Thinking about others: The neural substrates of social cognition

Jason P. Mitchell\textsuperscript{1,2}, Malia F. Mason\textsuperscript{1}, C. Neil Macrae\textsuperscript{1}, & Mahzarin R. Banaji\textsuperscript{2}

\textsuperscript{1} Department of Psychological and Brain Sciences, Dartmouth College
\textsuperscript{2} Department of Psychology, Harvard University

Address correspondence to:

Jason Mitchell, Ph.D.
Department of Psychology
Harvard University
William James Hall
33 Kirkland Street
Cambridge, MA 02138
Email: jmitchel@wjh.harvard.edu
Humans are set apart from other animals by a number of behaviors and the cognitive faculties that give rise to them. For example, *Homo sapiens* is the only species that has harnessed arbitrary symbols into languages capable of expressing an infinite number of ideas; that routinely develops and improves upon tools for augmenting our natural abilities (including mathematics); and that can suppress immediate desires or prepotent tendencies indefinitely in pursuit of abstract goals in the distant future (a fact to which any graduate student can attest). These faculties have no doubt contributed critically to the enormous global changes wrought by humans.

This list can safely be expanded to include another special feature of human behavior, namely, social behavior. Although a measure of sociability marks other animals throughout the phylogenetic tree – from our closest primate relatives to social insects such as bees and ants – the scale and complexity of human social abilities far outstrips those of even our closest primate relatives (Byrne & Whiten, 1988; Parker & Gibson, 1990; Ristau, 1991). Moreover, many of these abilities – such as recognizing oneself as a mental agent and inferring the psychological states of other such agents (even when their beliefs conflict with one’s own) – do not appear to have ready homologues among other animals (Gallup, 1985; Povinelli, Parks, & Novak, 1991; Tomasello & Call, 1997), suggesting that humans may have a unique adaptation for particular aspects of social cognition. Indeed, this apparent lack of comparable social-cognitive skills in other animals, along with the centrality of social interaction to human life, has prompted some observers to suggest that social cognition has represented the primary focus of evolutionary change in humans (Kamil, 2004; Tomasello, 1999).

Recently, researchers in the neurosciences have turned their attention to understanding the ways in which the brain gives rise to these remarkable social abilities. Of particular interest to both traditional social-cognitive psychologists and social cognitive neuroscientists is whether
the processes that give rise to social cognition are a subset of more general cognitive processes, or whether there are unique processes governing social cognition (Adolphs, 1999; 2003; Blakemore, Winston & Frith, 2004; Glick, 1978; Gelman & Spelke, 1981; Hoffman, 1981; Shields, 1978; Ostrom, 1984). As Blakemore, Winston, and Frith (2004) recently asked, are the “general cognitive processes involved in perception, language, memory and attention… sufficient to explain social competence or, over and above these general processes, are there specific processes that are special to social interaction?” (p. 216).

The notion that social cognition may rely on a set of unique mental processes is supported by observations that the cognitive challenges posed by social interaction appear to be distinct from those presented by physical (i.e., non-human) objects. That is, a successful encounter with another person demands certain kinds of cognitive skills that are not generally required in the rest of everyday life. For example, appropriate interaction with another person requires that we first recognize that the other entity is indeed another mental agent possessing internal psychological states not unlike our own. Having done so, we must accurately and rapidly intuit the motivations, feelings, and beliefs underlying their behavior (Baron-Cohen, 1995), while keeping in mind that in addition to moment-to-moment mental states, people possess stable dispositional characteristics (i.e., personalities) that influence their actions. Finally, we must compute how our own behavior will influence the other person, both in order to act in a socially appropriate manner as well as to manipulate his or her mental states and concomitant behavior (such as when attempting to convey a complicated idea or to persuade someone to act in a certain way).

