

# What the Social Brain Sciences Can Tell Us About the Self

Todd F. Heatherton, C. Neil Macrae, and William M. Kelley

*Department of Psychological and Brain Sciences, Dartmouth College*

---

**ABSTRACT**—*Social brain science is an emerging interdisciplinary field that encompasses researchers who use the approaches of evolutionary psychology, social cognition, and especially neuroscience to study human social nature. The advent of brain imaging and other cognitive neuroscience methods has provided researchers with new tools to explore the social mind. We describe how these methods can be used to explore the perplexing question of self, for example, resolving long-standing debates regarding theories of self-referential memory and providing novel insights into other aspects of self.*

**KEYWORDS**—*self; self-referential memory; self-recognition; fMRI; prefrontal cortex*

---

Social brain science is an emerging field that encompasses researchers who combine approaches of evolutionary psychology or social cognition with neuroscience to study the neural underpinnings of social behavior (Adolphs, 2003). From an evolutionary perspective, the brain is an organ that has evolved over millions of years to solve problems related to survival and reproduction. Those ancestors who were able to solve survival problems and adapt to their environments were most likely to reproduce and pass along their genes. For humans, some of the most pernicious adaptive challenges involve dealing with other humans. These challenges include selecting mates, cooperating in hunting and gathering, forming alliances, competing over scarce resources, and even warring with neighboring groups. Unlike many animal species, humans are not born precocious; they require substantial effort and resources from caregivers, who themselves are reliant on other group members for survival. Therefore, behaviors such as lying, cheating, or stealing are discouraged by social norms in all societies because they decrease survival and reproduction for other group members. It has even been proposed that humans have “cheater detectors” that are highly efficient at spotting individuals who violate social contracts (Cosmides & Tooby, 2000). This dependency on group living is not unique to humans, but the nature of relations among and between in-group and out-group members is especially complex in human societies.

---

Address correspondence to Todd F. Heatherton, Department of Psychological and Brain Sciences, 6207 Moore Hall, Dartmouth College, Hanover, NH 03755; e-mail: todd.f.heatherton@dartmouth.edu.

## THE SOCIAL BRAIN

Just as certain regions of the brain seem specialized for walking, talking, and breathing, the brain has developed specialized mechanisms for dealing with the social environment. Indeed, since the 19th century, it has been known that damage to certain brain regions (e.g., medial prefrontal cortex, or MPFC, the brain region behind the middle of your forehead) interferes with social competence while not affecting competence in other domains.

More recently, there has been a growing interest among social psychologists and cognitive neuroscientists in using brain imaging to study social cognition. Using these methods, researchers have identified a number of brain regions that appear to support highly specialized social capacities, such as theory of mind (e.g., understanding other people’s mental states), social emotions (e.g., empathy), recognition of faces and their emotional expressions, judgments of trustworthiness and attractiveness, and cooperation. Collectively, such studies suggest that people are often given privileged status by the brain as it processes objects in the environment. For example, recent work has shown that different brain regions are engaged when people make meaning-based judgments about people, as opposed to similar judgments made about other objects (Mitchell, Heatherton, & Macrae, 2002). Social brain science is providing new insights into long-standing questions regarding social behavior. In the remainder of this article, we discuss the use of neuroscience methods to study various aspects of the self.

## THE SELF

A unitary sense of self that exists across time and place is a central feature of human experience, at least for most people. Understanding the nature of self—what it is and what it does—has challenged scholars for many centuries. Although most people intuitively understand what is meant by the term self, definitions have tended toward the philosophical and metaphysical. Efforts at creating more formal definitions have largely been unsuccessful, as many features of self are empirically murky, difficult to identify and assess using objective methods. Yet the phenomenological experience of self is highly familiar to everyone. So, at issue is not whether the self exists, but how best to study it.

