

4 Thinking about Others: The Neural Substrates of Social Cognition

Jason P. Mitchell, Malia F. Mason, C. Neil Macrae, and Mahzarin R. Banaji

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 that may be realized only in a distant future (a fact to which any academic can attest). These faculties have no doubt contributed critically to the vast global changes wrought by humans.

This list can safely be expanded to include another special feature of human behavior, namely, social behavior. Although obvious examples of sociability are to be found in other animals throughout the phylogenetic tree, from other primates to social insects such as bees and ants, the scale and complexity of human social abilities far outstrip 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; Puvionelli, Parks, & Novak, 1991; Tomasello & Call, 1997), suggesting that humans may have their own adaptation for particular aspects of social cognition. Indeed, this apparent lack of comparable social-cognitive skills in other animals, together with the centrality of social interaction to human life, has prompted some observers to suggest that social cognition represents 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 human social abilities. Of particular interest to both traditional social-cognitive psychologists and cognitive neuroscientists is whether the processes that give rise to social cognition are a subset of more general cognitive processes, or whether specific social-cognitive processes exist (Adolphs, 1999,

2001; Blakemore, Winston, & Frith, 2004; Ostrom, 1984). As Blakemore et al. (2004) 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 specific mental processes is supported by observations that the cognitive challenges posed by social interaction appear to be distinct from those presented by physical (nonhuman) 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 that individual's behavior (Baron-Cohen, 1995), while keeping in mind that in addition to moment-to-moment mental states, people possess stable dispositional characteristics (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 the other's 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 special 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 proved critical to the resolution of similar theoretical debates. 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 from multiple processes? For example, ambiguities regarding the relation between implicit and explicit memory (whether implicit memory should be considered a degraded form of explicit memory or the product of an entirely different memory system) persisted even after years of accumulated behavioral data because they were explainable within a number of different theoretical frameworks. Resolution of this controversy in favor of the multiple-systems view came about after observations of patients with neuropsychological disorders (Gabrieli et al., 1995), and neuroimaging (see Schacter & Buckner, 1998) showed that the basis of implicit memory was both neuroanatomically and functionally distinct from explicit memory.

In a different controversy, neuroimaging data were crucial in showing that visual imagery relies on the very same neural mechanisms as actual visual perception (Kosslyn, 1994), thus providing compelling evidence against the opposing theoretical account that imagery relies on nonsensory, propositional knowledge (Pylyshyn, 1981, 2003).¹

Based on such results, we propose that data from neuroimaging and patient studies will also provide an efficient means for addressing the question of whether social cognition relies on its own set of mental processes or instead piggybacks on other, more general processes of memory, inference, executive planning, and so on. 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 most frequently implicated in social cognition is the medial prefrontal cortex (PFC), although research also suggests that a number of other regions contribute critically to social-cognitive processing, including the temporoparietal junction, orbitofrontal cortex, amygdala, superior temporal sulcus, and temporal poles (for reviews, see Adolphs, 1999, 2001; Blakemore et al., 2004; Frith & Frith, 1999; Gallagher & Frith, 2003).

Knowledge about the 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 consideration of their underlying motivations and feelings (intellectual curiosity, 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 mental 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 focused on dissociations among different classes of inanimate objects. An interesting, and somewhat unexpected, finding from this research is 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 (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 suggested that the brain's semantic representations of a class of object center around the features that are specific 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 gyms and superior temporal sulcus (Chao, Haxby, & Martin, 1999). Moreover, recent work suggests that regions of the motor cortex that support movement of various body parts (foot, arm, mouth) are also recruited when people read action words that are associated with these body parts, such as *kick*, *throw*, or *chew* (Hauk, Johnsrude, & Pulvermüller, 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 characteristics of other people simply rely on the same brain regions known to subserve semantic processing more generally (perceptual and functional representations)?

Using functional magnetic resonance imaging (fMRI), a series of studies 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 beside an adjective (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 inferior temporal 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 medial PFC, right temporoparietal junction, superior temporal sulcus, and fusiform gyms (figure 4.1). 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.

