wang, and Han (2009) showed Caucasian and Chinese subjects images of faces of both racial groups experiencing either painful or non-painful stimuli. When viewing their own racial group, subjects showed increased activations in pain processing- and empathy- related brain areas. This activity decreased significantly when viewing racial out-group faces. Such discrimination is also modulated by one's attitudes towards the out-group, with greater prejudice being associated with greater decrease in brain activation.

Again, the picture is not quite so simple—the division of in-group vs. out-group is more complex than it first appears. Tarrant, Dazeley, and Cottom (2009) found that empathic responses (measured via self-report) were higher for in-group members compared to out-group members. However, when subjects were primed with an in-group norm promoting empathy, they reported feeling more empathy towards out-group members. Avenanti, Sirigu, and Aglioti (2010) presented Black and White subjects images of black-, white- or violet-colored hands in painful and non-painful situations. All subjects tested exhibited implicit but not explicit racial in-group preferences. As expected and in line with the other studies presented here, when observing in-group hands receive pain (White subjects observing white hands, Black subjects observing black hands) subjects responded as if they felt the pain also. This was not the case when observing the other group's hands (White subjects viewing black hands, Black subjects viewing white hands). However, the response of both groups to the violet-colored hand was similar to their responses to their own colored hands. This opposite color versus violet effect was more pronounced in subjects who held stronger implicit racial biases; these results

suggest that we can and do respond with empathy to strangers, but our response is tempered by preexisting (negative) cultural biases. Similarly, a later study (Axt, Ebersole, and Nosek, 2014) looked at implicit racial biases of subjects reporting as one of four different races and also found that subjects tend to view their own race most positively; their implicit evaluation of other races follow a hierarchical pattern, with Whites evaluated most positively followed by Asians, Blacks, and Hispanics.

Using a more realistic/real-world design, Bruneau, Dufour, and Saxe (2012) found a similar pattern of results. They asked Arab, Israeli and South American subjects to consider the physical and emotional pain and suffering of individuals from each of the three groups. The Arabs and Israelis, groups in long-standing conflict with each other, both reported feeling significantly less empathy for the pain and suffering of the other group's members, but neither showed the same bias in their responses towards the South American targets. Instead, in both Arabs and Israelis, the brain regions that respond to others' suffering showed an in-group bias in response to the South Americans—the distant out-group—but not for the conflict out-group, especially in response to descriptions of emotional suffering. Here, as in Avenanti et al.'s experiment, out-group alone does not equate to reduced empathy. Preexisting cultural/historical biases are needed. Without those biases, out-group members are likely to be treated as would members of one's in-group.

People may also share common beliefs about a specific racial or other out-group, regardless of their own race, and these beliefs have an impact on empathic responses. For example, Trawalter, Hoffman, and Waytz (2012) found that both Black and White subjects assume that Blacks feel less pain than Whites, which affected all subjects'

empathy of Blacks' pain. In addition, this was not simply due to perceptions of race, but was connected to perceived status and privilege/hardship of target individuals. Trawalter and colleagues presented White, Black, and Nursing school students with pictures of White and Black individuals and had them rate the amount of pain these people would experience in response to various stimuli. Black faces were rated as significantly lower in pain response. In addition, when subjects recruited from Mechanical Turk viewed images of Black/White morphed faces, faces labeled as Black (though the same images as those labeled White) were rated significantly lower in experience of pain. However, when the experimenters controlled for the perceived level of status/privilege of target individuals, the effect of race was eliminated.

Gutsell and Inzlicht (2010) found that activation of motor areas in the brain that respond to observation of others' action occurred when the observed target individual was a member of the subjects' in-group but not when the target individual was a member of an out-group. In their study, subjects' (White Canadians) motor cortex activated both when subjects performed actions and when they observed other white individuals act. However, there was less spontaneous/ implicit mirroring activation when they observed out-group members (South Asians, Blacks, East Asians). Activation decreased as prejudice and dislike of out-groups increased. In a follow-up study presenting subjects with the same three racial groups, Gutsell and Inzlicht (2012) again found a similar pattern of results—empathy appears to be limited to in-group members. Subjects showed similar brain activation patterns both when feeling sad and when observing in-group members feeling sad. This activation was not seen when observing out-group members, and this effect was greater the more subjects were prejudiced towards the out-groups.

Cikara and Fiske (2011) studied how stereotypes of different (non-racial) out-groups modulate empathic responses to other's misfortunes. Their subjects viewed and rated nine positive, nine neutral, and nine negative events, each randomly paired with an image of an individual representing one of four target groups (pride, envy, pity, and disgust). They reported three key findings. First, compared to observing in-group members, subjects feel least empathy in response to observed misfortunes when the paired target is envied, and they feel the most empathy when targets individuals are members of a pitied group. Second, subjects are least willing to endorse harming pitied targets, with an exception: subjects who showed an increase in activation of the insula and middle frontal gyrus in response to pitied target/positive event pairs reported feeling worse about those events and were more willing to endorse harm to those targets. And third, subjects who showed an increase in activation in the bilateral anterior insula in response to positive events reported greater willingness to harm envy targets, but a decreased willingness to harm in-group targets.

Other factors that affect empathizing observers' gender and targets' previous behavior. Singer and colleagues (Singer et al., 2006) had subjects play cooperation (sequential Prisoners' Dilemma) games with confederates (one who played fairly and one unfairly), after which the subjects observed the confederates receive pain. Both male and female subjects showed empathy-related activation when they viewed the fair players receive pain. However, there was a sex difference in response to viewing the unfair players receive pain: Empathy-related activation was significantly reduced in male subjects. In addition, males also showed increased activation in reward-related brain areas, along with an expressed desire for revenge on the unfair players in the form of

physical punishment. Males' empathy is affected by their evaluation of others' social behavior.

The above studies suggest that learned adults' attitudes towards out-group members affect one's response to them. It also appears that sensitivity to out-group members emerges in adolescence and is the result of exposure. Telzer, Humphreys, Shapiro, and Tottenham (2013) found that while adults show a differential activation of the amygdala in response to faces of members of different races, this activation emerges over development and is not indicative of an inborn process (a finding in line with Kurzban, Tooby, and Cosmides' (2001) position—see Chapter 4). They presented children, ranging from 4 to 16 years of age, images of European American and African American faces during fMRI. The differential response to African American faces was not seen in the younger children, and did not appear until adolescence. However, in children who were raised in a more diverse environment, this differential activation to African American faces was reduced, suggesting that exposure reduces reaction to those faces as members of an out-group.

Mimicry is another way to reduce prejudice and increase empathy. Gutsell and Inzlicht (2010, 2012) demonstrated that prejudice reduces empathy, but found evidence that explicit mimicking of out-group members has the opposite effect (Inzlicht, Gutsell, and Legault, 2012). They had White subjects watch videos of actors repeatedly reaching for, picking up, and drinking from a glass of water in one of three experimental conditions: 1) passively watching Black actors, 2) watching and mimicking Black actors, and 3) watching and mimicking actors one's own in-group. Afterwards, subjects then completed implicit and explicit measures of racism and anti-Black prejudice. Inzlicht and

colleagues found that subjects who imitated Black actors showed similar implicit preferences for both Blacks and Whites; the other two groups preferred Whites over Blacks. The group mimicking Black actors also reported less explicit racism towards Blacks than those who mimicked in-group actors. They concluded that mimicking members of an out-group appears to reduce more general implicit (and possibly explicit) biases against that out-group.

In a related study, using a design that might better translate to the real world, Brannon and Walton (2013) studied the interactions of White subjects with Latino target individuals. They found that creating feelings of social connectedness with a member of an out-group can lead to a reduction of prejudice and foster interest in the out-group member's culture. Here, they found that even simple physical mimicking by the out-group member (rather than mimicking the out-group member as in Inzlicht et al.'s work) had such an effect on their subjects. But the biggest effect, persisting at a 6-month follow-up, occurred when non-Latino subjects were able to freely choose to participate with an Latino target (out-group member) in an activity that was culturally relevant to that individual.

These last several studies appear to collectively suggest that a default strategy of empathy towards out-group members, particularly when those individuals are part of a marginalized or lower class group, may be beneficial in fostering intergroup harmony. Yet another factor needs to be considered: does the out-group member want or need empathy? Vorauer and Sasaki (2012) found that attempting to be empathic in intergroup interactions can have a positive effect on the empathizer's subsequent behavior as long as the target describes significant hardships and expresses a desire for support. On the other

hand, when such pleas are not given, it has the opposite effect on subjects' behavior. The authors conclude that subjects' concerns regarding their own negative evaluation by the out-group members underlies this differential effect: Subjects feel if a target admits to experiencing hardship and they withhold support, it increases their likelihood of being negatively evaluated by the out-group member. This effect held in both lower- and higher-prejudice subjects.

Skorinko and Sinclair (2013) found a similar effect; taking another's perspective can confirm and increase stereotyping. While other work has suggested taking the perspective of an out-group member can reduce the possibility of stereotyping (e.g., Galinsky and Moskowitz, 2000; Vescio, Sechrist, and Paolucci, 2003), Skorinko and Sinclair found that the stereotypicality of the target needs to be considered. In this study, they chose to use the elderly and overweight as their out-groups. When subjects took the perspective of one of these out-groups' members, they were more likely to engage in stereotyping compared to non-perspective takers when the target was consistent with stereotype. However, when the target did not conform to stereotype, perspective-taking subjects were less likely to engage in stereotyping compared to non-perspective takers. Skorinko and Sinclair argue that this is because when stereotypes are salient, we are more likely to use them as a basis for taking the perspective of another. Or rather, when stereotypes are salient, they overshadow traits of the target individual, leading us to take the perspective of the stereotype. This holds only when negative information is relevant to stereotype. Irrelevant negative information did not have a similar effect on their subjects' judgments.

ToM and Mirroring

If ToM is viewed as a higher level, developmentally emergent cognitive mechanism, motor mirroring seems to be a separate ability. Hamilton et al.(2007) looked at imitation and action understanding in autistic patients, to test whether there is a concomitant impairment in these abilities, along with ToM. They hypothesized that if mirror neurons provide the basis for theory of mind, then imitation should also be impaired as well. However, they found no evidence of autism related impairments in mirror neuron system skills. In fact, subjects showed normal goal directed imitation and grasp planning skills, and even superior gesture recognition skills. Hamilton's results support the position that mirror neurons are not associated with ToM.

In one sense, this conclusion is not surprising, as the brain areas involved in ToM are not the brain areas where motor mirror neurons have been identified. However, one could argue that the subjects in Hamilton's study were mindlessly mimicking the actions of the experimenters with no understanding of the ostensible purpose of the movements. Still, as are other forms of empathy, ToM is just as much a perceptual ability as it is a cognitive ability. Perceiving or even anticipating the movements and facial expressions of a target are also necessary for making ToM attributions; observing the actions of a target are the basis for violation of expectation studies with infants. Although they may be separate abilities, the function of one (mirror neurons/motor empathy) can still inform us about the functions of the other (ToM) because mirror neurons are not considered in isolation, but as part of a mirroring system. All three types of empathic systems discussed (motor mirroring, emotional empathy and ToM) involve a similar simulation-based mechanism where the observed actions or inferred mental states are reproduced, or

mirrored, in the brain of the observer. And in fact, there is brain anatomical evidence for mirroring in ToM. Of the various areas implicated in ToM, the medial prefrontal cortex (mPFC), has been considered the primary location of ToM, showing the most consistent activation in ToM tasks (Frith and Frith, 2001). In addition, it has also been shown to be the most mirror-like of the ToM regions; it is activated when people think about their own mental states and when they think of others' mental states as well.

Variation in ToM

Between Individuals

ToM appears to be universally exhibited across humans, shows a similar developmental trajectory, and in terms of implicit understanding, shows some commonality across species (humans and chimpanzees), and therefore suggests a deep evolutionary history. However, even if we simply focus on humans, there is also a considerable amount of variation in ToM. It should not be viewed as a binary phenomenon; it is not something one either has or does not have.

There are few if any psychological traits that do not show such variation, some obvious examples include IQ and personality traits such as introversion/extroversion. Baron-Cohen has developed several tests that measure an individual's so-called autistic traits, in terms of an autism-spectrum quotient along two dimensions, empathizing and systemizing. In the context of these questionnaires, "autistic" should not necessarily be viewed as indicating pathology, but more accurately as "self-oriented."

The Autism Spectrum Quotient (Baron-Cohen et al., 2001) is a self-administered, 50 question personality inventory-type test that measures five areas: social skill, attention

switching, attention to detail, communication, and imagination (see Table 3.1 for sample questions). As its name suggests, it is designed to measure the presence of autistic traits in the test taker, with higher scores indicating a greater degree of autism. However, it has been used in non-clinical populations and results indicate differences in scores across groups. Men score higher than women. Scientists score higher than non-scientists, and mathematicians, physicists, computer scientists and engineers score higher than life scientists.

Reading the Mind in the Eyes Test (Baron-Cohen, Jolliffe, Mortimore, and Robertson, 1997) is a forced choice test that requires people to view images faces cropped to show only the area around the eyes. Subjects must make a choice as to which of two indicated emotions or cognitive states the individual in the photo is displaying. Adults with autism, even higher functioning Asperger syndrome show impairment on this test. Also, women score better than men. Baron-Cohen (2003) also has two other tests, the empathizing quotient (EQ) and systemizing quotient (SQ) questionnaires. Using these measures, Focquaert, Steven, Wolford, Colden, and Gazzaniga (2007) found that their data “strongly suggest that in the sciences versus humanities, both gender and major independently contribute to the assessment of an individuals’ systemizing and empathizing cognitive style. The main conclusions from our study are that on average (1) men are more systemizing than women, and (2) science students are more systemizing than humanities students" (p624).

Other personality traits, such as cooperative or Machiavellian traits are associated with increased or decreased ToM, respectively (Paal and Bereczkei, 2007). Other evidence of between-individual/group differences come from the cross-cultural studies

discussed in the Chapter 2. While there is remarkable similarity, they do not show a perfect synchrony in the emergence of explicit ToM/false belief understanding. Children in some groups began passing explicit false belief tasks later than Western children. Conventions of spoken Japanese appear to confer an advantage in false belief processing to younger children. And ToM/false belief understanding is also tied to the mastery of a grammatical structure, the syntactic complement.

Within Individuals

ToM, like motor and emotional empathy, also vary across situations and targets. It is an attempt to put oneself in the other person's shoes, and this implies differential success in doing so—there must be factors related to the situation or target individual that result in within individual differences, as well. Again, we see evidence of this in the research discussed in Chapter 2. As task demands are simplified, younger children, who were unable to pass the standard Displaced Object and related tasks, are able to pass explicit and implicit tests of false belief understanding. At the same time, when tasks are made more complex, adults who might find those same standard explicit tasks trivially easy, can make errors on false belief tasks similar to young children. Even higher-functioning autistic subjects appear able to respond correctly in certain experimental settings. Senju et al. (2009) found that subjects with Asperger syndrome appear unable to make spontaneous ToM attributions, however, they are able to consciously reason through the process and arrive at the correct answer.

Other sources of within-individual differences include time between observation and response (Cohen and German, 2009) and age of subject (Cavallini, et al., 2013).

Given a short enough duration between observation and response, Cohen and German found that false belief processing appeared to be automatic, even without instructions to attend to the target's beliefs. This suggests that in the immediacy of an interaction, we automatically attend to and utilize our understanding others' beliefs, but without an explicit need to remember, the awareness does not last. In adult subjects, ToM appears to be affected by age. In Cavallini et al's study, younger adults (age 20-30) outperformed older adults (age 59 and older) on an explicit measure of ToM, the Strange Stories Task. While this was a cross-sectional study and not a longitudinal one, it still suggests that as we age, our ability to make accurate mental state attributions will decline.

One final factor affecting individual performance in ToM related tasks is familiarity. As we have seen in this chapter, familiarity is a key factor in explaining individual differences in empathy, particularly emotional empathy. The same appears to be true for ToM. Liew et al.'s (2011) found that both same-race faces and familiar gestures activated brain areas implicated in ToM. In addition, this activation increased as subjects more strongly identified as members of their ethnic group. Similarly, a study of native Japanese and white American subjects found a same-race advantage in the Reading the Mind in the Eyes test (Adams, Rule, Franklin, Wang, Stevenson, Yoshikawa, et al., 2010). Personal familiarity with the target individual also affects ToM processing; people may be more willing to engage in mental state attribution when the target is liked (McPherson-Frantz and Janoff-Bulman, 2000). In exploring the other mental processes that may come into play in making mental state attributions, Rabin and Rosenbaum (2012) found that subjects utilized autobiographical memory, drawing on past personal experiences, when they reasoned about the mental states of target individuals who were

personally familiar to them. Alternately, when reasoning about mental states of unfamiliar target individuals, subjects would instead rely on semantic memory—scripts, schemas, and general knowledge of social situations.

Many of these sources of between- and within-individual differences are associated with traits internal to the individual (e.g., one's score on the Autism Spectrum Quotient, sex, or one's degree of identification with their ethnic group) and suggest that although variable, ToM might still be viewed as a fixed trait in a person. Yet others, particularly age, reveal that it is a trait that varies over time for a given person on a given type of task. Others are situational, such as the particular task one is presented with, either simple or complex. Still others, such as familiarity with a target individual, appear to be both—in a particular situation, a target individual will occupy a specific location on the unfamiliarity-familiarity continuum, yet familiarity can increase over time. By extension, so can one's ability to correctly infer the mental state of a given target individual. The next chapter explores this in more depth, reviewing the literature on how group membership and familiarity interact with (emotional) empathy. Doing so reveals that group distinctions are not necessarily static nor familiarity and the transition from out-group to in-group a relatively slow accumulative process. In fact, group membership is quite flexible and our empathic responses to others follows this flexibility. This should have implications for ToM as well.

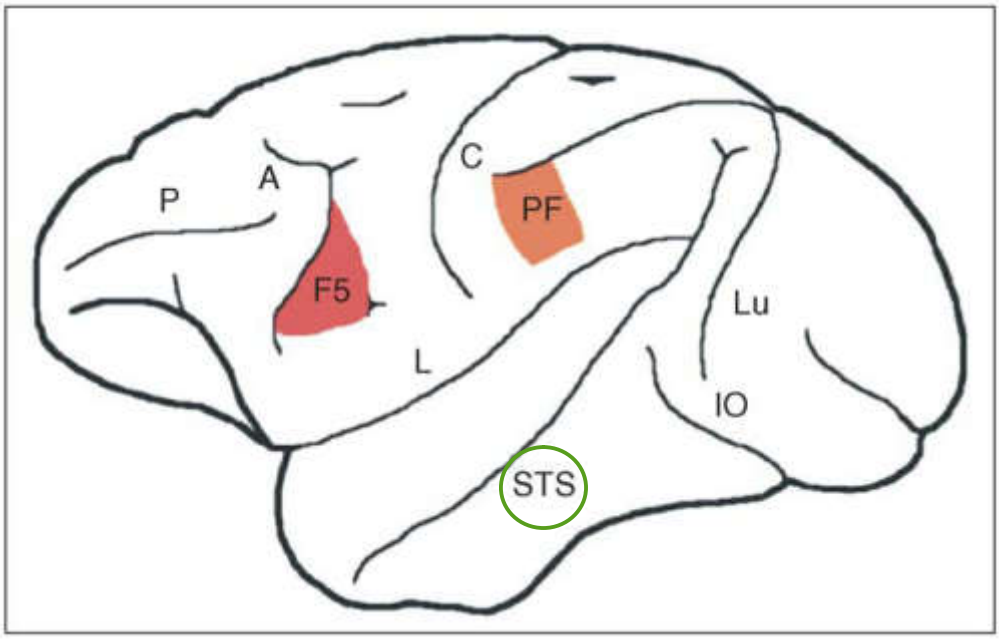


Figure 3.1: Brain areas in the macaque mirror neuron circuit. Area F5 (red) of the motor cortex and area PF (orange) of the inferior parietal lobule. PF receives projections from the superior temporal sulcus (STS), circled in green. (Adapted from Lyons et al., 2006, Figure 1.)

Table 3.1: Sample statements from the Autism-Spectrum Quotient questionnaire (Baron-Cohen et al.2001: 15-16). Response choices are “definitely disagree”, “somewhat disagree”, “somewhat agree”, and “definitely agree”.

I prefer to do things with others rather than on my own.
I usually notice car number plates or similar strings of information.
I am fascinated by dates.
I notice patterns in things all the time.
I would rather go to the theatre than a museum.
It does not upset me if my daily routine is disturbed.
I find it easy to read between the lines when someone is talking to me.
I usually concentrate more on the whole picture rather than the small details.
I am not very good at remembering phone numbers.
When I talk on the phone, I m not sure when it’s my turn to speak.
I am good at social chit-chat.
I enjoy meeting new people.
I find it very easy to play games with children that involve pretending.

Chapter 4: Coalitional Psychology

The classic Robber's Cave study (Sherif, Harvey, White, Hood, and Sherif, 1961) provides an early example of coalitional psychology and how easily group identity forms and shapes attitudes towards in-group and out-group members. Sherif and his colleagues followed two groups of boys at the same campsite at Robber's Cave State Part in Oklahoma. At the outset, neither group was aware of the other group. Both quickly found names for their respective groups, the “Eagles” and the “Rattlers.” As their 5-week long camping trip progressed, each group became aware of the other's presence and became concerned the other group was intruding on their grounds and began to insist that they meet in competition. When the researchers brought the Eagles and Rattlers together, the boys began insulting the other group and nearly came to physical violence. The researchers' attempts at engineering subsequent conciliatory meetings met with little success. It seems the boys had firmly established their groups, favoring their own and wanting nothing to do with the other. This study is of interest not only for the between-group conflict that arose, but also for the in-group cooperation. We are an especially social and cooperative species—what drives this tendency?

