The neurological correlates of the human mind-reading mechanism: An ERP investigation using the Prisoner's Dilemma Game

A Division III Project in Cognitive Science
by Bronwen Evans

Background and Introduction

Human culture is an unprecedented phenomenon, the single instance of this particular phenomenon in the known universe, and the study of human social cognition is a growing area of interest in cognitive science today. The singular traits underlying human social structure and culture are widely believed to be the result of recent and progressive evolution of the human cortex, allowing for enhanced information processing abilities (Donald, 1993; Adolphs, 2001). Our increased cognitive capacity sets us apart not only by facilitating our unique cultural and social world, but also in that we are the only species to whom increases in costly information processing proved an adaptive feature The earths truly dominant species, such as rats and cockroaches, are highly successful without the energy costs incurred by our developed cortex (Tomasello et al, 2004; Stevens and Hauser, 2004; Rilling et al, 2002). The aim of this paper is to explore the ways in which our progressively evolved cortex gives rise to our unique way of life by allowing us to act in our best interests despite the daunting complexity of our social environment.

The study of Theory of Mind (ToM), is one avenue of research that probes some of these issues. The ability to quickly form reliable inferences and predictions about the behaviors of our social partners has historically been considered the primary contribution of ToM to behavior (Bloom and German, 2001). ToM has been very popular in recent years, but a close examination of some of the underlying assumptions and research methods in current Theory of Mind research may reveal some important problems for the field. Many tasks employed in the research of ToM are measures of low level or passive judgments and reactions, and may not be a direct

measure of the cognitions which give rise to our remarkable ability to infer and interpret information about our social partners, at which we are so adept it is sometimes referred to as 'mind-reading' in the literature (e.g. Ames, 2004). This paper will focus on a task that requires actively participating in a social dilemma, a task of central importance to successful social interaction. A social dilemma is a concept from game theory, and can be defined as a mixed-motive two-person game with two choices—cooperate (be honest, truthful, or helpful) or defect (lie, cheat, or steal). Social dilemma's have traditionally been used in game theory to study evolutionary models of population dynamics, in order to help researchers understand how decisions of individuals affect their social partners and community (Macy and Flache, 2002; Nash, 1953).

Event related potentials will be recorded while subjects play the two-person prisoner's dilemma game, a game considered to be an example of a social dilemma, once against what they are told is a computer program and once against what they are told is a human opponent. Electrophysiological and behavioral differences in the data will reflect the difference between cognitive processes elicited by a social dilemma and by the same task in a nonsocial context. The task in both conditions is identical, and despite any other flaws in the paradigm it can be said that any differences in response pattern or EEG data must be caused by the perceptual difference between the two conditions (Colman, 2003; Gallagher et al, 2002; Howard, 1998; Kiesler et al, 1996). The data should reflect processing of game-relevant information, processes of attention, working memory, memory, option evaluation, and conflict resolution (Damasio et al, 2000; Colman, 2003; Gallagher and Frith, 2003). The way in which context, or the perceived type of dilemma, influences these processes is of primary interest to us here.

This paper will review the literature from ToM, game theory, decision making, and evolutionary psychology to support the relevance of the current task to a modern study of ToM and to bring into question some of the ways we traditionally think about the issues of central importance in ToM. A number of researchers have recognized and pointed out the same problems outlined here, and we should take note and push ourselves to ask more focused research questions in order to disambiguate these issues (Bloom and German, 2001; Tomasello et al, 2004; Leslie, 1994; Rilling et al, 2004).

Theory of Mind

One of the most important avenues of relevant research, and the one we will be primarily concerned with here, is the field of Theory of Mind (ToM), the study of how we use inferred mental state information to predict and interpret the behavior of other people (e.g. Leslie et al, 2004). The issues that ToM researchers are concerned with today are some of the oldest in the history of scientific investigation of the mind. Emotions, memories, ideas, beliefs, and many other elements central to ToM have been explored by philosophers of mind at least since 428 B.C. Sixteenth century scientist Descartes began to consider these issues in terms of a mind/brain, even going so far as to attempt to locate the 'self' within the brain (Dennet 1991). Even Darwin made attempts to better understand our capacity to represent internal states, and conducted empirical and observational investigations of self concept and perception of intentional action (Darwin, 1877; Gallup, 1970; Keenan, 2003). Psychologists and philosophers have continued to be fascinated with our ability to reflect on and infer internal states, and the modern field of ToM has integrated aspects from this rich history into the study of how we understand ourselves and our social partners.

Researchers of ToM believe that some representation of the internal mental

states of our social partners is used to appropriately mediate behavior (Gallagher et al, 2002; German, 2004; Sanna et al, 2003). We are able to infer and predict the contents of people's minds in a fairly reliable way, despite the paucity of information available during everyday social interaction. It is thought that by attributing to people mental states such as beliefs, goals, and intentions, we can reliably predict their likely actions and reactions in order to engage in appropriate social interaction (Walter and Adenzato, 2004). ToM research is becoming increasingly popular, but the way researchers are approaching this issue needs to be re-evaluated. The cognitions involved are ambiguous and difficult to measure, and researchers must be even more stringent than usual in asking focused questions and being open to new ways of approaching the problem.

Many difficult concepts are intrinsically entwined with what we refer to as ToM, but none is more daunting than consciousness. For the purposes of this discussion, consciousness will not be addressed, as modern science lacks the tools to measure it or an adequate theory of what it is. Similarly complicated issues such as the concept of 'self' will not be addressed. When the term self is used from this point on, it is used to refer to a unique individual with an immediate sense of self based on physical perspective in the environment and self-awareness; a limited sense of self such as we believe infants and chimpanzees possess (Keenan, 2003; Tomasello et al, 2003). We should abstain from any speculations about the phenomenological nature of this 'awareness' and its relationship to consciousness. This principle is extended into the discussion of the representation of the mental states of others, even though it may seem intuitive to think about mental state representation in terms of our own conscious experiences. While we may engage in conscious reflections regarding perceptions about the mental states of others, the relationship of this reflective

reasoning to ToM is beyond the scope of this paper. It follows that all future discussions of representations concern physical representations of information in the brain, in the hopes of explicating some of their underlying mechanisms and their relationship to behavior. The representation and manipulation of social information in the brain is an issue at the core of ToM, and what follows is an attempt to sum up and flesh out this central issue.

Mental Representation and ToM

It is one of the driving goals of cognitive science to understand how information is represented in the brain. Current theories of mental representation are largely informed by philosophical theory based on that which we do know something about; our world and our experience of reacting to and interacting with it (Dennett, 1987; MIT encyclopedia of cognitive science, 1999). These theories posit several intrinsic properties of mental representations: that they have "intentionality," and are about or refer to things; that the representation bearer represents aspects of the represented object by means of aspects of itself; and that this representation is available to some interpreter that interprets the representation in terms of its significance to the subject in such a way that the represented content can make a difference to the 'internal states' and behavior of the subject (MIT Encyclopedia of cognitive sciences, 1999). We know this much is true because we are in fact able to reliably interact with things in our world, sometimes based on very complex or subtle stimuli. Mental representations are also thought to have a constituent structure and syntax, consistent with a conception of mental representations as symbols, which by definition rely on conventions of meaning and syntactic type (Horst, 1996). The following are thought to be categories of things that can be represented in the brain:

concrete objects, sets, properties, events, the state of the world, possible worlds, fictional worlds, and abstract objects such as numbers. The occurrence, transformation, and storage of these information bearing structures constitutes an important part of cognition (MIT encyclopedia of cognitive sciences, 1999).

Being able to mentally represent elements of the immediate environment relative to the self is nothing to write home about, and is certainly present in animals with less developed cognition, for example dogs or horses (Field, 1978). Being able to represent past, future, hypothetical, and fictional states of the world, on the other hand, is an interesting ability without which fantasy, art, conception and execution of long term plans, and human social interaction would not be possible (Sperber, 2000; Diamond, 1994). Without access to representations of the world that don't correspond to reality, human experience of the world as continuous as well as the ability to learn from experience would not exist (Howe and Courage, 1997). Aside from representations of real and possible worlds, we can also represent representations and symbols, and just like representations of concrete objects, these representations are used to mediate our actions and reactions in the world (Field, 1978). Our ability to understand an infinite number of possible combinations of stimuli in terms of their unified greater significance is fundamental to language and to our social and cultural lives (Ferstle and von Cramon, 2002; Castelli et al, 2002). We are able to make tools because we can flexibly abstract information about object attributes in terms of their relevance to novel situations, allowing us to engage in creative problem solving and sophisticated tool use (Cosmides and Tooby, 2000). If you have ever used the sharp edge of a key to open a package, you can recognize how fundamental this ability is to the most mundane aspects of our lives. Our representational capacities allow us to generalize from the specific to gather information about the general. This ability has

proved a valuable asset to our species, and subserves on some level almost all of the behaviors that make us human.

While sketches and models of mental representation are highly speculative, the fact remains that humans do exhibit behaviors that demonstrate sensitivity to external events that requires some representation of those events. The problem of how representations of the mental states of our social partners are used to mediate social interaction is the central problem of ToM. Several problems need to be thought through, namely the nature of the mental entities in question, the validity of certain ToM tasks, and a strong definition of what ToM really is. Several researchers in the field have begun to address these problems, and their arguments are reviewed below.

The problem of ToM mental entities

ToM was originally postulated as a system for prediction and interpretation of social information (e.g. Baron-Cohen, 1998). As such, the representation of information about the internal mental states of others is one of the most visible focuses of the literature. It is more than likely that some representation of the contents of the minds of our social partners is used in appropriately mediating behavior in social situations, but exploring mechanisms correlated to "internal mental states" turns out to be problematic. It is hypothesized that we represent complicated things like rules of a logical system, concepts, scripts, analogies, images, perceptions, and ideas; but also empirically dubious mental entities such as beliefs, desires, and goals (e.g. Kiel, 1989). These familiar but evasive mental objects can be classified as concepts, of the most abstract kind (Barnaby and Marsh, 2002).

It has historically been a difficult task to think about ToM issues without invoking abstract concepts such as beliefs, desires, goals, and states of mind familiar

to us from our own experiences (e.g. Bloom and German, 2000). These concepts are foundational in psychology and philosophy, but are nearly impossible to define in terms of events physically realized in the brain. While these concepts do exist, and are certainly both cultural and mental entities, their role in social cognition may not be as transparent as it would seem from subjective experience. If we can refrain from interpreting experimental results in terms of these kinds of concepts, we may gain a clearer path to understanding the nature of the mechanisms and systems that give rise to the behaviors traditionally associated with ToM. It is an important task for future research to disambiguate these central issues, and to narrow its focus to the realm of that which is empirically testable and theoretically coherent. While a rough model of the elements that subserve a mature ability to infer and predict the mental states of others can rely on introspection regarding how our inferences seem to happen, a solid theory cannot rely on phenomenological abstractions.

Prevalent folk psychological accounts of ToM describe how people act in ways that make sense in light of the beliefs that they seem to be expressing. Concepts can be described as "elements from which propositional thought is constructed, providing a means of understanding the world, used to interpret the current experience by classifying it as a particular kind and relating it to prior knowledge," (MIT encyclopedia of cognitive sciences, 1999). On some level, mental entities such as beliefs must be represented in the brain and serve some purpose in interpreting and predicting events. This brings us back to what has now become a recurring theme: the use of representations to understand and interpret features of the world in terms of their relevance to the self. Familiar but elusive entities like beliefs no doubt play some functional role, but may be beyond the scope of this paper. While some theorists even define ToM as the process of attributing people intentions, beliefs, and

desires in order to predict and interpret their actions; functional mechanisms and perception of the causal role of beliefs probably don't quite match up, and addressing the question of how mental representations aide in reliably predicting and understanding the world need not be confused with accounts of how it seems to happen. The difficulty of classifying and defining these kinds of issues makes ToM research a difficult task, but this task can be facilitated by taking a close look at the way we go about researching it. It is difficult but particularly important to interpret data from this field as objectively as possible, and a good way to facilitate that process is to take a closer look at the tasks traditionally used to measure ToM.

The problem of ToM Measures

A number of cognitive mechanisms work together to give rise to the complex behaviors we attribute to ToM. We use inferred information to more effectively maneuver in a world in which our lives are not only in our hands but in the hands of our social partners. We predict our opponent's next move in a game of chess or our sibling's next ploy in an argument by incorporating information not present in the immediate environment and applying it to current events. (Atherton et al, 2003; Dasser et al, 1989; Gallagher et al, 2003). Everyday events such as understanding the significance of facial expression information, or the most likely mental state of a fictional character in a story, require that we have some representation decoupled from reality; specifically, a representation of inferred information about the mind of another (Dasser et al, 1989). It is not surprising then that everything from perception of intentional movements, story comprehension, the false belief task, and a plethora of other measures are employed in the name of theory of mind research (Fletcher et al, 1995; Baron-Cohen et al, 1998).

What we refer to as ToM has been approached from both too narrow a perspective and too broad a range of functions, and the problem needs to be reformulated. The current study will focus on the ability to form representations of the minds of others in social situations, and should consider this ability both within and outside of the context of traditional ToM research. Data from the decision making, game theory, morality, evolutionary, emotion, neuroeconomics, and social cognition literature will be reviewed to paint a more complete picture of the mechanisms that allow us to use inferred representations of mental states to infer and predict intentional action.