At a minimum, these informal observations of the unique problems posed by social behavior provide probable cause for hypothesizing that social cognition may rely on a distinct set of mental processes. How, then, does one proceed to gather evidence in support of this
possibility? Within several other fields of psychology, neuroimaging and neuropsychological research have proven critical to the resolution of similar theoretical debates, at least under a narrow set of conditions. In particular, these techniques have contributed significantly to several entrenched controversies when the central theoretical question could be specified in the following way: do two psychological phenomena result from a single set of mental processes or instead from multiple processes? For example, ambiguities regarding the relation between implicit and explicit memory (i.e., whether implicit memory should be considered a degraded form of explicit memory or the product of an entirely different memory system) persisted after years of cognitive studies had produced behavioral data that were explainable within a number of different theoretical frameworks. Resolution of this controversy in favor of the multiple-systems view came about only after observations from neuropsychological patients (Gabrieli, Fleischman, Keane, Reminger, & Morrell, 1995) and neuroimaging (see Schacter & Buckner, 1998) demonstrated that the neural basis of implicit memory was both neuroanatomically and functionally distinct from explicit memory. Conversely, neuroimaging data were crucial to demonstrations that visual imagery relies on the very same neural mechanisms as actual visual perception (Kosslyn, 1994) providing powerful evidence against opposing theoretical accounts that suggested that imagery instead makes use of nonsensory, propositional knowledge (Pylyshyn, 1981, 2003).

As such, we expect the use of neuroimaging and patient populations will provide an efficient means for addressing the question of whether social cognition relies on its own, unique set of mental processes or instead ‘piggybacks’ on other, more general processes of memory, inference, executive planning, etc. By demonstrating that social cognition relies on a discrete set of brain regions, extant neuroscience research has generally supported a view of social cognition as distinct from other types of mental processes. The brain region that has most frequently been
implicated in social cognition is the medial prefrontal cortex (PFC), although research has also suggested that a number of other regions contribute critically to social-cognitive processing, including the temporo-parietal junction, orbitofrontal cortex, amygdala, superior temporal sulcus, and temporal poles (for reviews, see Adolphs, 1999; Adolphs, 2001; Blakemore et al., 2004; Frith & Frith, 1999; Gallagher & Frith, 2003).

In this chapter, we review the body of neuroimaging and neuropsychological work that has identified differences in the representations of social and non-social knowledge. We conclude the chapter by speculating on the underlying computations that may give rise to some of our social-cognitive abilities, reviewing recent neuroimaging work that suggests a link between thinking about the minds of other people and thinking about one’s own self.

Knowledge about the unique characteristics of other people

In being governed by complex mental states, other people are a unique kind of stimulus. Whereas people’s behaviors are understood through a consideration of their underlying motivations and feelings (e.g., intellectual curiosity, love, teenage angst), objects such as jackhammers, avocados, and Jeeps – indeed everything other than people – are governed by external forces. Because appropriate social behavior is predicated on the recognition that other people can be described in such mentalistic terms, one fundamental social-cognitive challenge is to distinguish person-relevant from person-irrelevant semantic knowledge.

Until recently, studies of how the brain represents semantic knowledge have focused on dissociations among different classes of inanimate objects. An interesting, and somewhat unexpected, finding from this research has been that the brain appears to organize some types of semantic knowledge in a category-specific manner (Warrington, 1975). That is, knowledge about various kinds of objects (e.g., animals, tools) appears to be subserved by different brain regions. Although some controversy exists regarding the precise organizing principles
underlying such category-specificity, several influential theories have suggested that the brain’s semantic representations of a class of object center around the features that are unique to that class of object (Caramazza & Shelton, 1998). For instance, because most tools are defined by their function (and not some arbitrary physical property, such as color), motor regions – such as left premotor cortex – are involved in the representation of knowledge about tools (Martin, 2001). In contrast, because animals are differentiated from one another primarily on the basis of their visual features (rather than on the basis of function), semantic knowledge of animals appears to be represented by brain regions involved in the visual perception of animate objects and biological motion – lateral fusiform gyrus and superior temporal sulcus (Chao, Haxby, & Martin, 1999). Moreover, recent work has suggested that regions of the motor cortex that support movement of various body parts (e.g., the foot, arm, mouth) are also recruited when people read action words that associated with these body parts, such as kick, throw, or chew (Hauk, Johnsrude, & Pulvermuller, 2004).

Is semantic knowledge about other people also represented in such a category-specific manner? That is, given that people differ from other types of entities by virtue of having mental states, do discrete brain regions subserve knowledge about people as mental agents? Alternately, could our understanding about the unique characteristics of other people simply rely on the same brain regions known to subserve semantic processing more generally (i.e., perceptual and functional representations)?

