Social Cognition, Vol. 30, No. 4, 2012, pp. 367–379

SEARCHING FOR THE SELF IN MIND AND BRAIN

John F. Kihlstrom
University of California, Berkeley

Psychological theories of the self as a knowledge representation have important implications for finding the neural representation of the self or the brain module(s) that mediate self-referential thought. The view of the self as a concept suggests that we might have many context-specific selves, not just one. Similarly, there may be a number of modality-specific self-images. The autobiographical self is not a mere list of episodic memories, but more likely a narrative structure that includes the temporal and causal relations among remembered events. Although the mental representation of self is by definition unique, it is possible that its neural representation is not distinguishable from that of other persons; and that the brain modules that process information about the self also process information about other people, if not objects in general.

In his description of the stream of consciousness, James (1890/1980, p. 226) noted that "The universal conscious fact is not 'feelings and thoughts exist,' but '*I* think' and '*I* feel.'"

> The only states of consciousness that we naturally deal with are found in personal consciousnesses, minds, selves, concrete particular I's and you's [sic]…. It seems as if the elementary psychic fact were not *thought* or *this thought* or *that thought*, but *my thought*, every thought being owned.... Neither contemporaneity, nor proximity in space, nor similarity of quality and content are able to fuse thoughts together which are sundered by this barrier of belonging to different personal minds. The breaches between such thoughts are the most absolute breaches in nature. *On these terms the personal self rather than the thought might be treated as the immediate datum in psychology.* (last emphasis added; James, 1890/1980, p. 226)

I thank Jack Gallant, Stanley B. Klein, and John J. Skowronski for their comments on earlier versions of this paper.

Correspondence concerning this article should be addressed to John F. Kihlstrom, Department of Psychology, MC 1650, University of California, Berkeley, 3210 Tolman Hall, Berkeley, CA 94720-1650; E-mail: jfkihlstrom@berkeley.edu.

Certainly, personality and social psychologists would agree—just look in the subject index of any general textbook and note how many entries have the word *self* attached to them. But defining the self is another matter, as Allport admitted: "Who is the I who knows the bodily me," he asked, "who has an image of myself and sense of identity over time, who knows that I have propriate strivings?" (Allport, 1961, p. 61).

These days, we might also ask where: almost from its beginning (Stuss, 1991a, 1991b), the emerging field of social neuroscience has sought to find the neural representation of the self in the brain—or at least identify the brain modules, or systems, or networks that generate self-representations or mediate self-referential processing (e.g., Kircher & David, 2003). But unlike vision, where, what, and where appear to be independent questions (Palmer, 1999), it appears that, with regard to the self, the what question has to be answered before the where question. Without some sense of how the self and self-referential processes are represented mentally, at the psychological level of analysis, efforts to find the self in the brain will likely prove fruitless (Kihlstrom, 2010).

IMAGING THE SELF-REFERENCE EFFECT

A case in point is provided by the pioneering study of self-referential processing by Craik and his colleagues (1999), which capitalized on what is known as the self-reference effect in memory (SRE). Employing a variant on the levels-of-processing paradigm (LOP; Craik & Lockhart, 1972), Rogers, Kuiper, and Kirker (1977) found that incidental memory for items subject to self-referent processing was far superior to levels observed following a standard semantic processing task—leading them to conclude that the self is an exceptionally elaborate knowledge structure, and that self-referent processing is a distinctively rich and powerful encoding process. Employing a partial-least-squares analysis of PET imaging data, Craik et al. identified a specific pattern of activation in the right frontal lobe, corresponding to Brodmann's areas 9, 10, and 45, which appeared to be associated with self-referent processing, as opposed to phonological, semantic, or other-referent processing. These findings, in turn, were consistent with neuropsychological evidence that patients with right-hemisphere damage show a particular disturbance in self-awareness.