One of the most important avenues of research for ToM is development, a field which has always been a crucial avenue for investigating ToM,. From birth to the age of 6, children exhibit a characteristic development of ToM abilities, traditionally considered to be mature when a child can reliably pass the false-belief task (Baron-Cohen, 1998; Gopnik and Astington, 1988). The false belief task requires that a subject accurately predict the behaviors of actors based on knowledge about the intention of the actor and a false representation of the world. The classic example of this task is the "Smarties" experiment. A child of three will be shown a smarties box (a popular British candy) and asked to report what they believe it contains. Their answer, of course, "smarties". The child is then shown that the box actually contains pencils instead of the expected candy. Another child enters the room, and the first child is asked what they believe the second child thinks the smarties box contains, to which they inevitably reply, "pencils". This same experiment repeated with 4-5 year old children yields different results, and 5 year olds reliably recognize that their compatriot will assume the box contains candy based on their false representation of the world (Gopnik and Astington, 1988). This

difference is attributed to the emergence of more sophisticated representational capacities, namely the ability to represent false states of the world and the ability to represent the perspective of the other as separate from the perspective of the self (Iacombi et al, 2005; Leslie et al, 2004). The 4-5 year olds demonstrate an ability to represent the content of another person's mind, decoupled from reality and different from the content of their own mind. In current research, the false belief task more frequently involves making judgments about the actions of fictional characters (Bloom and German, 2000).

Around the time children begin to pass the false belief task, they characteristically begin to develop a number of other capacities. For example, this developmental landmark occurs around the time when children begin to produce more sophisticated personal narratives, to engage in symbolic play, and to demonstrate more elaborate representation of concepts and categories (Howe and Courage, 1997; de Villiers, 1984). Because so much cognitive development begins to manifest itself at this time, it has historically been difficult to identify and define the relationships of these different capacities to one another. Another large problem with this area of research is that most ToM tasks require language, and the data may be more indicative of children's ability to report ToM behavior than to engage in it (Gallup, 1997). One study testing 3-5 year olds 'moral' actions during the Prisoner's Dilemma game (PD game) indicates that this may be the case (Matsumoto et al 1986).

In Matsumoto et al (1986), children from 3-5 years of age played a simplified version of the Prisoner's dilemma game, a game historically used to study the emergence of other-oriented behavior. Historically, players engage in altruistic behavior even though the incentive to defect is more compatible with self interest, making this an excellent tool with which to test "morality," or sensitivity to the laws

of human social exchange (Matsumoto et al, 1986; Tomasello et al, 2004). Social exchange is very important to a discussion of ToM, and will be returned to later. Data from Matsumoto's study showed that these children did demonstrate 'moral' or other-regarding behavior. They engaged in a variety of actions that indicated an awareness of their partner and sensitivity to the equity or inequity of the division of rewards. These actions included cooperation, in which both partners received a moderate reward; betrayal, in which one partner violated a verbal agreement in order to gain more; and reparation, in which the player who is ahead agrees to make a losing play to allow the other child to catch up. Significant correlations were also found between the reactive emotional expressions of the children and their partner's following play. According to previous research, only about half of these children should be able to pass the false belief task, and yet all of them demonstrated significant sensitivity to information about the mental states of their partner (Matsumoto et al, 1986).

Despite certain shortcomings of this particular study, such as the small sample size (n=30) and the difficult nature of encoding facial expression data, it does illuminate two important points. First, it illustrates the problematic nature of relying on tasks that require verbal report, a problem long documented in the literature (Bloom and German, 2000; Tomasello et al, 2004; Gallup, 1970; Keenan, 2003; de Villiers, 1984). More importantly, it calls into question the definition of ToM. If these children show sensitivity to social contract information and facial expression information before they can pass a false belief task, are these two paradigms really measures of the development of the same mechanism? In theory, reacting appropriately to facial expression information implies a representation of the mind of another as being different and separate from the mind of the self, the exact thing we think passing the false belief task definitively demonstrates. There seem to be

multiple ways of interpreting these kinds of findings, and any data set that supports multiple interpretations is lacking in focus and control. However, the complexity of the behaviors involved and the blurry borders between proposed functional systems do not lend themselves to tight control of all variables. Additionally, language and ToM consistently activate similar areas in the brain, and additionally relevance theoretic pragmatics is founded on the idea that ToM is essential for pragmatic language use (Carston, 2003; Tomasello et al, 2004; Sperber and Wilson, 2002). One distinct possibility is that ToM and language are indeed impossibly entangled, co-evolved from a common origin.

Traditional measures of ToM exclude primates and pre or non linguistic humans. Only very few scientists still entertain the idea that other primates have something like a ToM, but perhaps we only lack the means to measure it (Gallup, 1997; Tomasello et al, 2003). Matsumoto and his colleagues suggest that actions speak louder than words, and that the behavior exhibited in their study indicates that the children could represent the mind of their social partner separated from their own. Additionally, the children appeared to be sensitive to 'fairness' (equity or inequity), and sensitive to their ability to change the outcome of their partner. These behaviors do indeed indicate that some representation of the mental world of the social partner is used to mediate behavior.

It is possible that this kind of inferential capacity, specific to social exchange situations, emerges earlier than the ability to make passive judgments about inferred mental states of actors or characters (Bloom and German, 2000; Sperber and Wilson, 2002; Whitehurst and Lonigan, 1998). It doesn't seem so strange that the ability to infer the mental states of social partners during natural social exchange should emerge before the ability to report false beliefs observed passively, frequently the false beliefs

of fictional characters. Competent pragmatic language, grounded in expressive conventions and extra-linguistic information such as is provided by context, emerges much earlier than the capacity to understand language out of context, such as fictional story comprehension or academic literacy (Whitehurst and Lonigan, 1998). Pragmatic language comprehension is additionally theorized to rely on ToM type processes, such as inference and interpretation of information about the communicative intent of a social partner. Decontextualized language comprehension, on the other hand, relies on sophisticated representational capacities requiring references to scripts and complex knowledge structures which can be applied to fictional characters and situations (Carston, 2003). Considering this, our ability to understand the actions of our social partners during natural social interaction seems like a very distant relative to the passive attribution of beliefs to a fictional character based on situational information.

While the example given above attacks the validity of one specific task, a great number of ToM studies are done using equally questionable measures. All ToM studies that employ measures that require the use of decontextualized language, passive judgments, and fictional or imaginary actors introduce the same problems as the false belief task (Bloom and German, 2000). Tasks involving perception of facial expressions are also commonly used to study ToM inference formation. While there is no doubt that in real world interactions such perceptual information is indeed used in making these inferences, the study of this kind of perception is separate from the study of inference formation. The cognitions studied in these and other related kinds of studies are indeed relevant to the study of ToM, but a real understanding of ToM cannot be expected to emerge until a clear definition of which cognitive events constitute ToM is firmly established and generally agreed upon.

A strong definition of ToM

The mechanism central to ToM is some mechanism or set of mechanisms used to form inferences and predictions about the intentional actions of our social partners (Gallagher et al, 2003; Baron-Cohen, 1998; Leslie et al, 2004). While this ability is reliant on perceptions of social cues, such as facial expressions, the cognitions foundational to successful social interaction are not perceptions but rather tasks of search, assessment, and option selection that allow us to interpret and act upon the information available to us in a time effective way (Marsh, 2002). This can only be described as the use of decision heuristics, or cognitive shortcuts used to make quick and reliably accurate judgments about real ecological problems (Ames, 2002; Marsh, 2002; Adolphs, 1999). Heuristics are cognitive shortcuts that enable us to make evaluations on the basis of one or a few simple rules or cues, allowing us to avoid the time and information processing costs of reviewing all the options by constraining our decision space in order to consider only the most appropriate options (Hastie and Dawes, 2001).

Our social world is rife with imperfect information, structural complexity, and limited time within which to make important decisions, as well as situations in which the ability to make quick and effective judgments can be particularly crucial. A specialized set of heuristics could effectively minimize the cognitive cost of decision and behavior monitoring processes involving social behavior (Colman, 2003; Ames, 2004; McCabe, 2000). The observed behavior that is attributed to ToM could be the result of a specialized set of heuristic mechanisms that have evolved to help us successfully navigate our social world (Stevens and Hauser, 2004; Donald, 1993). This theory is supported by evolutionary theories, a number of theories in social

cognition and social psychology, and relevant theories in the fields of attention, memory, emotion and decision making (Stevens and Hauser 2004; Fehr et al, 2004; Fischbacher et al, 2005, Hastie and Dawes, 2000; Ames, 2001). As such, ToM will be considered to be some mechanism that allows for the quick and effective evaluation and interpretation of social information. Now that we have outlined some flaws in current conceptions of ToM and how it should be researched, we can return to those parts of ToM research that can help us learn about this process of reflexively and accurately interpreting our complicated social world.

Folk Psychology

While we must stop relying on our subjective experiences of ToM to guide research hypotheses and interpretation of data, it is also important to understand those experiences in terms of what they really can tell us about how our brain works, and an understanding of what is referred to as "folk psychology" can give us real insight into the workings of our minds.

The ability to interpret and predict behavior is sometimes explained as a 'folk psychology', or a naive theory of the laws of human behavior. Similar phenomena, such as a naive 'folk physics', evolved to provide our ancestors with information to aid in navigating their physical world at the lowest cognitive cost. Our capacity for folk physics, evolved as a function of the adaptiveness of rough but reliable prediction of physical phenomena in the immediate environment, allows us to respond appropriately to physical phenomena without having any deep understanding of physical laws. We all know to reach towards the area below the point at which an object began to fall, and do so without any reference to the laws of gravity. It has been suggested that it is our innate tendency to perceive the physical world in this

particular way that can make academic physical theories seem so counterintuitive (Dennett, 1987). Although more difficult to concede, our interpretation of human behavior is similarly constructed on the principle of reliable prediction based on simplified principles.

Certain tendencies and sensitivity to historically reliable cues allow us to isolate significant themes and information well enough to understand our social partners based on minimal information. By observing folk psychological phenomena, we can form hypotheses about innate tendencies in human social perception and behavior. Two of these innate tendencies appear to be foundational to ToM; perception of causality and of agency. Humans exhibit a general tendency to perceive causal relationships, which appears to be reflexive (Leslie, 1995; Blakemore et al, 2001). An understanding of relationships between causes and effects is critical to making sense of our constantly changing physical world, and even greater sensitivity to abstract causal relations is required by the demands of our complex social organization. This is achieved by elaborating on available information to generate hypotheses about likely causal explanations for events, including self generated actions. Such causal attributions come naturally and automatically to normal adults, even when they do not appear to contribute to performance. A recent study demonstrated that the left hemisphere in split-brain patients generated a causal explanation for the behavior of the dissociated right hemisphere (Gazzaniga, 2000). In this instance, the left hemisphere had no access to the causal mental states of the right hemisphere and the generation of such a causal story did not contribute to performance on the task, in which subjects were asked to choose an image from an array of images presented to the right visual field, and then asked to explain their choice. The left hemisphere is dominant for language, and any causal explanation

subjects generated could have no access to the causal states of the right hemisphere. The researchers concluded that the generation of causal stories about why the subject chose the scene they did were constructed post-hoc, and that many of our causal stories are of a similarly constructed nature (Gazzinaga, 2000). This and similar findings, including research on the perception of causality while watching the movements of abstract shapes on a computer monitor, suggest that our causal stories are spontaneously generated to help us make sense of the world (Castelli et al, 2002; Gallagher and Frith, 2003).

Humans are similarly hard-wired to perceive agency in subject originating motion. We display an innate tendency to perceive contingencies between objects and people, both mechanical contingencies and more sophisticated intentional contingencies. The tendency to interpret events as agent originating triggered by the perception of the self-initiated movement of an agent, and this tendency is present from infancy (Leslie, 1995; Csibra et al, 1999). This kind of interpretation is similar to the proposed tendency to generate causal stories, but assumes some 'agent' capable of intentional action. Our reflexive interpretations of information regarding causality and contingencies between objects and people appear to be important foundations for ToM (Blakemore et al, 2003). Many researchers have studied these topics in depth, providing hard science answers to questions posed by naïve theories.

Investigation of folk psychological phenomena can provide us with useful frameworks from which to begin to investigate real cognitive mechanisms, and also provide us with puzzles that only a deeper understanding of the brain can help us solve.

Selective Impairment

One path to gaining a more objective understanding of the elusive underlying cognitions subserving ToM are instances in which ToM is selectively impaired, such as in patients with Autism spectrum disorder. The disorder is considered to be an impairment of social cognition, and it is interesting that patients often suffer language and symbol representation deficits as well (Castelli et al, 2002). Most importantly though, these patients are handicapped in their ability to interpret and predict behavior. Autistic patients have described their analysis of people in social situations as a mechanical, analytical, and slow process; frequently resulting in inappropriate conclusions (Ferstle and von Cramon, 2002). This ability is automatic and reliably accurate in normal adults (e.g. Gallagher et al, 2003).

Autistic children are believed to exhibit deficits in ToM, and fail the false belief task even when mentally retarded children of an equivalent or lower mental age pass it (Castelli et al, 2003). While this suggests that a general cognitive deficit cannot account for this effect, the high incidence of language impairment in Autistic subjects makes the false belief task, which relies on verbal report, inadequate (Gallup, 1997; Bloom and German, 2000; Keenan, 2003). Developmental data from studies of Autistic children documents deficits in other social mechanisms such as joint attention, the ability to attend to the object of a social partner's focus (Leslie, 2000). It is thought that sharing attention helps infants attend to appropriate stimuli crucial for development. Joint attention is also thought to be crucial to language learning (Howe and Courage, 1997; Tomasello and Farrar,1986). All of this data is testament to the inter-relatedness of language to theory of mind, and the ways this relationship confounds the task of empirical investigators. Selective impairment of ToM in deaf populations is of particular interest to those interested in how the development of language and ToM are inter-related.

There is an interesting body of literature documenting deficits in ToM in deaf individuals. Most deaf children fail to pass the false belief task before the age of 13-16, a shocking contrast to the normal age range of 4-5 (de Villiers et al, 1984). This effect is exciting to researchers of ToM because deaf subjects do not exhibit mental retardation or other cognitive deficits that could account for these experimental results. Patients with Autism spectrum disorder are more commonly used in studies of impaired ToM, but the high incidence of severe mental retardation and language deficits introduce uncertainty into any interpretation of the data. The deaf population provides a sample that is representative of normal brains developed in the absence of one important source of sensory input.