Using functional magnetic resonance imaging (fMRI), a recent series of studies has suggested that the brain may indeed represent semantic knowledge about the mental states of other people in a discrete manner. In an initial study, Mitchell, Heatherton, and Macrae (2002) presented participants with items from three different categories: people (denoted by common American forenames, e.g., John, Mary), fruits (banana, grape), and articles of clothing (mitten,
socks). Each item was presented alongside an adjective (e.g., curious, pitted, woolen), and participants were asked to indicate whether the adjective could ever be used to describe the target item (which it could on half the trials). Replicating earlier work on the neural basis of semantic processing, event-related fMRI analyses indicated that semantic judgments about inanimate objects engaged regions previously implicated in object knowledge tasks, specifically, left-lateralized inferotemporal cortex and ventrolateral PFC. In stark contrast, and despite the similarity between object and person trials, judgments about other people were associated with modulations in a qualitatively different set of brain regions. Dovetailing with earlier work on the potential neural basis of social cognition, these regions consisted of MPFC, right temporoparietal junction, superior temporal sulcus, and fusiform gyrus (Figure 1). In other words, making semantic decisions about characteristics of other people appeared to engage a qualitatively different set of brain regions – and, presumably, a concomitantly different set of cognitive processes – than did similar decisions about inanimate objects.

While suggesting that distinct neural representations may subserve knowledge about other people, this initial study raises several important questions about the precise nature of these representations. First, do the observed functional differences between person and object knowledge extend to all aspects of other people, or just to information about mental states? Because people differ from other stimuli by virtue of having such mental states, one might expect special mechanisms for representing these unique person characteristics, but not aspects of people shared with other stimuli, such as physical descriptors. Second, do these observed differences extend to knowledge about the psychological states of any stimulus or are they specific to understanding those of other people? Many nonhuman animals can be anthropomorphized as having mental states, such as curious or afraid; perhaps social-cognitive representations extend to the internal “mental” states of other animals. Most interesting,
however, would be the possibility that these mechanisms are specific to both the domain of
people and the process of understanding their mental states, that is, that the brain regions
engaged by social cognition may be engaged only when processing information about the mental
states of other people, but not other aspects of a person (such as his physical features) nor
knowledge about psychological states generally (e.g., for other animals).

In recent neuroimaging work, we have examined these questions directly. To examine
the extent to which the brain regions observed our earlier work (Mitchell et al., 2002) are
selective for understanding the mental states of other people, we scanned participants while they
made semantic judgments about two different kinds of targets, people and dogs (Mitchell,
Mason, Banaji, & Macrae, in prep). As in our earlier work, participants judged whether a word
could ever be used to describe the presented target. In this experiment, words could refer to one
of two aspects of the targets – either their potential psychological states (e.g., curious, frightened,
angry) or their unobservable physical parts (e.g., lung, heart, liver). An equal number of words
could not serve as potential descriptors (e.g., ionic, metallic). Critically, the words that could be
used to describe people were pretested to be equally applicable to dogs, and the same response
was made to each word regardless of whether the target was a person or a dog (i.e., curious
would require a “yes” response regardless of whether the target was a person or a dog). In
examining the brain response to each of these four types of trials, we observed that activity in the
medial PFC region previously implicated in person knowledge was selective for only one type of
judgment, namely, making decisions regarding the potential mental states of other people. That
is, medial PFC was not engaged when making judgments about the physical aspects of either
people or dogs, suggesting that its activity may be selective for understanding mental states.
However, additional activity in medial PFC was not observed during judgments of the
psychological states of nonhuman animals.
Another fMRI experiment addressed a similar question by examining the neural basis of action knowledge (Mason, Banfield, & Macrae, 2004). As in our earlier work, participants judged whether a word could ever be used to describe a person or a dog. Participants judged whether a series of actions (denoted by verbs such as run, sit, bite) could be performed by the target. As before, action words were pre-tested to ensure that they could apply equally to dogs and people. Results indicated that, whereas the action-related judgments about dogs were associated with activity in regions involved with mental imagery (occipital and parahippocampal gyri), identical judgments about people yielded activity medial PFC. Together, these data suggest that activity in some brain regions, such as the medial PFC, specifically subserves social-cognitive representations about the unique aspects of other people, including their mental and behavioral characteristics.

*Inferring the current mental states of another person*

Arguably the most important social-cognitive challenge is that of understanding the forces that govern other people’s behavior. Unlike inanimate objects, the behavior of people can often be attributed to unobservable mental states. According to Dennett (1987), perceivers understand other people and predict how they will act by adopting the “intentional stance” -- assuming that people are motivated by their current beliefs, desires, feelings, and goals. As such, a fundamental challenge to understanding other people is the ability to infer what these underlying mental states might be.