Unfortunately, as Craik et al. (1999) themselves recognized, the easy interpretation of the right medial prefrontal cortex (MPFC) as the seat of the self was compromised by ambiguity in the interpretation of the SRE. Shortly after the Rogers et al. paper, other investigators found that other-referent processing yielded an LOP effect roughly equivalent to the SRE—provided that the other was highly familiar to the subject (e.g., Bower & Gilligan, 1979; Keenan & Baillet, 1980). The comparison target in the Craik et al. study was a political figure who was known to the subjects only indirectly, by reputation, introducing a confound between self-reference and familiarity. Later research taking account of familiarity has generally found no difference between the brain areas activated when making judgments about oneself compared to judgments about other people who are close to oneself (e.g., Ochsner et al., 2005, Experiment 1; for a comprehensive review, see Gillihan & Farah, 2005).

The overall pattern of results might support an interpretation of the right MPFC as the neural substrate of social judgment, regardless of whether the target of the judgment is self or other. Unfortunately, another problem with the SRE calls even that conclusion into question (Klein & Kihlstrom, 1986). The self-referent (or, for that matter, the other-referent) task requires subjects to sort list items into two categories, depending on whether they are self- (or other-) descriptive or not. The comparison tasks, involving semantic, phonological, or orthographic processing, do not. As a result, self- (and other-) reference is confounded with organizational activity. When this confound is eliminated, the SRE disappears entirely. Put another way (and bluntly), the self-reference effect may have nothing to do with the self or self-reference. It is entirely possible that the MPFC is activated during self-referential processing simply because it invokes a high level of executive activity in general, rather than by virtue of self-referent or other social judgments in particular (Goldman-Rakic, Cools, & Srivastava, 1996; Miller & Cohen, 2001).

This analysis illustrates the general point (Hatfield, 2000; Kihlstrom, 2010) that the project of identifying the neural substrates of mental activity requires an accurate analysis of the subject's task at the psychological level of analysis. If the SRE is specific to self-reference, then its neural correlates will help us to answer the where question. But task performance reflects a more general set of social-cognitive processes—or does not involve social judgment at all—all bets are off. We have to know what before we can find where. If the task analysis is wrong at the psychological level, any neural correlate of task performance will be misleading.

A SELF CELL?

Logically, self-referent processing assumes the existence of a self which serves as the referent for the processing—and this consideration begs the question of how the self, itself (sorry), might be represented in the brain. There may have been a time when it seemed possible that, some day, we might identify the neuron, or more likely the cluster of neurons, which lie at the core of the neural representation of self. This localist possibility was initially inspired by the brain-stimulation studies of Penfield (1955), which seemed to indicate discrete locations in the temporal lobe that were associated with particular memories. Even before Penfield, Sherrington (1940) had discussed pontifical cells on which the neural representations of particular sensory scenes converged. However provocative it was, Penfield's proposal was initially overshadowed by Lashley's (1950) Law of Mass Action, and Hebb's (1949) notion of cell assemblies, both of which suggested that the neural representations of particular items of knowledge were distributed widely across cerebral cortex. But Penfield's idea was given new life by the discovery, through single-cell recording, of individual neurons that were responsive to particular features of a visual stimulus (Barlow, 1953; Hubel & Wiesel, 1959)—bug detectors and the like. After Hubel and Wiesel, Konorski (1967) proposed the existence of gnostic cells, or perhaps fields of adjacent cells, representing particular kinds of objects, the destruction of which could lead to category-specific agnosias. If neuroscientists could talk about mother cells and grandmother cells without embarrassment (see Lettvin's parable reprinted in Barlow, 1995; for an overview, see Gross, 2002), why not a self cell—a single unit, or a relatively small cluster of adjacent units, that is activated whenever a person thinks about himself, and which arguably would constitute the neural representation of the self?

Presumably, such a cell or field could be identified through the sorts of single-unit recordings that had made Hubel and Wiesel famous. At this point, history repeated itself, in the form of the first connectionist models of knowledge representation and processing (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986)—which held, much as Lashley and Hebb had argued, that individual items of knowledge are represented neurally as patterns of activation across the entire neocortex. Such a widely dispersed neural network would not be revealed by single-unit recording, but it might be visible to other techniques of brain-imaging.