Importantly, deaf children of deaf, fluent signers exhibit normal development and pass the false belief task within the average time frame. However, 90% of deaf children are born to hearing families with limited or absent proficiency in gestural communication (Woolfe et al, 2002). Parents of these children report frustration with trying to communicate about anything other than the immediate environment with their deaf children, and probably never communicate about abstract information such as mental states (de Villiers et al, 1984). Additionally, these children are deprived of information from the speech sounds of friends and family normally present during early development. This information is thought to be crucial to linguistic development, and may be crucial to development of ToM as well.

Most deaf children of hearing parents don't learn sign language until much later when they attend school with other deaf students, too late to ever have a normal language ability (de Villiers et al, 1984). Individuals who learn to sign in school will never become truly fluent in the way that their counter-parts who learn to sign from birth are. With normal stimulation, however, gestural language develops just as

spoken language would (Woolfe et al, 2003). Many researchers hypothesize that the failure of regular development of ToM is caused by the lack of communication about mental state information during development, caused by the difficulty of communicating about abstract entities in the absence of language.

While interpersonal communication certainly plays a large role in development of ToM, the difference observed between fluent and non-fluent signers on performance of the task suggests that there may be more to this story. These two groups of subjects have roughly equivalent language skills in terms of expressing communicative content (Woolfe et al, 2003). Additionally, at around 6-10 years of age the children would have entered school, an environment where they very well could be exposed to communication about mental states. These children begin to reliably pass the false belief task at about 15 years of age on average, with a great degree of variability (Woolfe et al, 2003; de Villiers, 1984). A study correlating degree of linguistic development and exposure to degree of ToM ability could aid in understanding how our earliest experiences with human language shape our ToM capacity. Aditionally, it is important that these findings are replicated using measures that do not rely on linguistic report to ensure that it is ToM and not linguistic ability that is being tested.

The degree to which language and ToM are intertwined may reflect a co-evolution of these two capacities (Tomasello et al, 2004). Many evolutionary stories posit a leading role for social structure and communication in recent human cortical evolution, and modern social structure gives us a few clues as to why that might have been. These evolutionary theories, coupled with data from the decision making and evolutionary game theory literature, begin to give us a good idea of how social cognition evolved and ways in which this manifests itself in what we refer to as ToM.

Evolutionary Theories

Evolutionary psychologists have hypothesized that ToM evolved as a response to changing social pressures of our ancestors, in which the phenomenon of 'social exchange' made ToM highly adaptive (Donald, 1993; Adolphs, 1999; Tomasello et al, 2004). The unique living conditions of our ancestors created a solitary instance in which increased cognitive capacity had the evolutionary edge. Evolutionary theories about these conditions could shed some light on the mechanisms underlying those abilities, allowing us to form testable hypotheses about these hard problems (Donald, 1993).

Many evolutionary psychologists attribute the rapid and recent development of the human brain to changing social conditions only 7 million years ago, when our ancestors diverged from those of the African ape (Donald, 1993). Some believe that the rapid evolution of the human cortex was a function of increasing social pressures as group size increased. Living in groups had numerous advantages to reproductive fitness: better predator protection, mate selection, and reliability of food and other resources, to name a few. Conversely, living in larger communities introduced more competition for mates and food. This increased the adaptiveness of more sophisticated social skills, introducing a situation in which cognitive capacity finally had the edge over physical prowess. A positive correlation between brain size and average group size supports this hypothesis (Ames, 2004; Stevens and Hauser, 2004). In human evolution, the phenomenon of social exchange behavior specifically is thought to have played a large role in the evolution of social cognitive capacities, including our exemplary inferential aptitude.

Social exchange is the behavior that allows us to trade items and services with

our social partners, relying on an unspoken contract that goods and services provided must be repayed with goods and services of an equivalent value. One of our predecessors skilled at foraging for food may have exchanged excess food to a dominant community member in exchange for physical protection from enemies. The social contract involved in social exchange behaviors requires a complex representation of values relative to the self and the other (Cosmides, 1989). The benefit of engaging in social exchange behavior stems from the ability to exploit situations in which you trade an item which has little value to you but a high value to your social partner. The ability to evaluate and represent the value of goods or services to another would be adaptive for both the community and the individual, providing the perfect circumstances to promote the reproductive fitness of individuals capable of these more intensive cognitions (Cosmides, 1989).

The ability to represent relative values in order to engage in successful social exchange requires representation of the mental states of the social partner, such as their unique world perspective. The adaptiveness of this ability in this specific situation describes one possible evolutionary origin for ToM. But what motivates us to live up to our social contract? Evolutionary game theory proposes that only evolutionarily stable strategies of social exchange (ESS) could persist (Nowak and Sigmund, 1993). Failure to honor the social contract, although adaptive in instances of one-time interaction, would not emerge as an evolutionarily stable strategy in a community with a stable set of social partners. Similarly, unconditional helping could never emerge as an evolutionarily stable strategy (Fehr et al, 2004). Studies in economics, game theory and decision making have frequently documented the emergence of cooperative behavior in situations where cooperation is not an economic decision in terms of individual gain. These data have been explained in

terms of the adaptiveness of conditional helping in social exchange. Evolutionary game theory predicts the dominance of conditional helping, and is supported by evidence from computer modeling, in which a tit for tat (TFT) strategy, equivalent to conditional helping, consistently outperforms other strategies in iterated instances of social interaction (Singer and Fehr, 2005).

Using the TFT strategy successfully requires representation of complex information about the mind of a social partner. It requires reliable and reflexive interpretation of intentional action, representation of goals and knowledge states, and sensitivity to the equity or inequity of an exchange (Nowak and Sigmund, 1993). Additionally, the ability to use this information to reliably predict and influence the actions of social partners would have been highly adaptive (Tomasello et al, 2004).

Based on this theory about increasing social pressures exacerbated by the emergence of social exchange behavior, it has been posited that humans have a mechanism specialized for cheater detection, in order to avoid engaging in social exchange with unreliable partners (Cosmides et al, 2002). People do appear to be very sensitive to equity, as illustrated in research using the dictator game, in which subjects consistently prefer receiving no money over an 'unfair' distribution of the money (Hoffman et al, 1996). While this mechanism does protect against exploitation, it doesn't constitute a cheater detection mechanism, only an ability to accurately judge the relative values of goods. Additionally, perceptual mechanisms involved in cheater detection, such as inferring information from facial expression and context, provide information used in hundreds of judgments other than cheater detection (Bechara et al, 2002). While cheater detection would have been a powerful tool in this kind of social environment, it seems more likely that the ability to detect cheating behavior may be subserved by a more general inferential ability which relies on sensitivity to

certain kinds of information. While cheater detection may have been adaptive to our recent ancestors, making inferences about the actions of our social partners must have contributed to reproductive fitness well before the emergence of such complex social structures.

Facial expressions have been used to communicate evaluative information to conspecifics since long before our ancestors diverged from the great apes, long before social exchange and any need for a cheater detection mechanism emerged (e.g. Call et al, 2003). In humans, this rudimentary system has developed in representational complexity along with the development of a more comprehensive collection of evaluative states. While it is probable that enhanced sensitivity to information useful in cheater detection was adaptive, it is only a small part of an intricate mechanism that aids us in reasoning about social information (Bechara et al, 2003). The ability to more effectively reason about social information may have been the real driving force in the evolution of ToM.

Decision Making

Reasoning about social information surely involves domain general reasoning mechanisms, but if we review the literature it becomes apparent that we have long been aware of differences between reasoning about social and nonsocial information (Howe and Courage, 1997; Bechara et al, 2004). Researchers of emotion have long been aware that social information elicits particularly salient emotional reactions (Bechara et al, 2000). Affective states allow us to learn from our experiences and make better decisions. Additionally, they influence a variety of judgments and responses, and are an important part of how we monitor our actions relative to our

goals (Schwartz and Clore, 1996). Complex tasks such as effective social interaction, in which outcomes are influenced by factors over which we have little or no control, could benefit greatly from such a monitoring system. Emotional states relevant only to social situations: guilt, shame, embarrassment, and jealousy, may have evolved specifically in order to mediate social behavior (Rilling at al, 1998).

A set of domain specific decision heuristics may do much to constrain the decision space in the instance of a social dilemma, but it also seems likely that humans are predisposed to engage in certain patterns of behavior, such as engaging in conditional helping. Evolutionary game theory proposes that conditional helping emerged as a dominant strategy because it contributed to the reproductive fitness of the individual, and that individuals predisposed to engage in conditional helping behavior would have been more likely to reproduce (Rilling et al, 1998). The evolution of ToM can be described as the evolution of a specialized and sophisticated decision making mechanism. Several recent studies have used games traditionally used in game theory in an attempt to isolate 'mentalizing', one term used to describe mental state representation. While a few researchers address the significance of game theory and decision making to ToM, the issue has not yet been fleshed out. The current study and this paper are aimed at illustrating the relevance of these fields to each other and the importance of communication across disciplines. The concept of an evolutionarily stable strategy in and of itself indicates that ToM was adaptive because it allowed humans to make decisions that increased reproductive fitness. Viewed from this perspective, it seems clear that cognitive scientific models of decision making and game theory can inform the study of ToM immensely.

Successfully engaging in social exchange can be described as engaging in good decision making strategies in situations where the actions and reactions of others

play a role in determining the eventual outcome (Colman, 2003). We represent states of the world disparate from reality, access information about past experiences, estimate future reactions to possible future states of the world, and have the ability to use this information to monitor performance relative to personal goals. The sum of these parts is a system that allows us to make reliably good decisions given our particular social world (Bechara et al, 2004). Mental states such as emotions appear to play a large role in how we make decisions, and it follows that inferring the emotional states of others could tell us a lot about what decisions they are likely to make. First, though, we must understand how representing affective states decoupled from reality could have been adaptive.

The ability to simulate possible mental states of both the self and a social partner can provide important information about possible outcomes of a decision. Decision making is a multiple step process of acquiring and processing information in order to initiate some specific action (e.g. Damasio et al, 1994). A reliable ability to predict the outcome of a particular decision and the value of that outcome relative to goals and preferences is an invaluable tool. Affective states play an important role in minimizing uncertainty in a number of ways. The average college student has probably experienced an instance in which they could choose to stay in bed an extra hour or get up and attend class. Affective information such as how pleasant it is to stay in bed, how unpleasantly cold it is outside, and how motivated the student is to perform well in the class is incorporated into making a decision in favor of one or the other option. There are certain uncertainties in making any decision; for example it may be more or less likely that the professor will even notice if the student misses class, or that today's lecture includes information that will be on the test. While people do not always weight risk and reward in a manner that elicits a rational

decision, their decisions are based on a cost-benefit analysis that yields appropriate behavior in normal individuals (Singer and Fehr, 2005; Colman, 2003).

Many models of decision making, such as Damasio's somatic marker hypothesis, postulate a central role for bioregulatory state information in decision making (Damasio, 1994). In these theories, emotions allow us to monitor our performance relative to personal goals by providing evaluative information about situation outcomes that can be used in learning about different kinds of experiences. In this way, affective states help us to constrain our decision space based on past and projected future experiences. It has long been observed that we preferentially allocate attentional resources to emotionally salient information, and that we tend to have more intense affective reactions to socially relevant information (Howe and Courage, 1997). This could be one component part of a system that allows us to process, remember and act on socially relevant information quickly and efficiently.

The ability to represent states of the world decoupled from reality is foundational to this view, as representations of past states and hypothetical future states as well as the relationship between the two inform the decision process (Bechara et al, 2002). In risk assessment tasks, abnormal bioregulatory reactions to risk have been correlated with poor performance (Verdejo-Garcia et al., 2006). In Verdejo-Garcia et al. (2006), lower than average changes in skin valence, a traditional indice of anxiety, was correlated with poorer than average performance on a task involving risk assessment. It has been hypothesized that this indicates a deficit in the ability to project possible future states of the self to use in appropriately weighting different options (Verdejo-Garcia et al, 2006). Emotional state information not only allows us to monitor our performance from instance to instance, but facilitates statistical learning about kinds of situations encountered multiple times, allowing our

performance in those circumstances to improve with experience. If our social performance is accompanied by more intense evaluative affective states, which effect preferential memory encoding, it follows that information relevant to social experiences may be more accessible than information about non-social experiences (Howe and Courage, 1997).

The ventromedial pre-frontal cortex is thought to be necessary for representing and accessing learned associations of situation classes with evaluative affective states; in order to apply information about which experiences are likely to elicit which internal states. Situations that have been previously experienced trigger reconstruction of the previously learned emotion/situation association, and the previously experienced state is re-enacted. Internal states with higher emotional valence have proportionately larger effects on behavior. This serves to efficiently constrain the decision space by selectively processing only the most relevant information (Bechara et al, 2002). Feedback is essential to this model because it allows for a comparison of the expected bioregulatory state with the bioregulatory state elicited by the decision outcome. This allows us to update stored information about types of situations with each new experience of that type.

While our human models of this feedback and reward process are theoretical, recent data indicates that in monkeys, midbrain dopaminergic neurons signal errors in reward prediction (McCoy and Platt, 2005). Larger disparities between predicted state and experienced state result in larger learning signals, and higher emotional and neural impact. The anterior cingulate cortex, or ACC, a region very close to areas consistently correlated with ToM, may be involved in this updating process by keeping track of rates of successes and failures based on evaluative feedback information. This mechanism may have evolved to aid in developing reliable

decision strategies based on statistical information in order to decrease outcome uncertainty (Damasio, 1994).