Many researchers (e.g., Henrich, 2004; Boyd and Richerson, 1985; Bowles and Gintis, 2003; Fehr and Fischbacher, 2003) argue that gene-culture co-evolution and cultural group selection lead to prosocial predispositions that now underlie the large-scale cooperation such as what we see in corporations, markets, and states. This was most certainly a key factor, but we do not always need to invoke such predispositions to explain individuals' motivations underlying this type of large scale cooperation (Cronk and Leech, 2013): Rather, gene culture co-evolution and cultural group selection are

perhaps more likely to lead to a flexible coalitional psychology. Consider the fact that prior to participation in the Sherif et al. study, none of the boys knew each other. Their ease in forming groups could be explained by prosocial dispositions. But group boundaries and coalitions are not always permanent things; in fact they may form and dissolve as need arises, and as can be seen in the Robber's Cave study, this can happen quite easily. It follows, therefore, that we should be attuned to cues that allow us to quickly identify who is and who is not a coalitionary partner, who is a member of our in-group and who is not. Furthermore, such cues are not always be fixed attributes, such as sex, accent or race, but also easily changeable factors such as clothing or other adornments, or simply being placed together in a group.

Chapter 3 reviewed literature which, taken together, reveals that the degree to which humans experience emotional empathy towards others is influenced by many factors. Often they are related to coalitions, including familiarity with the target individual and group membership (e.g., Cikara and Fiske, 2011; Gutsell and Inzlicht, 2010), subjects' gender and previous behavior of targets (Singer et al. 2006). Subjects also empathize more with in-group vs. out-group members (Xu, Zuo, Wang, and Han, 2009), but this is also modulated by attitudes towards the out-group (Avenanti, Sirigu, and Aglioti, 2010). Focusing on factors such as race and sex, as well as other cues such as accent (Kinzler, Corriveau, and Harris, 2011; Kinzler, Shutts, Dejesus, and Spelke, 2009) make it easy to conclude that groups are relatively permanent things, as one cannot easily change these attributes.

These feelings of group membership can be experimentally induced through different group induction methods falling collectively under what is known as the

minimal group paradigm. In the classic version, subjects are merely informed that their responses on a questionnaire place them in one of two categories (Tajfel, Billig, Bundy, and Flament, 1971). Other methods include randomly assigning subjects to groups under the pretense that their scores were too similar on a similar questionnaire (Brewer and Silver, 1978) or even having subjects memorize a list of names of members in a group (Pinter and Greenwald, 2004). The groups formed using these methods are “minimal” in that a given subject never meets the other group members (in fact, they may be non-existent), and they are arbitrary. Still, subjects show a clear pattern of in-group vs. out-group favoritism in subsequent tasks, without ever meeting their fellow group members.

Flexibility of Group Boundaries

Race

Race, in and of itself, may not be the element subjects are responding to in the studies described in Chapter 3. That is, race is not an automatic determinant of in-group/out-group status. Recall that Avenanti et al. (2010) found that presenting White subjects with a violet-colored hands in painful situations was not sufficient to produce the decrease in empathic response seen when viewing Black hands. Without preexisting beliefs and attitudes, or at least the awareness thereof, hands of a different color are treated similarly to one's own. In fact, Kurzban, Tooby, and Cosmides, (2001) argue that race is not plausible category for evolved cognitive "conceptual primitives" compared to other categories: In the environment of evolutionary adaptation (EEA), automatic processing of others' age or sex would have allowed observers to make a number of useful inferences about them. Our ancestors, however, were greatly restricted in the

distances they could travel, being limited to foot, making it highly unlikely that they would have encountered others who would have been different enough in appearance to be categorized as a different race. What we take to day to be an automatic encoding of race is rather a byproduct of adaptations for detecting coalitions. We use race as a proxy for group membership.

Cosmides, Tooby, and Kurzban (2003) describe three factors involved in our evolved coalition detection: First, in the environment of our evolutionary adaptation, coalition and alliance detection machinery should be sensitive to patterns of coordinated action, cooperation, and competition and other cues that help us make predictions about others' political affiliations/ group memberships. Second, it should be sensitive to fleeting cues about others' potential as coalitional partners that it can associate with other, longer lasting cues, assigning meaning to them. And third, it should recognize that no one cue is applicable to all situations—it is only worth consideration as long as it continues to make accurate predictions about others' coalitional memberships.

Further evidence that determining group membership is not innately based on race comes from the developmental literature. Kinzler and Spelke (2011) found that infants do not demonstrate any preference for members of their own race over members of another race, accepting toys equally from members of both groups. Two-and-a-half year olds showed a similar lack of preference in giving toys to same- or other-race individuals. However, in the same toy-giving scenario, the five-year-olds in their study did show a preference, favoring members of their own race, suggesting that distinguishing between races emerges at some point between 2.5 and 5 years of age. Taken with Telzer et al's (2013) study that showed amygdala sensitivity to other races emerges throughout

adolescence, this further suggests race increases in salience across development and that this is likely due to socialization factors rather than some type of innate racial processing.

Flexibility of Race as a Determinant of Group Membership

Kurzban et al. (2001) found that the use of race, a permanent category, as a cue for group membership could be overridden, or "erased" in favor of more fleeting and arbitrary cues such as, in their study, the color of target individuals' tee-shirts. Unlike age or sex, they found that race was not uniformly encoded, and that subjects easily grouped individuals of different races together in response to tee shirt color as a coalitional cue. While race may serve as quick heuristic for group membership, Kurzban et al. found that divisions along such lines can easily be overridden by other cues that are reliably associated with coalition membership, such as colored tee-shirts worn by others. This "alliance detection system" (Pietraszewski, Cosmides, and Tooby, 2014) is also sensitive to more meaningful categorizations, such as political affiliation (Pietraszewski, Curry, Peterson, Cosmides, and Tooby, 2015): such categorization reduced subjects tendency to categorize by race, but it did not affect conceptual primitives—sex and age.

Virtually putting one's self in another's shoes reduces racial bias (Peck, Seinfeld, Agliolti, Slater, 2013) and, it follows, its salience as a category for discrimination between groups. When Peck and colleagues' subjects were placed in a virtual reality setting, virtually embodied in a black avatar, it significantly reduced implicit attitude test (IAT) scores compared to pretest measures taken three weeks prior. This effect was not seen in the other conditions presented in the study, which included virtual embodiment in a light skinned avatar, virtual embodiment in an purple "alien-skinned" avatar, and a

black skinned, but non-embodied, avatar. Experiencing one's self as a member of an out-group, albeit here in a virtual setting, is capable of reducing one's implicit biases against members of that group. Full embodiment may not be necessary to achieve this effect, but it does provide a sense of ownership of the avatar. A similar reduction in negative implicit attitudes was also found in a study that elicited a sense of illusory ownership of an out-group hand constructed out of rubber (Farmer, Maister, and Tsakiris, 2014).

Factors external to the immediate task of group categorization can also make race more salient, as well. for example one's economic situation, or more broadly, the state of the economy. In one study, (Krosch and Amodio, 2014), subjects were primed to think in terms of economic recent scarcity (through presentation of zero-sum, Black vs. White outcomes such as “When Blacks make economic gains, Whites lose out economically”). The same subjects were then presented with images of faces that ranged from 100% White, morphing in 10% increments to 100% Black. The economic scarcity primes led subjects to more readily categorize the more ambiguous mixed-race images as Black. In addition, similar priming led subjects to favor white versus black faces in a resource allocation task where they were shown two pictures (one White, one Black) of people ostensibly in need, and asked to decide how to divide a sum of money between the two.

Contrary to Kurzban et al.'s (2001) findings, race may not be truly erased in the sense that word implies. Ratner, Kaul, and Van Bavel (2013) presented subjects with a task that required them to memorize which of two teams a series of photographed faces belonged to. Team membership cut across race, similar to Kurzban et al.'s study. When Ratner et al's subjects were placed in an fMRI scanner for a recall task, the researchers found areas in the visual cortex differentially responded to the race of the faces. Perhaps

race is not erased, but its salience is affected by situation. A recent study (Correll, Guillermo, and Vogt, 2014) lends support to this possibility. Subjects were presented with images of white and black faces in one of two conditions, control and goal. In the control condition, subjects showed a strong bias towards attending to black faces, looking at them much longer relative to the white faces. However, in the goal condition, subjects were tasked with locating a identifying the color of a dot on each image. This eliminated the gaze time bias for black faces. From the results of these two studies, it appears that race remains an important distinction for us, but it is not always necessarily the primary focus for making in-group/out-group distinctions. This conclusion is in line with Kurzban's (Kurzban et al., 2001, Cosmides, Tooby, and Kurzban, 2003) conceptualization of coalition detection.

Race can also be simultaneously erased and salient. Van Bavel and Cunningham (2009) randomly placed subjects in one of two groups and, showing photos of both groups' members, instructed them to memorize the members of each group. Both groups were mixed-race, half black and half white. In two tasks measuring automatic associations and conscious, controlled evaluations, subjects rated in-group blacks more positively than out-group blacks, a pattern that appeared to be due to more to in-group bias rather than out-group derogation. In addition, photographs of unaffiliated white and black faces elicited an automatic racial bias favoring whites. The authors conclude that self-categorization can override automatic racial bias and that automatic evaluation is sensitive to within- and between-group social contexts.

Taken as a whole, the studies reviewed in this section, while focusing on race, strongly suggest that we demonstrate a obvious flexibility in determining who qualifies as

an in-group or out-group member. Race, which at first glance appears to be a salient and readily used determinant of categorization is actually affected by many other features and situational factors.

Other Flexible Determinants of Group Membership

Shirts served as the stimuli in another study providing additional support for Kurzban et al's finding. Levine, Prosser, Evans, and Reicher (2005) devised a unique method to demonstrate the flexible nature of coalitions and how the boundaries of a coalition are sensitive to framing effects. They recruited Manchester United football fans to participate in a study that was ostensibly about soccer and soccer fandom. After completing a survey the subjects were asked to deliver the results to another building. On the walk over to the second location, subjects witnessed a runner fall and hurt themselves. The subjects were not aware that this was the actual experiment—they were observed to see whether they would assist the injured runner. Levine et al. found that group boundaries are malleable and subject to framing effects of the survey version they received, as well as the shirt worn by the runner. When the Manchester United fans took a version of the survey that focused on Manchester United and their fans, they were much more likely to aid the runner if he wore a Manchester United shirt, but not when he wore a Liverpool shirt (Manchester United's rival team) or an unbranded, non-football shirt. However, when given a survey version that primed them to think of themselves more broadly as a football fans, subjects aided the runner equally whether he wore Manchester United or Liverpool, but still not when he wore the unbranded shirt. These results suggest that the concept of an in-group is not a fixed thing, and determining at any given time

who is a fellow in-group member can depend on situational variables. In addition, the results of this study provide a behavioral measure to support the empathy studies previously described—rather than showing increased activity in associated brain areas, this study shows an actual increase in helping behavior.

Visible signals as cues to group membership are important elsewhere, too. McElreath, Boyd, and Richerson (2003) demonstrated their importance in a computer simulation study. In their simulation, virtual agents played a stag hunt coordination game. These agents were tagged with either a 0 or 1, their “ethnic marker.” When the agents were programmed with a preference for interacting with other agents with the same marker, unsurprisingly, two groups emerged. In addition, the markers served as reliable indicators of agents' strategy in the game. The authors suggest that any such signals of group membership may help solve coordination problems such as the stag hunt by making it clear who else shares the same assumptions regarding social interactions. In fact, they found that when groups were allowed to form in a virtual space, group differences were strongest at the boundaries, suggesting that when others are encountered, it is more important to clearly indicate one's own expectations and beliefs about interactions to avoid confusion or conflict. Moving from the virtual subject to human subjects, other researchers found similar results. When subjects playing a stag hunt game were given the option to sort themselves into groups each round with an arbitrary marker, they tended to form groups with consistent relationships between the markers and how the members play the game (Efferson, Lalive, and Fehr, 2008). We focus on markers that give clues to shared common knowledge of fellow in-group members.

Another recent paper explores mechanisms underlying group membership and identity. There appear to be two opposing ways in which people perceive a high degree of self-other correspondence with their fellow group members: social projection and self-stereotyping (Cho and Knowles, 2013). Social projection is the tendency to project one's own traits onto in-group members (the others are like me), and self-stereotyping is the opposite, the tendency for an individual to assume they share the traits of others in the group (I'm like the others). Cho and Knowles found that when they experimentally manipulated subject' self views, this led to subjects altering their judgments of a close in-group to be in line with those views. When they manipulated the apparent traits of this in-group, subjects revised their self-views to be consistent with the group's traits. Neither of these effects were seen with an out-group. Cho and Knowles' results reveal a flexibility in group membership, but not regarding the composition of the group. Rather, it is a flexibility identity that serves to maintain the group.

Circumstances under which a cultural group forms can subsequently affect group members' levels of parochialism (in-group favoritism/out-group discrimination). Pan and Houser (2013) had subjects work in teams to solve puzzles for time as part of a contest. Teams either worked on puzzles more conducive to independent work or to collaboration. Afterwards, individuals were paired with in-group or out-group members to participate in an economic trust game. Subjects who came from teams formed under the independent production scenario demonstrated high levels of parochialism, while those who came from cooperative production scenario showed reduced levels of parochialism.

Infants and Group Discrimination

While there does not appear to be an innate racial categorization ability present in infants (Kinzler and Spelke, 2011) (which provides further support that race is not a conceptual primitive) infants do demonstrate the ability to categorize others according along lines of group membership. This emphasizes the importance of such a skill and provides evidence of a fifth core knowledge system (see Chapter 2), one for representing social partners, coalitions, and in- versus out-group members (Spelke and Kinzler, 2007).

Infants as young as six months are able to distinguish between others who are helpers or hinderers (Hamlin, Wynn, Bloom, 2007) . Subjects in this study were shown a puppet show in which a red circle attempts to climb up a hill and is either aided by a yellow triangle or pushed back down by a blue square. When later presented with the triangle and square, infants demonstrated a preference for the triangle. Although this happens in the absence of any group membership cues, it suggests that at an early age, we begin to make judgments about who we would prefer to be near and thus suggests an early emerging mechanism for identifying potential in-group partners. This type of discrimination also appears to extend to taking others' perspectives and distinguishing between intentional and accidental actions. Choi and Luo (2015) presented 13-month-old infants with interactions between three puppets, A, B, and C. If B purposely hit C in the presence of A, infants expected A to exclude B in subsequent interactions, but not when the hit was accidental. However, if A was absent when B hit C, the infants expected A to continue to interact positively with B. Taken together, these two studies suggest that even in infancy, we are able to identify preferred social partners and also expect others to do the same.

Using a violation of expectation method, Powell and Spelke (2013) found that seven month old infants can distinguish whether actions performed by target individuals are consistent with those made by their fellow group members. After viewing two groups of animated shapes (e.g., stars and squares) performing actions unique to each group, infants looked longer when a lone member of one group performed the movement associated with the other group. Furthermore, this ability was limited to a social context; it did not transfer to a setting involving non-social agents. The ability to identify group-level traits and categorize others according to their group membership also emerges early in development.

Not only can infants categorize others into groups, they also show an early preference for others who are similar to them, ostensibly members of their own group (Mahajan and Wynn, 2012): After making a choice between two different foods (e.g., green beans vs. graham crackers), infants viewed two puppets who, in turn, each make their own choice between the two foods, with only one selecting the same food as the infant. Afterwards, the infants are presented with the two puppets. The puppet that chose the same food item is selected by the subjects much more often than the other. The same food choice method also reveals a further dimension of infants' preference—they prefer others who 1) help individuals who are similar to themselves, and 2) hinder dissimilar others (Hamlin, Mahajan, Lieberman, and Wynn, 2013).

Accent

While informative regarding our knowledge of social groups, some of the above studies rely on multi-step methods. We also use more immediate cues in making in-

group/out-group distinctions. One in particular is accent. Children prefer to befriend other children (Kinzler, et al., 2009) with whom they share a native accent. When presented with photographs and audio recordings of unfamiliar children, a group of five-year-olds chose to befriend children who were native speakers of their own language over children who spoke a foreign language or with a foreign accent (Kinzler, et al., 2009). If only pictures and no audio was available, the subjects preferred photos of children who were the same race as themselves. But when audio was available, accent was the basis of choice; subjects preferred other-race children with native accents over same-race children with foreign accents.

Accent also appears to be an important factor in trust and in selection of appropriate others to serve as sources of cultural learning (Kinzler, Corriveau, and Sarris, 2011). Kinzler and her colleagues presented native English-speaking children with videos of either native- or foreign-accented English speakers who spoke for 10 seconds then non-verbally demonstrated uses for various novel objects. They found their subjects preferred the uses demonstrated by the native speakers even though the speakers in the videos spoke in nonsense speech.

Unlike race, and more like conceptual primitives such as sex and age, accent appears to be a strong cue to group membership; it is a spontaneous and implicit dimension of social categorization (Pietraszewski and Schwartz, 2014a) that is not the result of other high- or low-level factors such as familiarity or acoustic differences. In addition, it has the same ability to “erase” race as a category for determining group membership (Kinzler, et al. 2009; Pietraszewski and Schwartz, 2014b).

Group Membership and ToM

The above findings have some potential implications for ToM. As discussed, we are more empathetic with in-group members, but it is possible that reliance on markers indicating shared group membership may serve as a shorthand—a rule of thumb about what others know or don't know rather than actual, active mind reading. This idea receives support from Rabin and Rosenbaum's (2012) study described in Chapter 3. The less familiar an individual is with a target, the more they rely on general social rules and scripts when attempting to make mental state attributions.

Shared Meta-Knowledge

This also suggests that ToM is an important factor to consider at the level of the cultural group. Culture can be defined simply and broadly as socially transmitted information (Boyd and Richerson, 1985; Cronk, 1999), or shared knowledge. For a group to successfully cooperate and coordinate actions with this shared knowledge, an additional level of sharing must occur, shared meta-knowledge—the mutual awareness that others share the same information (e.g., Chwe, 1998; Schotter and Sopher, 2003). For shared meta-knowledge to exist, a skill like ToM is necessary. The "I know that you know"/"You know that I know"/“I know that you know that I know” (or more perhaps more simply, “we all know that we all know”) structure of shared meta-knowledge relies on the syntactic complement, the same type of construction necessary to parse false belief: "I know that you (mistakenly) know that the toy is in the box where you left it." ToM itself may even be thought of as a culturally transmitted skill (Heyes and Frith, 2014). Again, ToM is an important skill not only for understanding others, but also for

coordinating interaction with them at the group level.

Chwe's work focuses on how this shared meta-knowledge is generated, or how shared knowledge becomes shared meta-knowledge. “Successful communication sometimes is not simply a matter of whether a given message was received. It also depends on whether people are aware that other people have received it...it is also about people knowing that other people know about it: 'metaknowledge of the message” (Chwe, 1998:49). The key point he identifies in this transformation is publicity (Chwe, 1998). He describes one particularly large scale example of this, Apple's 1984 Super Bowl add for their new Macintosh computer: “By airing the commercial during the Super Bowl, Apple did not simply inform each viewer about the Mackintosh; Apple told each viewer that many other viewers also know about the Macintosh” (p51). He connects this example to others by recognizing that watching the Super Bowl is an annual ritual. Rituals, specifically public rituals, are an effective means of creating shared meta-knowledge. The above example builds on the preexisting shared knowledge of the widespread viewing of the game.

On a much smaller scale, he notes that two people can accomplish this merely through eye contact. This is an important observation. Tomasello (Tomasello, et al., 2005; Tomasello, Melis, Tennie, Wyman, and Hermann, 2012) has argued that the ability to follow others' gaze and to engage in joint attention were likely first steps towards shared intentionality and ultimately, ToM[1]. Indeed, it has also been argued that the

[1] It appears that chimpanzees also understand that seeing leads to knowing and can thus recognize the internal state of knowledge in others. In a series of experiments on social problem solving in chimpanzees, Hare et al. (2000, 2001) found subordinate chimpanzees modified their behavior, based on what a dominant chimpanzee could see. They concluded that the subordinates knew what the dominants could/could not see, recognized that visual knowledge was specific to the observer, and used this knowledge to develop strategies in a competition over food.

morphology of the human eye—a dark iris against a visible, white sclera—makes it particularly easy for us to gaze follow (Kobayashi and Kohshima, 1997; 2001). Without shared intentionality or ToM, shared meta-knowledge is perhaps not possible, and coordinating actions and goals is much more difficult than with them.

Not only are eyes a salient cue in humans, the presence of images of eyes is enough to invoke cooperative behavior. Haley and Fessler (2005) had subjects play a Dictator Game on a computer. The monitor either displayed a control image or an image of stylized eyes. Presentation of the eyes during the game resulted in significantly increased generosity and probability that the subject would allocate money to another player. A later replication and meta-analysis (Nettle, Harper, Kidson, Stone, Penton-Voak, and Bateson, 2012) confirmed this finding, but with a revision: while presentation of eye images does increase the probability of allocation, it does not appear to affect generosity in terms of the amount given. Furthermore, eye images' effect does not extend to other behaviors such as individual choices but are limited to social interactions (Baillon, Selim, Van Dolder, 2013).

What about the vast space between two-person dyads and millions of football fans? There are other ways to create publicity in rituals and thus shared meta-knowledge. Chwe (1998) focuses on inward facing circles; eye contact with every other person may not be possible, but this configuration does allow each individual to easily see who is and who is not paying attention. When combined with rituals, then it is easy to see who is and who is not participating. Shared behaviors are an important part of how we categorize others as in-group or out-group members (see below).