A fair proportion of neuroimaging and neuropsychological research on social cognition has focused on understanding the brain mechanisms that subserve the capability to take another’s perspective or infer their mental states. This enterprise has generated an extraordinarily rich diversity of paradigms designed to manipulate the extent to which perceivers must infer the mental states of others. For example, in some of the first work on this topic, Goel and colleagues
(Goel, Grafman, Sadato, & Hallett, 1995) asked participants to indicate whether a historical figure (Christopher Columbus) would know the function of various artifacts, (such as a CD) and compared neural activity during this task to one in which participants considered semantic or visual knowledge about those objects. Around the same time, Fletcher et al. (1995) presented participants with stories that were understandable only if one considered the mental states of characters, as well as stories that instead required understanding physical causality. Similarly, participants in a later study by Gallagher et al. (2000) were presented with the same mental-state stories as well as cartoons that also required understanding the minds of the characters in them. Finally, more recent work has had participants playing interactive games that require second-guessing one’s opponent, such as in the children’s game “rock, paper, scissors” (Gallagher, Jack, Roepstorff, & Frith, 2002) or a decision-making “trust” game (McCabe, Houser, Ryan, Smith, & Trouard, 2001) and compared activations when subjects thought they were playing against a human opponent to activations when subjects thought they were playing against a computer.

Despite the wide diversity of tasks used to prompt mental state attribution – verbal stories, cartoons, competitive games, etc. – a remarkable empirical consensus has emerged regarding the underlying brain regions important for understanding the minds of another person. In each of these studies, greater activity was observed in MPFC during tasks that required participants to infer the mental states of another person. Of particular interest are the gaming studies of Gallagher et al., (2002) and McCabe et al. (2001), in which the only manipulation was whether participants believed they were playing against a human or a computer; despite identical visual input and task requirements, activity in MPFC differed as a function of whether or not mental state attribution was required.
Extracting and encoding ‘personality’ information

Although the rapid understanding of others’ transient mental states must surely be a central capacity of any successful social agent, human interaction can also be marked by repeated encounters with the same person, and, as such, the opportunity for perceivers to extract information about a person’s stable, idiosyncratic characteristics. For better or worse, we share an apartment with the same roommate, ride the elevator with the same colleagues at work, see the same family members on holidays, etc. Whether or not our social encounters with these individuals are successful hinges upon our ability to both detect moment-to-moment shifts in others’ beliefs or feelings and to consider how their stable traits may influence their behavior or reactions. Decades of research in social psychology suggest that dispositional aspects of another person can help provide a powerful framework for explaining the behavior of others. Indeed, humans appear to have a strong bias toward attributing the behavior of others to their stable dispositions, while often ignoring important additional influences on behavior, such as situational constraints (Gilbert & Malone, 1995).

Given the utility of extracting and remembering regularities that may guide another person’s behavior, as well as our tendency to explain others’ behaviors in terms of their dispositional traits (rather than situational constraints), it seems important to understand the underlying cognitive processes that give rise to such abilities. In the late 1970s, researchers began examining the mechanisms underlying perceivers’ abilities to infer the dispositional regularities that define other people (D. L. Hamilton, Driscoll, & Worth, 1989; D. L. Hamilton, Katz, & Leirer, 1980; Hastie & Kumar, 1979; Srull & Wyer, 1989; Wyer, Bodenhausen, & Srull, 1984). In such research, participants were typically presented with information about a series of novel target individuals. Some participants would be asked to use the information to form an impression of the target (i.e., to attend to the stable, dispositional aspects of the person being
described), whereas others would simply be asked to memorize the presented information. Subsequently, participants would be asked to retrieve all the information that had been presented. Intriguingly, researchers observed a number of dissociations in memory performance between the social-cognitive (impression formation) task and the nonsocial (memorize) task. First, and somewhat surprisingly, participants’ memory was typically better following impression formation than under conditions of intentional encoding. Second, patterns of memory performance were often different following impression formation than memorize. For example, participants directed to ‘form an impression’ were likely to recall items in clusters, suggesting that they had spontaneously organized information around implied traits (e.g., consecutively recalling many of the items that implied someone was honest, then recalling many of the items that implied he was intelligent, etc.). Moreover, impression formation often led to increased memory for information that was inconsistent with participants’ expectations of the target’s personality (Hastie & Kumar, 1979; Srull, 1981). For example, if a target was first described as honest, information that implied she was dishonest would be particularly well remembered, but only if the participant was trying to form an impression of the person (Hartwick, 1979; Hastie, 1981).

