Building a Social Brain

Todd F. Heatherton

Michael Gazzaniga had an immeasurable impact on my life and my career, as he had on those of the other contributors to this volume. Shortly after Mike returned to Dartmouth in the mid-1990s, he appeared at my office door and invited me to join him in producing an introductory psychology textbook. How could I know that agreeing to doing so would lead to so much more than coauthoring a book? During the next dozen or so years Mike and I enjoyed lively discussions over frequent lunches, sometimes talking academic politics, more often talking science. From these experiences I took from Mike at least three essential points about studying the mind:

• Psychologists have generally paid insufficient attention to the adaptive function of mind and behavior. According to Mike we need to be asking not “How does this aspect of the brain work?” but “What is this aspect of the brain for? What problem did it solve for our human ancestors?” Although I was always disposed to evolutionary accounts of behavior, Mike’s articulate theories on the evolution of complex behavior provided me with new ways of thinking about the issues I study. To understand the social brain requires thinking about how behavior has been constrained by social context over the long course of human evolution.

• The use of functional neuroimaging techniques, developed over the past two decades, provides researchers with the capacity to study the working brain in action, thus providing a new window for examining previously inaccessible mental states, including the phenomenological experience of self. In the late 1990s, Mike helped spearhead an effort to convince the administration to install an MRI (magnetic resonance imaging) machine in the basement of our new psychology building. Thus,
Dartmouth became the first psychology department in the world to have its own research-dedicated scanner. With Mike's encouragement and the guidance of some wonderful cognitive neuroscience collaborators, my research program has increasingly focused on using fMRI (functional magnetic resonance imaging) to study the social brain and its components.

Much of human experience occurs through the lens of interpretation. Not only do people lack direct access to the motives and thoughts of others, they have limited access to their own underlying cognitive and neural processes. Thus, much of human experience consists of making sense of events after they have occurred. The interpreter model of mind has many implications for understanding the social brain, as Mike so elegantly argued in his classic work from 1985, *The Social Brain*.

Another important lesson from Mike is that in science you should do what you like and like what you do. Mike's numerous contributions to the science of mind, such as founding journals and societies, launching new centers, and creating new fields, all capitalize on his enormous talents and the fact that he loves what he does. Moreover, Mike encourages those around him to keep focused on their big ideas and to take risks to pursue their academic goals. It is in this spirit that I present a model of the social brain that I have been developing over the past few years. It should come as no surprise that the author of *The Social Brain* has influenced much of my approach.

**Building the Social Brain**

My overall approach follows a social brain sciences perspective, which merges evolutionary theory, experimental social cognition, and cognitive neuroscience to elucidate the neural mechanisms that support social behavior (Heatherton, Macrae, & Kelley 2004). From an evolutionary perspective, the brain is an organ that has evolved over millions of years to solve problems related to survival and reproduction. Because we are a social species, humans have evolved a fundamental need to belong that encourages behaviors reflective of being good group members (Baumeister & Leary, 1995). From this perspective, the need for interpersonal attachments is a fundamental motive that has evolved for adaptive purposes. Effective groups shared food, provided mates, and helped care for offspring. As such, human survival has long depended on living within groups; punishment from the group was effectively a death sentence. Thus, the human brain is social at its core. Many of the cognitive, sensory, and perceptual systems—although not strictly social—are acutely attuned to social stimuli, such as the way in which people readily spot faces in clouds, inanimate objects, and grilled cheese sandwiches. Indeed, studies suggest that the brain gives "people" privileged status as it processes objects in the environment. For example, our work has shown that there is a distinct functional neuroanatomy for semantic judgments made about people compared to similar judgments made about other objects (Mitchell, Heatherton, & Macrae, 2002).

What do you need to make a social brain? Or what does the brain need to do to allow it to be social? Given the fundamental need to belong, there needs to be a social brain system that monitors for signs of social inclusion or exclusion and alters behavior to forestall rejection or resolve other social problems (Heatherton, in press; Heatherton & Kendi, 2009; Mitchell & Heatherton, 2009). Such a system requires four components, each of which is likely to have a discrete neural signature. First, people need self-knowledge—to be aware of their behavior so as to gauge it against societal or group norms. Thus, having a self serves an adaptive function for group living. Second, people need to understand how others are reacting to their behavior so as to predict how others will respond to them. In other words they need "theory of mind," the capacity to attribute mental states to others. This implies the need for a third mechanism, one that detects threat, especially in complex situations. Finally, there needs to be a self-regulatory mechanism for resolving discrepancies between self-knowledge and social expectations or norms, thereby motivating behavior to resolve any conflict that exists.