Scientific progress in understanding the nature of self was stifled by the inherent subjectivity and ambiguity that plagued much of the early research on the topic. Although social and developmental psychologists have made considerable advances in identifying behaviors that are related to the self, our focus here is on how the rise of cognitive

neuroscience has provided powerful tools, especially functional brain imaging (e.g., position emission tomography, or PET, and functional magnetic resonance imaging, or fMRI), that can be used to study questions that were previously outside the purview of scientific investigation. In this review, we hope to demonstrate the utility of applying a social-brain-sciences approach to understanding the nature and status of self-referential processing.

### SELF-REFERENTIAL MEMORY IS NOT ORDINARY

A fundamental question about the nature of self is whether information processed with reference to self has some privileged status in the brain, or whether processing such information is functionally equivalent to processing meaning-based (i.e., semantic) information about other classes of stimuli, such as animals, sports cars, or Commonwealth nations. Put simply, is self-referential processing special in any way?

The first line of evidence in favor of the view that self is special emerged from the pioneering work of Rogers, Kuiper, and Kirker (1977), who showed that memory for previously presented trait adjectives (e.g., *happy*) was better if they had been processed with reference to the self (e.g., “does *happy* describe you?”) than if they had been processed only for their general meaning (e.g., “does *happy* mean the same as optimistic?”). This self-referential effect in memory has been demonstrated many times, although there remains vigorous debate over precisely why it occurs. For the most part, controversy has centered on whether the self-reference effect in memory denotes a special functional role played by the self in human cognition.

There have been two major competing explanations for the self-reference memory effect. The first view, favored by Rogers, is that the self is a cognitive structure that possesses special mnemonic abilities, leading to the enhanced memorability of material processed in relation to self. The contrasting view is that the self plays no unique role in cognition (i.e., no distinct structure or neural process is dedicated to self-referential processing), and that the memory enhancement that accompanies self-referential processing can be interpreted as a standard depth-of-processing effect. That is, the wealth of personal information that resides in memory encourages the elaborative encoding of material that is processed in relation to self. In turn, this elaborative encoding enhances the memorability of self-relevant information. From this perspective, the self is quite ordinary; it just elicits a particularly high degree of elaboration during encoding.

So, who is right? A frustrating feature of these competing accounts is that they are difficult to evaluate using purely behavioral measures, as they make identical predictions (i.e., enhanced memory for self-relevant material). Herein lies the tremendous advantage of using brain imaging to study this central issue.

An initial attempt to examine the neural substrates of the self-reference effect used PET. Unfortunately, there is a limit to the number of trials that can be presented using PET, and perhaps because of this, the researchers did not obtain a statistically significant self-reference effect (Craik et al., 1999). Nonetheless, their results were intriguing in that during trials involving self-referential processing, they did find distinct activations in frontal regions, notably MPFC and areas of right prefrontal cortex.

Observing the limitations of PET, we chose to use fMRI in an attempt to identify the neural signature of self-referential mental activity (Kelley et al., 2002). In a standard self-reference paradigm, participants judged trait adjectives in one of three ways: *self* (“does

the trait describe you?”), *other* (“does the trait describe George Bush?”), and *case* (“is the trait presented in uppercase letters?”). These judgments produced the expected significant differences in subsequent memory performance (i.e., best memory in the self condition and poorest memory in the case condition). More important, however, we were able to test the competing explanations that have been offered for the self-reference effect in memory. Functional imaging studies have identified multiple regions within the left prefrontal cortex that are responsive to elaborate semantic encoding. Thus, if the self-reference effect simply reflects the operation of such a process, one would expect to observe elevated levels of activation in these left prefrontal areas when traits are judged in relation to self. If, however, the effect results from the properties of a unique cognitive self, one might expect self-referential mental activity to engage brain regions that are distinct from those involved in general semantic processing. Activation in the left prefrontal region, notable for its involvement in semantic-processing tasks, did not differentiate between self and other trials. Instead, the self trials were distinctive for their selective activity in areas of MPFC, suggesting that this region might be involved in the self-referential memory effect.