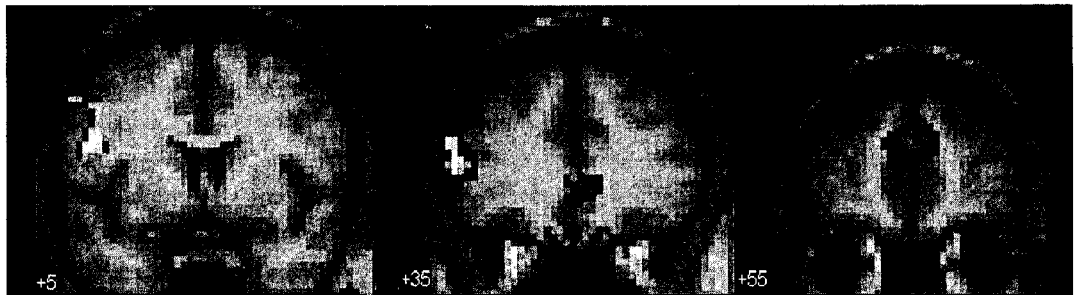


Figure 4.1

Regions of the prefrontal cortex that were differentially engaged by semantic judgments of objects and of people (Mitchell, Heatherton, & Macrae, 2002). Replicating earlier studies of 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, medial PFC (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.

While suggesting that distinct neural representations may subserve knowledge about other people, this study raises several important questions about the precise nature of these representations. First, do 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 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.

To examine directly the extent to which brain regions observed in 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, Rana, & Macrae, in press-a). 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—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., onic, metallic) and parts (nozzle, rudder). 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 judging words that referred to mental states, regardless of whether the target was another person or a dog. That is, the medial PFC was not simply engaged when participants made any judgment about a person (there was relatively low activity in this region when responding to "parts" trials). However, the medial PFC did appear to generalize to making mental state judgments regardless of whether the target of those judgments was another person or another mental agent (a dog).

Another fMRI experiment addressed a similar question by examining the neural basis of action knowledge (Mason, Banfield, & Macrae, 2004). Participants judged whether a series of actions (denoted by verbs such as run, sit, and bite) could be performed by a target, which again could be a person or a dog. As before, action words were pretested to ensure that they could apply equally to dogs and people. Whereas 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 in the medial PFC. Together, these data suggest that activity in some brain regions, such as the medial PFC, specifically subserves social-cognitive representations about specific 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 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 brain mechanisms that subserve the capability

to take others' perspectives 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 (1995) asked participants to indicate whether a historical figure (Christopher Columbus) would know the function of various artifacts (such as a compact disk), and compared neural activity during this task with that in 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. (2002) 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 the children's game "rock, paper, scissors" (Gallagher et al., 2002), and compared activations when subjects thought they were playing against a human opponent versus 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 mind of another person. In each of these studies, greater activity was observed in medial PFC during tasks that required participants to infer the mental state of another person. Of particular interest is the gaming study of Gallagher et al. (2002), 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 medial PFC differed as a function of whether or not mental state attribution was required.

Extracting and Encoding Personality Information

Although 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 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. Whether or not our social encounters with these individuals are successful hinges in part upon our ability to

form theories of their stable qualities. Decades of research in social psychology suggest that dispositional aspects of another person can help provide an efficient framework for explaining the behavior of others. Indeed, humans appear to have a 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 mechanisms underlying perceivers' abilities to infer the dispositional regularities that define other people (Hamilton, Driscoll, & Worth, 1989; Hamilton, Katz, & Leirer, 1980; Hastie & Kumar, 1979; Srull & Wyer, 1989; Wyer, Bodenhausen, & Srull, 1984). Participants were typically given information about a series of novel target individuals. Some participants would be asked to use the information to form an impression of the target (to attend to the stable, dispositional aspects of the person being described), whereas others would simply be asked to memorize the information. Subsequently, participants were asked to retrieve all the information that had been presented. Researchers observed a number of intriguing dissociations in memory performance between the social-cognitive (impression formation) task and the nonsocial (memorize) task. First, and somewhat a surprise, participants' memory was typically better after impression formation than after intentionally memorizing. Second, *patterns* of memory performance differed across the two conditions. 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 (consecutively recalling many of the items that implied someone was honest, then recalling many of the items that implied intelligence, 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 dishonesty would be particularly well remembered, but only if the participant was trying to form an impression of the person (Hartwick, 1979; Hastie & Kumar, 1979).