Increased group size and availability of resources introduced a new factor into the decision space; the goal oriented actions of other individuals motivated by their own self-interest (Donald, 1993). The ability to effectively predict, interpret, and manipulate this new space requires increasingly complex representations of past and future states of the world. By factoring information about the likely actions and reactions of social partners into a decision space, the uncertainty of a decision outcome value relative to personal goals is decreased. What we refer to as ToM may have evolved to constrain this decision space and reduce outcome uncertainties in our daily social interactions, and some process of representing the mental states of our social partners differentiates this from other kinds of decision making. Recent research has produced some interesting findings about differences and similarities between representing self-originating mental states and those of a social partner.

The Simulation Theory of Mind

Research on empathy and ToM indicates that during human interaction, the process of representing mental states of others is subserved by some of the same pathways that subserve representation of the states of the self. Specifically, the anterior cingulate cortex (ACC) is active while experiencing and while perceiving affective states. High levels of empathy have been positively correlated with conditional cooperation, indicating that representations of affective states of others may combine with our own affective state representations to bias behavior (Gallese, 2001). Bioregulatory states such as guilt, embarrassment, and jealousy are specific to

social situations, indicating that at least some mechanisms evolved specifically to provide evaluative information about performance in social situations (Gerrans, 2002). This account of mental state representation suggests that similar pathways represent both the mental states of the self and the other, a view that is in agreement with one of two prevalent theories about ToM, the “simulation theory” of ToM.

In cognitive science, two important theories of ToM have emerged. These two dominant theories are, "theory theory"(TT) and "simulation theory"(ST). TT posits that ToM is a folk psychological theory of human behavior developed over time as a product of experience. Theory theorists describe a process by which we formulate intelligent guesses about a set of laws that govern human behavior (Dennett, 1987). Simulation theory proposes that people attribute mental states by simulating the experience of the other in order to represent their perspective and corresponding affective states (Tirassa et al, 2005; Pelphrey et al, 2004; Ketelaar et al, 2003; Iacombi et al, 2005). This paper will consider ToM issues from the perspective of ST, as a review of recent literature provided indications that the processes of mental state inference is better described by ST than TT (Tirassa et al, 2005, Iacombi et al, 2005; Gallese and Goldman, 1998).

One of the most talked about findings in support of ST comes from non-human primate research. Recently, electrophysiological researchers have discovered a type of neuron with very interesting implications for ToM and ST. An unusual type of visuomotor neuron, referred to as 'Mirror Neurons' (MN), respond selectively to goal oriented actions. The literature indicates that perceiving intentional actions of others and executing intentional actions utilizes a common neurological substrate, exactly what one would expect if our understanding of others relied on simulating the experiences of others in the self. Researchers hypothesize that monkeys have a

reflexive translation mechanism that allows them to match perceptual information to the equivalent motor action. It is hypothesized that this execution/observation matching mechanism may be a precursor to modern man's theory of mind (Gallese, 2001 and 2005). It is possible that the MN matching system observed in monkeys originally evolved to promote learning about the self through imitation of conspecifics during development. While this conclusion is speculative, findings illustrating the significance of imitation for development of ToM provide additional support for the relevance of the mirror neuron findings to ST (Decety and Sommerville, 2003; Gallese, 2001 and 2005; Iacombi et al, 2005).

Developmental data suggests that human infants begin to imitate facial gestures of conspecifics within the first few days of life. Human infants do not imitate gestures of non-humans. This finding indicates that human infants are predisposed to attend to and replicate conspecific gestures, and have some translation mechanism that refers observed gestures into motor actions the self is capable of generating (Decety and Sommerville, 2003). Infants react significantly more to adults who are engaged in imitating their behavior (Leslie et al, 2003). Preferential attention to imitative behavior continues through development, and reciprocal imitative games with peers are thought to play an important role in development of social skills, such as turn taking and coordinated play. Imitation is thought to be an important tool for learning about the self, others, and how to interact with objects in the world (Tomasello et a, 2005; Hannah and Meltzoff, 1993; Nelson and Fivush, 2004).

Rizzolati, one of the leading researchers in the field, posits a theory of how the mirror neuron system could support imitative learning (Rizzolati, 2003). It is possible that the MN matching system observed in monkeys originally evolved to promote learning about the self through imitation, as many modern MN enthusiasts speculate.

However, not even primitive imitative behaviors have been observed in monkeys (Whiten et al, 2001). While the mirror neuron findings have been received with unprecedented enthusiasm, a theory of their role in ToM is at best speculative. While very intriguing, theoretical leaps from data about perception of gestures in non-human primates to theories of human ToM seem hardly justified.

Many theories have emerged about the significance of the intentional action motor/perception system to theory of mind and social cognition, and as exciting as these possibilities are, we should remember that mirror neurons have only been shown to be active during perception of crude motor actions such as grasping, and the distance between lower primate perception of grabbing and the ability to predict complex behaviors based on inferred information about abstract mental entities is relatively vast. However, there are some very strong indications that mirror neurons or not, we represent the mental states of others through simulation. For example, a recent ERP investigation recorded a component previously associated with error detection during both performed and observed errors, which indicates that there may be a shared biological reaction to both self and other originating errors (Bates et al, 2005). A mechanism like this would help us to not only understand the perspectives of our social partners, but to learn from their mistakes as well as our own. A review of the studies mentioned above and a plethora of similar findings provides solid support for a simulation theory of ToM.

While ST does give us some clues as to where and how ToM occurs in the brain, a comprehensive model of mechanisms for simulating the mental states of others is lacking. Functional imaging studies, as well as research on selective impairment of ToM, have begun to explicate the neuroanatomical basis for this complex cognitive process.

Neuroanatomical Correlates of ToM

Multiple cognitive mechanisms are thought to play a role in giving rise to ToM, and ToM is a necessary component part of many complex behaviors. Appropriately, a number of different methods have been employed in the attempt to identify neurological correlates of ToM. Due to the wide variety of cognitions of interest to ToM researchers, it can be difficult to interpret the data from these studies. The majority of these studies have used tasks that require passive judgments about and perceptions of ToM related stimuli, and are relevant only to those specific cognitions (Bloom and German, 2001). Both the measures employed and the results from these studies are highly variable. However, as each of these different tasks involves some task relevant to ToM, a review of the findings can begin to paint a picture of the neuroanatomical substrates of some mechanisms involved in forming inferences and predictions about the intentional actions of our social partners.

One avenue of research into the substrates of ToM is provided by selective impairment in patients with Autism spectrum disorder. Autistic patients typically have larger than average brains with a number of physical abnormalities. All lobes with the exception of the frontal lobe are oversized, although this is specific to male patients. It is postulated that excessive cell packing density in the hippocampus impairs the ability to integrate affective information about past events into a representation of ongoing events (Rapin and Katzman, 1998). Evaluative feedback information is further disturbed by amygdala dysfunction, which results in disordered assignment of evaluative information to incoming stimuli. The absence of normally reflexive motivation to attend to social information and interact with social partners has been attributed to dysfunction of the neuropeptide system involving oxytocin and

vasopressin, and the associated endogenous opiate system (Frith et al, 2002). Autism researchers stress that the disorder may not have a single biological cause, and that the diversity of symptoms in autism spectrum disorder may be the result of equally diverse neurobiological bases (Rapin and Katzman, 1998). Despite this caution, the Autism literature illustrates the importance of accurate affective feedback as well as preferential attention to domain specific (social) information to the development of normal ToM (Burnette et al, 2005).

Functional imaging data has also provided information about the neuroanatomical bases for ToM. A network of brain regions has been consistently activated during functional imaging ToM studies, spreading across multiple structures. The tasks employed by researchers in the field have been typically variable, but the most commonly activated areas are the anterior paracingulate cortex bi-laterally, the superior temporal sulcus (STS) bi-laterally, and the right temporo-parietal cortex (Frith and Frith, 2003). Some interaction of these regions subserves the perceptual, evaluative, and inferential processes involved in ToM.

The right temporo-parietal cortex has been prominent in many previous imaging studies (Baron-Cohen et al, 1999; Patel and Azzam, 2005), and has shown preferential activation during story comprehension requiring mental state inference and perception of biological motion (Grezes et al 1999). The common factor underlying studies that report temporo-parietal cortex activation is that they involve the explicit representation of behavior, either directly or by verbal description. This indicates that this region may be involved in processing behavioral signals that allow us to form hypotheses about internal mental states, but may not be directly involved in representing mental state information.

The superior temporal sulcus, or STS, is thought to play an important role in

processing of visual cues relevant to social cognition. Activation of this area has been correlated with perception of biological motion. Research on macaques indicates that the STS may be a high level processor of visual information in which motion, spatial, and object information are integrated. The STS additionally has strong reciprocal connections with the amygdala, which fit a story in which perception of the actions of others is monitored by evaluative emotional information (Allison et al, 2000). It has also been hypothesized that the STS is involved in retrieving and encoding autobiographical memories (Howe and Courage, 1997). Some patients with Autism spectrum disorder exhibit decreased connectivity between the STS and the extra-striate cortex, another area associated with mental state representation (Rapin and Katzman, 1998). In a system so reliant on integration of different kinds of information from different sources, a weak connection could cause massive dysfunction.

The anterior para-cingulate cortex has historically been activated by emotionally salient information, and is thought to play a role in processing emotional feedback (Bechara et al, 2001). Damage to this region sometimes results in a marked drop in motivation, possibly an indication that this area is involved in the relationship between decision making and motivation. Data from studies reviewed by Gallagher (Gallagher, 2003) implicate the border between Broadman's areas 32 and 9 as the location of mechanisms crucial to predicting and interpreting the actions of our social partners. This area is a cytoarchitecturally unique region at the anterior tip of the anterior paracingulate cortex, and is thought to be particularly important to tasks which involve actively engaging in mental state attribution.

In Gallagher et al (2003), one of the many functional imaging investigations of ToM, only the anterior paracingulate cortex showed a significant difference in

activation during the target mentalizing task. Researchers hypothesized that the failure to elicit activation in the temporo-parietal junction and STS reflects the difference between actively and passively attributing mental states. In the task of Gallagher et al (2003), subjects were actively engaging in mental state inference, whereas previous studies had focused on passive judgments about mental state information. Actively engaging in mental state inference or social interaction may be a powerful and relevant tool for ToM research, and such a task is employed in the current research. In light of this, the anterior paracingulate cortex and it's role in ToM, memory, and decision making are of particular interest in the current study.

The anterior paracingulate cortex is often considered to be part of the anterior cingulate cortex (ACC). The presence of spindle cells, an unusual type of projection neuron, suggests that the ACC has undergone recent changes in evolution (Uddin et al, 2004). In humans, these cells are not present at birth, but first appear at 4 months of age. Increased cortical folding in this region might also be indicative of recent evolution. The size and location of these cells is consistent with newly evolved neurons that provide strong connections between a widely distributed network of neural areas that give rise to complex behaviors (Uddin et al, 2004). Integrating information from different sources into a mental model would require such a system of pathways between information sources. However, Broadman's area 32 is cytoarchitecturally distinct from this bordering area, and the proposed recent evolution of the ACC may have no relationship to the system. However, the ACC is hypothesized to play a role in evaluating affective states to monitor decision strategies, as well as in processes of conflict resolution. The proposed link between ToM and decision making processes is reflected in the proximity of these two regions.

The Current Study

ToM can be defined as some mechanism or set of mechanisms that provides us with powerful tools for successfully engaging in interpersonal interaction. This ability is subserved at some level by the integration of information from different sources into our predictive mental models. Sophisticated and subtle integration of information from perception, working memory, representations of past-events, and inferences inform reasoning processes, and the ease with which we reason about complex social information supports some heuristic mechanism which allows for domain specific, cost-effective reasoning about social information (Bechara et al, 2001, Gallagher et al, 2003; Tomasello et al, 2001). A number of games traditionally used in game theoretic research can engage this domain specific reasoning mechanism. Most recently, researchers have used the ultimatum game and the prisoner's dilemma game to explore topics in social cognition (McCabe, 2004; Gallagher, 2001; Kiesler et al, 1996). A review of the methods and findings of these relevant studies has informed the design of the current study.

In Kiesler et al (1996), subjects exhibited significantly more cooperative behavior when playing the prisoner's dilemma game against a human than when playing against a computer program. This finding supports a theory that there is some predisposition to engage in other-oriented behavior in social but not nonsocial situations. In Rilling et al (2002), researchers attempted to isolate areas active during cooperative behavior by using fMRI to record neural activation while women played the PD game with other women. This functional imaging study identified the ACC and orbito-frontal cortex (OFC) as areas selectively engaged in cooperative behavior, and reported that these activational differences did not appear when subjects did not use cooperative strategies. The authors hypothesize that these areas may be involved

in motivating us to engage in cooperative behavior. In Gallagher (2002) the game rock, paper, scissors was played in two conditions; against what subjects were told was a computer program and what subjects were told was a human opponent. Both conditions employed identical tasks, elegantly isolating any differences due to the perceptual difference between conditions. The anterior para-cingulate cortex was preferentially activated when subjects believed they were playing a human opponent.

These studies are of particular interest because they all use a task in which the subject is actively engaged in reasoning about the actions of their social partner. It is interesting that unlike functional imaging studies of passive or perceptive ToM tasks, these studies do not report activation of the STS, and certain other areas correlated with different kinds of ToM tasks (e.g. Ito et al, 2004). Actively engaging in reasoning that relies on ToM is of primary interest in the current study, and the paradigm of Gallagher et al (2001) isolates the particular cognitions of interest (Gallagher et al, 2001). The prisoner's dilemma game is used instead of the game rock, paper scissors, which was employed in Gallager et al. (2001). The PD game is considered to be an example of a social dilemma, and involves cognitive processes relevant to day to day social interaction (Kiesler et al, 1996). Additionally, the PD game typically evokes cooperative and conditional helping behaviors, while there is little to no literature regarding the relevance of the game rock, paper, scissors to these elements central to an investigation of social cognition (Nash, 1953; Kiesler et al, 1996). The current study will use the paradigm of Gallagher et al (2001) modified to use the prisoner's dilemma game, in hopes of isolating mechanisms specific to reasoning about the mental states of a social partner during a social dilemma.