The powerful effect that creating common meta-knowledge has on subsequent

coordinated behavior can be seen in Schotter and Sopher's (2003, 2007; Chaudhuri, Schotter, and Sopher, 2009). In their first study, Schotter and Sopher (2003) had subjects participate in a multigenerational “battle of the sexes” coordination game (a two-person game in which the players receive a payoff only if they both choose the same of two alternatives, A or B; however, two additional conditions apply: 1) the players are not able to communicate with each other prior to making their choice, and 2) player 1 receives a higher payoff than player 2 if both choose option A, while player 2 receives a higher payoff than player 1 if they both choose B). Each generation of players played the game once and participants were allowed to make advice available to the subsequent generations on the best way to play the game to maximize payoff. This passing of advice was more effective in establishing coordination between players than when it was absent. They found similar results in a follow up study that used an intergenerational Ultimatum Game (Schotter and Sopher, 2007). However, in a third study (Chaudhuri, Schotter, and Sopher, 2009), this time using a “minimum effort game” (a stag-hunt game adapted for multiple players) they found that it is not simply the passing of advice to the next generation that effectively establishes coordination and maximizes payoff, but it depends on how it is presented. Specifically, the advice must be public and it must be made common knowledge. Privately shared strategies are not effective, but advice read aloud to all players together (as long as all are confident that the others were paying attention) leads to optimal playing (and maximized payoff) of the minimal effort game.

The effect of common meta-knowledge is not limited to the laboratory. Many real world examples exist as well. See Cronk and Leech (2013) for a discussion of two key examples, Alvard's (2003) study of whale hunting in the Lamalera community on the

island of Lembata in Indonesia, and Lansing's (Lansing and Kremer, 1993; Lansing and Miller, 2003) work on Balinese rice farmers' water rationing and pest control. Chwe (1998) offers an example of the use of circular structures in his description of Kivas found in the American Southwest.

ToM and Cooperation

ToM is an important, perhaps necessary, component in cooperative interactions. Takagishi, Kameshima, Schug, Koizumi, and Yamagishi. (2010) found a correlation between ToM and a preference for fairness. They first presented preschool aged children (approximately four to six years of age) with the Sally Anne (Displaced Object) task to test for ToM/false belief understanding. Following this task, they paired children up to play the Ultimatum Game. The children who passed the Sally Anne task showed a strong preference for fair offers (50%) compared to those who failed the task. As the ability to understand the mental states of others emerges, so too does an understanding of fairness.

But again, ToM is not a binary trait; it not simply present or absent. Tests such as Baron-Cohen's Autism Quotient test (Baron-Cohen et al. 2001) reveal there is a range of ToM ability across individuals. With this observation (and Baron-Cohen's questionnaire), Curry and Chesters (2012) designed a study to test their subjects' ability to solve coordination problems as a function of their level of ToM. Subjects sat at a computer and answered a series of 20 questions in which their goal was to give the same answer as an anonymous partner. After this they took the Autism Quotient test. There was a significant correlation between subjects' success on the coordination task and their Autism Quotient, but only on one subscale of the questionnaire: Understanding Others. The other

subscales, which measure non-ToM autistic traits were not related to the task.

Sylwester, Lyons, Buchanan, Nettle, and Roberts (2012) studied the relationship between ToM and cooperation. First, subjects ToM ability was measured via the Reading the Mind in the Eyes test (Baron-Cohen et al., 2001). Next, they viewed video clips from a game show involving a variation of the Prisoners' Dilemma. The clips lead up to (but not including) the two contestants' decision to either cooperate or defect. Sylwester and her colleagues found a small but significant positive correlation between ToM performance and identification of cooperators, as well as a negative correlation between ToM performance and identification of defectors. These results suggest that ToM is a key component in assessing cooperative intentions, and that it may interfere in identification of cheaters.

However, an earlier study (Lissek, Peters, Fuchs, Witthaus, Nicolas, Tegenthoff et al., 2008) found that brain areas implicated in ToM showed increased activation in response to both cooperative and competitive scenarios. In this study, subjects were presented with cartoon images depicting cooperation (two characters working together to achieve a common goal), competition (one character deceiving another), and a combination of the two (two characters working together to deceive a third). Subjects were then asked questions to assess their understanding of the true or false beliefs held by characters in the images. Lissek and her colleagues found that there was an overlap in activation—both cooperative and competitive scenarios resulted in activation in the temporoparietal junction (TPJ), precuneus, and posterior cingulate cortex. However, the competitive scenarios also resulted in the activation of additional areas, the prefrontal cortex, insula, and anterior cingulate cortex, suggesting that these areas respond to

mismatches between the intents of one person and the expectations of another.

Using the 2008 presidential election as a source for group membership, Falk, Spunt, and Lieberman (2012) also found evidence of differential activation of brain areas associated with ToM. They asked subjects to evaluate the degree to which Barack Obama and John McCain would agree or disagree with a series of statements about issues related to the election. When taking either perspective, posterior regions—areas typically associated with thinking about others' mental states—showed increased activation. However, the TPJ was more active in response to taking the opposing candidate's perspective, while the precuneus was more active in response to taking one's own candidate's perspective, as were frontal regions. Their results do not entirely match the patterns of activation found in Lissek et al.'s study, but do lend additional support for the claim that considering in-group versus out-group members activates different components of the ToM system.

The right TPJ also appears to be involved in another aspect of cooperation and coalitions. Using transcranial magnetic stimulation to temporarily disrupt TPJ functioning in their subjects, Baumgartner, Schiller, Reiskamp, Gianotti, and Knoch (2014) found reduced parochialism/in-group favoritism in third party punishment decisions. After presenting a minimal group induction, pairs of subjects (players A and B) played a simultaneous one-shot Prisoners' Dilemma game in which they were matched with either in-group or out-group opponents. A third subject (player C, who has also undergone the minimal group induction) reviews the outcomes of thirty such interactions, and given information about the group affiliations of each participant, must decide whether or not to punish player A. When subjects (in the role of player C) were exposed to TMS of the

right TPJ, it reduced their tendency to favor player A when player A's group affiliation matched player C.

If we rely on general social rules and scripts in making mental state attributions with unfamiliar others (Rabin and Rosenbaum, 2012), this suggests that we should make more errors as familiarity decreases, and we should not be good at making mental state attributions of out-group members. Frames, scripts, and schemata can all be viewed as useful common knowledge “cheat sheets” for social interactions (Cronk and Leech, 2013); many everyday situations often fall into categories that share similar elements such as making small talk, shopping at a grocery store, participating in meetings or classes: if you and others in your culture know the “rules” for making small talk, each interaction should run more smoothly—everyone knows what topics to bring up, which ones to avoid. In addition, the work of researchers such as Chwe and Sopher discussed above makes it clear that successful in-group coordination not only requires such scripts (or rituals) but that that they are presented in a way that ensures shared (meta-) knowledge of them. Members of one group do not necessarily use the same scripts or are even aware of the scripts used by other groups. However, it might also be argued that familiarity with other individuals may have the opposite effect on ToM as it does for emotional empathy. That is, the assumption that others share the same knowledge as one's self could lead to errors in ToM. If an individual assumes that others in his group share the same knowledge, it is possible that this might interfere with actually taking a target individual's perspective.

Yet we may have an (evolved) inclination to be more susceptible to the influence of social coordination norms than to other types of cultural traits, stemming from the

benefits our ancestors gained from engaging in coordinated social behavior. Cronk (2007; Cronk and Wasielewski, 2008) found that Maasai and American subjects readily alter their play in a trust game depending on context. In this two-person game, player one is given money and has the opportunity to offer any portion of it to player two. The experimenter applies a multiplier to this portion (increasing the amount player two has), and player two then has the opportunity to give any of this new portion back to player one. Cronk (2007) found Maasai players gameplay to be typical of the trust game when given an unframed version. However, when the game was presented in terms of a particular gift-giving relationship, *osotua,* in their culture, the Maasai players adjusted their offers and expectations of returns accordingly. More interestingly, after reading a short description of the *osotua* relationship, American subjects also readily altered their behavior when the trust game was labeled as an *osotua* game to be in line with that concept compared to the unframed version (Cronk and Wasielewski, 2008).

Other research on norm violation and negative stimuli also suggests that rather than being a matter of distinguishing between in-group and out-group behavior, it may come back to violation of expectation. Bell and Buchner (2012) note there is a large body of literature demonstrating that negative/threatening information and stimuli are more easily remembered. In the context of norm violations/cheater detection and the cheater detection module (Cosmides, 1989), this suggests that we may be focusing on the negativity of the interactions rather than cheaters. In fact, memory for cheating and disgusting contexts are similar in this regard (Bell, Giang, and Buchner, 2012). But it is not just negativity—information also needs to be threatening for it to be well remembered. Nor does it appear to be some sort of processing advantage of negative

information over positive information. The focus is on information that violates positive or negative expectations, which Bell and Buchner (2012) note, is consistent with findings that memory is enhanced for information that is emotionally incongruent with expectancies (e.g., Cook, Marsh, Hicks, 2003). And violation of expectation is a key component of nonverbal ToM and false belief tests. Perhaps we attend more closely to those who violate our expectations, and therefore make more accurate mental state attributions?

Saxe and Wexler (2005) presented subjects with stories about protagonists from either familiar or unfamiliar backgrounds who held either normal or norm-violating beliefs for members of their background. They found that the subjects attempt to form integrated impressions of the protagonists and resolve inconsistencies between the protagonists' social backgrounds and stated beliefs. They found that the temporoparietal junction (TPJ), one of the brain regions involved in ToM appears to be active when we are exposed to violations of expectation (Saxe and Wexler, 2005). The right TPJ focuses on whole percepts—if violations are detected, incongruities are reacted to and sent to the left hemisphere for further processing. Thus, a key part of the brain's ToM processing is attuned to spotting violations. But if violation of expectation draws our attention and is routed for additional processing, what is our default expectation of others behavior? Is there a default expectation of others? Does it differ if others are in-group or out group? As with much of the research covered in this paper, context is also a key factor. In cooperation games where most partners are cooperative, cheating is remembered better—it is a rare event, and so is the violation of expectation. On the other hand, if most partners are cheaters, cooperation becomes the violation of expectation, and will be

remembered better (Barclay, 2008; Bell, Buchner, Musch, 2010; Volstorf, Rieskamp, Stevens ,2011). Such a system makes sense. It allows groups to exclude cheaters and norm-violators when they are rare, and it allows cooperators to find each other and form groups when it is they who are rare.

And last, as discussed in Chapter 3, there is evidence that motivation is an important factor in emotional empathy (Duan, 2000). Kozak, Marsh, and Wegner (2000) summarize several studies in support of this claim: motivation appears to be the case for engaging in a broader set of mental state attributions as well; McPherson-Frantz and Janoff-Bulman (2000) found a positive relationship between subjects' liking of target individuals and their willingness to take the others' perspectives. In-group members receive more attributions of complex emotions than do out-group members, regardless of familiarity (Leyens, Paladino, Rodriguez, Vaes, Demoulin, Rodriguez, Gaunt, 2000), further suggesting that subjects are more motivated to consider in-group members' perspectives. And a propensity to view one's fellow in-group members as more human than out-group members may underlie this tendency (Cortes, Demoulin, and Rodriguez, 2005).

Another study utilizing a minimal group induction technique found an interaction between ToM, group affiliation and perceived humanness. Hackel, Looser, Van Bavel (2014) presented their subjects with a series of morphed photographs on a continuum between doll and human. Subjects were to rate the images on a scale from 1 (definitely has no mind/definitely not alive) to 7 (definitely has a mind/definitely alive). These images were labeled to indicate their group status relative to the subjects. Compared to in-group images, there was a higher threshold for out-group morphed faces in terms of

humanness in order to be rated as having a mind. In other words, for any given degree of morphing between human and doll, when subjects believed the image was of an out group member, it received a lower rating. Dehumanization is expected to occur most often in response to extreme out-groups, groups stereotypically considered both hostile and incompetent (Harris and Fiske, 2006). Less human targets would therefore have fewer mental states to consider. Again, this is in line with emotional empathy—we empathize less with out-group members only when that out-group is the subject of preexisting negative attitudes (e.g., Cikara et al., 2011). This final more human/less human dimension in considering in-group versus out-group members appears to be a less flexible one than some of the others discussed, but that should not imply that it is resistant to flexible coalition formation. Consider an example from sports fandom: we might dislike a star athlete on a rival team, but if he is traded to our hometown team, quite suddenly he will be regarded much more favorably.

Despite the numerous studies that demonstrate that ToM does indeed vary between and within individuals, as well as along lines of in-groups versus out-groups, there is one aspect of its potential variation that the literature does not discuss: Is our ability to make mental states about others, like emotional empathy, modulated by immediate coalitionary cues. That is, does it demonstrate the same flexibility?

Chapter 5: Experimental Methods

The experiment reported on herein relies on methods from two categories of previous studies. First, it is a ToM task designed to be used with adult subjects, and second, the presentation of the experimental conditions is accomplished via framing techniques. Studies utilizing these techniques have been described in the preceding chapters, but the focus was on their findings. Therefore, prior to describing the present research, a brief review of these methods is in order.

Given that by age five, normally developing children across various cultures can consistently pass tests such as the Displaced Object task (Wellman, Cross, and Watson, 2001), use of a similar test with adults runs the risk of a ceiling effect with all subjects answering all questions correctly. Thus, ToM and false belief tests administered to adult subjects must be more complex than those presented to children and infants, for example, requiring subjects to understand and distinguish between false beliefs, lies, jokes, double bluffs, sarcasm (Cavallini et al., 2013); determining whether sentences describing a target's belief and a related or unrelated statement about the reality of a setting correspond with a photograph of the setting (Apperly et al., 2008); and testing subjects' memory of several stories involving complex social interactions (Kinderman, Dunbar, and Bentall, 1998). Two adult studies that bear further description are Birch and Bloom's (2007) multiple location Displaced Object task and Keysar, Lin, and Barr's (2003) perspective taking/false belief task.

Birch and Bloom's (2007) task begins similarly to the standard version, but there are four containers, rather than two. In addition, the containers can change positions as well as the object. Scenarios are presented in story form with pictures. For example, a

young girl, Vicki finishes playing her violin and places it in a blue container before going outside to play. There are also three other containers in the room, red, green and purple. While Vicki is gone, her sister Denise places the violin in another container. She then rearranges the containers so that the red container is now where the blue container was. Subjects are then presented with one of three conditions: Denise moves the violin to another, unspecified container (ignorance condition); she moves it to the red container (knowledge-plausible condition); or she moves it to the purple container (knowledge-implausible condition). Last, for each container, subjects assign a percent probability that Vicki will look for her violin when she returns. These conditions were designed to vary the extent to which subjects would be inclined to rely on their own knowledge rather than focus on what the story's protagonist, Vicki knows. In particular, the knowledge-plausible condition leads subjects to construct a seemingly plausible reason for selecting the red (and wrong) container—it is in the original spatial location, while there is no similar reason for selecting the purple container in the knowledge-implausible condition.

Keysar, Lin, and Barr's (2003) method, which served as the basis for the present study, creates an ambiguity between what two people, a subject and a confederate, know. A subject may make a correct knowledge attribution and take the confederate's perspective, or else fail to suppress their own knowledge and make a response based on their own perspective, much in the same way that children younger than five years of age do prior to mastering the Displaced Object task. Each participant was assigned a specific role for the experiment—the confederate as the “director,” whose job was to instruct the subject, the “follower,” in the subsequent task. The pair was seated across from each other at a table with a vertical grid divided into 16 cubbies (four rows of four) placed on

it. The subject had an unobstructed view of all 16 squares, while five were occluded on the confederate's side of the table. For each round of the experiment, multiple objects would be placed in the squares and the director would instruct the subject to move them to various other locations. The subject's responses were recorded on video and with an eye tracker.

Each round in this experiment consisted of a test pair of objects, such as a roll of scotch tape and a cassette tape, that could both be described with the same word (here, “tape”). One would be placed such that it was visible to the director, while the subject was instructed to place the other in a bag and then place that bag on one of the occluded squares. Among the other instructions, the director/confederate would ask the subject to move the test object (e.g., “Move the tape.”). Keysar and his colleagues were interested in the extent to which their subjects would fail to take the confederate's perspective and select the hidden object visible only to themselves. As described in Chapter 2, their subjects did commit a significant number of two types of errors, either by reaching for the hidden object outright, or else by initially looking towards it (selecting it visually) before turning to the mutually visible object.

Framing, presenting tasks or choices in multiple ways, can have a profound outcome on the actions or decisions we make (Tverskey and Kahneman, 1981). Such techniques have been used to great effect in empathy- and coalition-related studies discussed in Chapters 3 and 4. Yet in contrast to the level of complexity needed to present ToM tasks to adults, framing techniques can be quite simple. Turner, Brown, and Tajfel (1979) demonstrated how a simple framing can lead to a strong in-group bias with their minimal group paradigm. In their study, the framing technique had subjects view

pairs of images of abstract artwork and select their preference in each pair. After this they were told that, according to their preferences, they fell into one of two groups: shape people or color people. In fact, these assignments were arbitrary. However, they lead to subjects making economic decisions that favored others in their own group even though they never met any face to face.

Levine et al.'s (2005) “English football study” also used a simple framing technique to affect their subjects' perception of their in-group boundaries. By simply asking subjects to answer questions about either their Manchester United team fandom or football fandom more broadly, they achieved a differential response when subjects later witnessed an accident involving a person wearing a Manchester United team shirt, a Liverpool team shirt, or unbranded sport shirt.

Both of these methods do still rely on techniques that first ask the subject about themselves. Similar effects can be achieved by having subjects learn about others. Cronk and Wasielewski (2008) were able to induce their study participants to adopt an unfamiliar cultural norm, the Maasai gift-giving relationship osotua, by having them read a short passage about it and having them subsequently play an economic trust game. Labeling it as an “*osotua* game” evoked responses consistent with that cultural norm. While they argue that humans are particularly adept at attending to cultural cooperative norms, this study still illustrates that as long as one is familiar with a concept, merely renaming a task will evoke a different response.

And even simpler, changing one word in the instructions presented to subjects can be enough to alter their behavior on a subsequent task. Burnham, McCabe, and Smith, 2000) performed this manipulation in their study designed to explore the existence of a

preconscious “friend-or-foe” mental mechanism that might be involved in evaluating the intentions of other. To do this they had pairs of subjects participate in an economic trust game, and rather than use the neutral term “counterpart,” they referred to the other person in the pair as either a subject's “partner” or “opponent.” This resulted in different patterns of play in the game, with partners responding with more positively and with more trust compared to opponents, suggesting that we are quite sensitive to friend or foe/in-group versus out-group membership distinctions.

Hypotheses

While all the framing methods in the preceding section are simple, the last is particularly relevant to looking at the effect that a flexible coalitional psychology might have on ToM in that it highlights just how weak a frame can be and still result in significant differences in subjects' performance on a task. Taken together, the ToM and framing methods discussed above suggest that a ToM sophisticated enough for adult subjects paired with a simple framing task would indeed be a useful combination to test the sensitivity of ToM to cues of group affiliation.

I predicted that alternate framings of the instructions for a ToM task that present it as a neutral, cooperative, or competitive task with another individual would result in a group membership-dependent effect as for other types of empathy. That is, a subject's ability to make accurate mental state attributions on a ToM task would be sensitive to cues of coalition (cooperation versus competition) with a second individual involved in the task. Specifically:

1. Since cooperation/cooperativeness are positively associated with ToM (Elliott et

al., 2006; Paal and Bereczkei, 2007), when presented cues of shared group membership with the confederate, subjects will make more accurate attributions of confederates' mental state. This would manifest itself in a lower number of errors on a ToM task in response to a cooperative frame relative to a neutral control frame or a competitive frame.

2. Presentation of cues of opposing group membership will result in less accuracy in subject making mental state attributions of the confederate. Presentation of cues of opposing group membership would result in less accuracy in subject making mental state attributions of the confederate, resulting in a greater number of errors in response to a competitive frame relative to a neutral control frame or cooperative frame.

There are two additional predictions worth considering, though neither are directly related to the hypotheses above.

3. There should be a sex difference in the number of errors made on the task. Previous research has shown that females tend to outperform males on ToM and perspective taking tasks (e.g., Baron-Cohen et al., 2001; Focquaert et al., 2007; Ibanez, Huepe, Gempp, Gutierrez, Rivera-Rei, and Toledo, 2013), and if the task in this study is measuring the same underlying skill as these other ToM tasks, a similar result is predicted here.
4. And second, the two confederates who participated in the study were male and female. This results in four different subject/confederate pairings: female/female, male/female, female/male, and male/male. It is possible that this may represent

different in-group and out-group pairings, as sex is a conceptual primitive (Kurzban, Tooby, and Cosmides, 2001; Cosmides, Tooby, and Kurzban, 2003), a dimension of person perception that leads to automatic categorization equally across social situations. Thus, the sex of the confederate may affect subjects' perceptions of shared/opposing group status and so subjects may respond differentially to the two confederates. Specifically, in-group pairings (female/female and male/male) should result in fewer errors on the task, while out-group pairings (female/male and male/female) should result in more errors.

In testing these hypotheses, subjects were paired with a research confederate to work on a task (described in Methods). Each subject-confederate pair participated in one of three conditions: control (no framing), in-group prime, or out-group prime.