But how could we know that this brain activity was responsible for the increase in memory for material encoded with reference to self? That is, activity in MPFC accompanied self-referential processing, but did this activity contribute to the formation of memories in the brain? To investigate this possibility, we measured brain activity while participants judged the relevance of a series of personality characteristics. Afterward, memory for the items was tapped in a surprise recognition task. By contrasting brain activation elicited by items that were later remembered and by those that were later forgotten, we could identify brain regions that predicted successful recognition. We found that the level of activity in MPFC during self-referential judgments predicted which items would be remembered on the surprise memory test (i.e., the greater the MPFC activity, the more likely an item was to be remembered; Macrae, Moran, Heatherton, Banfield, & Kelley, 2004). Thus, not only does activity in MPFC track with self-referential processing, but it also contributes to the formation of self-relevant memories.

The observation that MPFC plays a critical role in self-referential processing is supported by evidence from a number of sources, such as other imaging studies in which individuals make judgments about people or engage in introspection (see Macrae, Heatherton, & Kelley, in press, for a review). Moreover, the available neuropsychological evidence also confirms that MPFC plays a prominent role in self-referential processing. Impairments in the ability to self-reflect, introspect, and daydream have long been associated with damage to areas of prefrontal cortex (Stuss & Levine, 2002). We speculate that the self-reference effect in memory depends on an intact ability to be self-reflective and that neural activity in MPFC reflects the operation of just such a process. Our major point, however, is that brain imaging allowed us to test competing hypotheses that could not be discriminated by standard behavioral testing.

### SELF-RECOGNITION MAY RELY ON LEFT-HEMISPHERE PROCESSES

If selves are special, then people should be especially good at recognizing stimuli that are relevant to themselves. Indeed, the well-known cocktail-party effect in attention demonstrates that people attend to their names even when conversation is otherwise masked by

noisy surroundings. Moreover, research in social psychology has demonstrated that people show an evaluative bias in favor of objects that are related to themselves, even preferring their own initials to other letters. One physical feature that is critical to a sense of self is one's face, which plays an important role in personal identity. To operate effectively in the world, people must be able to distinguish "me" from "not me." So, do the neural structures that support face recognition also support recognition of one's own face?

The accumulated evidence from many imaging studies, as well as studies of patients with brain damage, is that face recognition relies on structures in the right cerebral hemisphere, such that damage to these areas impairs people's ability to recognize others. But is the right hemisphere similarly specialized for self-recognition? One study appeared to support this idea. Keenan, Nelson, O'Connor, and Pascual-Leone (2001) examined patients who were undergoing a clinical procedure known as the Wada test. While a cerebral hemisphere was anesthetized with sodium amytal, patients were shown a face that was a morph of their own face with that of a famous person. After the test, patients were asked whether they had been shown a picture of themselves or of someone famous. If the left hemisphere had been anesthetized, the patients were more likely to report that they had observed their own face, but following right-hemisphere anesthesia, patients were more likely to report that they had observed the famous face. The authors concluded—on the basis of this finding and other behavioral data from their lab—that the right hemisphere appears to be critical for self-recognition.

By contrast, a series of imaging studies of self-recognition (e.g., Kircher et al., 2002) showed a different pattern of results. In these studies, self-recognition was characterized by increased activity in an area of prefrontal cortex in the left hemisphere. Thus, evidence from imaging tentatively implicates the left hemisphere.

We conducted a study to test the hypothesis that recognition of familiar people (other than the self) involves primarily the right hemisphere, whereas self-recognition involves the left cerebral hemisphere (Turk et al., 2002). To investigate this possibility, we assessed the ability of a split-brain patient (a person whose hemispheres had been surgically disconnected) to recognize himself or other familiar people. The stimuli were morphed facial images of the patient and a person familiar to him. Split-brain patients afford an ideal test of potential hemispheric differences in person recognition, as morphed facial images can be presented in such a way that they are processed separately by either the left or the right hemisphere of the disconnected brain. In this particular study, we found that whereas the patient's right hemisphere showed a bias toward recognizing morphed faces as a familiar other, as would be expected from the face-processing literature, his left hemisphere displayed the opposite pattern, that is, biased recognition in favor of self. The findings from our study indicate that although self recognition can be accomplished by both hemispheres, the left hemisphere plays an expanded role in the execution of this process. The results of this experiment are theoretically important as they suggest that self-recognition may be functionally dissociable from general face processing, a finding that has important implications for contemporary models of social cognition.