In making sense of such differences, researchers generally suggested that social-cognitive processing prompts deeper, more elaborative encoding of the sort that generally supports episodic memory (Craik & Lockart, 1972), such as the generation of schemas (Cohen & Ebbesen, 1979; Rumelhart, 1980) or the formation of a particularly rich network of inter-item associations (Hastie, 1980; Srull, 1981). However, a second possibility suggests that rather than simply engaging deeper processing attempts to form an impression of someone actually engages different processing operations. In other words, introducing social-cognitive goals such as impression formation may prompt the deployment of qualitatively distinct cognitive processing
mechanisms. Functional brain imaging may shed light on these issues by distinguishing between the circumstances under which (a) processing information about social objects, such as people, overlaps with nonsocial objects as well as (b) the conditions under which these processes diverge and unique neural networks are recruited for social and nonsocial cognition.

Some recent neuroimaging research suggest that impression formation does indeed prompt distinct kinds of cognitive processing (Mitchell, Macrae, & Banaji, 2004). In one study, participants were scanned while incidentally encoding information that described a series of unfamiliar people. Each person (denoted by photographs of his face) was paired with 10 descriptive statements that described various activities ostensibly performed by the person (e.g., “studied for his calculus final on the flight home for the holidays”). Each pair was accompanied by one of two cues (Form Impression, Remember Order) that indicated which of two orienting tasks was to be performed on that trial. For ‘Impression Formation’ trials, participants were instructed to use the statement to infer the personal characteristics and traits of the target person, integrating across the entire set of statements for each person. For ‘Sequencing’ trials, participants were instructed to encode the order in which statements are paired with each face (although earlier cognitive work has generally compared impression formation to explicit attempts to memorize the information, the cognitive processes engaged during memorization may vary from participant to participant, or even from trial to trial, making intentional encoding too underconstrained for event-related neuroimaging). Subsequently, participants performed an associative retrieval task during which they were asked to match statements to the face with which it was originally presented.

Initial fMRI analyses revealed the expected neural differences between Impression Formation and Sequencing. Specifically, greater activation was observed for Impression Formation > Sequencing along a wide extent of the medial PFC. Of course, given that
participants’ task during Impression Formation was very different than that during Sequencing, the observation of brain differences between the two tasks is somewhat unsurprising. However, the particular design of this experiment allowed an additional analysis based on conditionalizing encoding data as a function of subsequent memory. Specifically, encoding trials were retroactively conditionalized (“binned”) on the basis of both (a) what orienting task had been performed (Impression Formation, Sequencing) as well as (b) subsequent memory success (i.e., whether an item went on to be correctly remembered (hits) or to be forgotten (misses)), resulting in four types of trials: impression hits, impression misses, sequencing hits, sequencing misses. Results indicated that, for trials encoded as part of the Impression Formation task, only a single region – dorsomedial PFC – demonstrated higher activity for subsequent hits than for subsequent misses. Importantly, no significant difference was observed between hits and misses in this region for items that were initially encoded as part of the Sequencing task. That is, whereas encoding activity in dorsomedial PFC was greater for impression-hits than impression-misses, activity in this region did not differentiate between sequencing-hits and sequencing-misses. In contrast, for trials encoded as part of the Sequencing task, subsequent memory success was correlated only with activity in the right hippocampus. Importantly, no significant difference was observed between impression-hits and impression-misses in this region. (Figure 2) That is, encoding activity in right hippocampus was selectively correlated with subsequent memory for Sequencing, but not Impression Formation, trials. By showing the distinct neural basis of impression formation, these results suggest that not only are unique cognitive processes engaged by impression formation tasks, but that memory differences following social and nonsocial tasks result from the operation of these separate cognitive processes.