Methods

I chose to model my ToM task on the underlying concept of Keysar, Lin, and Barr's (2003) method because 1) while this is a laboratory study, a face to face task would more representative of a real-life interaction than reading stories (Cavallini et al., 2013), testing memory (Kinderman, et al., 1998), making judgments about sentence and photo pairs (Apperly et al., 2008) or the somewhat convoluted multiple location displaced object scenario task used by Birch and Bloom (2007). Also, 2) this method would draw more on subjects' automatic reactions to the task (similar to violation of expectation or anticipatory looking tasks described in Chapter 2), rather than a conscious reasoning task and 3) this set-up allowed for ease in presenting versions of the instructions specific to

each condition. As for the framing method, a subtle and minimal framing would appear to run the least risk of subjects intuiting the manipulation. It would also pair well with a task designed to test implicit ToM reactions. For that reason, I chose to model my framing method after Burnham, McCabe, and SMith's (2000) and present subjects with a neutral control condition along with cooperative and competitive conditions. While not as strong as some of the other manipulations, observing an effect using this technique would provide the best evidence for the sensitivity of ToM to group affiliations.

Pilot Study

In May, 2014, I conducted a short pilot study on a group of 6 students (mean age = 21.8 years, 4 female) at Rutgers University in New Brunswick, NJ. The test was loosely based on the method used by Keysar, Lin and Barr (2003). My goal for both the pilot and the full study was not to replicate their methods exactly, but to present a task that would present my subjects with the same type of ambiguity of stimulus and thus explore ToM in a comparable manner. The purpose of this pilot work was to do some basic testing of the methods to I planned to employ in the full study and to determine whether any changes to the procedure would be necessary.

Materials

Objects to be placed on table between subject and confederate: three wooden frogs (one small, one medium, and one large); two beanbags (one yellow, one red); one pair of castanets; white abstract figurine; metal interlocking gear puzzle; wooden abacus; rectangular river stone; black eyeglass case; wooden box with dragon design; box of

bandages

Extra objects remaining in a bag on the table: metal interlocking C-clamp puzzle; wooden ring puzzle; small wooden box.

Two cameras were set up in the room, one behind and to the right of each participant. One, an iPhone, was secured to a microphone stand with an iKlip and placed behind the subject's chair, aimed at the confederate (see Appendix for diagram of layout). This was a dummy camera and was not actually be powered on or recording. The second, a Nikon J1, on a tripod, was placed similarly behind the confederate's chair and aimed at the subject, zoomed in so the subject and items on table filled the majority of the viewing area.

Procedure

Due to the small number of subjects available, only a control condition with no framing was presented: In the instructions, the subject and confederate are referred to as Subject 1 and Subject 2, respectively, and the task is described as an interaction task.

Prior to the arrival of the subject and confederate, the experimenter arranges 13 objects on a table (see Appendix). A bag containing additional objects is placed to the side of this arrangement. Upon the subject's arrival, the experimenter invites him/her in and closes the door. The subject is asked to sit in the appropriate chair, and given the consent forms to read and sign. The experimenter explains that another participant is due to arrive and surreptitiously texts the confederate, who is waiting nearby, to come wait outside the door. This 1) ensures that the confederate arrives after the subject has come in

and the manipulation is complete, and 2) helps create the illusion that the confederate is actually another subject come to participate.

The experimenter then checks placement of the cameras to be used during the trial. While making camera adjustments, the experimenter makes a show of counting objects on table and explains that there should only be twelve, not thirteen. Included among the thirteen objects placed on the table are three wooden frogs, one small, one medium, and one large. While making a show of adjusting the cameras, the experimenter looks over the table and asks the subject, "It looks like I have too many items out on the table, could you put the large wooden frog back in the bag with the other extra things?" This is the manipulation that creates a similar ambiguity as in the Keysar et al. (2003) study: from the subject's perspective, the large frog has been put away, but when the confederate arrives she only see two wooden frogs, and from her perspective, the large frog is the subject's medium frog. When asked by the confederate to move the large frog during the experiment, whose perspective does the subject take? He/she may look to or reach into the bag where their large frog is, or he/she may look to and move the medium frog—what the confederate would see as the large frog.

After the manipulation is complete, the experimenter checks the door for the confederate (there is a sign outside asking participants to not knock when the door is closed). She is directed to the appropriate chair, and is also given the consent forms to read and sign. Next, subject and confederate are given a demographic information/personal identity phrase[1] (PIP) form to complete. After the paperwork is collected, the experimenter should introduce the task:

[1] A three word phrase, randomly selected by the subject to serve in place of a subject ID generated by the experimenter.. The phrase is used to link the video of the subject to their demographic information.

> "Thank you for your participation today. We will begin with a simple interaction task. Subject 1 [point to subject] will take on the roll of follower, whose job will be to carry out a list of 12 instructions read by subject 2 [point to confederate], who will take on the role of director."

The experimenter then hands the instruction list (see Table 5.1) to the confederate and starts the cameras[2]. The participants are instructed to hold up their PIP to the cameras for 5 seconds. The experimenter then indicates to the confederate that she should begin reading the instructions when ready. The cameras are stopped when task is finished. Upon completion, the experimenter reveals the true role of the confederate, who then leaves while subject is debriefed. The subject receives and signs a debriefing form, and the experimenter explains that there were two instances of deception used in the design (the true role of the confederate was hidden, as well as the full nature of the research question), and answers any questions.

Results and Discussion

With the limited number of subjects used and the use of only one condition, no meaningful statistical analyses can be performed from this study. However, that was not the goal of this study. Again, the purpose here was to assess the set up and procedures and to identify any potential issues needing correction prior to undertaking the full scale study.

Observational data indicated that some subjects did make errors similar to those noted by Keysar et al. (2003), including pauses, asking questions (to the confederate and the experimenter), and looking towards or reaching for the large frog hidden in the bag. However, three potential methodological issues emerged that needed to be addressed in

[2] Only the camera facing the subject is actually started, the experimenter simply mimes turning on the second one.

the full study:

1. Use of unfamiliar objects: Several subjects were unfamiliar with the names of two of the objects used, a small abacus, and a pair of castanets. As a result, they did not know which objects to move according to the instructions, suggesting that more commonly known/easy to describe objects should be used instead . The abacus and castanets were not used in the full study.
2. Presentation of instructions: The confederate read the list of instructions from a preprinted list. This was done to provide a consistent set of instructions across subjects without requiring the confederate to memorize the list. However, during the some of the trials, this recitation created an unintended ambiguity during the test question—rather than making judgments about the confederate's mental state or intent, some subjects saw the confederate simply as the medium of delivery. On hearing the test question ("Place the large frog next to the eyeglass case"), some subjects assumed I, the researcher, had made a mistake earlier in having them put away the large wooden frog. One subject looked at the experimenter and said, "We put the large frog away, so I guess it would be this one." This observation was confirmed with other subjects during debriefing. To prevent this occurrence in the full study, confederates improvised a list of instructions during each trial, consistent with Keysar, Lin, and Barr (2003).
3. Gaze direction as experimental data: Though looking errors were detectable during the course of each trial, they were difficult to detect/confirm on the video recordings taken of each subject, especially if they were limited to eye movements only. Without the availability of eye tracking equipment or alternate

video equipment, it was not possible to consistently and accurately extract looking errors from the videos in the full study. Instead, latency in object selection (operationalized as pauses/hesitations in a response) was selected as a substitute: in cases where subjects needs to consider which object to select, there is a hesitation in subjects' object selection.

4. Last, steps used to create the illusion that the confederate was participating as a subject were needlessly complex. The same effect could be achieved without surreptitious texting or a second, prop camera. It was decided to use a single camera in the full study and to simply have the experimenter "decide" on the roles each participant would play when handing out consent forms and explain that due to the nature of the task, only one of the participants needed to be recorded.

Full Study

Subjects

A total of 122 subjects participated in this study, between December 2014 and May 2015. Three were excluded from analysis due to incomplete data, for a revised total of119 (control condition n = 65; In-group manipulation n = 28; Out-group manipulation n = 27). Subjects were asked to provide age, sex and race/ethnicity: the mean age was 20.1 years; 62.3% were female (n = 74); and self-reported race fell into the following categories: 44.5% Caucasian (n = 53); 22.7% Asian (n = 27); 14.3% African-American (n = 17); 7.5% Indian (n = 9); 6.7% Hispanic (n = 8); and 4.2% other/mixed (n = 5). See Table 5.2 for the full list of self-reported racial categories. Subjects were all undergraduate students at Rutgers University in New Brunswick, NJ, recruited from large

lecture classes in the Anthropology and Philosophy departments. They were compensated with either a $10 Visa gift card or extra course credit for their participation.

Materials

- Distractor objects (see image 5.1): Three plastic cups (one red, one blue, one green), two beanbags (one yellow, one red), white abstract figurine, metal interlocking gear puzzle, metal interlocking C-clamp puzzle, wooden ring puzzle, rectangular river stone, large wooden box, small wooden box, wood block, Altoids tin, purple medium-sized binder clip, deck of playing cards in a plastic case, plastic egg (Silly Putty container), old-fashioned iron key, red wooden bowl, and two stacks of three 2x4 Lego bricks (one gray, one yellow)
- Test objects (see Image 5.2): Small stack (single 2x4 brick), medium stack(two 2x4 bricks), and large stack (three 2x4 bricks) of blue Lego bricks; a small, medium, and large black binder clip; and a small, medium, and large paper clip.
- A white foam core board measuring 20 inches by 21.63 inches (50.8 x 54.93 cm) divided into a 4x4 grid marked with black lines, each cell measuring 5 inches by 5.41 inches (12.7 x 13.7 cm).
- A white cardboard blinder measuring 19.7 inches high by 30.75 inches wide (50 x 78 cm), with a vertical crease allowing it to stand upright, reducing its width to approximately 26 inches (66 cm).

A single digital video camera was set up to record each trial for the purpose of collecting and later scoring subjects' responses. All trials were filmed using a Nikon J1

camera and a 1 Nikkor lens (10-3-mm 1:3.5-5.6 VR ϕ40.5). The camera was mounted on a Vantage Commander V tripod, extended to full height (50 inches/127 cm tall), behind and to the right of confederate.

Procedure

For the present study, the experimental set up was as follows:

As with the pilot study, this experiment was set up as a paired interaction between two individuals: one subject and one experimental confederate posing as a second subject. Two Rutgers undergraduates, one female and one male, were recruited to serve in this role, taking turns participating in blocks of trials throughout the duration of data collection. Each confederate participated in approximately equal numbers of trials in total (n_{female} = 58; n_{male} = 61) and within each experimental condition (control: n_{female} = 32, n_{male} = 32; cooperative: n_{female} = 13, n_{male} = 15; competitive: n_{female} = 13, n_{male} = 14). Prior to participating in trials with the subjects, the confederates received training in the methods that follow; it was their responsibility to ask the key test question in each round with a subject that would generate the responses being measured.

At the beginning of each trial, the confederate and researcher waited together in the study room for the subject to arrive. Upon their arrival, both the subject and confederate were given a consent form to sign in order to create the illusion that the confederate was participating as an actual subject and not an assistant. In the event a subject arrived early, before a previous trial has ended, some other explanation for the confederate's presence was given, such as that due to the nature of the study, the researcher has scheduled some volunteers to participate in multiple rounds. The

researcher selected one of the two participants—always the true subject—to be video recorded during the experiment and had him/her fill out additional consent paperwork, demographic information and, as described in the pilot study methods above, create a personal identity phrase (PIP).

The participants were then seated across from each other at a table upon which the empty white foam core grid was placed, and the experimenter introduced the experiment as follows:

> “Thank you for your participation today. I'm going to have you engage in a [manipulation type: interaction/cooperative/competitive] task that I've designed to test exploring how different ways of interacting with others affects our ability to infer their thoughts and beliefs. The way this will work is that I'm going to have each of you adopt a role throughout the study: one of you [point to confederate] will take on the role of "director" whose job will be to come up with a list of simple instructions for moving around various random objects I'll place on the grid. For example, “Move the red box to the space in front of the white figurine,” or “put the yellow beanbag here [point to location]”. And the other [point to subject] will take on the roe of "follower" whose job will be to carry out those instructions. The experiment will consist of 5 rounds, and for each round, I'd like you [indicating confederate] to come up with eight different instructions. We'll do them one at a time, alternating between an instruction and [indicating subject] your response. And last, as I set up the grid at the start and between each round, I'll place this blinder up to block your [indicating confederate] view until we're ready to start the round. Any questions?"

After these instructions were given and questions answered, the researcher placed the blind to block the confederate's view and set up the grid while the subject watched. For each round, a total of ten objects were removed from a bag and placed on the grid: four distractor objects, a set of three test objects, and three plastic cups. Two cups were placed upright, and the third was placed upside down over the smallest of the three test objects. (See Image 5.3 for an example set-up with test object exposed, as would be presented to subjects and Image 5.4 for the set up as it would appear to the confederate.)

Last, the blinder was removed, exposing the layout to the confederate. The camera was started, subject's PIP shown to the camera, and the pair was then instructed to begin.

Placing the smallest test object under one of the cups creates the same ambiguity between the perspectives of the two participants as in the pilot study and Keysar et al.'s (2003) study. At the beginning of each round the subject knew there were, for example, a small, medium, and large binder clip on the table. Since this happened out of view of the confederate, he or she would, from the subject's perspective, incorrectly believe there are only two binder clips on the table, one small and one large. This created an ambiguity regarding which clip is the small clip—to the subject it was the clip hidden under the cup, but to the confederate it was the visible, medium sized clip.

The confederate, however, having received previous training regarding the scenario, was aware of the three sets of test objects, and seeing two binder clips, two blue Lego brick stacks, or two paper clips among the other objects, and knew that the third (the smallest) was hidden under the cup. In addition, the confederates were instructed to include one instruction each round that exploits this ambiguity between perspectives; this instruction always referred to the *small* test object (e.g., "Move the *small* binder clip one square forward."). And, importantly, confederates were told to avoid asking subjects to move the overturned cup, as this would expose the presence of the hidden object.

This is a test of subjects' ability to take the perspective of others, and, while not a False Belief task in the strictest sense, subjects must take confederates' ostensible erroneous belief into account in order to respond to this prompt correctly and so is a test of false belief understanding and ToM. Does the subject take the confederate's perspective, and look at or reach for the *medium* binder clip in plain view? Or do they

take their own perspective and select the *small* binder clip under the cup?

Task manipulations

Each subject was exposed to one of three frames during the course of their trial: a control condition, a cooperative condition, or a competitive condition. All subjects heard the same set of instructions (see above), with the exception of only one word changing in the second sentence across conditions. These conditions were designed to subtly prime subjects to view the confederate as either a neutral party, a coalitional partner (in-group member), or an opponent (out-group member):

1. *Control frame:* This frame represents a neutral condition in which the task was referred to in the instructions above as an "interaction task."
2. *Cooperative frame*: In this version of the task, the instructions remain identical to the control condition except for one change; the task was referred to as a "cooperative task."
3. *Competitive frame:* Again, the instructions remain the same except for the task name, In this condition it was referred to as a "competitive task."

Debriefing

The task manipulations above all involved deception in that subjects were led to believe the confederate is also a legitimate subject. It was necessary to keep subjects ignorant about the confederate's status in order to assign the subject to the "follower" role without arousing suspicion or adversely affecting the outcome of the trials. In addition, the instructions ad framing techniques were also considered deceptive, in that they were intended to hide the experimental hypothesis from the subjects, and cause the subjects to

think about the task, their relationship to the confederate, and the roles of director and follower in different ways without explicit instruction to do so. This was necessary; foreknowledge of the hypothesis may lead subjects to (un-)consciously perform in a way that is expected, rendering the data invalid. After each subject has completed their participation, they were debriefed as to the exact nature of the research question, as well as the deception involved. This was presented in the form of a written statement that each subject read and signed. Once the deception was explained, subjects were asked to indicate on the form whether or not they wish to withdraw from the study and have all records of their involvement removed from the principal investigator's files. No subjects chose to withdraw.

Coding/Scoring

Each subject participated in one experimental trial. Each trial consisted of five rounds of eight instructions. Within each round, one test question was administered, for a total of five per trial. Subjects were video recorded for later review and scoring of their performance on the task (scoring/coding of responses was not undertaken during the trial). Each subject received three error scores, summed across all five rounds: Response errors, Hesitation errors, and a Total error score, which were defined as follows:

1. Response error: Since this test was a measure of perspective taking, a response error occurred whenever a subject selected and moved the hidden, small-sized test object, rather than the visible, medium-sized test object in a given round, whereas a response was coded as correct when a subject selected and moved the visible object. Responses were coded as 0 for correct and 1 for incorrect response,

resulting in a Response Error Score ranging from 0 to 5.

2. Hesitation error: A *hesitation* error occurred whenever the ultimate response (selection of either the visible or hidden object) was preceded by a pause/hesitation relative to the speed of response to the non-test questions. Such pauses also included mid-choice redirections (initially reaching for the hidden object but ultimately selecting the visible object, and vice versa). Hesitations were similarly recorded as 0 (for no hesitation in a subject's response) or 1 (hesitation before responding), resulting in a Hesitation error score also ranging from 0 to 5.
3. Total error: This was calculated simply as the sum of Response and Hesitation error scores, and ranged from 0 to 10.

These two error types were chosen to reflect the two error types used in Keysar, Lin, and Barr's (2003) study: reaching errors and looking errors. They argued that it was useful to consider these two types because they measured both overt response (reaching error) as well as a more subtle/implicit mistake (initially looking towards a hidden object) that would otherwise be missed. Here, the two error types may be thought of similarly. A response error represents an explicit error, automatically taking one's own perspective. Hesitations are not necessarily entirely analogous to Keysar, Lin, and Barr's looking errors, but they do indicate an uncertainty on the part of the subject; whether a correct or incorrect response is made, the subject is momentarily confused about which object should be selected.

The hesitation error score was initially intended to be a measure of decision/response duration, measuring the time it took subjects, from the confederate's

completion of the instruction, to select an item to move and perform the instructed action. The assumption was that rather than being a binary score, hesitation, or response latency, would be a continuous variable, with longer hesitations indicating more response confusion. Due to limitations of the available equipment and methods (see Chapter 7 for a discussion), this was not possible. However, the Hesitation errors as recorded still reveal a more subtle error than the pure Response errors, in that they indicate a difficulty in selecting a perspective to take even in the case of selecting the correct test object. In order to avoid overestimation of Hesitation errors, only errors that were directly attributable to indecision were counted. Hesitations due to mishearing an instruction or asking the confederate to repeat the instruction were not considered (both of these hesitations also occurred in response to non-test questions), nor were hesitations that were consistent with similar pauses prior to responding on the other, non-test questions in a given round.

Analysis

All data was analyzed using R version 3.0.2 (2013, The R Foundation for Statistical Computing) and RStudio version 0.98b (2015, The Foundation for Open Access Statistics).

To begin, two-tailed t-tests were run to test for any overall sex differences in each error type (Response, Hesitation, and Total), as predicted in hypothesis 3.

The main hypotheses (1 and 2) tested involve comparison of three different conditions: Control, Cooperative, and Competitive. Testing for differences in the mean number of errors made across conditions calls for a one-way ANOVA. This was done for

each error type (Response, Hesitation, as well as Total). In addition, to test for any interactions between subject sex and condition, three 2 (subject sex) x 3 (condition) ANOVAs were run, one for each error type. And last, to test for further interactions attributable to the confederates, including the predicted subject/confederate pairing effects (hypothesis 4), three 2 (subject sex) x 2 (confederate) x 3 (condition) ANOVA were run, again for each error type.

A significance level of $p < 0.05$ was selected for all analyses. However, given the relatively small sample sizes of the Cooperative ($n = 28$) and Competitive ($n = 27$), results falling between $p < 0.1$ and 0.05 were considered for discussion as well. Though they fall short of the standard significance cut-off, they may indicate a trend that is worthy of follow up in future studies.

Table 5.1: Confederate's list of instructions for Pilot Study

1. Turn the stone upside down
2. Stand the abacus upright
3. Rotate the eyeglass case 90 degrees
4. Put the castanets in front of the white figurine
5. Put the red beanbag next to the stone
6. Place the large frog next to the eyeglass case
7. Turn the dragon box upside down
8. Turn the white figurine to face the opposite direction
9. Place the eyeglass case in front of you
10. Place the Band-aids in the dragon box
11. Place the gear puzzle in front of me
12. Place the yellow beanbag on top of the red beanbag

Table 5.2: Subjects' Self-Reported Racial Categories

Category			**n**	**%**
Caucasian/White			53	44.5
Asian	Asian	16		
	South Asian	2		
	East Asian	1		
	Chinese	5		
	Filipino	3		
	Asian Total		27	22.7
Black/African-American			17	14.3
Indian			9	7.5
Hispanic	Hispanic	5		
	Latino	2		
	Dominican	1		
	Hispanic Total		8	6.7
Other	Pakistani	3		
	Asian-White	1		
	Latino-White	1		
	Other Total		5	4.2
Total			119	

Image 5.1: Full set of Distractor items and cups used in full study.