### SUMMARY

There is an ongoing revolution in psychological science, a revolution characterized by increased emphases on evolutionary principles and

on biological and genetic aspects of psychological activity. The field of social brain sciences reflects this new interdisciplinary and dynamic approach to studying the mind. The problems that are studied are those that have intrigued social psychologists for decades, but the methods and theories that are used reflect recent discoveries in neuroscience. Coupled with the functional view that social psychological mechanisms serve important adaptive functions, social brain sciences are providing new insights into long-standing social psychological questions.

The social brain sciences are in their infancy, with scholars from widely diverse areas (e.g., social psychology, neuroscience, philosophy, anthropology) working together and across levels of analysis to understand fundamental questions about human social nature. At the same time, there has been rapid progress in identifying the neural basis of many social behaviors. We anticipate that this approach will continue to grow in popularity as scientists attempt to examine some of the most fascinating and previously intractable aspects of the essential social nature of human life.

---

### Recommended Reading

- Adolphs, R. (2003). (See References)  
 Gallagher, H.L., & Frith, C.D. (2003). Functional imaging of 'theory of mind.' *Trends in Cognitive Sciences*, 7, 77–83.  
 Macrae, C.N., Heatherton, T.F., & Kelley, W.M. (in press). (See References)  
 Ochsner, K., & Lieberman, M. (2001). The emergence of social cognitive neuroscience. *American Psychologist*, 56, 717–734.
- 

**Acknowledgments**—Preparation of this manuscript and the research described herein were supported in part by grants from the National Science Foundation (BCS 0072861) and the National Institute of Mental Health (MH59282 and MH6672).

### REFERENCES

- Adolphs, R. (2003). Cognitive neuroscience of human social behavior. *Nature Reviews Neuroscience*, 4, 165–178.  
 Cosmides, L., & Tooby, J. (2000). The cognitive neuroscience of social reasoning. In M.S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 1259–1270). Cambridge, MA: MIT Press.  
 Craik, F.I.M., Moroz, T.M., Moscovitch, M., Stuss, D.T., Winocur, G., Tulving, E., & Kapur, S. (1999). In search of the self: A positron emission tomography study. *Psychological Science*, 10, 26–34.  
 Keenan, J.P., Nelson, A., O'Connor, M., & Pascual-Leone, A. (2001). Self-recognition and the right hemisphere. *Nature*, 409, 305.  
 Kelley, W.T., Macrae, C.N., Wyland, C., Caglar, S., Inati, S., & Heatherton, T.F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, 14, 785–794.  
 Kircher, T.T.J., Brammer, M., Bullmore, E., Simmons, A., Bartels, M., & David, A.S. (2002). The neural correlates of intentional and incidental self processing. *Neuropsychologia*, 40, 683–692.  
 Macrae, C.N., Heatherton, T.F., & Kelley, W.M. (in press). A self less ordinary: The medial prefrontal cortex and you. In M.S. Gazzaniga (Ed.), *New cognitive neurosciences III*. Cambridge, MA: 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, 647–654.

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, USA*, 99, 15238–15243.

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, 677–688.

Stuss, D.T., & Levine, B. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review of Psychology*, 53, 401–433.

Turk, D.J., Heatherton, T.F., Kelley, W.M., Funnell, M.G., Gazzaniga, M.S., & Macrae, C.N. (2002). Mike or me? Self-recognition in a split-brain patient. *Nature Neuroscience*, 5, 841–842.