The prisoner's dilemma game was originally used in game theory research, a branch of applied mathematics that studies strategic situations where players choose

different actions in an attempt to maximize their returns. The prisoner's dilemma game introduces a situation in which every rational move is to defect. In this game, as in many others, it is assumed that the primary concern of each individual is to maximize his own advantage. The PD game results in a Nash Equilibrium, or a situation in which neither player can change game outcome by switching strategy, over a sufficient number of iterated plays (Nash, 1953). Since in every circumstance playing defect is more rational than cooperating, all rational players will defect. Needless to say, this is not the behavior that typically emerges. Subjects generally exhibit levels of cooperation far higher than would be supported by an economical evaluation of costs and benefits (Singher and Fehr, 2005; Cosmides, 1989; Colman, 2003). Subjects typically begin by cooperating and continue to do so until one player defects, after which tit for tat (TFT) behavior emerges, and punishment for defecting remains the dominant play for the remainder of the game. Previous research on emotional reactions to this kind of situation reports that positive affective states are elicited by cooperation, but that even stronger positive affective states are elicited by punishment of a defector, indicating a strong bias to engage in reciprocal helping behavior (e.g. Fischbacher et al, 2005).

Subjects will play the iterated PD game on a computer in two conditions, a condition in which the subject is told they are playing against a computer program governed by a set of rules, and a condition in which the subject is told they are playing against the experimenter. In reality, participants will play the same computer program in both conditions. Each condition will consist of 5 lead in rule governed computer plays followed by a series of 10 randomly generated computer plays, and concluded by a lead-out sequence of 15 rule governed plays. This paradigm has been used in previous imaging studies with good results, and should effectively isolate the

target cognitions. Previous imaging studies using this paradigm and a variety of games used fMRI and PET to measure neural activity (Gallagher et al, 2001; McCabe et al, 2001). The previous studies were able to corroborate accounts of activation over a distributed network of regions during mental state representation; however, the increased temporal accuracy of ERP measures may tell us more about the nature of the mechanisms involved. If we can get a rough look at the spatio-temporal pattern of electrical activity during the task we may gain some insight into how the interactions of a distributed network of brain regions gives rise to ToM.

While many studies have implicated general regions, a model of how these areas interact and the specific mechanisms they subserve is lacking. These studies have consistently implicated the same network of regions, but cannot definitively provide any information about what this constellation means. By pulling together information from the research explored above, additional functional imaging data, and a careful analysis of what we do know about cognitive mechanisms necessary for the current task, some light may be shed on the issue. There are a number of good reasons as to why we know so little about the particulars of this kind of cognition, one of which is that this kind of strategic decision making process is a higher-order cognition that involves subtle interactions of mechanisms difficult to measure with current technology. I do not propose a solution, but rather suggest that we continue to chip away at the problem slowly with the tools available to us.

While fMRI and PET may be more appropriate than ERP for capturing these kinds of cognitions, which can be reflective and temporally distributed, the more temporally specific information provided by ERP could provide information about how different areas act in sequence or conjunction during specific processes. In the current study, patterns of activation will be recorded during three steps of the multiple

step process. ERP data will be recorded during play selection, perception of information regarding the previous move of the opponent, and consideration of the significance of the total game score for both players. During game play, subjects should evaluate options, form inferences and predictions about the probable future play of the opponent, monitor performance, maintain a changing model of the current situation in working memory, and choose plays based on the information these processes make available (Bechara et al, 2001; Joyce and Kutas, 2005; Hasselton and Buss, 2000). This process should be the same in both conditions, with the exception of the effects of the perceptual difference. Findings should reflect the effect of the perceived context on decision making processes.

The neurological correlates of the human mind-reading mechanism: An ERP investigation using the Prisoner's Dilemma Game

Abstract:

It has long been observed that people are capable of accurately and flexibly interpreting the behavior of those around them. The ability to infer information about the causal mental states of our social partners is central to this everyday act, and it is remarkable that we are able to do this effortlessly based on the ted information available to us. This study will attempt to isolate processes uniquely involved in decision making in social dilemma's. It is hypothesized that social and non-social information is processed and acted upon differently. In the iterated prisoner's dilemma game subjects interpret order of play information to predict future plays of their opponent. Subjects will play the iterated PD game in two conditions, one in which they believe they are playing a human opponent and one in which they believe they are playing a computer program. Behavioral and ERP data from this task will effectively isolate differences in cognition caused by this perceptual difference. I believe this subtle manipulation will be enough to cause activational differences in the broad network of brain regions thought to be involved in this kind of task, supporting previous work on the neural correlates of this task and providing more temporally specific information about this process(Kiesler et al, 1996; Gallagher et al, 2003). Future work in the fields of social cognition and decision making can elaborate on this preliminary investigation, which is designed to provide some precursory information for future ERP investigations of decision making and social cognition.

Introduction:

We appear to be able to infer the intentions, desires, and knowledge states of those around us and to use this information to predict and manipulate their behaviors and reactions. In cognitive science, this ability to reason about the states of other people's minds is attributed to a faculty referred to as "Theory of Mind" (ToM), which allows us to infer and predict the contents of people's minds through some system that inputs observable behaviors and interprets them (Leslie et al, 2004). This ability comes naturally, if somewhat imperfectly, to normal adults, and several explanations of this phenomenon have been put forward.

ToM, as it will be addressed in this study, will be considered to be some mechanism(s) used to form inferences and predictions about the intentional actions of

our social partners (Bloom and German, 2000; Leslie et al, 2004; Matsumoto et al, 1986). The hypothesized mechanism integrates information from diverse sources including perception, working memory, representations of past events, and general world knowledge to form inferences regarding the intentions and actions of our social partners. The ease with which we engage in behaviors reliant on this kind of potentially computation intensive process suggests that we exploit some heuristic mechanism or set of heuristic mechanisms which allows for cost effective reasoning about social information (Bechara et al., 2001; Tomasello et al., 2001). The hypothesis of the current study is that the human 'mind-reading' ability is the result of a specialized set of heuristic mechanisms that have evolved to help us successfully navigate our social world (Stevens and Hauser, 2004, Adolphs, 1999; Marsh, 2002). Such a system should effectively minimize the cognitive cost of decision and behavior monitoring processes related to social partner information (Colman, 2003; Ames, 2004; McCabe, 2000; Marsh, 2002).

To test differences between decision making processes engaged during situations involving one individual or situations involving a social partner the prisoner's dilemma game was used. This game has historically been considered to be a social dilemma; a mixed-motive two-person game with two choices—cooperate (be honest, truthful, or helpful) or defect (lie, cheat, or steal) (Howard, 1998; Macy and Flache, 2002; Nash, 1953). In the prisoner's dilemma game, pairs of individuals can either cooperate or defect in a given trial. If both partners choose to cooperate, both benefit moderately over the long term. If one partner cooperates and the other defects, the defector receives greater individual benefits, while the cooperator is harmed. If both players defect, both lose (Ostrom et al, 2000; Rachlin et al, 1998). Previous research on the prisoner's dilemma game has reported higher rates of cooperative

behavior when playing against a human opponent as opposed to a non-human opponent, and this different pattern of behavior suggests that different cognitive processes underlie the two tasks (Kiesler et al, 1996). In this study, participants played against a computer in two conditions of interest; one in which they were informed that they were playing against a computer program, and one in which they were informed that they were playing against a human. In previous functional imaging studies using variants of this paradigm, different patterns of activation have been reported between the two conditions (Gallagher, 2001; McCabe et al, 2001; Gallagher et al, 2003).

The task of the current study will involve a number of cognitive mechanisms, but the tight control provided by the paradigm ensures that any observed differences are a result of the treatment. Any differences in behavior, spatial distribution of activation, temporal distribution of activation, and component amplitude and latency can be attributed to differences in the perceptual experience of the subject (Gallagher et al, 2003; Gallagher, 2001). If some set of heuristics facilitates reasoning about social but not nonsocial dilemmas, reaction time and electrophysiological data should reflect a faster, less taxing reasoning process in the human opponent condition. Human opponent activation should be lower generally, as the use of heuristics such as reliance on cultural norms and stereotyping reduce the costs of accurately interpreting the relevant data (Marsh, 2002; Ames, 2004). Given the lack of a hypothesis regarding specific event related components, more triggers should have been included in order to more thoroughly explore the time course of the series of tasks. While data from this kind of study cannot be strongly conclusive, this kind of study is not inappropriate given the lack of data regarding the time-course of these processes.

A large body of data implicates a network of areas in ToM, decision making,

stochastic learning, and memory function; all tasks relevant to the current study (Haruno et al, 2004; Gallagher, 2002; Bechara et al, 1997). Most importantly, BA 32, an area of the anterior para-cingulate cortex bordering the anterior cingulate-cortex, has been consistently correlated with relevant ToM events (Uddin et al, 2004; Frith and Frith, 2003).

Activation of areas near BA 32 and the anterior-paracingulate cortex is expected, although the lack of temporal data makes it inappropriate to speculate as to which processes during this multiple-step task will elicit differences. Some previous data does indicate that activity may be lateralization right during the human opponent condition, however primarily bi-lateral activation is expected (Gallagher et al, 2003; McCabe et al, 2001).

As the task in both conditions is identical, both target conditions should be reliant on similar kinds of cognitive mechanisms; however some thing or set of things causes us to treat these two identical situations differently (Kiesler et al, 1996). Popular models of decision making describe a system in which current perceptual and contextual information, stored information about similar past situations, and inferred information about possible future states of the world interact to inform decision making (Damasio, 1994; Verdejo-Garcia et al, 2006). It is thought that socially relevant information is attended to and stored preferentially, and it would be interesting to know how informational content influences knowledge structure and behavior. Differences in components associated with the storage and manipulation of this kind of representation during option evaluation or prediction accuracy feedback steps would be particularly interesting to find.

The behavioral differences elicited by these conditions have been robustly exhibited and replicated, and higher rates of cooperation in the human opponent

condition are expected (McCabe et al, 2001, Shamay-Tsoori et al, 2005; Kiesler et al, 1996). As little to no information regarding the time-course of this kind of cognition is currently available, the hypothesis regarding specific ERP components is an open-ended one. Although not the most stringent of scientific techniques, the aim of this study is largely exploratory and the primary goal is to provide some provisional clues to aid in the development of more sophisticated methods of investigating this kind of decision making.

Methods:

Subjects for the study were 32 undergraduate students at Hampshire college, 17 male and 15 female. Subjects were all right handed and had normal or corrected to normal vision. Subjects were recruited through posters put up around campus and received $10 for their participation in the study.

Participants were told by the experimenter that they would be playing the two person prisoner's dilemma game in three different conditions. The first condition would be a practice round, in which both the subject and their opponent chose plays at random. In the following two conditions, the subject would be playing against either a computer program that chose a play based on some rule regarding the previous play of the subject, or against the experimenter who would be playing from the main lab area next door. Both target conditions, the computer opponent and the human opponent condition, occurred first in half of all trials, to ensure no effects of practice or conditioning.

The premise of the game was explained to the participants as follows:

"You and a partner are planning a crime, and are apprehended by the police. You are both brought to the police station for questioning but are brought to separate

interrogation rooms. You are both offered the same deal, which is that if you give your partner up, but your partner refuses to give you up, you will walk and your partner will go to jail. If neither of you talks, you will both walk, but if both of you talk, you will both go to jail. In the game, you will make this decision by either choosing the play 'cooperate', which means cooperating with your partner, or 'defect', which means giving your partner up to the police."

The participants were then told that they would see two boxes appear on the computer monitor, labeled 'cooperate' and 'defect', and would select their play with the mouse. After play selection, they would be informed of what their partner's play was and what the eventual outcome of the round was before playing again. Each condition lasted approximately 5 minutes, relative to how quickly the subject selected plays, and consisted of making 30 plays which determined the final score of both the subject and their opponent. Stimuli were presented on a computer monitor comfortably within the subject's field of vision. While subjects were told that the last two conditions were different, they were in fact identical and all behavioral and electrophysiological differences in the data must be a result of a perceptual difference between the two conditions. The plays of the computer during the first condition were randomly generated, but the two conditions of interest conditions consisted of 5 lead in trials in which the computer cooperated, followed by 10 trials in which the computer selects the same response as the subjects last play, followed by 10 random computer plays, followed by 5 trials in which the computer alternately cooperated or chose the same play as the subject's last play. This pattern is roughly modeled after that used in Gallagher et al (2003), but it was necessary to make some changes as the game employed in the previous research, rock, paper scissors, has three possible plays while the PD game has only two.

ERP data was recorded at three points; at the screen where subjects were informed of their opponent's last play; at the screen where subjects were informed of the total game score before choosing their next play, and at the screen where subjects were prompted to choose a play by selecting their move with the mouse. Behavioral data was recorded to determine if there was a difference in response patterns and reaction time between the two conditions, and a brief survey consisting of three open-ended questions asking participants to describe their experience during each of the conditions was administered to gage whether the perceptual difference of interest occurred.

Electroencephalogram (EEG) data was recorded with a 32 channel tin Electro-cap. Electrodes were referenced to the mastoids and impedances were kept below 5k ohms for all participants. Data was recorded using a Synampse Amplifier and Scan software and was digitized at the rate of 500 Hz using a bandpass filter of .1 to 30 Hz. ERP data was recorded at three points during the course of a trial, and several peaks were analyzed from each of these three stages. As no previous ERP studies of this kind were available, peak windows were selected based on averaged files and pilot data. A visual inspection of grand mean files revealed a number of peaks at each trigger. At the screen informing participants of their opponent’s last move, an N250 and a P300 were identified; at the screen displaying total game score, an N60, N200 and P400 were identified; and at the screen prompting subjects to choose between the two plays, an N100, P200, P300 and N450 were identified. Peak amplitude and peak latency were analyzed using repeated measures multivariate analysis of variance.