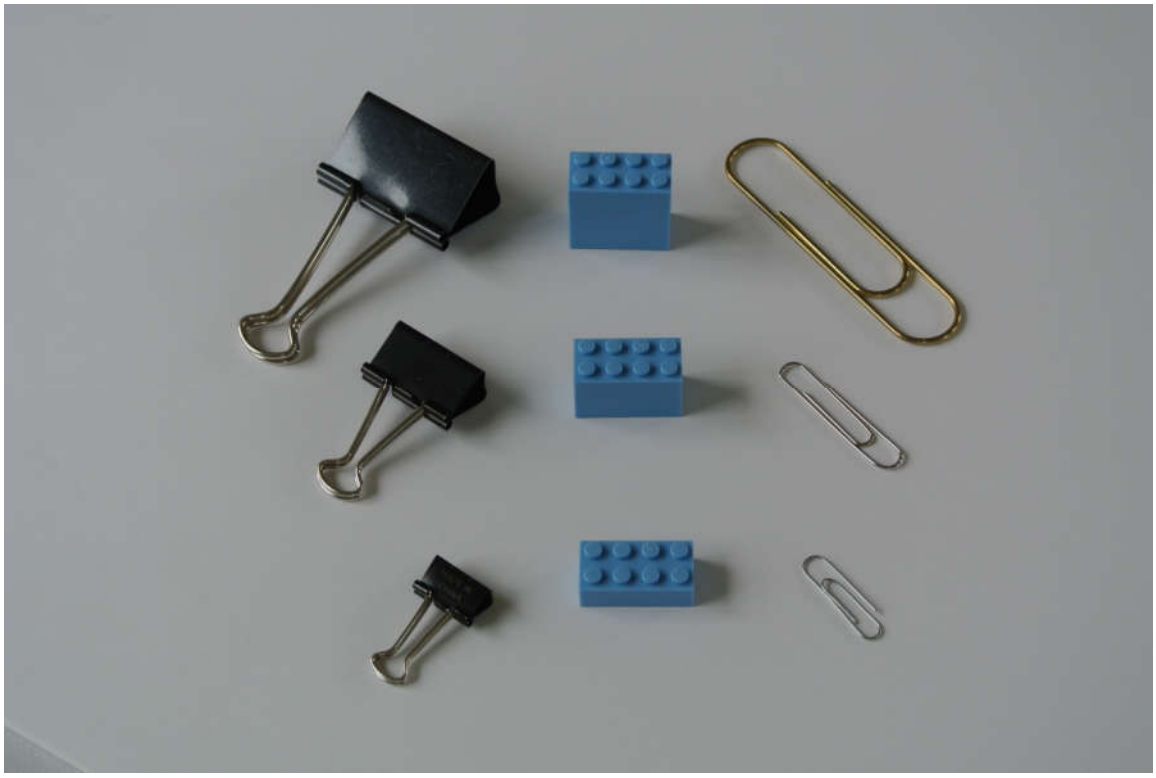


Image 5.2: The three sets of small, medium, and large test items used in the full study.

Image 5.3: Example of initial set-up for full study, but with small test item (binder clip) shown next to green cup.

Image 5.4: Example of initial set-up for full study with small test item (binder clip) hidden under green cup.

Chapter 6: Results

Sex Difference in Error Scores

Restricting comparison only to male versus female subjects (Table 6.TK), independent of experimental condition, two-tailed *t*-tests show that the two groups performed differently on the task, with female subjects committing significantly fewer Total errors (Response + Hesitation) than male subjects (M_{Female} = 1.919*;* M_{Male} = 3.178*; df* = 80.619 *; t* = -2.588; *p* = 0.011). This was due to Response errors, which were also significantly different, again with female subjects committing significantly fewer errors than males (M_{Female} = 0.851*;* M_{Male} = 1.178*; df* = 74.204 *; t* = -2.491*; p* = 0.015). There was no significant difference in Hesitation errors between sexes (M_{Female} = 1.068*;* M_{Male} = 1.4*; df* = 83.97 *; t* = -1.392*; p* = 0.168).

Error Scores and Experimental Frames

If the framing method used in this study affected subjects' ToM across condition, we should see a difference in the mean number of errors made on the task. Specifically, the number of errors (response, hesitation or total) made in the cooperative condition should be lower than either the control or competitive conditions and higher in the competitive condition. Three one-way ANOVAs were run to separately compare Response errors, Hesitation errors, and Total errors within each manipulation (Control, Cooperative, and Competitive).

Framing condition did not have an effect on either Total errors ($F(2, 116) = 1.234$, $p = 0.29$, see Table 6.TK) or Response errors ($F(2, 116) = 0.289$, $p = 0.75$, see Table 6.TK). However, there was a potentially significant effect on Hesitation errors ($F(2, 116)$

= 2.666, significant at $p = 0.074$, see Table 6.TK); though it does not conform to the $p < 0.05$ level set for true significance. Further analysis of the Hesitation error data with Tukey's HSD reveals that the difference lies between the Cooperative and Competitive frames ($p_{adj} = 0.067$), still not technically significant, but suggestive nonetheless. Thus, while neither the Cooperative nor Competitive conditions differed significantly from the Control condition, they may differ from each other. This provides only partial confirmation of the experimental hypothesis: the ability to make accurate mental state attributions is sensitive to cues of coalition as measured by Hesitation errors in this perspective-taking task.

Error Scores, Experimental Frames, and Subject Sex

Given that there were significant sex differences in the number of errors made (independent of condition), further analysis adjusting for this difference is necessary. To address this, three separate two-way ANOVAs were subsequently run in order to examine the effect on Total, Response, and Hesitation errors by sex of subject and experimental frame within each error type.

The ANOVAs for all three error types (see Tables 6.TK, 6.TK, and 6.TK) confirmed the sex difference seen in the initial *t*-test. Likewise, the Hesitation ANOVA confirmed the Cooperative – Competitive frame difference from the one-way ANOVA described above. Both of these results were as expected. However, turning to possible interaction effects between subject sex and experimental frame, none of the ANOVAs indicated any significant subject sex: experimental frame interaction.

Error Scores, Frames, Subject Sex, and Confederate

Last, one final set of ANOVAs were run to take into account a possible third variable and examine the mediating role of sex of confederate relative to the subject. Here, each error type was analyzed in terms of subject sex, frame, and now additionally, confederate.

Total errors: Considering only new interactions, there were two significant results to report. As predicted, there was a significant interaction between subject sex and confederate sex ($F(1, 107) = 8.199$, $p = 0.005$, see Table 6.TK). Tukey's HSD reveals that the subject sex-confederate interaction was likely due to three interactions: male:female – female:female ($p_{adj} = 0.001$); female:male – female:female ($p_{adj} = 0.091$, though technically not significant); and male:male – female:female ($p_{adj} = 0.097$, again, technically not significant). That is, 1) male and female subjects significantly differed in their responses (males made more errors) to the female confederate, 2) female subjects may have differed in their responses to a male vs. female, making more errors in response to the male confederate and 3) the male subject paired with the male confederate may have made more errors than the female subjects paired with the female confederate.

In addition, the three-way interaction between subject sex, experimental frame and confederate sex was significant ($F(2, 107) = 4.105$, significant at $p = 0.02$, See Table 6.TK). Tukey's HSD reveals only two significant triads. First, the male:female – female:female difference appeared to be restricted to the control condition, with male subjects making more errors than females when paired with the female confederate in the control condition only ($p_{adj} = 0.042$). This difference did not extend to either the Cooperative or Competitive conditions. Second, the female:control:female –

male:competitive:female interaction was not technically significant, but still fell within the $p < 0.1$ range ($p_{adj} = 0.084$): female subjects in the control condition with the female confederate made fewer errors than male subjects in the competitive frame with the female confederate. However, neither of these interactions provides support to the experimental hypothesis. The first is descriptive of the control group, telling us nothing about either the Cooperative or Competitive conditions. However, an alternate way to look at this is that this sex difference disappears in the Cooperative and Competitive frames. The second seems to be difficult to interpret meaningfully as it addresses one sex in one condition to the other sex in a second condition.

Response errors: As with Total errors, there were two new Response error interactions to report. There was a significant interaction between sex of subject and confederate ($F(1, 107) = 4.514$, significant at $p = 0.036$, see Table 6.TK). Tukey's HSD shows this to be due to the same three interactions: male:female – female:female ($p_{adj} = 0.01$); female:male – female:female ($p_{adj} = 0.073$); and male:male – female:female ($p_{adj} = 0.034$), with the second of the three not achieving the $p < 0.05$ criterion. In each case, the subjects in the first pairing made significantly more Response errors than the subjects in the second pairing.

The three-way interaction (subject sex – frame – confederate) was marginally significant ($F(2, 107) = 2.847$, significant at $p = 0.062$, See Table 6.TK), with Tukey's HSD showing this to be due to only one barely significant triad: male:control:female – female:control:female ($p_{adj} = 0.1$). Male subjects made more Response errors in the Control condition with a female confederate than did female subjects in the same. Again, this is a comparison that does not address the experimental hypothesis in any way as it is

limited to the control condition.

Hesitation errors: Last, with Hesitation errors, the same two interactions as Total and Response errors were significant: the two-way subject sex – confederate ($F(1, 107) = 6.073$, $p = 0.015$, see Table 6.TK) and the three-way subject sex – frame – confederate interaction ($F(2, 107) = 2.969$, $p = 0.056$, see Table 6.TK), though the three-way interaction is just short of true significance at the $p < 0.05$ level. Tukey's HSD identifies two likely interactions for the first: male:female – female:female ($p_{adj} = 0.024$) and male:male – male:female ($p_{adj} = 0.093$). Male subjects made significantly more Hesitation errors in response to the female confederate than did female subjects. In addition, they made fewer errors (not quite significant) in response to the male confederate than to the female confederate. In the three-way interaction, two groups may have contributed to the significance, female:control:female – male:competitive:female ($p_{adj} = 0.064$) and female:cooperative:female – male:competitive:female ($p_{adj} = 0.057$), though neither quite reached the $p < 0.05$ level. While the confederate remains constant, these interactions are again comparing one sex in one condition to another sex in a second condition, making meaningful interpretation difficult.

Table 6.1: Mean Total Errors by Subject sex

Sex	n	Response	Hesitation	Total
Female	74	0.851	1.068	1.919
Male	45	1.778	1.400	3.178

Table 6.2: Mean Total Errors by Type

Condition	*n*	*Response*	*Hesitate*	*Total*
Control	65	1.266	1.141	2.406
Cooperate	28	0.964	0.893	1.857
Compete	27	1.296	1.630	2.926

Table 6.3: ANOVA Effect of Frame on Total Errors

	Df	SumSq	MeanSq	F	Significance
Frame	2	15.7	7.860	1.234	0.295
Residuals	116	738.7	6.368		
Total	118	754.4			

Table 6.4: ANOVA Effect of Frame on Response Errors

	Df	SumSq	MeanSq	F	Significance
Frame	2	2.1	1.041	0.289	0.749
Residuals	116	417.1	3.596		
Total	118	419.2			

Table 6.5: ANOVA Effect of Frame on Hesitation Errors

	Df	SumSq	MeanSq	F	Significance
Frame	2	7.85	3.923	2.666	0.074 (< 0.1)
Residuals	116	170.71	1.472		
Total	118	178.56			

Table 6.6: ANOVA Effect of Sex and Frame on Total Errors

	Df	SumSq	MeanSq	F	Significance
Sex	1	44.3	44.35	7.230	0.008 (<0.01)
Frame	2	11.7	5.84	0.952	0.389
Sex:Frame	2	5.3	2.65	0.433	0.650
Residuals	113	693.1	6.13		
Total	118	754.4			

Table 6.7: ANOVA Effect of Sex and Frame on Response Errors

	Df	SumSq	MeanSq	F	Significance
Sex	1	24.0	24.017	6.925	0.010 (<0.01)
Frame	2	0.9	0.435	0.125	0.882
Sex:Frame	2	2.4	1.197	0.345	0.709
Residuals	113	391.9	3.468		
Total	118	419.2			

Table 6.8: ANOVA Effect of Sex and Frame on Hesitation Errors

	Df	SumSq	MeanSq	F	Significance
Sex	1	3.09	3.092	2.086	0.151
Frame	2	7.23	3.614	2.438	0.092 (<0.1)
Sex:Frame	2	0.74	0.368	0.248	0.780
Residuals	113	167.50	1.482		
Total	118	178.56			

Table 6.9: ANOVA Effect of Sex, Confederate, and Frame on Total Errors

	Df	SumSq	MeanSq	F	Significance
Sex	1	44.3	44.35	7.962	0.006 (< 0.01)
Frame	2	11.7	5.84	1.048	0.354
Confed	1	3.6	3.57	0.640	0.425
Sex:Frame	2	5.3	2.65	0.476	0.623
Sex:Confed	1	45.7	45.67	8.199	0.005 (< 0.01)
Frame:Confed	2	2.2	1.10	0.198	0.821
Sex:Frame: Confed	2	45.7	22.86	4.105	0.019 (< 0.05)
Residuals	107	596.0	5.57		
Total	118	754.7			

Table 6.10: ANOVA Effect of Sex, COnfederate, and Frame on Response Errors

	Df	SumSq	MeanSq	F	Significance
Sex	1	24.0	24.017	7.318	0.008 (<0.01)
Frame	2	0.9	0.435	0.132	0.876
Confed	1	7.1	7.139	2.175	0.143
Sex:Frame	2	2.3	1.144	0.349	0.706
Sex:Confed	1	14.8	14.814	4.514	0.036 (<0.05)
Frame:Confed	2	0.2	0.096	0.029	0.971
Sex:Frame: Confed	2	18.7	9.344	2.847	0.062 (<0.1)
Residuals	107	351.2	3.282		
Total		118	419.2		

Table 6.11: ANOVA Effect of Sex, Confederate, and Frame on Hesitation Errors

	Df	SumSq	MeanSq	F	Significance
Sex	1	3.09	3.092	2.220	0.140
Frame	2	7.23	3.614	2.594	0.079 (<0.1)
Confed	1	0.61	0.614	0.441	0.508
Sex:Frame	2	0.71	0.354	0.254	0.776
Sex:Confed	1	8.46	8.461	6.073	0.015 (<0.05)
Frame:Confed	2	1.10	0.548	0.393	0.676
Sex:Frame: Confed	2	8.27	4.136	2.969	0.056 (<0.1)
Residuals	107	149.08	1.393		
Total	118	178.55			

Chapter 7: Discussion

The two secondary predictions discussed above were both confirmed: First, there is an overall sex difference in error scores between female and male subjects, with males committing more Total and Response errors than females. This is line with previous research that has also found such a sex difference on ToM and perspective taking tasks. Second, there was an interaction effect between the sex of subjects and confederates. Looking at both the Total errors and Response errors, the same three sex pairings were significant: 1) When paired with a female confederate, male subjects commit more errors (i.e., they take their own perspective more) than females, 2) Female subjects take their own perspective more often when faced with a male confederate than they do when facing a female, and 3) males facing males take their own perspective more than do females facing females. Turning to Hesitation errors, males again tend to take their own perspective more often than female subjects when facing a female. In addition, males made more Hesitation errors in response to the male confederate than they did facing the female confederate. Since these outcomes are exclusive of experimental frame, they cannot be interpreted in those terms with certainty, but overall, females paired with another female commit fewer errors than the other pairings, while male-male or mixed sex pairings result in subjects taking their own perspective more. However, this may be related to an evolutionary context of aggression by out-group males (e.g., Navarrete, McDonald, Molina, and Sidanius, 2010; Yuki and Yokota, 2009; McDonald, Navarrete, and Van Vugt, 2012), whose appearance would have been a threat to females (and males) of a group. Such males would not make appropriate choices for coalitionary partnerships. Strange males, or at least unknown/unfamiliar males, may yet represent a threat that

might be difficult to overcome in the context of the simple cooperative/competitive framing used in the present study. However, this is speculative.

Turning to the primary predictions (Hypotheses 1 and 2), only partial support was found. The ability to make accurate mental state attributions is sensitive to cues of coalition in terms of cooperation versus competition framing of the perspective taking task. This was true only for Hesitation errors, which as discussed above, represent a potentially more sensitive measure. While the prediction was that both the Cooperative and Competitive conditions would significantly differ from the Control condition, this was not the case. However, the Cooperative and Competitive conditions were different from each other, and in the expected direction: Subjects in the Cooperative frame made fewer perspective taking errors than subjects in the Competitive frame. While these results were shy of the standard significance cut-off of $p < 0.05$, the fact, however, that they were obtained merely through the changing of only one word in the instructions, without subjects being explicitly directed to that change must be remarked on. That such a subtle framing difference has an effect on subjects suggests that other, more involved framing techniques may result in a stronger effect (see Future Directions, below). Furthermore, it demonstrates that our attention to coalitionary cues is quite sensitive.

Further breaking down the conditions by incorporating additional explanatory variables revealed additional significant results, but none that clearly lent support to Hypotheses 1 or 2. Given the overall sex difference between females and males, one might expect that to have an effect on the framing conditions; perhaps the lower error scores of the female subjects reduced the effect the frames had when considering both sexes together. However, there were no interactions to note. Likewise, given that there

was a significant interaction between sex of subject and sex of confederate independent of frame, perhaps this too might be expected to affect the framing conditions. Again, however, while in this case, there were some significant interactions, none were directly related to the hypothesis under consideration.

Potential Issues

There are few meaningfully significant results to discuss in terms of the hypothesis being tested in this study. Is this due to issues related to the hypotheses or to methodological limitations? It could be the case that ToM is simply not responsive to the type of situational coalitional cues that were presented. It could also be the case that the framing method used was not effective enough and failed to prime subjects to behave in accordance with the different coalitional cues presented. In selecting a framing method, I did knowingly choose the least overt technique, which unfortunately carried the greatest risk for negative results. Another factor to consider is the relatively small sample sizes of the Cooperative and Competitive framing groups compared to the Control group; increasing the number of participants in both of those conditions could result in stronger statistical effect sizes as well as reveal additional significant interactions that may have been suppressed.

Given the data, there is no way to be certain which is the case. However, the relative values of each error type do show a non-significant trend in the predicted directions (see Table 6.TK). Fewer errors, Response, Hesitation, and Total, were made in the Cooperative condition compared to the Control and the Competitive conditions. Likewise, subjects in the Competitive condition made more Hesitation and Total errors.

As these are not significant results, caution must be used in drawing any conclusions, but given that their directionality, I choose to be hopeful and continue to pursue my overall hypothesis as it is, and instead focus on potential methodological issues that could explain the current lack of results, leading to its confirmation in the future.

Despite the changes made in response to the pilot study, there were still several issues that appeared in the full study that need to be addressed in future work. The first and foremost of these is related to the necessarily subjective method used in measuring hesitation errors. This may have led to an underestimate of true hesitation errors as well as inclusion of hesitations not directly related to the experimental design.

In regards to the former, Keysar, Lin, and Barr (2003) were able to achieve precise much more measurements of their subjects' looking (hesitation) errors with the aid of eye tracking equipment. This allowed them to accurately capture hesitations and glances measuring fractions of seconds—durations much too brief to record given the equipment limitations of the present study. An example of the latter would be a visual comparison of the two visible test objects to confirm which is the smaller. Subjects' gaze was often obscured from the camera making it impossible to distinguish between this type of comparison and a hesitation arising from deciding between a visible and hidden object. Without the aid of eye tracking equipment, this variability obscures brief errors of hesitation and gaze direction.

An initial attempt was still made to measure subjects' reaction times whether their response to the test question was to select the visible or hidden test object. Two different intervals were considered: the duration between the end of the instruction and first contact with the selected item or the duration between the end of the instruction and

completion of the instructed movement. However, other factors led to a decision to abandon both of these as a potential source of meaningful data. One was the variability in the pacing of instructions by the confederates between rounds and trials. An inadvertent pause (e.g., “Put the small… binder clip on the wooden box”) could potentially give subjects time to orient to one object or the other giving the appearance that there was no hesitation error. Different tasks take different amounts of time, such as a subject moving an object directly in front her one square to the left, vs. picking something up and putting it in a cup on the other side of the grid.

Subjects also exhibited other timing issues unrelated to the decision, sometimes waiting for an instruction to be completed, other times selecting an object in anticipation of the second half of the instruction. Alternately, sometimes subjects made errors on non-test instructions, such dropping items, or pausing to ask clarifying questions regarding the destination, confirming they heard the instruction correctly.

Another issue confounding timing data was related to the design of the experiment. In each round of each trial, all objects—distractors, cups and test—were placed randomly on the grid. This means that at times the hidden and visible test objects could be, one or both closer to or further away from the subject, leading to artificially lengthened or shortened response times. To avoid this in the future, the placement of objects will need to be identical each round across subjects. And last, subjects also encountered difficulty picking up some of the smaller test items, especially the paper clips, which also artificially extended their reaction times.

Many of these issues are easily addressed and can be corrected in future studies: inclusion of new objects (both distractor and test) that are easier to manipulate,

standardization of grid maps across trials, training confederates to be more fluent in their instruction dictation, and acquiring additional funding for eye tracking or other equipment that allows for more precise measurements.

Despite these issues, subjects still committed obvious hesitation errors in response to the test questions that ranged from quick pauses to longer episodes of thinking through their decisions, to asking questions to the confederate such as “The small one or the smallest one?” or even stating “There's a smaller one.” And while these issues did not apply to the other class of error, that of the actual selection, a comment made by one subject reveals a potential confound in the response data, “I felt I needed to keep my responses the same, from his perspective.” This suggests that once the initial response was made in round 1, this (and perhaps other) subjects made a decision to be consistent in their response and make the same choice on subsequent rounds. Seventy-nine percent of subjects maintained the same response to the test question throughout their entire trial, with most of those (eighty-two percent) taking the confederate's perspective. To guard against this possibility in the future, the instructions given by the experimenter can be amended to give permission to change one's mind.

A final, and perhaps most important, methodological issue that must be addressed in future work is related to the actual coding of errors from the video recordings. The principal investigator was responsible for all scoring, and so was not obtained in an optimally blinded condition. Steps were taken, however, to minimize potential biases in scoring. To begin, there was no indication of which experimental condition a subject was participating in on any of the videos. This was kept separate through the use of Personal Identity Phrases (PIP) which were displayed on the initial round of each subject's

participation; experimental condition and Personal Identity Phrase were noted on subjects' demographic information sheet and the conditions were not linked to the scoring data until all videos had been coded. In addition, subjects' videos were shuffled and recorded in random order to prevent any memory on the part of the investigator of the day or condition a given subject participated in. Despite these measures, some bias in coding errors, particularly the more subjective Hesitation errors, may have been present. To avoid this possibility in the future, all scoring should be done by naive coders.