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 & Lockhart, 1972), such as generation of schemas (Cohen & Ebbesen, 1979) or formation of a particularly rich network of inter-

item associations (Hastie & Kumar, 1979; Srull, 1981). However, a second possibility suggests that rather than simply engaging deeper processing, attempts to form an impression actually engage different processing operations. In other words, introducing social-cognitive goals such as impression formation may prompt deployment of qualitatively distinct cognitive mechanisms. Functional brain imaging may shed light on these issues by distinguishing between the circumstances under which (1) processing information about social objects, such as people, overlaps with processing information about nonsocial objects, as well as (2) conditions under which these processes diverge and distinct neural networks are recruited for social and nonsocial cognition.

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 ten 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. 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 were paired with each face. (Although earlier cognitive work generally compared impression formation with 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 memory task during which they were asked to match statements to the face with which it was originally presented.

Initial fMRI analyses revealed expected neural differences between impression formation and sequencing; specifically, greater activation for impression formation along a wide extent of the medial PFC. Of course, given that participants' task during impression formation was very different from that during sequencing, brain differences between the tasks are somewhat unsurprising. However, the 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 what orienting task had been performed (impression formation, sequencing) as well as subsequent memory success; that is, 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-had higher activity for subsequent hits than for subsequent misses. Of importance, 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. It was again important that no significant difference was observed between impression hits and impression misses in this region (figure 4.2). That is, encoding activity in right hippocampus was selectively correlated with subsequent memory for sequencing, but not for impression formation. By showing the distinct neural basis of impression formation, these results suggest that not only are specific cognitive processes engaged by impression formation tasks, but

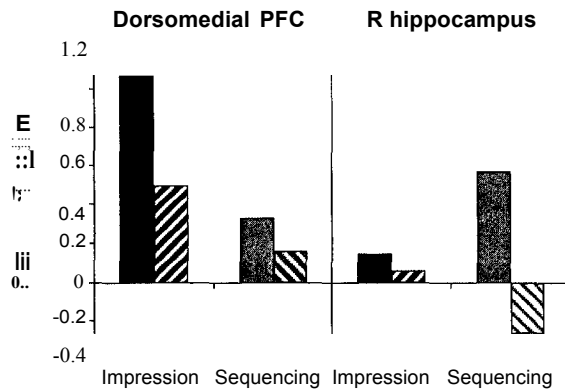


Figure 4.2

Participants encoded face-statement pairs in either a socially relevant (impression formation) or irrelevant (sequencing) manner (Mitchell, Macrae, & Banaji, 2004). 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 PrC (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.

memory differences after social and nonsocial tasks result from the operation of these separate cognitive processes.

However, just as for research on the neural systems that subservise semantic knowledge about other mental agents, the 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 press*). As before, participants saw a series of statements that ostensibly described an associated target. However, only half the targets were other people; the remaining half were inanimate objects (cars, 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, and for the other half they 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 (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 the medial PFC was generally engaged by attempts to form an impression, regardless of 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 medial PFC specifically subserves social-cognitive processing, results revealed 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

Although 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) 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 lecture?), speech (how do I effectively communicate a complex scientific concept to a novice?), and reciprocal responses (is it a good idea to snort derisively at the search committee member's uninformed question?).

However, relative to the amount of extant work on the intrapersonal aspects of social cognition, little neuroscience research has examined the demands of interpersonal social behavior. This state of affairs is due at least in part to the strict constraints 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 real time (Montague et al., 2002), 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 in patients with neuropsychological disorders. Damage to the orbitofrontal cortex and medial aspects of the PFC, especially to its more ventral aspects, is 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 orbitofrontal-medial PFC damage typically come to the attention of neurologists because of marked changes in social behavior, including personality changes, lack of empathy, and inappropriate social interactions, no detailed account of their behavioral deficits has yet emerged. Perhaps the fullest sketch of social changes after damage to these regions can be found in the description of a patient known as Elliot (Damasio, 1994). After 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, he became unable to hold a job, lost his life savings in a series of bad business ventures, and was divorced by his wife; he married another woman of whom his friends and family disapproved, and divorced a second time.