In a monumental achievement of computational neuroscience, Kay, Gallant, and their colleagues were able to identify, from a subject's pattern of brain activity, which particular image he was looking at (Kay, Naselaris, Prenger, & Gallant, 2008). Capitalizing on the retinotopic organization of visual cortex, they measured the pattern of activation in areas V1-V3 while subjects viewed 1,750 photographs of natural objects. This data then generated a quantitative receptor-field model that correlated the activity of each voxel in the brain image with the image that the subject had been viewing. This model was then used to identify which of 120 novel photographs the subject was viewing. Across two subjects, the model was successful in an average of 82% of the instances (based on 13 repetitions of each test stimulus), and on 42% of the instances when there was only a single trial. Both figures are obviously greater than would be expected by chance alone (0.8%). When the test set was increased to 1,000 images, the performance of the model declined only slightly. Given sufficient computational power, the same sort of model could conceivably identify the specific pattern of brain activity, distributed across the entire cerebral cortex, associated with thinking about oneself, as opposed to someone else.

Considerations of power and neural plausibility shifted neuroscientific opinion decisively in favor of the distributed codes of connectionism. However, recent work by Quian Quiroga and his associates has given new life to localist ideas about memory representation (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). During exploratory procedures prior to surgical treatment of intractable epilepsy, these investigators recorded the activities of single units (or very small groups of adjacent units) in the medial temporal lobe while patients viewed pictures of famous people, objects, animals, and landmarks that had been pre-selected on the basis of interviews. They discovered, in one patient, a single unit—corresponding to a single neuron or a small, dense cluster of adjacent neurons—that responded specifically to pictures of Jennifer Aniston. In another patient, a somewhat larger unit responded specifically to pictures of Halle Berry (including one of Berry dressed in her role as Catwoman), as well as to the spelling of the actress' name. In yet another patient, pictures of the Sydney Opera House (including a picture of a Baha'i temple which the patient had previously identified as the Opera House), as well as its name. Within the limits of the clinical testing situation, these invariant neural representations appeared to be both highly specific and abstract, in that the units involved responded to different views of the object, to line drawings as well as photographs, and to names as well as pictures.

Quian Quiroga et al. (2008) did not actually argue for the existence of *Jennifer Aniston* or *Halle Berry neurons* in the brain (Bowers, 2009; Quiroga & Kreiman, 2010a, 2010b). Instead, they argued for a sparse version of connectionism, in which objects are represented by patterns of activation involving a relatively small number of units (e.g., Olshausen & Field, 2004). Still, if there is anything like a Jennifer Aniston neuron in the brain, whether localized or as a sparsely distributed cluster of units, then there is no reason not to think that there is a self neuron as well—a neuron, or a small group of neurons, that is active whenever one thinks about oneself, as opposed to one's grandmother (or Jennifer Aniston) or people in general. And if we are going to find the self in the brain, we must first have a good idea what the self looks like at the psychological level of analysis.

THE SELF AS A KNOWLEDGE REPRESENTATION

From a cognitive point of view, the self can be construed as one's mental representation of oneself—in principle, no different than one's mental representation of any other person (Kihlstrom, 1993; Kihlstrom, Beer, & Klein, 2002; Kihlstrom & Cantor, 1984; Kihlstrom et al., 1988; Kihlstrom & Klein, 1994, 1997). Put another way, the self represents one's knowledge of oneself. Accordingly, when we ask what the self looks like, from a psychological point of view, we naturally turn to psychological theories of knowledge representation to find out what knowledge representations look like. For present purposes, three of these are most relevant: self as concept, self as image, and self as memory.

THE SELF-CONCEPT

The term *self-concept* is fairly prominent in ordinary language, mostly referring to self-evaluation, but it has rarely been connected to the scientific understanding of conceptual structure offered by modern cognitive psychology. This understanding has itself undergone considerable evolutionary change, beginning with Aristotle's definition of categories as proper sets, followed by the prototype, exemplar, and theory or explanation views of conceptual representation (Murphy, 2002; Smith & Medin, 1981). For social cognition, certainly the most popular of these views was the prototype view of concepts as sets of objects related through a principle of family resemblance, and represented by a summary list of the features that are characteristic of category members (Cantor & Mischel, 1979; Rosch, 1975). Such a solution works well for natural categories such as birds and furniture, as well as for social categories like extravert and schizophrenic; but when applied to the self, it immediately brings to mind a further question: how do we abstract a summary prototype based on family resemblance when there is only one instance in the category—namely, the individual's own, personal, unique self?