In this chapter I briefly explore the functional neuroanatomy associated with these four components of the social brain: self-awareness, theory of mind, threat detection, and self-regulation. Unlike many other aspects of cognition, most of what we know about the social brain has been uncovered in the last decade and a half. Fortunately, the emergence of social neuroscience has been both rapid and far-reaching, and thus, despite its infancy, this approach has resulted in a substantial number of reliable empirical findings about the social brain.
Self-Awareness

Survival in human social groups requires people to monitor their behaviors and thoughts in order to assess whether they are in keeping with prevailing group (social) norms. Social neuroscience has made excellent strides in identifying brain regions that are involved in processing information about the self (Heatherton, Macrae, & Kelley, 2004). Both neuroimaging and patient (lesion) research has identified various regions of the prefrontal cortex as being crucial for the normal functioning of self. For instance, a series of imaging studies conducted over the past ten years has documented a substantial role of the medial region of the prefrontal cortex (MPFC) in processing self-relevant information (Craik et al., 1999; Heatherton et al., 2006; Johnson et al., 2002; Kelley et al., 2002; Macrae, Moran, Heatherton, Banfield, & Kelley, 2004; Moran, Heatherton, & Kelley, 2009; Moran, Macrae, Heatherton, Wyland, & Kelley, 2006; Schmiz, Kawahara-Baccus, & Johnson, 2004; Ochsner et al., 2004). This region is more active when, for example, people report on their personality traits, make self-relevant judgments about pictures, or retrieve autobiographical memories of past events. The issue of whether the self is somehow "special" is somewhat contentious (see Gillihan & Farah, 2005), but the imaging literature is quite clear regarding tasks that involve self-awareness; imaging studies (Gusnard, 2005) show that they activate the MPFC. It is important to note that converging evidence from patient research indicates that frontal lobe lesions, particularly to the MPFC and adjacent structures, have a deleterious effect on personality, mood, motivation, and self-awareness. Patients with frontal lobe lesions show dramatic deficits in recognizing their own limbs, engaging in self-reflection and introspection, and even reflecting on personal knowledge.

I hasten to add that there is no specific "self" spot of the brain, no single brain region that is responsible for all psychological processes related to self. Rather, psychological processes are distributed throughout the brain with contributions from multiple subcomponents determining discrete mental activities that come together to give rise to the human sense of self (Turk, Heatherton, Macrae, Kelley, & Gazzaniga, 2003). From this perspective, then, the sense of self is an emergent conscious experience of ongoing neural activity that occurs in a social world.

Studying people with brain damage, who are often unaware of their deficits, supports this idea. For instance, people who have eye injuries notice that they have vision problems because visual areas of the brain notice that something is wrong. But if you damage a part of the brain responsible for vision, such as occurs for blindsight, then there is no output from that region to consider and nothing is noted as being wrong. Cooney and Gazzaniga (2003) explain this phenomenon by arguing that a left-hemisphere interpreter can make sense only of information that is available, so even though we might find the behavior of the hemi-neglect patients bizarre, they see the state of their world as perfectly normal. Studies like these show us that the experience of the brain injury patients often does not include awareness of the deficit, which supports the idea that consciousness arises as a result of the brain processes that are active at any point in time. Critically, Gazzaniga (2000) has argued that the interpreter may play a prominent role in the experience of self: "Insertion of an interpreter into an otherwise functioning brain creates many by-products. A device that begins by asking how one thing relates to another, a device that asks about an infinite number of things, in fact, and that can get productive answers to its questions cannot help but give birth to the concept of self" (page 1320).

My view of self has been highly influenced by Gazzaniga’s interpreter model. Various psychological processes in discrete brain regions are active, depending on various environmental triggers, such as food, and internal body states, such as hunger. Other cognitive processes that are also active—such as autobiographical and prospective memory—include our goals, dreams, and aspirations. Although speculative, I believe it is possible that the MPFC operates by binding together various physical experiences and cognitive operations that have implications for the self. The prefrontal cortex receives input from all sensory modalities, and is therefore the brain region where inputs from internal sources conjoin with information received from the outside world. This region may act in a metacognitive fashion to monitor all stimuli, whether internal or external, so that our conscious sense of self at any particular moment reflects a workspace determined by which brain regions are most active.