Of interest, as is typical in cases of individuals with damage to orbitofrontal cortex and medial PFC, Elliot had 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 its basic rules, often despite relative sparing of abilities in other domains. Of interest, a disorder known as Williams syndrome appears in many ways to present the converse pattern of deficits: these individuals often appear hypersocial and verbal, but are profoundly impaired in nonsocial domains (Tager-Flusberg, Boshart, & Baron-Cohen, 1998).

In another interesting line of research, patients with frontal variant frontotemporal dementia (FTD) were examined on a series of mental tasks of varying degrees of difficulty (Gregory et al., 2002). FTD is a progressive disorder that results in degeneration to areas of frontal and temporal lobe. Individuals with the frontal variant of the disorder typically experience various changes in social behavior similar to those in patients with lesions to medial PFC, including personality changes, lack of empathy, and socially inappropriate behavior. Meta-analytic procedures over large numbers of individuals diagnosed with frontal variant FTD (Salmon et al., 2003) suggested that neural degeneration is most severe in a circumscribed region of PFC highly similar to medial regions observed in neuroimaging studies of social cognition. Gregory et al. compared performance of such patients with that of individuals with Alzheimer's disease (AD) on a series of mental tasks: first-order false belief, second-order false belief, reading the minds in the eyes (Baron-Cohen et al., 2001), and faux pas detection. Compared with AD controls, patients with frontal variant FTD were impaired on all the tasks, including the ability to detect socially inappropriate behavior on the faux pas task.

One question regarding such results is whether the deficits of social behavior of patients with neuropsychological disorders are a direct result of impairments in the ability to mentalize about the minds of other people. Future research is necessary to examine whether separate systems exist for such 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.

Thinking about Oneself and Others

Although the data reviewed above suggest that social cognition may indeed rely on distinct mental processes, a central question remains regarding precisely of what those processes consist. Given the fact 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 suggests that knowledge of one's own mind can be used successfully to help infer the mind of another person (Davis & Stone, 1995a, b). This account, broadly known as simulation theory, proposes that one valuable source of information about the thoughts, feelings, or potential behavior of another person is a first-person prediction about what / 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, (emotional expression, direction of eye gaze, folk theories about how other people work), simulation could provide particularly useful information in this regard, especially in complex or novel social situations.

A good deal of overlap appears to exist between 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 medial PFC appears to be an integral component of tasks that require participants to assess one's own qualities or current feelings (Johnson et al., 2002; Kelley et al., 2002; Macrae et al., 2004; Schmitz, Kawahara-Baccus, & Johnson, 2004; Zysset et al., 2002). 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 described their own personality, described the personality of current president George W. Bush, or appeared in uppercase or lowercase letters. Results indicated that self-judgments were associated with additional activity in medial PFC compared with other or case judgments, suggesting a role for the structure in self-relevant processing (Kelley et al., 2002).² Later research revealed that activity in this region of medial PFC also correlates with subsequent memory for items that were encoded in a self-referential manner (Macrae et al., 2004). These results dovetail with those of Gusnard and colleagues (2001), 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 medial PFC 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. Mitchell et al. (in press-b) performed an fMRI study to test two predictions that derive from such simulation accounts of social cognition. First, one

should engage in simulation only 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 it will be applicable to the person in question. If a person thinks very differently (perhaps because of cultural or interpersonal differences), it is unclear that understanding his mental states or predicting his behavior can be achieved through consideration of one's own.

During event-related fMRI scanning, participants saw a series of faces for which they were asked to perform either a mentalizing or nonmentalizing task. For half the faces, participants were asked to mentalize about the target's internal states by judging how pleased the person seemed to be to have his or her photograph taken. For the other half, participants were asked to make judgments that did not include a mentalizing component, namely, indicating how symmetrical (left to right) each face appeared. After 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 conditionalize items retroactively as a function of the task performed during the initial presentation (mentalize, nonmentalize) as well as how similar each participant felt he or she was 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 medial PFC similar to that in our earlier work on impression formation. More important for simulation accounts of social cognition, activity in medial PFC differentiated between similar and dissimilar targets, but only for those for whom the mentalizing task was performed. Specifically, for faces in the mentalizing task, activity in the medial PFC was higher 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 occur only when a target is similar enough to the perceiver to make simulation an appropriate basis for understanding that person's mind.