One solution to this problem (in advance) was offered by symbolic interactionism. Cooley's (1902) *looking-glass self* implies that we each have as many selves as we have significant others in our social world. Similarly, Mead argued that each of us harbors as many selves as there are social roles for us to play (Mead, 1934). Perhaps the prototypical self is abstracted from these *empirical selves* (Sarbin, 1952). Alternatively, in a nod to the competing exemplar view of conceptual structure, perhaps there is no prototypical self at all—just a family of context-specific empirical selves, each represented independently in memory (Rosenberg & Gara, 1985). There may well not be a single unitary, monolithic self-concept; but even if there is, a focus on it may lead investigators to miss a number of context-specific self-concepts that may well be just as important to the individual's personality and social interaction—if not more so.

No matter how many self-concepts we possess, the general idea is that each of them can be represented by a list of features—which begs the next question: which features? One can describe oneself in terms of a whole host of trait adjectives—nearly 18,000, by Allport and Odbert's (1936) famous count. But not all of these are equally relevant to personality description, much less the subject's self-concept. Allport suggested that personality description be confined to a relatively small number of cardinal and central traits, and the same point might be made of the features that comprise the self-concept. Markus (1977) included in the self-schema only those traits that a subject identified as both highly self-descriptive and important to the subject's own self-concept. In fact, an argument could be made for abandoning self-descriptiveness altogether, and including in the self-concept only those attributes which subjects identify as really central to how they think about themselves (Burke, Kraut, & Dworkin, 1984). Alternatively, subjects might be asked to identify those traits that most distinguish themselves from other people (McGuire, 1984; McGuire & McGuire, 1981). Rather than asking subjects to respond to a list of adjectives provided by the experimenter, subjects might be asked to spontaneously generate their own descriptors (McGuire & Padawer-Singer, 1976).

It was for just such a purpose that PERSPACE was introduced (Kihlstrom & Cunningham, 1991; Kihlstrom, Marchese-Foster, & Klein, 1997). Inspired by Kelly's Role-Construct Repertory Test (Kelly, 1955; see also Maher, 1968; Walker & Winter, 2007), PERSPACE is a highly flexible, idiographic, computer-based procedure for mapping an individual's personal space. In one application of the program, subjects are asked to list the people who are important in his or her life. These significant others presumably provide the context in which the subject experiences him- or herself. In a second phase, these targets are presented as cues, and the subject is asked to freely describe what he or she is like with each of these individuals. In a third phase, each combination of target and descriptor is presented to the subject for a self-rating. The ratings entered into this target × descriptor similarity matrix (which, of course, can be very large) can then be subjected to a procedure like cluster analysis to reveal the subject's context-specific self-concepts (Kaufman & Rousseeuw, 2005). Whatever features these context-specific selves may have in common may be all there is to a single, unitary, monolithic self. But in any event, researchers looking for the self in the brain need to be prepared for the possibility that there is not one such self, but several.

THE SELF-IMAGE

In addition to a self-concept (or, perhaps, a family of self-concepts), we also have a self-image—not, as the term so often designates, an overall positive or negative evaluation of ourselves, but an actual mental image of what we look and sound like (etc.). The distinction between the self-concept and the self-image is, essentially, the difference between meaning-based and perception-based knowledge representations (Anderson, 2000). The self-concept, as a meaning-based representation, is essentially verbal in nature, and can be assessed in terms of trait adjectives and the like. By contrast, the self-image would represent the one's own physical features and the spatio-temporal relations among them.