One consequence of this model is that our sense of self is limited to functioning brain circuits that support psychological activity. So, for
example, until the frontal lobes are sufficiently developed there is only a minimal sense of self-awareness. Likewise, people often are unable to understand their personal failings and inadequacies, perhaps because people cannot recognize their own incompetence because they lack the capacity and expertise to identify and recognize competent behavior (Dunning, Johnson, Ehrlinger, & Kruger, 2003). Likewise, those who are socially anxious may have a dysfunctional connection between subcortical limbic structures associated with emotion, especially the amygdala, and prefrontal regions that normally regulate them, with a resulting sense of self that is overwhelmed by negativity. Finally, those who have frontal injuries have social deficits that they are not aware of because the very brain regions necessary for theory of mind and threat detection are unavailable to inform whether the self is living up to societal or group norms, and therefore such people lack insight and have impaired social emotions (Beer, Heerey, Keltner, Scabini, & Knight, 2003).

Hence, just as Gazzaniga’s (1985, 1989) proposed interpreter gathers available evidence and tries to make sense of the world, I propose that the self serves a similar function, although it is biased to make sense of the world in a way that casts the self in a positive light (Baumeister, 1998). Indeed, substantial evidence points to a strong motivation for self-enhancement that is supported, in part, by cognitive biases and illusions. The adaptive significance of a positivity bias in an interpretive structure remains an open question, as do other questions such as why only some neurological activity contributes to a unitary experience of self.

**Theory of Mind**

In addition to recognizing our own mental states, living harmoniously in social groups requires that we be able to interpret the emotional and mental states of others (Heatherton & Krendl, 2009). For example, social emotions require that we be able to draw inferences about the emotional states of others (even if those inferences are inaccurate). For instance, to feel guilty about hurting a loved one, people need to understand that other people have feelings. Similarly, interpersonal distress results from knowing that people are evaluating you (thereby giving rise to emotions such as embarrassment), which at its core means recognizing that other people make evaluative judgments. The ability to infer the mental states of others is commonly referred to as mentalizing, or having the capacity for theory of mind (ToM). ToM enables the ability to empathize and cooperate with others, accurately interpret other people’s behavior, and even deceive others when necessary. The rapidly emerging neuroimaging literature on theory of mind has consistently implicated the MPFC as a central component of the neural systems that support mentalizing (Amodio & Frith, 2006).

Interestingly, neuroimaging research has demonstrated that the ability to mentalize relies heavily on similar neural networks engaged in processing self-relevant information, notably the MPFC. The area of greatest activity in the MPFC tends more often to be the more dorsal region in theory-of-mind studies than in self-reference studies, where the activity tends to be more ventral. Sometimes overlap between ventral and dorsal MPFC is observed when perceivers are asked to infer the mental states of targets—other people—who are most similar to them (Mitchell, Banaji, & Macrae, 2005). This finding suggests the possibility that mental simulation is engaged during theory-of-mind tasks, posing the question, “What would I do if I were that person?” This points to the possibility that the MPFC plays a similar role in both self-awareness and theory of mind. Although activity in other brain regions has been observed during ToM tasks—notably the superior temporal sulcus, the temporoparietal junction, and, less often, the amygdala—the dorsal MPFC appears to play a central role in the ability to make mental state attributions about other people. Indeed, this area reliably differentiates between people—people and people-computer interactions, even when the pairs are engaging in the same tasks. That is, “people” are given privileged status by the dorsal medial region of the prefrontal cortex as it processes information coming from the environment (see Mitchell, Heatherton, & Macrae, 2002).

**Detection of Threat**

Over the course of human evolution, a major adaptive challenge to survival was other people. Put simply, other people can be dangerous. There are two basic social threats: those from the in-group and those from the out-group. The nature of these threats is distinctly different; the major threat from the in-group is social exclusion. As mentioned earlier, humans
have a fundamental need to belong, because during the course of evolutionary history being kicked out of the group was a potentially fatal sentence. By contrast, out-group members are threatening because they want to take your group’s resources or they may even want to kill you. Thus, the social brain requires threat mechanisms that differentiate in-group from out-group, or that are differentially sensitive to the nature of the social threat. A variety of brain regions have been identified as relevant to the detection of threat, but the two most prominent regions are the amygdala and the anterior cingulate cortex. Both regions have been implicated in social cognition.

Let’s start with the out-group threat. In the social neuroscience literature, the most common area identified as relevant to threat from out-group members is the amygdala. For instance, studies have associated amygdala activity with negative response by whites to African Americans (Phelps et al., 2000; Richeson et al., 2003). People who possess stigmatizing conditions that make them seem less than human, such as the homeless, also activate regions of the amygdala (Harris & Fiske, 2006). We also have found amygdala responses to the physically unattractive or people with multiple facial piercings (Krendl et al., 2006). Considered together, it is clear that evaluating out-group members involves activity of the amygdala. So, what does the amygdala do in the social context? It has long been thought to play a special role in responding to stimuli that elicit fear (LeDoux, 1996). From this perspective, affective processing in the amygdala is a hard-wired circuit that has developed over the course of evolution to protect animals from danger. For example, much data support the notion that the amygdala is robustly activated in response to primary biologically relevant stimuli such as faces, odors, tastes, and so forth, even when these stimuli remain below the subjects’ level of reported awareness (Whalen et al., 1998). Although there are other stimuli that elicit amygdala activity—biologically relevant positive objects such as food or sexual stimuli—the key role of the amygdala in learning what to fear may explain its involvement in detection of out-group threats.