Future directions

In addition to correcting the potential data collection issues discussed above, future work on this topic should include additional framing techniques and subject-confederate pairing types.

Framing methods

There was nothing particularly cooperative or competitive about the task presented in this study. It was best described by the control condition as an interaction. This was regarded as necessary, however, in order to use the same task across conditions. And, as noted above, the framing technique used was among the least overt methods available. Again, this was done to test how sensitive ToM would be to cues of group membership—would simply telling subjects they were engaging in a cooperative task or competitive task be sufficient to create a differential ToM response? However, the pattern of results seen here may not be generalizable to other situations or scenarios, but only to this particular method. Therefore, other cooperative and competitive framing techniques

should be considered.

Rather than only indirectly hint at their relationship by referring to a cooperative task or a competitive task in the instructions, the next step up would be to make the relationship explicit by referring directly to the roles subject and confederate take relative to each other, partners or opponents, as Burnham, McCabe, and Smith (2000) did in their study. An even more direct approach would be to use a framing method based on Tajfel et al.'s (1971) minimal group paradigm. The subject and confederate both answer a short (fake) questionnaire on an unrelated topic (such as evaluating pieces of art) and are informed that their responses place them into one of two categories (e.g., a preference for shape vs. color). In the cooperative condition, both would be identified as belonging to the same category, while in the competitive condition, they would be placed into different groups. This does differ from the original in that the subject is actually meeting another participant face to face, whereas in the original subjects were simply made aware of their group status. This also represents another step up in potential strength of framing effect; in this method, rather than playing a cooperative or competitive game, or being arbitrarily labeled as partners or opponents, subjects are completing a task that "earns" them a place in a particular group. However, a recent comparison of minimal group induction methods suggests that an additional step would be warranted: having subjects memorize the names of members of one group lead to stronger implicit preferences for and identification with that group compared to other methods (Pinter and Greenwald, 2010).

One last method to consider came out of conversations with my confederates during the course of this study and may have the potential to evoke strong feelings of group membership similar to the minimal group paradigm. But rather than have subjects

and confederates be placed into groups irrelevant to the task at hand, the framing could actually make use of the roles identified in the instructions, directors and followers. In the cooperative condition, the subject-confederate pair is told their performance is being compared to other such dyads. The pair in the room is the in-group, and they are competing together against the other pairs who compose the out-group. And in the competitive condition, the pair is told that the performance of the followers will be compared to that of the directors.

Subject-Confederate Pairings

It is true that any given subject might come to the task with a racial or sexist bias towards the confederate. In addition, there may be a significant systematic bias to overcome in that all subjects were Rutgers undergraduates and so share a salient group membership. At best, subjects may have held a neutral opinion of the confederate—just another student here to participate in the same task. But regardless of which frame is used, with all else being held constant, there was no systematic, preexisting "cultural baggage" between subject and confederate pairs intentionally built into this study. What is needed in future studies is the ability to test the flexibility and sensitivity of ToM to immediate cues using any (or all) of the framing techniques described above, but with preexisting groups that have the cultural baggage that the subjects in this study lacked. Several additional hypotheses could be tested, and if ToM demonstrates a response to immediate coalitional cues regardless of long standing group dynamics, it would be a much more compelling result and it would provide a much more compelling argument for considering ToM as a type of empathy.

The mutual hostility that exists between Native and Non-Native people in the northern Great Plains suggests these groups may provide a testing ground for these additional hypotheses (see below). Many Native American people continue to live in extreme poverty on reservations surrounded by rural non-Native communities, and much of the animosity they feel is based in a history of genocidal warfare and assimilationist policies that served to outlaw their traditional religious practices and mandated attendance at abusive missionary boarding schools (Brave Heart, 1998). Their Non-Native neighbors have stereotypically viewed Native Americans as untrustworthy, dangerous, and holding on to grievances that are no longer relevant.

One of my graduate student colleagues at Rutgers, Michelle Night Pipe, is studying one of these Native American groups, the Lakota, and the effect that the Annual Dakota 38 Memorial Ride may have on reducing historical trauma among the Lakota and fostering coalitional realignment and the reduction of tensions between Non-Native and Lakota. The Ride is a memorial to the execution of 38 Dakota that took place in 1862 in Mankato, Minnesota. Afterwards, the Dakota were brought to the Lower Brule Reservation in South Dakota, where they still remain (Chomsky, 1990).

While our shared goal is to shed light on the flexible nature of coalitional psychology through the relationships that exist between Native and Non-Native cultures, Night Pipe is taking a more cultural anthropological/experiential approach, looking at the effect the Ride has on the communities at large. The group dynamics between the Non-Native population and the Lakota could also be an ideal testing ground for testing the flexibility of ToM in response to situational coalitional cues.

A future study could include groups of both non-Native and Native subjects. In

order to help compare results with the present study (which would serve as a Non-Native/Non-Native pairing condition), subjects would be recruited from pools of college student volunteers at a large university campus such as the University of South Dakota in Vermillion, SD to form two pairing types, Native/Native and Non-Native/Native.

This research, like the present study, would seek to understand the role of target individuals' group affiliation in our ability to make accurate mental state attributions about them. Pairing Non-Native and Native individuals on the same task will not only provide a strong, long-standing out-group condition, but also allow the testing of how susceptible such an established group membership will be to more immediate, short term cues of shared group membership on a ToM task. It would also allow the testing of additional hypotheses:

1. *Hypothesis 1:* In the control condition, accuracy of mental state attribution will be greater when a subject is paired with a target that is a cultural in-group member (Non-Native/Non-Native and Native/Native pairings) relative to a cultural out-group member (Native/Non-Native pairing).
2. *Hypothesis 2*: Relative to the control condition, subjects in Non-Native/Non-Native and Native/Native pairing combinations will show *greater* accuracy in mental state attribution when presented with an additional situational in-group prime for the task (e.g., partners).
3. *Hypothesis 3:* Relative to the control condition, subjects in Non-Native/Non-Native and Native/Native pairing combinations will show *reduced* accuracy in mental state attribution (similar to the Non-Native/Native control condition) when presented with an additional situational out-group prime for the task (e.g.,

opponents).

4. *Hypothesis 4:* Given previous work that suggests immediate coalitional cues can override long-standing racial prejudice (Kurzban, Cosmides, and Tooby, 2001), in the Non-Native/Native pairing, exposure to situational coalitional cues will result in subjects' accuracy in making mental state attributions exceeding the control condition for this pairing, approaching levels in the control condition for Non-Native/Non-Native and Native/Native pairings.
5. *Hypothesis 5:* Given that accuracy of mental state attribution is already predicted to be reduced in the Non-Native/Native control condition, I predict there will be no change when this pairing is presented with an additional situational out-group prime.

Incorporating such groups and pairing opposing group members on a task such as this also has the added benefit of moving beyond the stereotypical "WEIRD" sample (Henrich et al., 2010) sample that the present study draws upon. Despite the inclusion of both female and male subjects from a diverse range of ethnic and racial backgrounds (16 self-reported racial groups in all, which were collapsed into 5 broad categories for coding purposes), this sample was drawn from a major US university. At the same time, it can be argued that the perspective taking test used here, coding errors the way it does, addresses implicit ToM/perspective taking in addition to explicit ToM. As such, it is measuring an automatic process which is likely free of conscious and cultural/linguistic influences seen in purely explicit tasks discussed in Chapters 2 and 3. Still top-down processes can have some influence in many areas, such as perception of emotion (Gendron, Lindquist,

Barsalou, and Barrett, 2012), color (Winawer, Witthoft, Frank, Wu, Wade, and Boroditsky, 2007), musical pitch (Dolscheid, Shayan, Majid, and Casasanto, 2013), as well as ToM (Matsui, et al., 2009). To overcome this issue, first, refinement of the present methodological issues as discussed above is necessary. Once a suitable framing technique and implicit response measure are found, expanding the study cross-culturally will result in more broadly applicable results.

Concluding Remarks

As discussed in Chapter 1, the ability to make mental state attributions about others' beliefs is a key factor in our ability to form complex and extensive cooperative social relationships, religion, and large scale societies. Yet it is also important on a much smaller scale, for coordinating one-on-one interactions. Without ToM, we might lack the capacity to believe in gods. Without gods (particularly high gods, we might not be able to form large corporations, cities, or states. Without ToM, we would not be able to pass along the huge amount of complex cultural information that we do and would be greatly limited in the number of tools at our disposal and in turn, the number of environments we could survive and thrive in. We might even lack the ability to cooperate effectively at the level of dyads, for without an understanding of what a potential cooperative partner needs, wants, thinks, or believes to guide our own actions and responses, coordinating actions meaningfully becomes a challenge. As such, it is an important topic for the field of anthropology, for these topics represent large areas of study in this field. Without groups, cooperation, religion, or culture, anthropology would be a much smaller field of inquiry.

ToM functions as a type of empathy (Blair, 2005; Preston and de Waal, 2002) allowing us to view the world from others' perspectives. Both motor and emotional empathy are affected by many factors, including race differences between observer and target (Xu et al., 2009), observers' pre-existing attitudes towards the target (Avenanti et al., 2010), and target individual's past behavior (Singer et al., 2006). At the same time, simple cues of coalition membership can override some of these other factors affecting ability to empathize (e.g., Kurzban, Cosmides, and Tooby, 2001). And so it is important to understand how similar factors come into play during ToM. ToM can and does vary *across* individuals (e.g., Baron-Cohen, et al., 2001; see Chapter 3 for additional references).

The goal of the present study was to add to this literature by testing the extent to which ToM varies *within* individuals. We know that task demands affect both children and adults (e.g., Clements and Perner, 1994; Birch and Bloom, 2007), as do other factors such as time between observation and response (Cohen and German, 2009), and familiarity with the target individual (Rabin and Rosenbaum, 2012). But given our flexible coalitional psychology, ToM should, like other forms of empathy, be sensitive to situational cues regardless of other factors. Unfortunately, the data in the present study did not fully support this prediction. However, it does suggest that such an interaction between ToM and coalitions exists, and that further work as outlined above may reveal this interaction more fully.

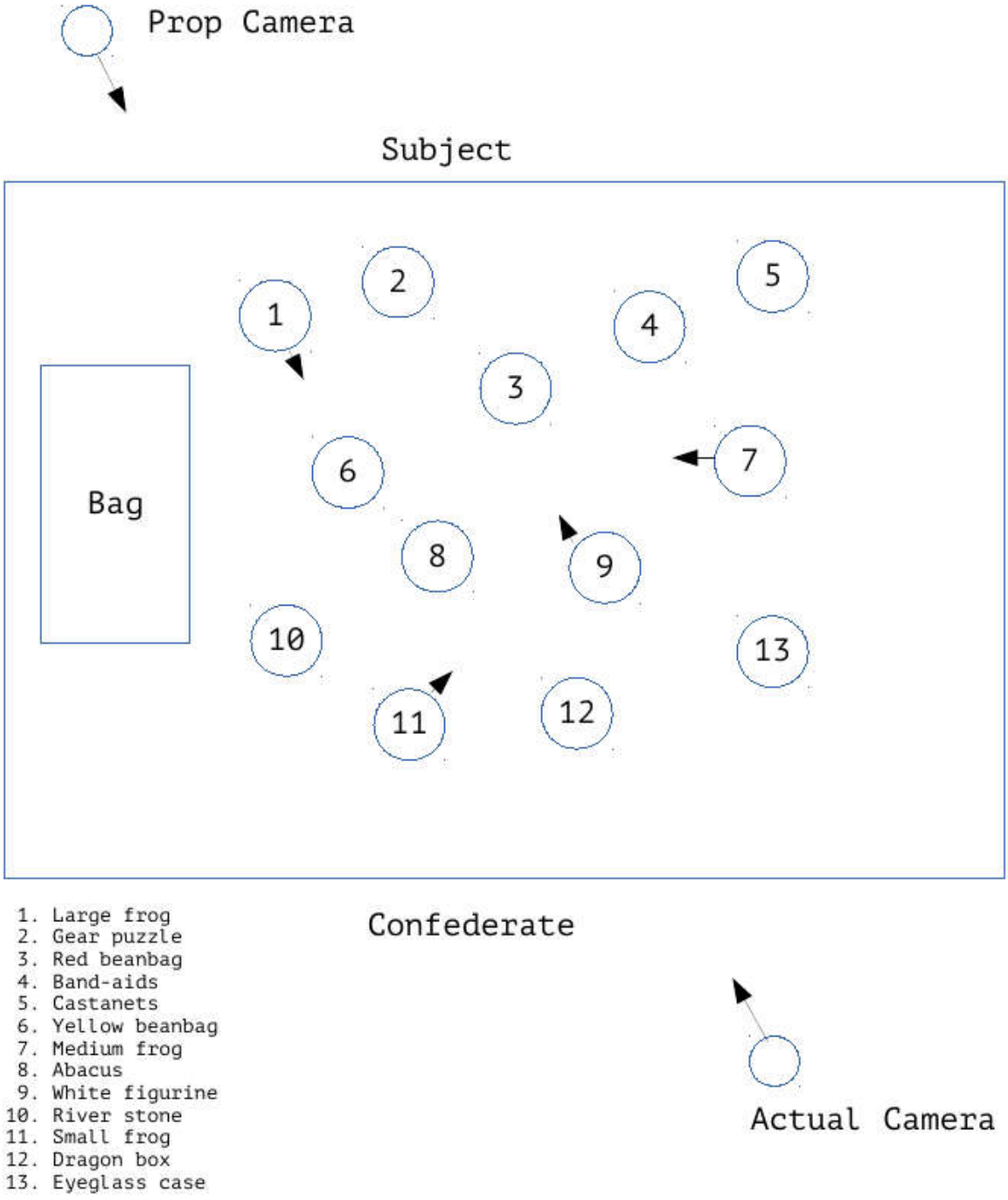
From an anthropological perspective, understanding ToM will help illuminate our understanding of religion, cooperation, and culture. Knowing how ToM may be influenced by the cues of coalition and familiarity that affect emotional empathy is an

important part of gaining that understanding. In the present study, interactions between two people were sensitive to framing those interactions in terms of cooperation or competition. Rather than view ToM as an "either/or" type of skill that an individual either has or does not, or as a skill that can (and does) vary across individuals, this study shows that subtle, small changes in the framing of an interaction can affect the extent to which one person is able to take the cognitive perspective of another. Given the role of ToM in religion, cooperation, and culture, small changes at the individual level could propagate upwards and have far reaching effects.

This project and future studies building on it have two main implications for broader social issues. First, it will lead to a better understanding of how ToM, the ability to make accurate inferences about others' mental states, responds to characteristics of the target individual and situational factors. While there is a large body of work exploring how emotional empathy varies in response to attributes of a target individual, ToM research has instead focused on other areas, including developmental emergence in humans, the deficits exhibited by autistic individuals, and the extent to which it is a skill exhibited in non-human primates. If accurate inference of others' mental states *is* dependent on whether we perceive them, for example, as friend or foe, it is important to learn the full extent—and how easily—those coalitional cues can affect ToM. This is because, second, this work could also point to ideas for increasing accurate ToM among individuals and groups who might otherwise be antagonistic towards each other. Group boundaries are malleable, and coalitions form and dissolve as the need arises. Like "us vs. them" categorizations, our underlying evolved coalitional psychology—tuned to detect and act upon cues of group membership--is flexible also. Engaging in simple cooperative

tasks or being presented with subtle primes and cues of shared group membership may not only affect how we respond emotionally to others, but also how easily we may come to understand their inner cognitive worlds, whether they are friends or members of opposing group in conflict. This could lead to new approaches for more effective peace-making or reconciliation techniques for disputes between neighbors to long-standing animosities between cultural, political or racial groups.

Appendix: Pilot Study Layout



1. Large frog
2. Gear puzzle
3. Red beanbag
4. Band-aids
5. Castanets
6. Yellow beanbag
7. Medium frog
8. Abacus
9. White figurine
10. River stone
11. Small frog
12. Dragon box
13. Eyeglass case

Arrows indicate direction of camera or facing of objects.

References

Abell, F., Happé, F., and Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16.

Adams, R. B., Rule, N. O., Franklin, R. G. Wang, E., Stevenson, M. T., Yoshikawa, S., Nomura, M., Sato, W., Kveraga, K., and Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience, 22(1),* 97-108.

Adolphs, R., Tranel, D., Damasio, H., and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669–72. doi:10.1038/372669a0

Ahearn, L. M. (2011). *Living Language: An Introduction to Linguistic Anthropology*. Malden, MA: John Wiley and Sons. Retrieved from http://books.google.com/books?id=ma4u7Airp5UCandpgis=1

Alvard, M. S. (2003). Kinship, lineage, and an evolutionary perspective on cooperative hunting groups in Indonesia. *Human Nature*, *14*(2), 129-163.

Alvard, M. S. (2003). The adaptive nature of culture. *Evolutionary Anthropology*, *12*(3), 136–149. doi:10.1002/evan.10109

Apperly, I. A., Back, E., Samson, D., and France, L. (2008). The cost of thinking about false beliefs: evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *106*(3), 1093–108.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., and Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841–4. doi:10.1111/j.1467-9280.2006.01791.x

Apperly, I. A., Samson, D., and Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental psychology*, *45*(1), 190–201.

Ashwin, C., Baron-Cohen, S., Wheelwright, S., O'Riordan, M., and Bullmore, E. T. (2007). Differential activation of the amygdala and the "social brain" during fearful face-processing in Asperger Syndrome. *Neuropsychologia*, *45*(1), 2–14. doi:10.1016/j.neuropsychologia.2006.04.014

Astington, J. W., and Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental psychology*, *35*(5), 1311.

Atance, C. M., Bernstein, D. M., and Meltzoff, A. N. (2010). Thinking about false belief: It's not just what children say, but how long it takes them to say it. *Cognition*, *116*(2), 297–301.

Atran, S., and Norenzayan, A. (2004). Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavioral and brain sciences*, *27*(06), 713-730.

Avenanti, A., Sirigu, A., and Aglioti, S. M. (2010). Racial bias reduces empathic sensorimotor resonance with other-race pain. *Current Biology*, *20*(11), 1018–22.

Axt, J. R., Ebersole, C. R., and Nosek, B. A. (2014). The rules of implicit evaluation by race, religion, and age. *Psychological Science*, *25*(9), 1804–1815. doi:10.1177/0956797614543801

Baillargeon, R, Scott, RM, and He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*, 110–8.

Baillon, A., Selim, A., and Van Dolder, D. (2013). On the social nature of eyes: The effect of social cues in interaction and individual choice tasks. *Evolution and Human Behavior*, *34*(2), 146-154.

Balslev, D., Nielsen, F. A., Paulson, O. B., and Law, I. (2005). Right temporoparietal cortex activation during visuo-proprioceptive conflict. *Cerebral Cortex, 15*(2), 166–9. doi:10.1093/cercor/bhh119

Baptista, L. F., and Petrinovich, L. (1984). Social interaction, sensitive phases and the song template hypothesis in the white-crowned sparrow. *Animal Behaviour*, *32*(1), 172-181.

Baptista, L. F., and Petrinovich, L. (1984). Social interaction, sensitive phases and the song template hypothesis in the white-crowned sparrow. *Animal Behaviour*, *32*(1), 172-181.

Barclay, P. (2008). Enhanced recognition of defectors depends on their rarity. *Cognition*, *107*, 817–828.

Barkow, J. H. (1989). *Darwin, Sex, and Status: Biological Approaches to Mind and Culture*. Toronto, ON: University of Toronto Press. Retrieved from http://books.google.com/books?id=0u-AAAAAMAAJandpgis=1

Baron-Cohen, S, Leslie, AM, and Frith, U (1985). Does the autistic child have a “theory of mind”? *Cognition*, *21*, 37–46.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., and Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry*, *38*(7), 813–22. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9363580

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and*

Developmental Disorders, *31*(1), 5–17. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11439754

Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., et al. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B: Biological Sciences*, *280*.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in cognitive sciences*, *4*(1), 29-34.

Barrett, J. L., and Lanman, J. A. (2008). The science of religious beliefs. *Religion*, *38*(2), 109-124.

Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology letters*, *2*(3), 412-414.

Baumgartner, T., Schiller, B., Rieskamp, J., Gianotti, L. R. R., and Knoch, D. (2014). Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction. *Social Cognitive and Affective Neuroscience*, *9*(5), 653–660. doi:10.1093/scan/nst023

Behne, T., Carpenter, M., Call, J., and Tomasello, M. (2005). Unwilling versus unable: infants' understanding of intentional action. *Developmental Psychology*, *41*(2), 328–37. doi:10.1037/0012-1649.41.2.328

Bell, R., and Buchner, A. (2012). How Adaptive Is Memory for Cheaters? *Current Directions in Psychological Science*, *21*(6), 403–408. doi:10.1177/0963721412458525

Bell, R., Buchner, A., and Musch, J. (2010). Enhanced old–new recognition and source memory for faces of cooperators and defectors in a social-dilemma game. *Cognition*, *117*(3), 261-275.

Bell, R., Giang, T., and Buchner, A. (2012). Partial and specific source memory for faces associated to other-and self-relevant negative contexts. *Cognition and emotion*, *26*(6), 1036-1055.

Beller, S., and Bender, A. (2008). The limits of counting: Numerical cognition between evolution and culture. *Science*, *319*(5860), 213-215.

Belot, M., Crawford, V. P., and Heyes, C. (2013). Players of Matching Pennies automatically imitate opponents' gestures against strong incentives. *Proceedings of the National Academy of Sciences*, *110*(8), 2763–2768.