However, just as for research on the neural systems that subserve semantic knowledge about other mental agents, a demonstration that distinct processes are engaged for social
cognition requires that tasks be held constant across targets that are both social and nonsocial. Perhaps these earlier results reflect the particular demands of forming an impression about any stimulus, whether another person or an inanimate object. To address this possibility, a follow-up study examined the functional neuroanatomy associated with impression formation for both other people as well as inanimate objects (Mitchell, Macrae, & Banaji, in prep). As before, participants saw a series of statements that ostensibly described an associated target. However, in this study, only half the targets were other people; the remaining half of the targets were inanimate objects (cars and computers). For half of the people and half of the inanimate objects, participants were once again asked to use the statement to form an impression of the target whereas, for the other half of the items, participants were asked to remember the order in which statements were paired with the target. When brain regions were defined in the same way as in the initial study (i.e., Impression Formation > Sequencing for person trials only), a very similar region of dorsomedial PFC was observed, providing a direct replication of earlier findings. However, of critical interest in this experiment was whether or not this MPFC region was generally engaged by attempts to form an impression, regardless of the social aspects of the targets. To address this question, the neural response to forming an impression of inanimate objects was examined. Consistent with the notion that MPFC specifically subserves social-cognitive processing, results demonstrated a weak response in this region when participants formed an impression of inanimate objects, similar to the response observed for the Sequencing task. In other words, although the dorsomedial PFC was significantly activated during attempts to form an impression of other people, its activity was not increased by putatively similar attempts to form an impression of inanimate objects. Activity in this region appeared to track specifically with the social-cognitive demands of the orienting task.
Guiding social behavior

While understanding certain aspects of the minds of other people is generally necessary for appropriate social behavior, the ultimate output of a system for social cognition must be the direct guidance of such behavior. As Fiske (1992) has pointed out, “thinking is for doing”; a system that apprehends something (such as the state of another’s mind) but is unable to act appropriately based on that information will have little adaptive value. As such, social cognition must be able to direct social behavior, including the selection and initiation of appropriate actions (is it permissible to start the wave during a particularly inspiring departmental colloquium?), speech (how do I effectively communicate a complex scientific concept to a student?), and reciprocal responses (is it a good idea to snort derisively at the search committee member’s uninformed question?).

However, relative to amount of extant work on the intrapersonal aspects of social cognition, little neuroscience research has examined the unique demands of interpersonal social behavior. This state of affairs is at least partially due to the strict contraints of neuroimaging techniques, including restriction of movement, limit to one participant at a time (although recent attempts have been made to use coordinated scanners to image two or more people as they interact in realtime), and difficulty recording verbal output (at least for fMRI, where scanner noise is often prohibitively loud). As such, what little is known about the systems that guide social behavior has tended to come from research with neuropsychological patients. Damage to the orbitofrontal cortex (OFC) and medial aspects of the PFC, especially to its more ventral aspects, has been associated with a wide range of social deficits, including selective failure on theory of mind tasks (including second-order false belief and faux pas tasks). Although individuals with MPFC/OFC damage typically come to the attention of neurologists because of marked changes in social behavior – including personality changes, lack of empathy,
inappropriate social interactions – no detailed account of the behavioral deficits acquired in such patients has yet emerged. Perhaps the fullest sketch of social changes following damage to these regions can be found in Antonio Damasio’s description of a patient known as Elliot (Damasio, 1994). Following resection of an area of his frontal lobes to remove a tumor (mainly confined to orbitofrontal cortex and ventromedial aspects of the medial PFC), Elliot suffered severe changes to his ability to negotiate the social world. Formerly a reliable professional with a rewarding family life, Elliot became unable to hold a job; lost his life savings in a series of bad business ventures; and was divorced by his wife, subsequently married to another woman of whom his friends and family disapproved, and divorced a second time.

Interestingly, as is typical in cases of individuals with damage to orbitofrontal cortex and medial PFC, Elliot demonstrated unimpaired intelligence, language, and working memory. Such cases provide an important clue that social behavior results from the operation of cognitive processes that are distinct from those guiding behavior in other domains. Despite severe social deficits, patients with orbitofrontal and medial PFC damage are often unimpaired on tasks outside of the social domain, even when those tasks are quite challenging.