Results:

Data from 17 of the 32 subjects was analyzed. Data was excluded from

analysis if fewer than 20 artifact free trials were available for analysis, and in the case of three subjects based on information provided on the background questionnaire that excluded them from analysis. The study elicited a high percentage of eye artifacts, and the visual setup of the experiment should be revised to reduce this problem.

Electrophysiological data was reduced by calculating the latency to peak, and peak amplitudes were calculated across the latency windows specified above. Data from the three different stimulus triggers will be discussed separately, as each trigger should be representative of different cognitive events.

Behavioral data:

A repeated measures ANOVA revealed that significantly different response patterns were elicited in the two conditions ($F(1,16)=3.71$ $p<.01$), and subjects cooperated significantly more in the human opponent condition. Additionally, the surveys administered after the experiment indicate that the subjective experience of the two target conditions was in fact different, and no participants indicated that they had been aware of the ruse. This supports the validity of the experimental design, however it does not provide any definitive evidence regarding underlying cognitions of interest. A repeated measures ANOVA revealed no significant differences in reaction times between conditions, as the nature of the task made this data far too variable. Another study with tighter control over subject response time would be very interesting, as reaction time data could be very relevant to the investigation of proposed heuristic mechanisms.

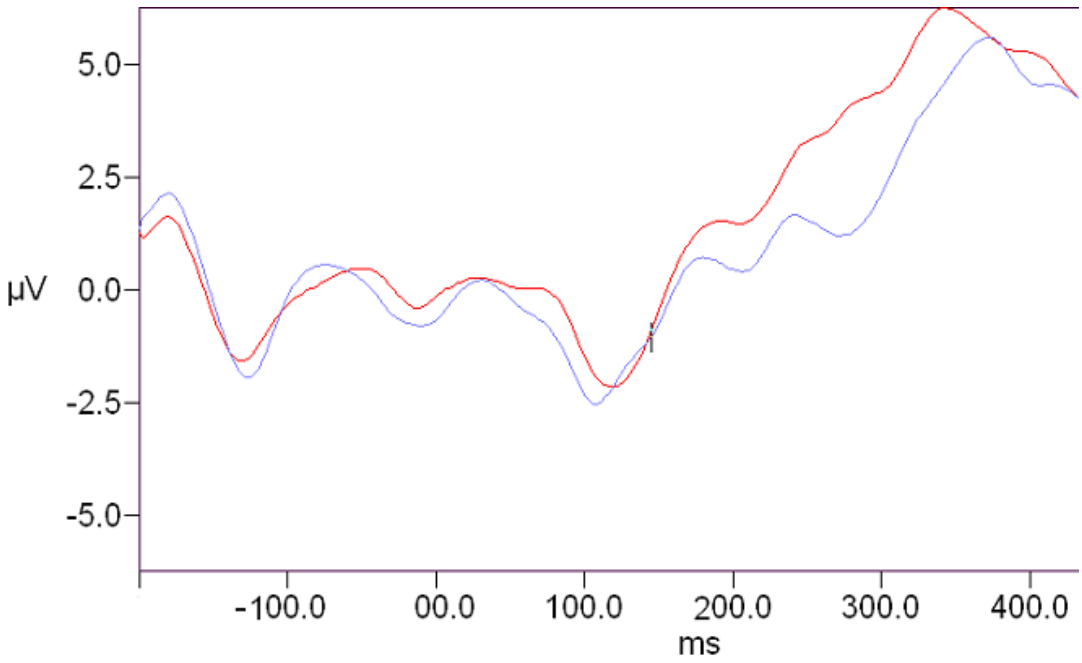

Figure A. The N250 and P300 components elicited by opponent play information recorded at CP3. Red wave represents data from the computer opponent condition and the blue line represents data from the human opponent condition

Reaction to order of play information:

The most interesting differences measured in this study reflect differences in information processing when presented with information about the play of the opponent. A screen informing the subject of their opponents last play was presented, and activity elicited by this stimulus is thought to reflect a performance evaluation and context updating process, in which the subject compares the predicted outcome to the actual outcome of the play and is incorporated into a mental model. ERP activity elicited by this screen was significantly different between conditions. A negative component 250 ms after stimulus onset was significantly larger in the human opponent condition in a large left hemisphere area strongest at CP3 ($f(1,16)=18.811$, $p<.001$).

A later positive component at around 400ms after stimulus onset was

significantly smaller in the human condition in an even broader area centered at TP7 ($f(1,17)=9.599$, $p=.007$). The pattern of activation observed for this component indicates that it is an example of the classical P300 component, which can occur from 300 to 900ms post stimulus and has a characteristic pattern of activation originating in the temporo-parietal junction similar to the pattern of activation seen here (Patel and Azzam, 2005).

Reaction to total game score report:

The second trigger coincided with a screen reporting the total game score of both players, and activity elicited by this screen is assumed to be representative of some updating of mental models and reflection on order of play information. The data from this trigger is of the least interest, and as this task is more reflective it is unclear which cognitive processes are reflected by the ERP data.

The activity elicited by this screen does reflect significant differences between conditions. A negative component at around 60ms after stimulus onset was significantly larger in the computer opponent condition in a predominantly right fronto-temporal region with the strongest differences at electrode T8 ($f(1,17)=160.099$, $p=.006$). An N200 component was significantly larger in the computer opponent condition, in a very broad but predominantly right hemispheric area centered around F4 ($f(1,17)=10.065$, $p=.006$).

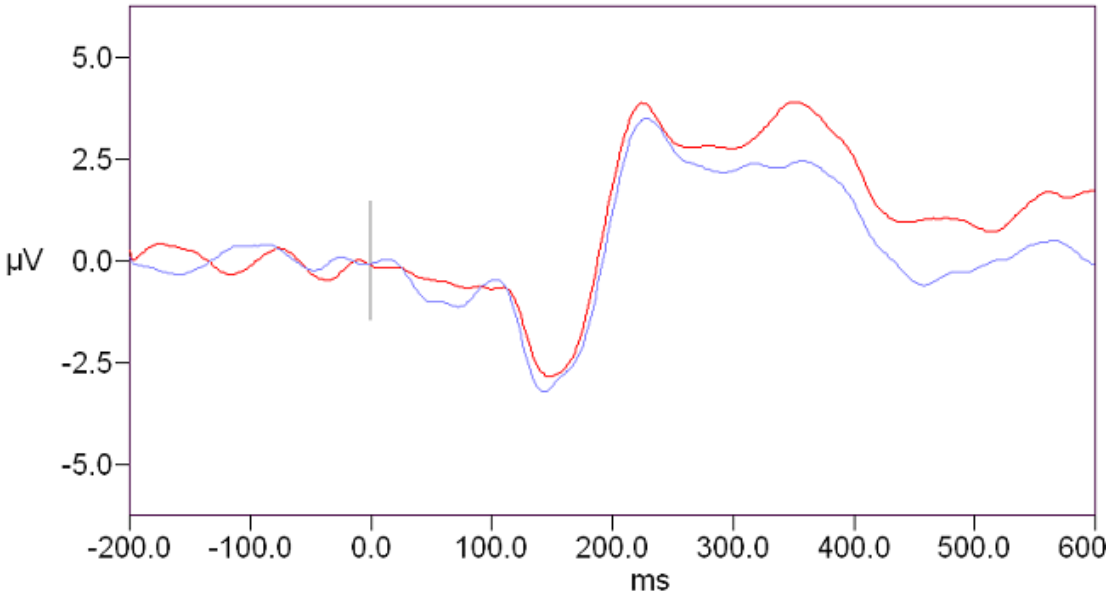

Figure B. N100, P200, P300, and N450 components elicited by the prompt to choose between plays recorded at PZ. The red line represents data from the computer opponent condition and the blue line represents data from the human opponent condition

Prompt to choose between plays:

The participants were prompted to choose to cooperate or defect by the appearance of a screen displaying the two options. Activity evoked by this stimulus is thought to be representative of some process of option evaluation, which could include reasoning, accessing long term memory, updating mental models, and processes involved in conflict resolution. The ERP activity elicited by this screen showed significant differences in the amplitude of the P300 and N450 component.

The P300 component was significantly higher in the computer opponent condition at FP1 ($f(1,17)=3.989$, $p=.04$), and an analysis of activity over an averaged area including FP1, FC3 and F3 (left frontal electrode sites) compared to the activity in FP2, FC4, and F4 (the corresponding right frontal electrode sites) indicated that a higher P300 in the left-frontal lobe was significantly larger across conditions in addition to being significantly larger in the computer opponent condition than in the

human opponent condition($f(1,17)=6.92$, $p=.018$). However, the data from this trigger is marginal and can do no more than corroborate previous evidence that the left frontal lobe plays an important role during option evaluation.

The later N450 component elicited by this stimulus was significantly larger in the human opponent condition in a central parietal region centered at PZ ($f(1,17)=9.643$, $p=.007$). The N450 component, which originates in the ACC, is traditionally considered to be an indice of response conflict and may reflect conflict resolution processes engaged by this task (Swift and Turken, 2002).

Discussion:

N250:

The screen reporting the play of the participant's opponent elicited a significantly stronger N250 component in the left hemisphere, particularly in the temporo-parietal region. While the N200 component can occur within a similar time frame, the location of the component observed in this study is more consistent with the N250 (Riba et al, 2005). Previous research has reported findings correlating amplitude of the N250 to the salience of specific kinds of social information, such as perception of in-group vs. out-group faces, making this finding particularly interesting to researchers of social cognition (Ito et al, 2004). The amplitude of the N250 has been positively correlated with explicit recognition after the interval of a week, and the amplitude of the N250 is sensitive to stimulus rates, indicating that this component may reflect some process involved in stochastic learning and memory (Riba et al, 2005; Joyce and Kutas, 2005). The N250 may be an indice of the strength of a learning signal in stochastic learning processes. It has been postulated that this component may be correlated to stimulation of the locus coerulus, which has

noradrenergic connections to the ACC, an area bordering BA 32 and of interest to researchers of decision making and ToM (Riba et al, 2005). The observed difference in N250 amplitude could indicate a stronger learning signal elicited by social partner plays as opposed to the plays of a computer program. These results could indicate that there is preferential learning about social partner information.

P300 at 400ms after stimulus onset:

The N250 is followed by a positive component 400ms after stimulus onset, in a pattern of activation consistent with the classical P300 (P3) component (Patel and Azzam, 2005). This component is significantly different at TP7, and is larger in the computer opponent condition. The P3 is considered to reflect broad recognition and memory updating processes, specifically the degree of match/mismatch between the current stimulus and a consciously maintained memory trace (Patel and Azzam, 2005). Specifically, the P3 is thought to be an indice of "context updating," a cognitive routine supporting the formulation of an internal model in which a specific stimulus can be evaluated (Costa et al, 2000; Patel and Azzam, 2005; Falkstein et al, 2000; Friedman and Johnson, Jr., 2000). This popular theory is consistent with the events postulated for this task, in which information about the opponent's play is evaluated and integrated into a mental model of the current situation. The amplitude of the P3 is thought to be an index of stimulus salience, and a lower amplitude in the human opponent condition could indicate that information relevant to a social dilemma is more easily accessible than information relevant to the same task in a nonsocial context, or that mental models used to mediate behavior in social tasks require less intensive updating than models of equivalent nonsocial tasks. The interpretation of this finding is speculative, but the pattern of activation strongly

suggests that this component can be considered to be an example of the classical N3, and that some process of context updating does occur at this stage.

N60:

When game score totals for both parties were displayed, a negative component 60ms after stimulus onset was significantly larger in the computer opponent condition in a predominantly right fronto-temporal region strongest at electrode T8. The N60 has been correlated with response error, and is referred to as the error related negativity, or ERN, in the literature (Gehring et al, 1993). The ERN is traditionally associated with a mismatch signal thought to be an indice of the difference between an actual response and the correct response, and is thought by some to be an indication of the process of comparing actual and correct response (Falkstein et al, 2000). However, this slide does not follow any response and this component cannot be representative of a response monitoring process. It is possible that this component is representative of some comparison between expected and experienced outcome, although there is little to substantiate this theory. Further investigation of the nature of the N60 and it's significance to this task could be very interesting.

N2c:

A widespread right hemispheric area exhibited a significantly stronger N200 component in the computer opponent condition, especially in the right frontal region. The N2 is typically considered to indicate a deviation in form or context, and the location of the component indicates that it may be an example of the N2 sub-component the N2c, historically elicited by incompatible stimuli in the flanker task (Linden et al, 1999; Patel and Azzam, 2005). A larger N2c may indicate greater

cognitive effect of the stimulus in the computer opponent condition, consistent with a view that some heuristic mechanism minimizes the cognitive effort in the human opponent condition. It is ambiguous what mental tasks are being engaged in at this slide, and the ERP data recorded during it must be regarded with that in mind. This finding is reported here because despite the ambiguity of the task, the areas activated have historically been linked to ToM; the two components occur near the right anterior cingulate and anterior para-cingulate cortex.

P300:

Data elicited by the prompt screen, during which subjects are presented with their two choices, was not strongly significant and can only suggest that the left frontal cortex plays some role in option evaluation processes. The left frontal P300, similar to the more parietal P3 component observed earlier, is thought to indicate some process of updating or refreshing the contents of working memory, allowing us to maintain an ongoing representation of our world (Baharami et al, 1997; Linden et al, 1999). The frontal P3, sometimes referred to as the P3a, is thought to reflect more passive comparisons than the more parietal classical P3 component evoked by the screen informing participants of their opponent's last play (Patel and Azzam, 2005). This is consistent with the different natures of the two tasks, as the decision prompt slide should evoke more passive or reflective updating of mental models.