Beran, M. J., and Beran, M. M. (2004). Chimpanzees remember the results of one-by-one addition of food items to sets over extended time periods. *Psychological Science*, *15*(2), 94. Retrieved from http://pss.sagepub.com/content/15/2/94.short

Birch, S. A. J., and Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*(5), 382–6. doi:10.1111/j.1467-9280.2007.01909.x

Birch, S. A., and Bloom, P. (2003). Children Are Cursed An Asymmetric Bias in Mental-State Attribution. *Psychological Science*, *14*(3), 283-286.

Birch, S. A., and Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in cognitive sciences*, *8*(6), 255-260.

Blair, R. J. R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition*, *14*(4), 698–718.

Bloom, P. (2007). Religion is natural. *Developmental science*, *10*(1), 147-151.

Bloom, P., and German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, *77*(1), B25–31.

Bonnie, K. E., Horner, V., Whiten, A., and de Waal, F. B. (2007). Spread of arbitrary conventions among chimpanzees: a controlled experiment. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*(1608), 367-372.

Bowerman, M., and Choi, S. (2003). Space under construction: Language-specific categorization in first language acquisition. In D. Gentner and S. Goldin-Meadows (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 387–427). Boston: MIT Press.

Bowles, S., and Gintis, H. (2003). Origins of human cooperation. *Genetic and cultural evolution of cooperation*, *2003*, 429-43.

Boyd, R., and Richerson, P. J. (1985). *Culture and the Evolutionary Process.* Chicago: University of Chicago Press.

Brannon, T. N., and Walton, G. M. (2013). Enacting cultural interests: How intergroup contact reduces prejudice by sparking interest in an out-group's culture. *Psychological Science*, *24*(10), 1947–1957.

Brave Heart, M. Y. H. (1998). The return to the sacred path: Healing the historical trauma and historical unresolved grief response among the Lakota through a psychoeducational group intervention. *Smith College Studies in Social Work*, *68*(3), 287–305.

Breugelmans, S. M., and Poortinga, Y. H. (2006). Emotion without a word: shame and guilt among Rarámuri Indians and rural Javanese. *Journal of Personality and Social Psychology*, *91*(6), 1111–22. doi:10.1037/0022-3514.91.6.1111

Brewer, M. B., and Silver, M. (1978). Ingroup bias as a function of task characteristics. *European Journal of Social Psychology*.

Bruneau, E. G., Dufour, N., and Saxe, R. (2012). Social cognition in members of conflict groups: behavioural and neural responses in Arabs, Israelis and South Americans to each other's misfortunes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1589), 717–730.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V.,Seitz, R. J., Ziles, K., Rizzolatti, G., and Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, *13*, 400–404. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1460-9568.2001.01385.x/full

Burnham, T., McCabe, K., and Smith, V. L. (2000). Friend-or-foe intentionality priming in an extensive form trust game. *Journal of Economic Behavior and Organization*, *43*(1), 57–73.

Butterworth, B., Reeve, R., Reynolds, F., and Lloyd, D. (2008). Numerical thought with and without words: Evidence from indigenous Australian children. *Proceedings of the National Academy of Sciences*, *105*(35), 13179–84. doi:10.1073/pnas.0806045105

Buxbaum, L. J., Kyle, K. M., and Menon, R. (2005). On beyond mirror neurons: internal representations subserving imitation and recognition of skilled object-related actions in humans. *Cognitive Brain Research*, *25*(1), 226–39.

Caldwell, C. A., and Millen, A. E. (2009). Social learning mechanisms and cumulative cultural evolution: Is imitation necessary? *Psychological Science*, *20*(12), 1478–83. doi:10.1111/j.1467-9280.2009.02469.x

Call, J., and Tomasello, M. (1999). A nonverbal false belief task: the performance of children and great apes. *Child Development*, *70*(2), 381–95. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10218261

Call, J., and Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–92. doi:10.1016/j.tics.2008.02.010

Call, J., Hare, B., Carpenter, M., and Tomasello, M. (2004). "Unwilling" versus "unable": chimpanzees' understanding of human intentional action. *Developmental Science*, *7*(4), 488–98. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15484596

Callaghan, T., Rochat, P., Lillard, A., Claux, M. L., Odden, H., Itakura, S., Tapanya, S., and Singh, S. (2005). Synchrony in the onset of mental-state reasoning: evidence from five cultures. *Psychological Science*, *16*(5), 378–84.

Cantlon, J. F., and Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*(5), 401–406. Retrieved from http://pss.sagepub.com/content/17/5/401.full

Cantlon, J. F., and Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biology*, *5*(12), e328. doi:10.1371/journal.pbio.0050328

Capaldi, E. J., and Miller, D. J. (1988). Counting in rats: Its functional significance and the independent cognitive processes that constitute it. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*(1), 3–17. doi:10.1037//0097-7403.14.1.3

Castelli, F., Frith, C., Happé, F., and Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, *125*(Pt 8), 1839–49.

Cavallini, E., Lecce, S., Bottiroli, S., Palladino, P., and Pagnin, A. (2013). Beyond false belief: Theory of Mind in young, young-old, and old-old adults. *International Journal of Aging and Human Development*, *76*(3), 181–98.

Chaudhuri, A., Schotter, A., and Sopher, B. (2009). Talking ourselves to efficiency: Coordination in inter-generational minimum effort games with private, Almost Common and Common Knowledge of Advice. *Economic Journal*, *119*(534), 91–122. doi:10.1111/j.1468-0297.2008.02207.x

Cho, J. C., and Knowles, E. D. (2013). I'm like you and you're like me: Social projection and self-stereotyping both help explain self-other correspondence. *Journal of Personality and Social Psychology*, *104*(3), 444–56.

Choi, Y., and Luo, Y. (2015). 13-Month-Olds ' Understanding of Social Interactions. *Psychological Science*, *26*(3), 274–283. doi:10.1177/0956797614562452

Chomsky, C. (1990). The United States-Dakota war trials: A study in military injustice. *Stanford Law Review*, 13-98.

Chwe, M. S.-Y. (1998). Culture, circles, and commercials: Publicity, common knowledge, and social coordination. *Rationality and Society*, *10*(1), 47–75.

Cikara, M., and Fiske, S. T. (2011). Bounded empathy: Neural responses to outgroup targets' (mis)fortunes. *Journal of Cognitive Neuroscience*, *23*(12), 3791–803.

Clements, W. A., and Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*, 377–395. Retrieved from http://www.sciencedirect.com/science/article/pii/0885201494900124

Cohen, A. S., and German, T. C. (2009). Encoding of others' beliefs without overt instruction. *Cognition*, *111*(3), 356–63. doi:10.1016/j.cognition.2009.03.004

Cohen, A. S., and German, T. C. (2010). A reaction time advantage for calculating beliefs over public representations signals domain specificity for "theory of mind." *Cognition*, *115*(3), 417–25. doi:10.1016/j.cognition.2010.03.001

Cook, G. I., Marsh, R. L., and Hicks, J. L. (2003). Halo and devil effects demonstrate valenced-based influences on source-monitoring decisions. *Consciousness and cognition, 12*(2), 257-278.

Correll, J., Guillermo, S., and Vogt, J. (2014). On the flexibility of attention to race. *Journal of Experimental Social Psychology*, *55*, 74-79.

Cortes, B. P., Demoulin, S., Rodriguez, R. T., Rodriguez, A. P., and Leyens, J. P. (2005). Infrahumanization or familiarity? Attribution of uniquely human emotions to the self, the ingroup, and the outgroup. *Personality and Social Psychology Bulletin, 31*(2), 243-253.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*(3), 187-276.

Cosmides, L., Tooby, J., and Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences, 7*(4), 173–179.

Cronk, L. (1995). Is there a role for culture in human behavioral ecology? *Ethology and Sociobiology, 16*(3), 181–205. doi:10.1016/0162-3095(95)00001-2

Cronk, L. (1999). *That Complex Whole: Culture and the Evolution of Human Behavior*. Boulder, CO: Westview Press. Retrieved from http://books.google.com/books?id=ftrL4eFFe5oCandpgis=1

Cronk, L. (2007). The influence of cultural framing on play in the trust game: A Maasai example. *EHB*, *28*, 352–358.

Cronk, L., and Leech, B. L. (2012). *Meeting at Grand Central: understanding the social and evolutionary roots of cooperation*. Princeton University Press.

Cronk, L., and Wasielewski, H. (2008). An unfamiliar social norm rapidly produces framing effects in an economic game. *Journal of Evolutionary Psychology*, *6*(4), 283–308.

Curry, O., and Chesters, M. J. (2012). “Putting Ourselves in the Other Fellow's Shoes": The Role of “Theory of Mind” in Solving Coordination Problems. *Journal of Cognition and Culture*, *12*(1), 147–159.

De Bruin, L. C., and Newen, A. (2012). An association account of false belief understanding. *Cognition, 123*(2), 240–59. doi:10.1016/j.cognition.2011.12.016

De Villiers, J. G. (2000). Language and theory of mind: What are the developmental relationships? In S. Baron-Cohen, H. Tager-Flusberg, and D. Cohen (Eds.), *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience* (2nd ed., pp. 83–123). Oxford: Oxford University Press.

De Villiers, J. G., and Pyers, J. E. (2002). Complements to cognition: a longitudinal study

of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, *17*(1), 1037–1060. doi:10.1016/S0885-2014(02)00073-4

De Waal, F. B. M., and Ferrari, P. F. (2010). Towards a bottom-up perspective on animal and human cognition. *Trends in Cognitive Sciences*, *14*(5), 201–207. doi:10.1016/j.tics.2010.03.003

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, *1*(04), 568–570.

Dennett, D. C. (1989). *The Intentional Stance*. Cambridge, MA: MIT Press. Retrieved from http://books.google.com/books?id=Qbvkja-J9iQCandpgis=1

Dennett, D., 1996. *Kinds of Minds: Toward an Understanding of Consciousness*. Basic Books: New York, NY.

Dindo, M., Thierry, B., de Waal, F. B. M., and Whiten, A. (2010). Conditional copying fidelity in capuchin monkeys (Cebus apella). *Journal of Comparative Psychology*, *124*(1), 29–37. doi:10.1037/a0018005

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception*, *22*, 15-15.

Dittrich, W. H., Troscianko, T., Lea, S. E., and Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727-738.

Dolscheid, S., Shayan, S., Majid, A., and Casasanto, D. (2013). The Thickness of Musical Pitch Psychophysical Evidence for Linguistic Relativity. *Psychological Science*, *24*(5), 613-621.

Duan, C. (2000). Being empathic: The role of motivation to empathize and the nature of target emotions. *Motivation and Emotion*, *24*(1), 29–49.

Efferson, C, Lalive, R, and Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science*, *321*, 1844–9.

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotions. In J. Cole (Ed.), *Nebraska Symposium on Motivation* (pp. 207–283). Lincoln, NE: University of Nebraska Press.

Elliott, R., Völlm, B., Drury, A., McKie, S., Richardson, P., and Deakin, J. F. W. (2006). Co-operation with another player in a financially rewarded guessing game activates regions implicated in theory of mind. *Social Neuroscience*, *1*(3-4), 385–95. doi:10.1080/17470910601041358

Fadiga, L., and Craighero, L. (2006). Hand actions and speech representation in Broca's area. *Cortex*, *42*, 486–490.

Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611.

Falk, E. B., Spunt, R. P., Lieberman, M. D., and Robert, P. (2012). Ascribing beliefs to ingroup and outgroup political candidates: neural correlates of perspective-taking , issue importance and days until the election. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 731–743.

Farmer, H., Maister, L., and Tsakiris, M. (2014). Change my body, change my mind: the effects of illusory ownership of an outgroup hand on implicit attitudes toward that outgroup. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.01016

Fehr, E., and Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785-791.

Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*(7), 307-314.

Ferguson, H. J., Apperly, I. A., Ahmad, J., Bindemann, M., and Cane, J. (2015). Task constraints distinguish perspective inferences from perspective use during discourse interpretation in a false belief task. *Cognition*, *139*, 50–70. doi:10.1016/j.cognition.2015.02.010

Ferrari, P. F., Gallese, V., Rizzolatti, G., and Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, *17*(8), 1703-1714.

Focquaert, F., Steven, M. S., Wolford, G. L., Colden, A., and Gazzaniga, M. S. (2007). Empathizing and systemizing cognitive traits in the sciences and humanities. *Personality and Individual Differences*, *43*(3), 619–625.

Frank, M. C., Everett, D. L., Fedorenko, E., and Gibson, E. (2008). Number as a cognitive technology: evidence from Pirahã language and cognition. *Cognition*, *108*(3), 819–24. doi:10.1016/j.cognition.2008.04.007

Frith, C. D., and Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–4.

Galef, B. G., and Laland, K. N. (2005). Social learning in animals: empirical studies and theoretical models. *Bioscience*, *55*(6), 489-499.

Gallese, V, Fadiga, L., Fogassi, L, and Rizzolatti, G (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.

Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the cognitive sciences*, *4*(1), 23-48.

Gallese, V. (2007). Before and below “theory of mind”: embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 659–69. doi:10.1098/rstb.2006.2002

Gallese, V., and Goldman, A. I. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501.

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*(9), 396–403. doi:10.1016/j.tics.2004.07.002

Gangitano, M., Mottaghy, F. M., and Pascual-Leone, A. (2001). Phase-specific modulation of cortical motor output during movement observation. *NeuroReport*, *12*(7), 1489–92. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11388435

Gendron, M., Lindquist, K. A., Barsalou, L., and Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, *12*(2), 314.

Gergely, G., Nádasdy, Z., Csibra, G., and Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*(2), 165–93.

Gleitman, L. R., and Papafragou, A. (2005). Language and thought. In K. J. Holyoak and R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 633–661). Cambridge: Cambridge University Press.

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

Goldman, A. I., and Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition*, *94*(3), 193–213. doi:10.1016/j.cognition.2004.01.005

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*(1), 1–14. Retrieved from http://journals.cambridge.org/abstract_S0140525X00028636

Gopnik, A. (1996). The scientist as child. *Philosophy of Science*, *63*, 485–514. Retrieved from http://www.jstor.org/stable/188064

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*(5695), 496–499. Retrieved from http://www.sciencemag.org/cgi/content/abstract/1094492

Grèzes, J., Armony, J. L., Rowe, J., and Passingham, R. E. (2003). Activations related to “mirror” and “canonical” neurones in the human brain: an fMRI study. *NeuroImage*, *18*(4), 928–937. doi:10.1016/S1053-8119(03)00042-9

Gutsell, J. N., and Inzlicht, M. (2010). Empathy constrained: Prejudice predicts reduced mental simulation of actions during observation of outgroups. *Journal of Experimental Social Psychology*, *46*(5), 841–845.

Gutsell, J. N., and Inzlicht, M. (2012). Intergroup differences in the sharing of emotive states: neural evidence of an empathy gap. *Social Cognitive and Affective Neuroscience*, *7*(5), 596–603.

Hackel, L. M., Looser, C. E., and Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology*, *52*, 15–23. doi:10.1016/j.jesp.2013.12.001

Hale, C. M., and Tager-Flusberg, H. (2003). The influence of language on theory of mind: a training study. *Developmental Science*, *6*(3), 346–59.

Haley, K. J., and Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. Evolution and Human Behavior, 26, 245–256.

Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature*, *406*(6792), 147–50. doi:10.1038/35018000

Hamilton, A. F. D. C., Brindley, R. M., and Frith, U. (2007). Imitation and action understanding in autistic spectrum disorders: how valid is the hypothesis of a deficit in the mirror neuron system? *Neuropsychologia*, *45*(8), 1859–68.

Hamlin, J. K., Mahajan, N., Liberman, Z., and Wynn, K. (2013). Not like me = bad: Infants prefer those who harm dissimilar others. *Psychological Science*, *24*(4), 589–594. doi:10.1177/0956797612457785

Hamlin, J. K., Wynn, K., and Bloom, P. (2008). Social evaluation by preverbal infants. *Pediatric Research*, *63*(3), 219. doi:10.1203/PDR.0b013e318168c6e5

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of autism and Developmental disorders*, *24*(2), 129-154.

Hare, B., and Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour*, *68*(3), 571–581. doi:10.1016/j.anbehav.2003.11.011

Hare, B., Call, J., and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, *61*(1), 139–151. doi:10.1006/anbe.2000.1518

Hare, B., Call, J., Agnetta, B., and Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771–785. doi:10.1006/anbe.1999.1377

Harris, L. T., and Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science*, *17*(10), 847–853.

Heider, F., and Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 243-259.

Henrich, J. (2004). Cultural group selection, coevolutionary processes, and large scale cooperation. *Journal of Economic Behavior and Organization, 53,* 3-35.

Henrich, J. (2014). *The Secret of Our Success.*

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–135. doi:10.1017/S0140525X0999152X

Hermans, E. J., Putman, P., and van Honk, J. (2006). Testosterone administration reduces empathetic behavior: a facial mimicry study. *Psychoneuroendocrinology*, *31*(7), 859–66.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., and Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *science*, *317*(5843), 1360-1366.

Heyes, C. M., and Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, *344*(6190), 1243091–1243091. doi:10.1126/science.1243091

Hill, J. H., and Mannheim, B. (2012). Language and world view. *Annual Review of Anthropology*, *21*(1992), 381–406.

Hogrefe, G.-J., Wimmer, H., and Perner, J. (1986). Ignorance versus False Belief: A Developmental Lag in Attribution of Epistemic States. *Child Development*, *57*(3), 567.

Holmes, H. A., Black, C., and Miller, S. A. (1996). A Cross-Task Comparison of False Belief Understanding in a Head Start Population. *Journal of Experimental Child Psychology*, *63*(2), 263–85. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8954606

Horner, V., and de Waal, F. B. (2009). Controlled studies of chimpanzee cultural transmission. *Progress in brain research*, *178*, 3-15.

Horner, V., and Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal Cognition*, *8*(3), 164–81. doi:10.1007/s10071-004-0239-6

Iacoboni, M. (2005). Neural mechanisms of imitation. *Current Opinion in Neurobiology*, *15*(6), 632–7.

Ibanez, A., Huepe, D., Gemp, R., Gutierrez, V., Rivera-Rei, A., and Toledo, M. I. (2013). Empathy, sex, and fluid intelligence as predictors of theory of mind. *Personality and Individual Differences, 54(5),* 616-621.

Inzlicht, M., Gutsell, J. N., and Legault, L. (2012). Mimicry reduces racial prejudice. *Journal of Experimental Social Psychology*, *48*(1), 361–365. doi:10.1016/j.jesp.2011.06.007

Izard, C., and Buechler, S. (1980). Aspects of consciousness and personality in terms of differential emotions theory. In R. Pluchik and H. Kellerman (Eds.), *Emotion: Theory, Research, and Experience, Vol. 1: Theories of Emotion* (pp. 165–187). New York: Academic Press.

Jaakkola, K., Fellner, W., Erb, L., Rodriguez, M., and Guarino, E. (2005). Understanding of the concept of numerically "less" by bottlenose dolphins (Tursiops truncatus). *Journal of Comparative Psychology*, *119*(3), 296–303. doi:10.1037/0735-7036.119.3.296

Jackson, P. L., Meltzoff, A. N., and Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage*, *24*(3), 771–9.

Jackson, P. L., Meltzoff, A. N., and Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *NeuroImage*, *31*(1), 429–39. doi:10.1016/j.neuroimage.2005.11.026

Jern, A., and Kemp, C. (2015). A decision network account of reasoning about other people's choices. *Cognition*, *142*, 12–38. doi:10.1016/j.cognition.2015.05.006

Justus, T., and List, A. (2005). Auditory attention to frequency and time: an analogy to visual local-global stimuli. *Cognition*, *98*(1), 31–51. doi:10.1016/j.cognition.2004.11.001

Kaufman, E. L., Lord, M. W., Reese, T. W., and Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology*, *62*(4), 498–525.

Keysar, B., Lin, S., and Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

Kinzler, K. D., and Spelke, E. S. (2011). Do infants show social preferences for people differing in race? *Cognition*, *119*(1), 1–9. doi:10.1016/j.cognition.2010.10.019

Kinzler, K. D., Corriveau, K. H., and Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science*, *14*(1), 106–111. doi:10.1111/j.1467-7687.2010.00965.x

Kinzler, K. D., Shutts, K., Dejesus, J., and Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, *27*(4), 623–634.

Kobayashi, H., and Kohshima, S. (1997). Unique morphology of the human eye. *Nature*, *287*(1992), 767–768.

Kobayashi, H., and Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, *40*(5), 419–35. doi:10.1006/jhev.2001.0468

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297*:846–48.

Kovács, A. M., Teglas, E., and Endress, A. D. (2010). The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults. *Science*, *330*(6012), 1830–1834.

Kozak, M. N., Marsh, A. A., and Wegner, D. M. (2006). What do I think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology*, *90*(4), 543–55. doi:10.1037/0022-3514.90.4.543

Krachun, C., Carpenter, M., Call, J., and Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–35. doi:10.1111/j.1467-7687.2008.00793.x

Krosch, A. R., and Amodio, D. M. (2014). Economic scarcity alters the perception of race. *Proceedings of the National Academy of Sciences, 111*(25), 9079–9084. doi:10.1073/pnas.1404448111

Kurzban, R., Tooby, J., and Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*(26), 15387–92. doi:10.1073/pnas.251541498

Laland, K. N., and Hoppitt, W. (2003). Do animals have culture?. *Evolutionary Anthropology: Issues, News, and Reviews, 12*(3), 150-159.

Laland, K. N., and Plotkin, H. C. (1990). Social learning and social transmission of foraging information in Norway rats (Rattus norvegicus). *Animal Learning and Behavior*, *18*(3), 246-251.

Laland, K. N., and Williams, K. (1997). Shoaling generates social learning of foraging information in guppies. *Animal Behaviour*, *53*(6), 1161-1169.