The distinctiveness of social behavior is also underscored by research on the autistic syndrome (Baron-Cohen, 1995). Although little is understood about the brain basis of the disorder, one characteristic of autism is a profound disengagement from the social world. Autistic individuals may shun social interaction and appear unable to learn the basic rules of social interaction, often despite relative sparing of abilities in other domains. Interestingly, a disorder known as Williams syndrome appears, in many ways, to present the converse pattern of deficits: individuals with Williams syndrome often appear hyper-social and verbal, but are profoundly impaired in nonsocial domains (Tager-Flusberg, Boshart & Baron-Cohen, 1998).
In another interesting line of research, Gregory et al. (2002) examined patients with frontal variant frontotemporal dementia (fvFTD) on a series of mentalizing tasks of varying degrees of difficulty. FTD is a progressive disorder that results in degeneration to areas of frontal and temporal lobe. Interestingly, individuals with the frontal variant of the disorder typically present to neurologists with various changes in social behavior similar to those observed in patients with lesions to MPFC, including personality changes, lack of empathy, and socially inappropriate behavior. Meta-analytic procedures over large numbers of individuals diagnosed with fvFTD (Salmon et al., 2003) have suggested that neural degeneration is most severe in a circumscribed region of PFC highly similar to the medial regions observed in neuroimaging studies of social cognition. Gregory et al. compared performance of fvFTD patients to that of individuals with Alzheimer’s disease (AD) on a series of mentalizing tasks: first-order false belief, second-order false belief, Reading the Minds in the Eyes (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), and faux pas detection. Compared to their AD controls, fvFTD patients were impaired on all the mentalizing tasks, including the ability to detect socially inappropriate behavior on the faux pas task.

One question regarding such demonstrations is whether the deficits of social behavior demonstrated by neuropsychological patients are a direct result of impairments in the ability to mentalize about the minds of other people. Future research will be needed to examine whether separate systems exist for the guidance of social behavior, or whether social interaction is predicated exclusively on the intact ability to understand the minds of those with whom one is engaging.

The link between self and social cognition

Although the data reviewed above suggests that social cognition may indeed rely on distinct mental processes, a central question remains regarding precisely of what those processes
consist. Given the demonstration that understanding people is not the same as understanding inanimate objects, how is it that one goes about making sense of the mental states and behavior of other people? One influential theory of how we come to understand the minds of other people suggests that knowledge of one’s own mind can be successfully used to help infer the mind of another person (Davis & Stone, 1995a, 1995b). This account – broadly known as “simulation theory” – suggests that one valuable source of information about the thoughts, feelings, or potential behavior of another person is a first-person prediction about what I myself might think, feel, or do in a similar situation. Although clues about what is going on in another’s mind can certainly be gleaned from a variety of sources, (e.g., emotional expression, direction of eye gaze, folk theories about how other people work), simulation could provide particularly useful information about other people, especially in complex or novel social situations.

Intriguingly, there appears to be a good deal of overlap between the brain regions that subserve thinking about other people and those that subserve thinking about oneself. Specifically, in addition to its general role in social cognition, the MPFC appears to be an integral component of tasks that require participants to assess one’s own qualities or current feelings. For example, a recent study examined the neural basis of the self-relevance effect in memory, whereby participants typically demonstrate enhanced episodic memory for information that has been related to oneself (Rogers, Kuiper & Kirker, 1977). Participants incidentally encoded a series of adjectives by making one of three judgments about each: whether the word (a) described their own personality, (b) described the personality of current president George W. Bush, or (c) appeared in uppercase or lowercase letters. Results indicated that self judgments were associated with additional activity in MPFC compared to other or case judgments, suggesting a role for the MPFC in self-relevant processing (Kelley et al., 2002). Subsequent research has demonstrated that activity in this region of MPFC also correlates with subsequent
memory for items that were encoded in a self-referential manner (Macrae, Moran, Heatherton, Banfield, & Kelley, 2004). These results dovetail with those of Gusnard, Akbudak, Shulman, and Raichle (2001), who demonstrated that a region of dorsomedial PFC was activated during judgments of photographs in a way that required participants to refer to their own affective experience.

Together, these observations that MPFC appears to subserve thinking about self as well as thinking about others provide initial support for simulation accounts of social cognition. By suggesting that social cognition and self-referential thought may rely on a common set of cognitive processes, these disparate lines of research converge on the notion that understanding oneself is an integral component in understanding other people. In a recent fMRI study, we sought to test two predictions that derive from such simulation accounts of social cognition. First, one should only engage in simulation when one’s task is to infer the current mental states of another person. When interacting with a person in a way that does not require mental state attribution (e.g., when looking for a friend’s familiar face in a crowded bar), one need not simulate the minds of others based on one’s own. Second, simulation should be useful only when one has reason to believe that a simulation will be applicable to the person in question. If a person thinks very differently – perhaps because of cultural or interpersonal differences – then it is unclear that an understanding of his mental states or a prediction of his behavior can be accurately inferred from consideration of one’s own.