This component is also thought to be important to decision making, particularly inhibitory control and accurate risk assessment, and a lower frontal P300 has been associated with a prevalence for poor lifestyle decision patterns such as alcoholism and drug abuse (Miller et al, 1999; Bolla et al, 2003; Haybeych et al, 2005; Costa et al, 2000; Justus et al, 2001). The smaller amplitude of the P300 during

the human opponent condition does support the theory that processing information relevant to social dilemmas relies on heuristic mechanisms that minimize cognitive effort. This conclusion is highly speculative and based on marginally significant data, and further investigation is in order before any such conclusions can be drawn.

N450:

The N450 component that follows the P300 showed a stronger pattern of significance, and was larger in the human opponent condition. The peak amplitude of the N450 has been positively correlated to degree of response conflict (Swick and Turken, 2002). This component is thought to originate from a source in the anterior cingulate cortex, an area consistently associated with both ToM and conflict resolution and of specific interest in this study (Liotti et al, 2000). The component is typically strongest around electrodes CZ, CPZ and PZ, and here is strongest at electrode PZ, indicating that this is a clear example of the N450 component. The N450 observed in this study was significantly larger in the human opponent condition, indicating a greater degree of response conflict in this condition. If we are predisposed to engage in pro-social behavior in social dilemmas, the introduction of a human opponent into a mixed-motive game should result in a greater motivation to cooperate, eliciting greater response conflict and a larger N450 component (Colman, 2003). These findings suggest that there may be significantly less motivation to cooperate in the absence of a real social partner.

The prompt to choose a play is assumed to evoke some process of option evaluation and conflict resolution, and the late arriving N450 may reflect some stage of this conflict resolution process; however as reaction times were highly variable it is difficult to conclude anything concrete about the time course of this particular stage of

play. Future research on the effect of context (social vs. nonsocial) on option evaluation and conflict resolution should devise a more appropriate task, in which more constraints are placed over the time course of subject responding.

Human opponent condition Computer opponent condition

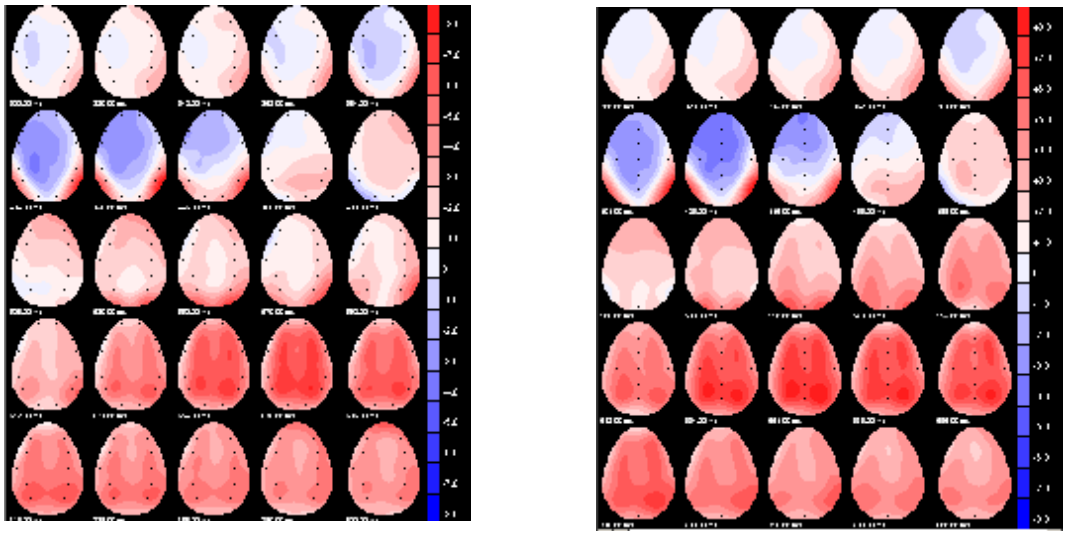

Figure C. 2D map of activation in response to information about opponent's last play

Conclusion:

The three triggers examined in this study represent three distinct neural events relevant to high-order cognitive processes that occur during game play. The specific components elicited by each stage of the task describe two processes of attention, learning, option evaluation and conflict resolution that vary in degree of engagement and not in their underlying mechanism.

All interpretation of the data of the current study must be considered to be speculative in light of the fact that the study posited no directional hypothesis regarding specific components. While the data can be considered to describe significant differences between the two conditions, it cannot be considered to significantly support any specific hypothesis. These findings should be considered

primarily as a source of temporal information for future research regarding the cognitions elicited by the task, and hopefully support for further research on decision heuristics specialized for social dilemmas. Without further research into the real significance of each of these components to the proposed mechanism, no satisfactory conclusion can be reached. However, this data can support a strong difference between conditions, and does indicate a significant cognitive difference between the two tasks.

Preferential processing of information relevant to our social and emotional life has long been hypothesized by researchers of emotion and memory, and this tendency is clearly demonstrated here (Norris et al, 2000; Becharia et al, 2004). The findings of the current study could indicate that a stronger learning signal is elicited by social and emotional information, which aids in stronger memory encoding of that type of information. This could be one of many mechanisms preferentially engaged when encountering social information, allowing for more spontaneous interpretation of social information than for other kinds of information. The pattern of stronger attentional and learning signals coupled with weaker cognitive search and context updating signals in the human opponent condition indicates that just such a bias could account for the differences in behavior elicited by the two tasks.

A more thorough and complete account of what heuristics may facilitate social performance, how these heuristics develop, and the degree of their specialization is an interesting and important area for future research. There is additionally a need for research regarding the interaction of working memory, error monitoring, and retrieval and updating of relevant information in long term memory in decision making; both social and otherwise. Facilitated storage of and access to socially relevant information could account for some of the cognitive shortcuts involved in social

decision making, as multiple sources of relevant information could be easily called upon to effectively constrain decision spaces (Bolla et al, 2003; Ames, 2004; Howe and Courage, 1997), however the findings of the current study indicate that this is not the only heuristic mechanism engaged specifically in social contexts.

A set of domain specific decision heuristics may do much to constrain the decision space in the instance of a social dilemma, but it also seems likely that humans are predisposed to engage in certain patterns of behavior, such as engaging in conditional helping.The findings regarding the N450 implicate at least one additional heuristic mechanism at work during social reasoning tasks. Specifically, this finding could support theories consistent with evolutionary game theory, which postulates that successful social exchange behavior relies on a predisposition to engage in conditional helping, or tit for tat, behavior in social situations. While this study demonstrated only that there may be greater motivation to cooperate when engaged with a human opponent as opposed to a computer opponent, an effect long documented in the behavioral literature (e.g. Howe and Courage, 1997), the difference in amplitude across conditions in this late arriving component may shed some new light on how an innate pre-disposition plays a role in the relevant option evaluation and decision making processes. Above all, this finding provides support for future focused investigations of the degree to which this kind of pre-disposition and economic reasoning processes interact to inform behavior.

While my data certainly doesn't tell as grand a story as I do, it doesn't disconfirm my story either. And while it doesn't explain how mental representations affect behavior, it does suggest some mechanisms for acquiring, storing, and manipulating mental representations in a way that could appropriately mediate behavior. Further explication of the interactions between the posited network of brain

areas is certainly needed, and is a worthy although daunting task for future researchers. ERP studies investigating the cognitive events that are the component parts of the multiple step behavior described above can benefit from the preliminary information provided by the current study. Specifically: error monitoring, attention, memory encoding, memory retrieval, and the interaction between working and long term memory appear to be important parts of the complex process that results in what we call Theory of Mind. This process of manipulating different kinds of mental representations to guide behavior is one of the most interesting problems for cognitive scientists.

One of the most interesting aspects of this issue is the way in which the domain of a particular stimulus can influence the way it is represented and manipulated in the brain. The findings of this study clearly indicate that this is in fact the case, and this issue specifically will certainly be the focus of future research efforts. The literature reviewed above offers a good case for shifting the way we think about mental state representations in social cognition. We need to consider the nature of the measures traditionally employed in ToM, and consider the data generated by these measures accordingly. While it may seem that the broad spectrum of ToM tasks all share certain qualities, closer inspection will reveal that they could indeed be very disparate kinds of cognitive tasks associated with a number of possibly confounding variables. Researchers of ToM must begin asking more focused research questions, and push to disambiguate important fundamental issues in the field. Tasks that require active participation in social interactions requiring a ToM are an important tool for investigating natural ToM mechanisms, and provide a promising avenue of research for those specifically interested in human social cognition. It could be argued that active reasoning about the mental states of others is the most

important task subserved by ToM.

Researchers of decision making, emotion, evolutionary game theory, linguistics, and a wealth of other disciplines stand to gain a lot by joining forces to better understand some of the most important aspects of human life and cognition. Recent interest in social cognition and all that it entails promises to promote these kinds of important cross-pollinations, and we should look forward to seeing new directions and perspectives on these issues emerge in the upcoming years.

Bibliography

Adolphs, R., (1999) Social Cognition and the human brain, Trends in Cognitive Science, Vol. 3, No. 12

Adolphs, R., (2001) The neurobiology of social cognition, Cognitive Neuroscience

Ahn, Ostrom, Schmidt, Shupp, Walker, (2000) Cooperation in Prisoner's Dilemma Games: Fear, Greed, and History of Play, Public Choice, Vol. 106, No's 1-2, pp. 137-155

Allison, Puce, McCarthy, (2000) Social Perception from visual cues: The role of the STS region, Trends in Cognitive Sciences, Vol. 4 No. 7

Ames, Daniel R.,(2004) Inside the Mind Reader's Tool Kit: Projection and Stereotyping in Mental State Inference Journal of Personality and Social Psychology Vol.87, No.3, pp340-353

Atherton, Zhuang, Bart, Hu, He, (2003) A functional MRI study of high-level cognition: The game of chess, Cognitive Brain Research, Vol. 16, pp. 26-31

Barnaby, Marsh, (2002) Heuristics as Social Tools, New Ideas in Psychology, Vol. 20, pp. 49-57

Barrett, Rugg, (1990) Event Related Potentials and the semantic matching of pictures, Brain and Cognition, Vol. 14, No. 2, pp. 201-212

Bates, Patel, Liddle, (2005) External Behavior Monitoring Mirrors Internal Behavior Monitoring, Journal of Psychophysiology, Vol.19, No. 4, pp. 281-288

Bechara, Damasio, Damasio, (2000) Emotion, Decision Making, and the orbitofrontal cortex, Cerebreal Cortex, Vol.10, pp. 295-307

Beuzeron-Mangina, Mangina, (2000) Event related potentials to memory workload and 'Analytical-specific perception' in patients with early Alzheimers, International Journal of Psychophysiology, Vol.37, No. 1, pp. 55-69

Blakemore, Boyer, (2003) The detection of contingency and animacy from simple animation, Cerebreal Cortex, Vol.13, No.8, pp.837-844

Bloom, German, (2000) Two reasons to abandon the false belief task as a test of Theory of Mind, Cognition, Vol. 77, pp. 25-31

Burnette, Mundy, Meyer, Sutton, Vaughn, Charak, (2005) Weak Central Coherence and it's relationship to Theory of Mind and anxiety in Autism, Journal of Autism and Developmental Disorders, Vol.35, No.1, pp.63-73

Carston, (2000) The relationship between generative grammar and relevance theoretic pragmatics, Language and Communication, Vol. 20, pp. 87-103

Castelli, Frith, Happe, Frith, (2002) Autism, Asperger syndrome, and brain mechanisms for the attribution of mental states to animated shapes, Brain, Vol.125, No. 8, pp. 1839-1849

Colman, Andrew, (2003) Cooperation, psychological game theory, and limitations of rationality in social interaction Behavioral and Brain Sciences Vol.26, pp. 139-198

Cosmides, (1989) The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task., Cognition, Vol. 31, No. 3, pp.187-276

Cosmides, Tooby, (2000) Consider the source: The evolution of adaptations for decoupling and metarepresenting, from Metarepresentation, edited by Dan Sperber

Costa, Bauer, Kuperman, Porjesz, O'Connor, Hesselbrock, (2000) Frontal P300 decrements, alcohol dependence, and anti-social personality disorder, Biological Psychiatry, Vol.47, No. 12 , pp. 1064-1071

Csibra, Gergely,(1999) Goal attribution without agency cues, Cognition, Vol.72, pp.237-267

Damasio, Descartes error: emotion, reason, and the human brain, (1994), Grossett/Putnam, New York, NY

Darwin, (1877) A Biological Sketch of an Infant, Mind, Vol. 2, No. 7, pp.285-294

Dasser, Ulbaek, Premack, (1989)The Perception of Intention Science, Vol. 243, No. 4889 365-367

Decety, Sommerville, (2003) Shared representations between the self and other: A social cognitive neuroscience view, Trends in Cognitive Sciences, Vol.7, No.12

de Villiers, Bates, Volterra, (1984) Gestural Communication in Deaf Children: The Effects and Noneffects of Parental Input on Early Language Development Monographs of the Society for Research in Child Development, Vol. 49, No. 3/4, pp. 1-151

Dennett, D. The Intentional Stance, (1987), MIT Press, Boston, Ma, USA

Dunn, Dunn, Lanquist, Andrews, (1998) The relation of ERP components to complex memory processing, Brain and Cognition, Vol. 36, No. 3, pp. 355-376

Egan, (1995) Folk Psychology and Cognitive Architecture**,** Philosophy of Science, Vol. 62, No. 2, pp. 179-196

Ehlis, Herman, Bernhard, Fallgatter, (2005) Monitoring of Internal and External Error Signals, Journal of Psychophysiology, Vol.19, No. 4

Fadiga, Fogassi, Gallese, Rizzolatti, (2000) Visuomotor neurons: ambiguity of the discharge or 'motor' perception?, International Journal of Psychophysiology, Vol. 35, No. 2, pp. 165-177

Falkstein, Hoorman, Christ, Hohnsbein, (2000) ERP components on reaction errors and their functional significance: a tutorial, Biological Psychology, Vol. 51, pp. 87-107