How about threat from in-group members? If humans have a fundamental need to belong, then there ought to be mechanisms for detecting inclusionary status (Leary, Tambor, Terald, & Downs, 1995). Put another way, given the importance of group inclusion, humans need to be sensitive to signs that the group might exclude them. Thus, it was perhaps not surprising that a recent study implicated brain regions commonly associated with physical pain as crucial for the experience of social pain. Specifically, Eisenberger, Lieberman, and Williams (2003) found that a region of the dorsal anterior cingulate cortex (dACC), which is well established in the literature as involved in the experience of physical pain, was responsive during a video game designed to elicit feelings of social rejection when virtual interaction partners suddenly and surprisingly stopped cooperating with the research participant. Although these findings are intriguing, they clash with prior research and theorizing on the anterior cingulate cortex. In numerous prior studies, the dACC has been most closely associated with cognitive conflict, such as occurs when expectancies are violated (Bush, Luu, & Posner, 2000) whereas activity in the ventral anterior cingulate cortex (vACC) is more typically associated with social and emotional processes. Thus, one complication in interpreting the Eisenberger, Lieberman, and Williams findings is whether the method used to induce social rejection also likely violated research participants’ expectations. Put simply, the participants expected to participate. When this did not happen, it violated expectations, producing cognitive conflict. The experiment left unanswered whether the activation patterns they observed in that study were produced by cognitive conflict or social rejection.

We sought to address this issue by designing studies that allowed for an independent examination of the neural underpinnings of social rejection and expectancy violation (Somerville, Heatherton, & Kelley, 2006). Using an elaborate cover story we led subjects to believe that they had been evaluated by others as likable or not and we asked people to make similar judgments about those who had evaluated them, without their knowing what judgments the others had made. This approach permitted a factorial analysis that examined neural responses to feedback as a function of expectancy violation (when feedback matched or did not match participants’ first impressions) and social feedback (when feedback was negative and when positive). Results revealed a double dissociation between the dorsal and ventral ACC regions. The dorsal ACC was uniquely sensitive to expectancy violations, with greater response when the fictitious feedback was inconsistent with participants’ impressions. This was true regardless of whether the feedback was a rejection
or an acceptance. Conversely, a region in the ventral ACC was uniquely sensitive to social feedback, with significantly greater response to negative feedback than positive feedback, irrespective of expectancy violations.

Of course, social rejection is a complex phenomenon that includes, for most people, a violation of the central human expectation of social inclusion. As Leary's theory dictates (Leary et al., 1995), any situation in which people act in ways that permit the possibility of social exclusion should produce cognitive conflict, in part to signal people that they need to alter their behavior to avoid rejection. Various views of the anterior cingulate cortex have proposed that it helps resolve conflict by instigating other executive processes (Botvinick, Cohen, & Carter, 2004). This suggests that the anterior cingulate cortex should be important for self-regulation, a topic I consider shortly. What is apparent is situations that elicit threats of exclusion produce a cascade of neural responses associated with negative affect and cognitive conflict, which may promote behaviors that forestall social rejection. Given the role of the ventral ACC in processing the valence of self-descriptive terms, its abnormalities in depression, and its involvement in social rejection, further exploration of this region will be especially valuable for exploring the affective basis of self.

Self-Regulation

People who defy group norms—such as by cheating, lying, or being incompetent—often experience social emotions that indicate that something is wrong. We feel embarrassed when we goof, guilty when we harm, and ashamed when we get caught. Likewise, encounters with out-group members can leave us wary or even afraid, even if we can ultimately override our prejudices and treat them fairly. The important point is that emotions that arise from social interactions serve as guides for subsequent behavior. This is what makes something like feeling guilty adaptive (Baumeister, Stillwell, & Heatherton, 1994). Feeling socially excluded, which threatens the need to belong, motivates behavior to repair social relationships. Feeling ashamed about considering cheating on our partner helps reign in temptations. In other words, social emotions promote self-regulation, which allows us to alter our behavior, make plans, choose from alternatives, focus attention on pursuit of goals, inhibit competing thoughts, and regulate social behavior (Baumeister, Heatherton, & Tice, 1994).