During event-related fMRI scanning, participants saw a series of faces for which they were either asked to perform a mentalizing or nonmentalizing task. For half the faces, participants were asked to mentalize about the target’s internal states by judging how happy or pleased the person seemed to be to have his or her photograph taken. For the other half of the faces, participants were asked to make judgments that did not include a mentalizing component,
namely, indicating how symmetrical (left to right) each face appeared. Following scanning, participants saw each of the faces a second time and were asked to judge each for how similar they felt the person was to themselves. The event-related nature of the design allowed us to retroactively conditionalize items as a function of the task performed during the initial presentation (mentalyze, nonmentalyze) as well as how similar each participant felt they were to each target (similar, dissimilar). As predicted on the basis of extant research on social cognition, the contrast of mentalizing > nonmentalizing yielded a region of MPFC, similar to that observed in our earlier work on impression formation. More importantly for simulation accounts of social cognition, activity in MPFC differentiated between similar and dissimilar targets, but only for those for which the mentalizing task was performed. Specifically, for faces in the mentalizing task, higher activity in the MPFC was observed for faces that participants judged to be similar than for faces judged to be dissimilar; no such dissociation was observed for faces in the nonmentalizing task. These results are consistent with the prediction that simulation should only occur when a target is similar enough to the perceiver to make simulation an appropriate basis for understanding that person’s mind.

**Conclusions**

We began this chapter by pointing out a number of ways that social cognition – thinking about the minds of other people – poses a set of distinct challenges that may not have ready parallels in the physical world. Other individuals are complex, dynamic entities that have properties – mental states – that are not possessed by any other class of stimuli. To make sense of another’s behavior we must accurately identify these mental states and consider how, in combination with stable dispositional traits, they may be influencing how people interact with the world around them.
The centrality of these abilities to human life is demonstrated most clearly by individuals who have selective social deficits, such as in autism or after particular kinds of brain damage. Indeed, the fundamental importance of human sociability both to everyday life as well as to the cultural achievements of our species has prompted some observers to suggest that social cognition may have been one of the primary engines of human evolution. For example, Tomasello (1999) has argued convincingly that what sets *Homo sapiens* apart from other primates is the ability represent the mind of conspecifics.

Given both the uniqueness of challenges posed by the social world and the importance of sociability to human life, one might expect distinct cognitive processes to be dedicated to thinking about and interacting with other individuals. Armed with neuroimaging and neuropsychological methods, researchers have begun to provide triangulating empirical support for the notion that separate systems do indeed subserve social and nonsocial cognition. The promise of future research in this domain is to begin to specify the precise computations that allow us to perform the remarkable feats of social gymnastics of which humans are routinely capable.
Footnotes

1 Of course, the ability to use neuroimaging and neuropsychological patients to address issues of theoretical relevance in this way relies on the following two assumptions: (a) that two different brain regions cannot give rise to precisely the same cognitive process and (b) that no single brain region instantiates multiple such processes. Although neither of these assumptions is compelled philosophically or empirically, neither too has any serious challenge been posed to these two foundations of the cognitive neuroscience enterprise.
References


Figure legends

**Figure 1.** Regions of the prefrontal cortex that were observed by Mitchell, Heatherton, and Macrae (2002) to be differentially engaged by semantic judgments of objects and of people. Replicating earlier studies of the neural systems that subserve semantic knowledge, object judgments engaged an extensive region of left ventrolateral PFC (red-orange scale). In contrast, person judgments were associated with activity in qualitatively distinct regions of the prefrontal cortex, specifically, MPFC (blue-green scale). T-maps from comparisons of object and person judgments are overlaid on coronal slices (y values = 5, 35, and 55, respectively) of participants’ mean normalized brain.

**Figure 2.** Participants in a study by Mitchell, Macrae, and Banaji (2004) encoded face-statement pairs in either a socially relevant (Impression Formation) or irrelevant (Sequencing) manner. Distinct neural correlates of subsequent memory success were observed as a function of the orienting task performed during encoding. For Impression Formation trials, activity in dorsomedial PFC (left panel) was higher for subsequent hits (solid black bars) than misses (striped black bars); no significant difference was observed between hits and misses for Sequencing trials (gray solid and striped bars) in this area. In contrast, for Sequencing trials, activity in right hippocampus (right panel) was higher for subsequent hits than misses, and no difference was observed in this region for Impression Formation trials.
Figure 1
Figure 2

Dorsomedial PFC  R hippocampus

Parameter estimate

Impression  Sequencing  Impression  Sequencing