Conclusion

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 who 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 influence 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 autism or particular kinds of brain damage. Indeed, the fundamental importance of human sociability both to everyday life as well as to cultural achievements of the species prompted some observers to suggest that social cognition may have been one of the primary engines of human evolution. For example, Tomasello (1999) argued convincingly that what sets *Homo sapiens* apart from other primates is the ability to 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. Using neuroimaging and neuropsychological methods, researchers have begun to provide triangulating 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.

Notes

1. Of course, the ability to use neuroimaging and patients with neuropsychological disorders to address issues of theoretical relevance in this way relies on the following two assumptions: first, that two different brain regions cannot give rise to precisely the same cognitive process, and second, that no single brain region instantiates multiple such processes. Although neither of these assumptions is compelled philosophically or empirically, nor has serious challenge been posed to these two foundations of the cognitive neuroscience enterprise.

2. A somewhat counterintuitive finding from this study was that the medial PFC was not engaged by judging the personality characteristics of another person (George W. Bush), despite consistent observations that this region accompanies similar social-cognitive tasks. One possibility is that the personality of famous figures can be judged in an abstract, semantic manner that does not involve self-referencing. For instance, one may know that a well-known politician is dishonest or unintelligent in much the same way that one has semantic knowledge that grapes grow on vines and can be pressed into wine. Recent research suggested that the region of medial PFC implicated in self-referencing is also engaged when making comparable judgments about the personality of a significant other (Schmitz et al., 2004).

References

Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Science*, 3(12), 469-479.

- Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, 11, 231-239.
- Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge: MIT Press.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42(2), 241-251.
- Blakemore, S. J., Winston, J., & Frith, U. (2004). Social cognitive neuroscience: Where are we heading? *Trends in Cognitive Science*, 8(5), 216-222.
- Byrne, R. E. & Whiten, A. (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford: Oxford University Press.
- Caramazza, A. & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, 10, 1-34.
- Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, 2, 913-919.
- Cohen, C. E. & Ebbesen, E. B. (1979). Observational goals and schema activation: A theoretical framework for behavior perception. *Journal of Experimental Social Psychology*, 15(4), 305-329.
- Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 617-684.
- Damasio, A. R. (1994). *Descartes' Error*. New York: Grosset/Putnam.
- Davis, M. & Stone, T. (1995a). *Folk Psychology: Readings in Mind and Language*. Oxford: Blackwell.
- Davis, M. & Stone, T. (1995b). *Mental Simulation: Readings in Mind and Language*. Oxford: Blackwell.
- Dennett, D. C. (1987). *The Intentional Stance*. Cambridge: MIT Press.
- Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from daguerreotype to laser-photo. *Journal of Personality and Social Psychology*, 63(6), 877-889.
- Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, 57(2), 109-128.
- Frith, C. D. & Frith, U. (1999). Interacting minds-A biological basis. *Science*, 286, 1692-1695.
- Gabrieli, J. D. E., Fleischman, D. A., Keane, M. M., Reminger, S. L., & Morrell, F. (1995). Double dissociation between memory systems underlying explicit and implicit memory in the human brain. *Psychological Science*, 6, 76-82.
- Gallagher, H. L. & Frith, C. D. (2003). Functional imaging of "theory of mind." *Trends in Cognitive Science*, 7(2), 77-83.

- Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia*, 38, 11-21.
- Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage*, 16(3 Pt 1), 814-821.
- Gallup, G. G., Jr. (1985). Do minds exist in species other than our own? *Neuroscience and Biobehavioral Reviews*, 9(4), 631-641.
- Gilbert, D. T. & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117, 21-38.
- Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *Neuroreport*, 6(13), 1741-1746.
- Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., et al. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. *Brain*, 125(Pt 4), 752-764.
- Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Sciences*, 98, 4259-4264.
- Hamilton, D. L., Driscoll, D. M., & Worth, L. T. (1989). Cognitive organization of impressions: Effects of incongruity in complex representations. *Journal of Personality and Social Psychology*, 57, 925-939.
- Hamilton, D. L., Katz, L. B., & Leirer, V. O. (1980). Cognitive representation of personality impressions: Organizational processes in first impression formation. *Journal of Personality and Social Psychology*, 39(1 suppl 6), 1050-1063.
- Hartwick, J. (1979). Memory for trait information: A signal detection analysis. *Journal of Experimental Social Psychology*, 15(6), 533-552.
- Hastie, R. & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology*, 37(1), 25-38.
- Hauk, O., Johnsrude, I., & Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41(2), 301-307.
- Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., & Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain*, 125(Pt 8), 1808-1814.
- Kamil, A. C. (2004). Sociality and the evolution of intelligence. *Trends in Cognitive Science*, 8(5), 195-197.
- Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, 14, 785-794.
- Kusslyn, S. M. (1994). *Image and Brain: The Resolution of the Mental Imagery Debate*. Cambridge: MIT Press.