Lansing, J. S., and Kremer, J. N. (1993). Emergent properties of Balinese water temple networks: coadaptation on a rugged fitness landscape. *American Anthropologist*, 97-114.

Lansing, J. S., and Miller, J. H. (2003). Cooperation in Balinese rice farming. *Santa Fe, NM: Santa Fe Institute*.

Leslie, A. M. (2000). "Theory of Mind" as a mechanism of selective attention. In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2nd ed., pp. 1235–1247). Cambridge, MA: MIT Press.

Leslie, A. M., Friedman, O., and German, T. P. (2004). Core mechanisms in "theory of mind." *Trends in Cognitive Sciences*, *8*(12), 528–33. doi:10.1016/j.tics.2004.10.001

Leslie, A. M., German, T. P., and Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cognitive Psychology*, *50*(1), 45–85. doi:10.1016/j.cogpsych.2004.06.002

Levine, M., Prosser, A., Evans, D., and Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, *31*(4), 443–53.

Levinson, S. C., Kita, S., Haun, D. B. M., and Rasch, B. H. (2002). Returning the tables: language affects spatial reasoning. *Cognition*, *84*(2), 155–88. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12175571

Leyens, J. P., Paladino, P. M., Rodriguez, R. T., Vaes, J., Demoulin, S., Rodriguez, A. P., Gaunt, R. (2000). The emotional side of prejudice: The role of secondary emotions. *Personality and Social Psychology Review*, *4*, 186–197.

Li, P., and Gleitman, L. R. (2002). Turning the tables: language and spatial reasoning. *Cognition*, *83*(3), 265–94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11934404

Liew, S.-L., Han, S., and Aziz-Zadeh, L. (2011). Familiarity modulates mirror neuron and mentalizing regions during intention understanding. *Human Brain Mapping*, *32*(11), 1986–97.

Lillard, A. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin*, *123*(1), 3–32. Retrieved from http://faculty.virginia.edu/early-social-cognition-lab/reprints/reprints/ethnopsych-4c.pdf

Lissek, S., Peters, S., Fuchs, N., Witthaus, H., Nicolas, V., Tegenthoff, M., Jukel, M., and Brüne, M. (2008). Cooperation and deception recruit different subsets of the theory-of-mind network. *PloS One*, *3*(4), e2023. doi:10.1371/journal.pone.0002023

Lucy, J. (1992). Grammatical Categories and Cognition. Glasgow, Scotland: Cambridge University Press.

Lyons, D. E., Santos, L. R., and Keil, F. C. (2006). Reflections of other minds: how primate social cognition can inform the function of mirror neurons. *Current Opinion in Neurobiology*, *16*(2), 230–4. doi:10.1016/j.conb.2006.03.015

Mahajan, N., and Wynn, K. (2012). Origins of "'Us'" versus "'Them'": Prelinguistic infants prefer similar others. *Cognition*. doi:10.1016/j.cognition.2012.05.003

Matsui, T., Rakoczy, H., Miura, Y., and Tomasello, M. (2009). Understanding of speaker certainty and false-belief reasoning: a comparison of Japanese and German

preschoolers. *Developmental Science*, *12*(4), 602–13. doi:10.1111/j.1467-7687.2008.00812.x

McCabe, K. A., Smith, V. L., and LePore, M. (2000). Intentionality detection and "mindreading": Why does game form matter? *Proceedings of the National Academy of Sciences*, *97*(8), 4404–9. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=18254andtool=pmcentrezandrendertype=abstract

McDonald, M. M., Navarrete, C. D., and Van Vugt, M. (2012). Evolution and the psychology of intergroup conflict: the male warrior hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1589), 670-679.

McElreath, R., Boyd, R., and Richerson, P. J. (2003). Shared norms and the evolution of ethnic markers. *Current Anthropology*, *44*(1), 122–130.

McPherson Frantz, C., and Janoff-Bulman, R. (2000). Considering Both Sides: The Limits of Perspective Taking. *Basic and Applied Social Psychology*, *22*(1), 31–42. doi:10.1207/S15324834BASP2201_4

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iaconboni, M., and Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology, 20(8),* 750-756

Navarrete, C. D., McDonald, M. M., Molina, L. E., and Sidanius, J. (2010). Prejudice at the nexus of race and gender: an outgroup male target hypothesis. *Journal of personality and social psychology*, *98*(6), 933.

Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., and Bateson, M. (2013). The watching eyes effect in the Dictator Game: It's not how much you give, it's being seen to give something. *Evolution and Human Behavior*, *34*(1), 35-40.

Newton, A. M., and de Villiers, J. G. (2007). Thinking while talking: adults fail nonverbal false-belief reasoning. *Psychological Science*, *18*(7), 574–9. doi:10.1111/j.1467-9280.2007.01942.x

Nickerson, R. S. (1999). How we know--and sometimes misjudge--what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737–759.

Nielsen, M., and Tomaselli, K. (2010). Overimitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological Science*, *21*(5), 729–36. doi:10.1177/0956797610368808

Norenzayan, A., Gervais, W. M., and Trzesniewski, K. H. (2012). Mentalizing deficits constrain belief in a personal God. *PloS One*, *7*(5), e36880. doi:10.1371/journal.pone.0036880

Nuñez, J. M., Casey, B. J., Egner, T., Hare, T., and Hirsch, J. (2005). Intentional false

responding shares neural substrates with response conflict and cognitive control. *NeuroImage*, *25*(1), 267–77. doi:10.1016/j.neuroimage.2004.10.041

Onishi, K. H., and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–8. doi:10.1126/science.1107621

Onishi, K. H., Baillargeon, R., and Leslie, A. M. (2007). 15-Month-Old Infants Detect Violations in Pretend Scenarios. *Acta Psychologica*, *124*(1), 106–28.

Paal, T., and Bereczkei, T. (2007). Adult theory of mind, cooperation, Machiavellianism: The effect of mindreading on social relations. *Personality and Individual Differences*, *43*(3), 541–551. doi:10.1016/j.paid.2006.12.021

Pan, X. (Sophia), and Houser, D. (2013). Cooperation during cultural group formation promotes trust towards members of out-groups. *Proceedings of the Royal Society B: Biological Sciences*, *280*.

Panksepp, J. (2000). Emotions as natural kinds within the mammalian brain. In M. Lewis and J. Haviland (Eds.), *The Handbook of Emotions* (2nd ed., pp. 139–156). New York: Guilford Press.

Papafragou, A., and Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition*, *86*(3), 253–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12485740

Paukner, A., Anderson, J. R., Borelli, E., Visalberghi, E., and Ferrari, P. F. (2005). Macaques (*Macaca nemestrina*) recognize when they are being imitated. *Biology Letters, 1(2),* 219-222.

Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and Cognition*, *22*(3), 779–787. doi:10.1016/j.concog.2013.04.016

Peers, P. V., Ludwig, C. J. H., Rorden, C., Cusack, R., Bonfiglioli, C., Bundesen, C., Driver, J., Antoun, N, and Duncan, J. (2005). Attentional functions of parietal and frontal cortex. *Cerebral Cortex*, *15*(10), 1469–84. doi:10.1093/cercor/bhi029

Penn, D. C., and Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a “theory of mind.” *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 731–44. doi:10.1098/rstb.2006.2023

Pepperberg, I. M. (1994). Numerical competence in an African gray parrot (Psittacus erithacus). *Journal of Comparative Psychology, 108*(1), 36–44. Retrieved from http://psycnet.apa.org/journals/com/108/1/36/

Perner, J., Leekam, S. R., and Wimmer, H. (1987). Three-year-olds’ difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental*

Psychology, *5*(2), 125–137.

Peterson, C. C., and Siegal, M. (2000). Insights into Theory of Mind from Deafness and Autism. *Mind and Language*, *15*(1), 123–145.

Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503. doi:10.1126/science.1102085

Pietraszewski, D., and Schwartz, A. (2014a). Evidence that accent is a dedicated dimension of social categorization, not a byproduct of coalitional categorization. *Evolution and Human Behavior*, *35*(1), 51–57. doi:10.1016/j.evolhumbehav.2013.09.005

Pietraszewski, D., and Schwartz, A. (2014b). Evidence that accent is a dimension of social categorization, not a byproduct of perceptual salience, familiarity, or ease-of-processing. *Evolution and Human Behavior*, *35*(1), 43–50. doi:10.1016/j.evolhumbehav.2013.09.006

Pietraszewski, D., Cosmides, L., and Tooby, J. (2014). The content of our cooperation, not the color of our skin: An alliance detection system regulates categorization by coalition and race, but not sex. *PLoS ONE*, *9*(2), e88534. doi:10.1371/journal.pone.0088534

Pietraszewski, D., Curry, O. S., Peterson, M. B., Cosmides, L., and Tooby, J. (2015). Constituents of political cognition: Race, party politics, and the alliance detection system. *Cognition*, *140*, 24–39.

Pineda, J. O. A., and Oberman, L. M. (2006). What goads cigarette smokers to smoke? Neural adaptation and the mirror neuron system. *Brain Research*, *1121*(1), 128–35. doi:10.1016/j.brainres.2006.08.128

Pinter, B., and Greenwald, A. G. (2004). Exploring implicit partisanship: Enigmatic (but genuine) group identification and attraction. *Group processes and intergroup relations*, *7*(3), 283-296.

Pinter, B., and Greenwald, A. G. (2010). A comparison of minimal group induction procedures. *Group Processes and Intergroup Relations*, *14*(1), 81–98.

Pinto, J., and Shiffrar, M. (1999). Subconfigurations of the human form in the perception of biological motion displays. *Acta Psychologica*, *102*(2), 293-318.

Ponseti, J., Bosinski, H. A., Wolff, S., Peller, M., Jansen, O., Mehdorn, H. M., Buchel, C., and Siebner, H. R. (2006). A functional endophenotype for sexual orientation in humans. *NeuroImage*, *33*(3), 825–33. doi:10.1016/j.neuroimage.2006.08.002

Povinelli, D. J., and Bering, J. M. (2000). Toward a science of other minds: escaping the argument by analogy. *Cognitive Science*, *24*(3), 509–541. doi:10.1016/S0364-

0213(00)00023-9

Povinelli, D. J., and Bering, J. M. (2002). The Mentality of Apes Revisited. *Current Directions in Psychological Science*, *11*(4), 115–119. doi:10.1111/1467-8721.00181

Povinelli, D. J., and Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends in Cognitive Sciences*, *7*(4), 157–160. doi:10.1016/S1364-6613(03)00053-6

Powell, L. J., and Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences*, *110*(41), E3965–E3972. doi:10.1073/pnas.1304326110

Prather, J. F., Peters, S., Nowicki, S., and Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature*, *451*(7176), 305–10. doi:10.1038/nature06492

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

Preston, S. D., and de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(1), 1–71.

Rabin, J. S., and Rosenbaum, R. S. (2012). Familiarity modulates the functional relationship between theory of mind and autobiographical memory. *NeuroImage*, *62*(1), 520–9.

Ratner, K. G., Kaul, C., and Van Bavel, J. J. (2013). Is race erased? Decoding race from patterns of neural activity when skin color is not diagnostic of group boundaries. *Social Cognitive and Affective Neuroscience*, *8*(7), 750–755. doi:10.1093/scan/nss063

Repacholi, B., Slaughter, V., Pritchard, M., and Gibbs, V. (2003). Theory of mind, Machiavellianism, and social functioning in childhood.

Rizzolatti, G., and Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–194. doi:10.1016/S0166-2236(98)01260-0

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–92. doi:10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–41.

Robertson, L. C. (1996). Attentional persistence for features of hierarchical patterns. *Journal of Experimental Psychology: General*, *125*(3), 227–49. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8751819

Robertson, L. C., Lamb, M. R., and Knight, R. T. (1988). Effects of lesions of temporal-

parietal junction on perceptual and attentional processing in humans. *Journal of Neuroscience*, *8*(10), 3757–3769. Retrieved from http://www.jneurosci.org/cgi/content/abstract/8/10/3757

Russell, P. A., Hosie, J. A., Gray, C. D., Scott, C., Hunter, N., Banks, J. S., and Macaulay, M. C. (1998). The development of theory of mind in deaf children. *Journal of Child Psychology and Psychiatry*, *39*(6), 903–10.

Samson, D., Apperly, I. A., Chiavarino, C., and Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, *7*(5), 499–500. doi:10.1038/nn1223

Santiesteban, I., White, S., Cook, J., Gilbert, S. J., Heyes, C., and Bird, G. (2011). Training social cognition: From imitation to Theory of Mind. *Cognition*, *122*(2), 228–235. doi:10.1016/j.cognition.2011.11.004

Saxe, R. R., and Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842. doi:10.1016/S1053-8119(03)00230-1

Saxe, R. R., and Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, *17*(8), 692–9. doi:10.1111/j.1467-9280.2006.01768.x

Saxe, R. R., and Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, *43*(10), 1391–9. doi:10.1016/j.neuropsychologia.2005.02.013

Saxe, R. R., Carey, S., and Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, *55*, 87–124. doi:10.1146/annurev.psych.55.090902.142044

Schick, B., de Villiers, P., de Villiers, J. G., and Hoffmeister, R. (2007). Language and theory of mind: a study of deaf children. *Child Development*, *78*(2), 376–96.

Schotter, A., and Sopher, B. (2003). Social Learning and Coordination Conventions in Intergenerational Games: An Experimental Study. *Journal of Political Economy*, *111*(3), 498–529. doi:10.1086/374187

Schotter, A., and Sopher, B. (2007). Advice and behavior in intergenerational ultimatum games: An experimental approach. *Games and Economic Behavior*, *58*(2), 365–393. doi:10.1016/j.geb.2006.03.005

Schürmann, M., Hesse, M. D., Stephan, K. E., Saarela, M., Zilles, K., Hari, R., and Fink, G. R. (2005). Yearning to yawn: the neural basis of contagious yawning. *NeuroImage*, *24*(4), 1260–4. doi:10.1016/j.neuroimage.2004.10.022

Scott, R. M., and Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs

about object identity at 18 months. *Child Development*, *80*(4), 1172–96.

Senju, A., Southgate, V., White, S., and Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science*, *325*(5942), 883–5.

Shariff, A., Norenzayan, A., and Henrich, J. (2010). The birth of high gods. *Evolution, culture, and the human mind*, 119-136.

Shatz, M., Diesendruck, G., Martinez-Beck, I., and Akar, D. (2003). The influence of language and socioeconomic status on children's understanding of false belief. *Developmental Psychology*, *39*(4), 717–729.

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., and Sherif, C. W. (1961). Intergroup cooperation and conflict: The robbers cave experiment. *Norman, OK: University of Oklahoma Book Exchange*.

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews*, *30*(6), 855–63. doi:10.1016/j.neubiorev.2006.06.011

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, *439*(7075), 466–9.

Skorinko, J. L., and Sinclair, S. A. (2013). Perspective taking can increase stereotyping: The role of apparent stereotype confirmation. *Journal of Experimental Social Psychology*, *49*(1), 10–18.

Southgate, V., Senju, A., and Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–92.

Spelke, E. S. (2003). What makes us smart? Core knowledge and natural language. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 277–311). Cambridge, MA: MIT Press.

Spelke, E. S., and Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. doi:10.1111/j.1467-7687.2007.00569.x

Stroop, J. R. (1992). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General*, *121*(1), 15–23. Retrieved from http://psycnet.apa.org/journals/xge/121/1/15/

Stürmer, S., Snyder, M., Kropp, A., and Siem, B. (2006). Empathy-motivated helping: The moderating role of group membership. *Personality and Social Psychology Bulletin*, *32*(7), 943–956. doi:10.1177/0146167206287363

Surian, L, Caldi, S., and Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological science*, *18*, 580–6.

Sylwester, K., Lyons, M., Buchanan, C., Nettle, D., and Roberts, G. (2012). The role of Theory of Mind in assessing cooperative intentions. *Personality and Individual Differences, 52*(2), 113–117.

Tajfel, H., Billig, M. G., Bundy, R. P., and Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. doi:10.1002/ejsp.2420010202

Takagishi, H., Kameshima, S., Schug, J., Koizumi, M., and Yamagishi, T. (2010). Theory of mind enhances preference for fairness. *Journal of Experimental Child Psychology*, *105*(1-2), 130–7. doi:10.1016/j.jecp.2009.09.005

Tarrant, M., Dazeley, S., and Cottom, T. (2009). Social categorization and empathy for outgroup members. *British Journal of Social Psychology*, *48*(3), 427–446. doi:10.1348/014466608X373589

Telzer, E. H., Humphreys, K. L., Shapiro, M., and Tottenham, N. (2013). Amygdala Sensitivity to Race Is Not Present in Childhood but Emerges over Adolescence. *Journal of Cognitive Neuroscience*, *25*(2), 234–44.

Teufel, C., Fletcher, P. C., and Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *Trends in Cognitive Sciences*, 1–7. doi:10.1016/j.tics.2010.05.005

Thornton, I. M., and Knoblich, G. (2006). Action perception: Seeing the world through a moving body. *Current Biology*, *16*(1), R27–9.

Tomasello, M. (2011). Human culture in evolutionary perspective. *Advances in culture and psychology, 1*, 5-51.

Tomasello, M., Call, J., and Hare, B. (2003). Chimpanzees understand psychological states – the question is which ones and to what extent. *Trends in Cognitive Sciences*, *7*(4), 153–156. doi:10.1016/S1364-6613(03)00035-4

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–91; discussion 691–735. doi:10.1017/S0140525X05000129

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., and Herrmann, E. (2012). Two Key Steps in the Evolution of Human Cooperation. *Current Anthropology*, *53*(6), 673–692.

Trawalter, S., Hoffman, K. M., and Waytz, A. (2012). Racial bias in perceptions of others' pain. *PLoS One, 7*(11), e48546.

Tremoulet, P. D., and Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, *29*(8), 943-952.

Turner, J. C., Brown, R. J., and Tajfel, H. (1979). Social comparison and group interest in ingroup favouritism. *European Journal of Social Psychology*, *9*(February 1978), 187–204.

Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.

Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition*, *110*(1), 124–9. doi:10.1016/j.cognition.2008.10.008

Uller, C., Jaeger, R., Guidry, G., and Martin, C. (2003). Salamanders (Plethodon cinereus) go for more: rudiments of number in an amphibian. *Animal Cognition*, *6*(2), 105–12. doi:10.1007/s10071-003-0167-x

Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., and Rizzolatti, G. (2001). I know what you are doing. A neurophysiological study. *Neuron*, *31*(1), 155–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11498058

Van Bavel, J. J., and Cunningham, W. a. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin*, *35*(3), 321–335. doi:10.1177/014616720832774

van de Waal, E., Borgeaud, C., and Whiten, A. (2013). Potent social learning and conformity shape a wild primate's foraging decisions. *Science*, *340*(6131), 483-485.

Vinden, P. G. (1996). Junin Quechua Children's Understanding of Mind. *Child Development*, *67*(4), 1707–1716.

Vinden, P. G. (1999). Children's Understanding of Mind and Emotion: A Multi-culture Study. *Cognition and Emotion*, *13*(1), 19–48.

Vinden, P. G. (2001). Parenting attitudes and children's understanding of mind A comparison of Korean American and Anglo-American families. *Cognitive Development*, *16*(3), 793–809. doi:10.1016/S0885-2014(01)00059-4

Vinden, P. G. (2002). Understanding minds and evidence for belief: A study of Mofu children in Cameroon. *International Journal of Behavioral Development*, *26*(5), 445-452.

Volstorf, J., Rieskamp, J., and Stevens, J. R. (2011). The good, the bad, and the rare: Memory for partners in social interactions. *PloS one*, *6*(4), e18945.

Vorauer, J. D., and Sasaki, S. J. (2012). The pitfalls of empathy as a default intergroup interaction strategy: Distinct effects of trying to empathize with a lower status outgroup member who does versus does not express distress. *Journal of Experimental Social Psychology*, *48*(2), 519–524.

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., and Østergaard, L. (2005). To musicians, the message is in the meter pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage*, *24*(2), 560–4. doi:10.1016/j.neuroimage.2004.08.039

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, *72*(3), 655–84. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11405571

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutun, C .E. G., Wrangham, R. W., and Boesch, C. (1999). Cultures in chimpanzees. *Nature*, *399*(6737), 682-685.

Whiten, A., Horner, V., and De Waal, F. B. (2005). Conformity to cultural norms of tool use in chimpanzees. *Nature*, *437*(7059), 737-740.

Whorf, B. L. (2001). The Relation of Habitual Thought and Behavior to Language. In A. Duranti (Ed.), *Linguistic Anthropology: A Reader* (pp. 197–215). Malden, MA: Blackwell.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron*, *40*(3), 655–64. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14642287

Wierzbicka, A. (1986). Human Emotions: Universal or Culture-Specific? *American Anthropologist*, *88*(3), 584–594. doi:10.1525/aa.1986.88.3.02a00030

Wilson, D. S., Near, D. C., and Miller, R. R. (1998). Individual differences in Machiavellianism as a mix of cooperative and exploitative strategies, evolution and human behavior. *Behavioral and Brain Sciences*, *19*(2), 203-212.

Wilson, M., and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, *131*(3), 460–73. doi:10.1037/0033-2909.131.3.460

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–5. doi:10.1073/pnas.0701644104

Xu, X., Zuo, X., Wang, X., and Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *The Journal of Neuroscience*, *29*(26), 8525–9.

Yamamoto, S., Humle, T., and Tanaka, M. (2012). Chimpanzees' flexible targeted helping based on an understanding of conspecifics' goals. *Proceedings of the National Academy of Sciences*, *109*(9), 3588–3592.

Yuki, M., and Yokota, K. (2009). The primal warrior: Outgroup threat priming enhances intergroup discrimination in men but not women. *Journal of Experimental Social Psychology*, *45*(1), 271-274.