Ferstle, von Cramon (2002) What does the frontomedian cortex contribute to language processing: Choherence or Theory of Mind?, NeuroImage, Vol.17, No. 3, pp. 1599-1612

Field, H., (1978) Mental Representation, Springer Netherlands, Vol. 13, No. 1, pp. 9-61

Friedman, Johnson Jr., (2000) Event Related Potential studies of memory encoding and retrieval: A selective review, Neuroimaging of Memory, Vol. 51, No.1, pp. 6-28

Gallagher, Frith (2003) Functional Imaging of Theory of Mind, Trends in Cognitive Science, Vol.7, No. 2

Gallagher, Jack, Roepstorff, Frith, (2002) Imaging the Intentional Stance in a Competitive Game NeuroImage, Vol.16, pp. 814-821

Gallese, Goldman (1998) Mirror neurons and the Simulation Theory of Mindreading, Trends in Cognitive Science, Vol.2, No. 12

Gallese, (2001) The "shared manifold" hypothesis: from mirror neurons to empathy, The Journal of Consciouness

Gallup, (1970) Chimpanzees: Self-recognition, Science, Vol. 167, pp. 86-87

German, (2004) Neural Correlates of detecting pretense: automatic engagement of the intentional stance, Journal of Cognitive Neuroscience, Vol.16, pp. 1805-1817

Gerrans, P. (2002) The Theory of Mind module in evolutionary psychology, Journal of Biology and Philosophy, Vol.17, No. 3, pp. 305-321

Gopnik, Astington, (1988) Children's understanding of representational change and it's relation to the understanding of false belief and the Appearance/Reality distinction, Child Development, Vol. 59, No.1, pp. 26-37

Grezes, Frith, Passingham, (2004) Brain Mechanisms for Inferring Deceipt in the

Actions of Others, The Journal of Neuroscience, Vol. 24, No. 24, pp. 5500-5505

Hanna, Meltzoff, (1993) Peer Imitation by Toddlers in Laboratory, Home, and Day-Care Contexts: Implications for Social Learning and Memory, Developmental Psychology, Vol. 29, pp.701-710

Haruno, Kuroda, Doya, (2004) A Neural Correlate of Reward-Based Behavioral Learning in Caudate Nucleus: A functional Magnetic Resonance Imaging Study of a Stochastic Decision Task, The Journal of Neuroscience, Vol. 24, No.7, pp.1660-1665

Haselton, Buss, (2000) Error Management Theory: A new perspective on biases in cross-sex mind reading, Journal of Personality and Social Psychology, Vol. 78, No.1, pp. 81-91

Hastie, R., Dawes, R., Rational Choice in an Uncertain World, (2001) Sage Publications, Thousand Oaks, CA, USA

Hoffman, McCabe, Smith, (1999) Social distance and other-regarding behavior in dictator games, The American Economic Review, Vol. 89, No. 1, pp. 340-341

Howard, (1998) Cooperation in the Prisoner's Dilemma Game, Theory and Decision, Vol.24, No. 3, pp.203-213

Howe, Courage, (1997) The emergence and early development of autobiographical memory, Psychological Review, Vol. 104, No. 2, pp. 499-523

Hugdal, Northby, (1991) Hemisphere Differences in Conditional Learning, Cortex, Vol. 27, No. 4, pp. 557-570

Iacombi, Molnar-Szakcs, Gallese, Rizzolatti, (2005) Grasping the intentions of others with one's own mirror neuron system, Vol. 3, No. 3

Ito, Thompson, Caccioppo, (2004) Tracking the timecourse of Social Perception: The effects of racial cues on event related brain potentials, PSPB, Vol.30, No.10, pp.1267-1280

Jeannerod, Marc Who needs emotions? The brain meets the robot, (2005) Oxford University Press, New York, NY

Joyce, Kutas, (2005) Event related potential correlates of long-term memory for briefly presented faces, Journal of Cognitive Neuroscience, Vol.17, pp. 757-767

Justus, Finn, Steinmetz, (2001) P300, Disinhibited Personality, and Early onset Alcohol problems, Alcoholism: Clinical and Experimental Research, Vol. 25, No.10, pp. 1457-1466

Kampe, Frith, and Frith, (2003) "Hey John":Signals Conveying Communicative Intention toward the Self Activate Brain Regions Associated with "Mentalizing," Regardless of Modality The Journal of Neuroscience Vol. 23, No.12, pp. 5258-5263

Kane, Picton, Moscovitch, Winocur, (2000), Event related potentials during conscious and automatic memory retrieval, Cognitive Brain Research, Vol.10, pp. 19-35

Kay, A., Ross, L., (2003) The Perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the Prisoner's Dilemma Journal of Experimental Social Psychology Vol. 39 pp 634-643

Keenan, J. Gallup, G., Falk, D. The Face in the Mirror, 2004, HarperCollins Publishers, Inc., New York, NY

Keltner, Kring, (1998) Emotion, Social Function, and Psychopathology, Review of General Psychology, Vol. 2, No. 3, pp. 320-342

Ketelaar, T., Wing Tung Au, (2003) The effects of feelings of guilt on the behavior of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction Cognition and Emotion Vol.17, No. 3, pp 429-453

Kiesler, Waters, and Sproull, (1996) A Prisoner's Dilemma Experiment on Cooperation With People and Human-Like Computers Journal of Personality and Social Psychology , Vol. 70, No. 1, pp47-65

Kotchoubey, Jordan, Grozinger, Wesphal, Korhuber, (1996)Event related potentials in a varied-set memory search task, Psychophysiology, Vol. 33, No. 5, pp. 530-540

Langton, Vicki, (2000) You must see the point: Automatic Processing of Cues to the Directions of Social Attention, Journal of Experimental Psychology, Vol. 26, No.2 pp.747-757

Leslie et al, (2004) Core Mechanisms in Theory of Mind, Trends in Cognitive Science, Vol. 8, No.12

Leslie, (1994) Pretending and believing: Issues in the theory of TOMM, Cognition, Vol. 50, pp. 211-238

Leslie, (1995) A theory of agency, Causal cognition: a multidisciplinary debate

Liotti, Woldorff, Perez III, Mayberg, (2000) An ERP study of the temporal course of the stroop color-word interference effect, Neuropsychologia, Vol. 38, pp.701-711

Linden, Prvulovic, Formisano, Vollinger, (1999) The functional neuroanatomy of target detection: An fMRI study of visual and auditory oddball tasks, Cerebral Cortex, Vol. 9, pp. 815-823

Macy, Flache, (2002) Learning dynamics and social dilemmas, PNAS, Vol.99, No.53

Marsh, (2002) Heuristics as Social Tools, New Ideas in Psychology, Vol. 20, pp. 49-57

Matsumoto, Haan, Yabrove, Theororou, (1986) Preschoolers' moral actions and

emotions in the Prisoner's Dilemma, Developmental Psychology, Vol. 22, No. 5, pp. 663-670

McCabe, Houser, Ryan, Smith, Trouard, (2001) A functional imaging study of cooperation in two-person reciprocal exchange, PNAS, Vol. 98, No. 20

McCabe, Smith, and LePore, (2000) Intentionality detection and "mindreading":Why does game form matter? PNAS Vol. 97, No. 8, pp. 4404-4409

McCoy and Platt, (2005) Expectations and outcomes: decision making in the primate brain, Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, Vol. 191, No. 3, pp. 201-211

Meltzoff, (1999) The Origins of Theory of Mind, Cognition, and Communication, Journal of Communicative Disorders, Vol. 32, No. 4, pp.251-269

Nash, (1953) Two person cooperative games, Econometria, Vol. 21, No. 1, pp. 128-140

Nowak, Sigmund, (1993) A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game, Nature, Vol. 364, pp.56-58

O'Laughlin, Malle, (2002) How People Explain Actions Performed by Groups and Individuals Journal of Personality and Social Psychology, Vol.82, No.1, pp 33-48

Patel, Azzam, (2005) Characterization of the N2 and P3: Selected studies of the event related potential, International Journal of Medical Science, Vol. 2, No. 4, pp. 147-154

Paulus, Hozack, Zhuang, Zauscher, McDowell, (2001) Prefrontal, parietal, and temporal cortex networks underlie decision-making in the presence of uncertainty, NeuroImage, Vol. 13, pp. 91-100

Pelphrey, Morris, McCarthy, (2004) Grasping the Intentionality of Others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception, Journal of Cognitive Neuroscience, Vol. 16, No. 10, pp. 1706-1716

Rapin, Katzman, (1998) Neurobiology of autism, Annual Review of Neurology, Vol. 43, No. 1, pp. 7-14

Riba, Rodriguez-Furnells, Morte, (2005) Noradrinergic stimulation enhances human action monitoring, Journal of Neuroscience, Vol.17, No. 25, pp. 4370-4374

Rilling, Gutman, Zeh, Pagnoni, Berns, Kilts, (2002) A Neural Basis for Social Cooperation, Neuron, Vol. 35, pp. 395-405

Rilling, Sanfey, Aronson, Nystrom, Cohen, (2004) The neural correlates of Theory of Mind within interpersonal interactions, NeuroImage, Vol. 22, pp.1694-1703

Sabbagh, Moulson, and Harkness, (2002) Neural Correlates of Mental State Decoding in Human Adults: An Event-related Potential Study Journal of Cognitive Neuroscience Vol. 16, No. 3, pp 415-426

Sanna, Parks, Chang, (2003) Mixed-motive conflict in social dilemmas: Mood as input to competitive and cooperative goals, Group Dynamics, Vol.7, No.1, pp. 26-40

Sara, Gordon, Krauhin, Coyle, Howson, Mears, (1994) The P300 ERP component: An index of cognitive dysfunction in depression?, Journal of Affective Disorders, Vol.1, No.1, pp.229-238

Saxe and Kanwisher, (2003) People thinking about thinking people: The role of the temporo-parietal junctionin "theory of mind" NeuroImage, Vol.19, pp. 1835-1842

Saxe, Wexler Making sense of another mind: the role of the temporo-parietal junction In Press at Neuropsychologia

Shamay-Tsoory, Tomer, and Ahron-Peretz, (2005)The Neuroanatomical Basis of Understanding Sarcasm and Its Relationship to Social Cognition Neuropsychology Vol.19, No. 3, pp. 288-399

Singer, Fehr, (2005) The neuroeconomics of mind-reading, IZA discussion paper distributed by IZA

Silverstein, Cross, Brown, Rachlin, (1998) Prior experience and patterning in a prisoner's dilemma game, Journal of Behavior and Decision Making, Vol.11, No. 2, pp. 123-138

Sperber, Dan. Metarepresentations: a multidisciplinary perspective(2000) Oxford University Press, Oxford, UK

Sperber, D., Wilson, D., (2002) Pragmatics, Modularity and Mind Reading, Mind and Language, Vol. 17, No. 1-2, pp. 3-23

Steinmetz, Roy, Fitzgerald, (2000) Attention modulates synchronized neuronal firing in primate somatosensory cortex, Nature, No. 404, pp. 187-190

Stevens, Hauser, (2004) Why be nice? Psychological constraints on the evolution of cooperation, Trends in Cognitive Science, Vol. 8, No. 2

Strelny, K. The representational Theory of Mind (1990) Blackwell Publishers, Oxford, UK

Stuart, Pleffer, Barry, Clarke, Smith, (2005) Development of Inhibitory Processing during the go/nogo task, Journal of Psychophysiology, Vol.19, No.1, pp. 11-23

Swick, Turken, (2002) Dissociation between conflict detection and error monitoring in the human anterior cingulate cortex, PNAS, Vol. 99, No. 25

Tomasello, Carpenter, Call, Behne, Moll, (2004) Understanding and sharing intentions: The origins of culture and cognition, in press Behavioral and Brain Sciences

Tomasello, Call, Hare, (2003) Chimpanzees understand psychological states- the question is which ones and to what extent?, Trend in Cognitive Sciences, Vol.7, No.4

Tomasello, Farrar, (1986) Joint Attention and Early Language
Child Development, Vol. 57, No. 6, pp. 1454-1463

Urretavizcaya, Moreno, Benlloch, Cardoner, (2003) Auditory event related potentials in 50 melancholic patients, Journal of Affective Disorders, Vol. 74, No. 3, pp. 293-297

Verdejo-Garcia, Perez-Garcia, Bechara, (2006) Emotion, Decision-Making and Substance Dependence: A Somatic-Marker Model of Addiction, Current Neuropharmacology, Vol. 4, No. 1, pp.17-31

Vogeley, Bussfeld, (2001) Mind Reading: Neural Mechanisms of Theory of Mind and Self-Perspective NeuroImage Vol.14, pp. 170-181

Walet, Adenzato, (2004) Understanding Intentions in Social Interaction: The Role of the Anterior Paracingulate Cortex, Journal of Cognitive Neuroscience, Vol.16, No.10

Wilson, R., Keil, F., 1999. The MIT Encyclopedia of the Cognitive Sciences. The MIT Press, Cambridge, MA.

Whitehurst, Lonigan, (1998) Child Development and Emergent Literacy
Child Development, Vol. 69, No. 3, pp. 848-872

Woolfe, Want, Siegal, (2002) Signposts of Development: Theory of Mind in Deaf Children, Child Development, Vol. 73, No. 3, p. 768

Woolfe, Want, Siegal, (2003) Siblings and Theory of Mind in Deaf Native Signing Children, Journal of Deaf Studies and Deaf Education, Vol. 8, No. 3, pp.340-347

Wu, Coulson, (2005) Meaningful gestures: Electrophysiological indices of iconic gesture comprehension, Psychophysiology, Vol. 42, No. 6, p. 256

Yeung, Cohen, Botvinivk, (2004) The neural basis of error detection: Conflict monitoring and the error-related negativity, Psychological Review, Vol. 111, No. 4, pp. 931-959