Neuroscience research indicates that various regions of the prefrontal cortex are responsible for the human capacity for self-regulation (see the review by Banfield, Wyland, Macrae, Mante, & Heatherton, 2004). For instance, functional neuroimaging studies have implicated the anterior cingulate cortex in decision monitoring, initiating the selection of an appropriate novel response from several alternatives, performance monitoring, action monitoring, detection or processing of response conflict, and internal cognitive control (Wyland, Kelley, Macrae, Gordon & Heatherton, 2003). More recently, we found an important role for the ACC in efforts to suppress unwanted thoughts (Mitchell et al., 2007). What we observed was that the ACC was transiently engaged following the occurrence of unwanted thoughts, whereas dorsolateral prefrontal cortex was most active during ongoing efforts to suppress those thoughts. This finding is in keeping with the important role of prefrontal regions in executive functions more generally, all of which are necessary for successful self-regulation. Since the days of Phineas Gage we have known that the damage to certain prefrontal regions is associated with poor impulse control and self-regulatory difficulties more generally.

More recently, we have been using fMRI to study self-regulatory failures such as those that occur with smoking or dietary relapse. For example, laboratory research indicates that providing high-calorie foods to chronic dieters leads them to eat a great deal more than they would if their diets were intact. By contrast, nondieters eat less because they are filled by the food. We have recently used fMRI in a food-cue-reactivity paradigm (showing nondieters and dieters pictures of tempting foods) and found substantial differences in nucleus accumbens activity as a function of the high-calorie preload (Demos, Kelley, & Heatherton, in preparation). We found that somehow dieters are able to view attractive food cues without activating reward circuitry while their diets are intact (although how they do this is currently unknown). In sharp contrast, dieters who have just drunk a large milkshake that should have induced satiety, and that eliminated a reward response among nondieters, showed much greater reward-related food cue reactivity. Studies such as these begin to provide information relevant to people's
efforts to regulate their thoughts and actions, which are key aspects of the self-regulatory component of the social brain.

Summary

In this chapter I have proposed that building a social brain requires four components, each of which involves distinct functional brain regions. First, people need self-awareness—to be aware of their behaviors so as to gauge them against societal or group norms. The available evidence indicates that ventral MPFC is especially important for the experience of self. Second, people need to have a theory of mind—to understand how others are reacting to their behavior so as to predict how others will respond to them. This capacity for theory of mind has been most closely associated with a region of the medial prefrontal cortex that is more dorsal than that observed for self-referential processing. Third, they need to be able to detect threats. Threat detection involves at least the amygdala and the anterior cingulate cortex, although the precise nature of their roles in threat detection remains somewhat unclear. For instance, the amygdala may be especially important in ambiguous situations, such as when people are anticipating negative social judgments, whereas the anterior cingulate cortex may be more important once negative feedback has been received. The fourth component is the ability to self-regulate. This involves a number of prefrontal brain regions, including the anterior cingulate cortex, the lateral prefrontal cortex, and the ventra-medial prefrontal cortex. It is possible that these areas play different roles in self-regulation failure, depending on whether the failure is related to an impaired sense of self (ventral-medial prefrontal cortex), impaired theory of mind (dorsal prefrontal cortex), or failure to detect threat or conflict (anterior cingulate cortex).

There is much yet to discover about the social brain. Three decades ago, when Mike, with Joe LeDoux, published The Integrated Mind (Gazzaniga & LeDoux, 1978), and later in his important volume The Social Brain (Gazzaniga, 1985), Mike foreshadowed current efforts to use the methods of neuroscience to understand what it means to be a member of our social species. An important contribution from Mike’s approach was his development of the interpreter concept, which has gained broad support from many areas of scientific inquiry. The interpreter idea is central to my thinking about the social brain, as it helps to reveal the interpretive nature of social interaction, for it is necessary to make sense not only of one’s own behavior but also that of other people. In closing I express my gratitude to Mike for the many lunches and the many opportunities to see his social brain in action.

Acknowledgments

The author wishes to acknowledge his collaborators in social brain sciences at Dartmouth College and recognizes the support of grants NIMH 59282, NIDA 022582, and NSF BCS 0354400.

References


Mind Matters

Horn of the head

what do you want from me?
From us? From we?
Who can I be

but you—whom I (must) deny.
Do you why?
Are you one or many?
Any that has a different voice?
Any that can offer an alternative choice?
Any that wait for me to decide
whether to automatically react, thoughtfully enact, or merely hide somewhere inside?
Any that can help clear up the confusion emerging from the neural collusion
that leads me not only to feel
that “I” am real
but to insist
that “I” exist!