- Macrae, C. N., Moran, J. M., Heatherton, T. F., Banfield, J. F., & Kelley, W. M. (2004). Medial prefrontal activity predicts memory for self. *Cerebral Cortex*, 14(6), 647-654.
- Martin, A. (2001). Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of Functional Neuroimaging of Cognition* (pp. 153-186). Cambridge: MIT Press.
- Mason, M. F., Banfield, J. F., & Macrae, C. N. (2004). Thinking about actions: The neural substrates of person knowledge. *Cerebral Cortex*, 14(2), 209-214.
- Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (in press-a). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *NeuroImage*.
- Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (in press-b). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*.
- Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences*, 99, 15238-15243.
- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2004). Encoding specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience*, 24(21), 4912-4917.
- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (in press). Forming impressions of people versus inanimate objects. Social-cognitive processing in the medial prefrontal cortex. *NeuroImage*, 26, 251-257.
- Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., et al. (2002). Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage*, 16(4), 1159-1164.
- Ostrom, T. M. (1984). The sovereignty of social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of Social Cognition* (vol. 1, pp. 1-37). Hillsdale, NJ: Erlbaum.
- Parker, S. T. & Gibson, K. R. (1990). *"Language" and Intelligence in Monkeys and Apes: Comparative Developmental Perspectives*. Cambridge: Cambridge University Press.
- Povinelli, D. J., Parks, K. A., & Novak, M. A. (1991). Do rhesus monkeys (*Macaca mulatta*) attribute knowledge and ignorance to others? *Journal of Comparative Psychology*, 105(4), 318-328.
- Pylyshyn, Z. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, 88(1), 16-45.
- Pylyshyn, Z. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Science*, 7(3), 113-118.
- Ristau, C. A. (1991). *Cognitive Ethology: The Minds of Other Animals: Essays in honor of Donald R. Griffin*. Mahwah, NJ: Erlbaum.
- Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 35(9), 677-688.

- Salmon, E., Garraux, G., Delbeuck, X., Collette, F., Kalbe, E., Zuendorf, G., et al. (2003). Pre-dominant ventromedial frontopolar metabolic impairment in frontotemporal dementia. *Neuroimage*, 20(1), 435-440.
- Schacter, D. L. & Buckner, R. L. (1998). Priming and the brain. *Neuron*, 20, 185-195.
- Schmitz, T. W., Kawahara-Baccus, T. N., & Johnson, S. C. (2004). Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *Neuroimage*, 22(2), 941-947.
- Srull, T. K. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning and Memory*, 7(6), 440-463.
- Srull, T. K. & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review*, 96, 58-83.
- Tager-Flusberg, H., Boshart, J., & Baron-Cohen, S. (1998). Reading the windows to the soul: Evidence of domain-specific sparing in Williams syndrome. *Journal of Cognitive Neuroscience*, 10(5), 631-639.
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.
- Tomasello, M. & Call, J. (1997). *Primate Cognition*. Oxford: Oxford University Press.
- Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, 27(4), 635-657.
- Wyer, R. S., Bodenhausen, G. V., & Srull, T. K. (1984). The cognitive representation of persons and groups and its effect on recall and recognition memory. *Journal of Experimental Social Psychology*, 20, 445-469.
- Zysset, S., Huber, O., Ferstl, E., & von Cramon, D. Y. (2002). The anterior frontomedian cortex and evaluative judgment: An fMRI study. *Neuroimage*, 15(4), 983-991.