The self-image mediates visual self-recognition in infants (Bertenthal & Fischer, 1978) as well as adults' preferences for mirror-reversed photographs of themselves (Mita, Dermer, & Knight, 1977). That the self-image extends below the neck is suggested by the body-image-ratings of eating-disordered patients, which are often highly discrepant from reality (e.g., Fallon & Rozin, 1985; Rozin & Fallon, 1988). The existence of such self-images is also suggested by the delusions of patients with body dysmorphic disorder (Phillips, 1991) and other forms of somatization disorder (Kihlstrom & Canter Kihlstrom, 1999). Nor is the self-image exclusively visual. We recognize our own voices as well as our own faces (Holzman & Rousey, 1966; Holzman, Rousey, & Snyder, 1966; Hughes & Nicholson, 2010). Recognition of our own voices may be somewhat less accurate than recognition of others' voices, and we may be distressed at the recorded sound of our own voices. After all, the perception-based mental representation of our own voice is a product of both air-conduction and bone-conduction, and so may differ from a recorded stimulus that comes from an external source. Something like an olfactory self-image presumably mediates the priming effect of one's own body odor on mirror self-recognition (Platek, Thomson, & Gallup, 2004). Autotopagnosia, anosognosia, phantom limbs, and other neurological disorders also suggest the existence of a postural or body schema that is, essentially, a mental representation of one's own body, and its constituent parts, in space—in the words of Head and Holmes (Head & Holmes, 1911), an "organized model of ourselves" (p. 189; see also Gallagher, 2005; Haggard & Wolpert, 2005).

The point is that just as there may be many context- or role-specific self-concepts there may be many modality-specific self-images. We may also possess memory-based self-images, which record our representations of ourselves in childhood or adolescence, before a weight gain or loss, or before old age set in; and aspirational" self-images, which represent what we wish we looked, or sounded like, each of which may be mediated by different brain modules or networks.

THE AUTOBIOGRAPHICAL SELF

The self-concept and the self-image are both relatively abstract and static—even when they are bound by time or place: "This is what I was like before I got married"; "This is what I looked like before I went bald." As such, the self-concept and self-image may be thought of as represented in semantic memory as meaning-based or perception-based representations of oneself. But the self also includes the record of one's actions and experiences: "I saved the child from drowning"; "A bully kicked sand in my face at the beach." In fact, John Locke famously identified the self with memory: arguing that a person's identity includes whatever of his past he can remember (Locke, 1690/1959, Book II, Chapter 27). For Locke, self-knowledge—like all knowledge—is derived a posteriori from experience. For Hume (1739–1740/1978), the self includes not only the events and experiences that we remember, but also the events that we know or believe happened, based on inferences from what we do remember. Either way, a person who remembered nothing of his past literally would have no identity—no sense of self beyond that of in immediate experience (the Cartesian "I think, therefore I am").

We now know that, in at least some important sense, Locke and Hume were wrong. Amnesic patients, who cannot remember events that occurred after their brain damage, nonetheless retain a sense of identity: they know who they are, where they came from, and what they are like even if they have no postmorbid memories. Of course, some of this self-knowledge may be based on their recollections of premorbid experiences, which are not always degraded by the amnesic process. But some self-knowledge reflects information acquired post-morbidly: even H.M. had a sense of himself as an amnesic patient, and would freely identify himself as such—he knew that he could not remember (Hilts, 1995). More formal studies, as of patients K.C. (hippocampal amnesia; Tulving, 1993) and W.J. (temporary traumatic retrograde amnesia; Klein, Loftus, & Kihlstrom, 1996), confirm that individuals can acquire new self-knowledge from experience, even though they cannot consciously remember those experiences. Such findings lend support to the idea, based initially on research employing priming paradigms, that episodic (autobiographical) and semantic (trait) self-knowledge are represented independently of each other (Klein & Loftus, 1993; see also Klein, Robertson, Gangi, & Loftus, 2008). They also show that amnesia does not affect all declarative knowledge, but only episodic knowledge; and that amnesic patients can acquire new declarative knowledge as well as new procedural knowledge.

Autobiographical memory is often tapped by some variant of Galton's technique (Crovitz, 1970; see also Robinson, 1976; Thompson, Skowronski, Larsen, & Betz, 1996), in which subjects are asked to recall specific events related to familiar words presented as probes. But there is more to the autobiographical self than a set of disconnected episodic memories (Kihlstrom, 2009). It is, first and foremost, a chronological narrative—the story of our lives (McAdams, 1993; Neisser & Fivush, 1994; Singer & Salovey, 1993). It includes the chronological relations among events, as well as their causal relations—what caused an event to occur, and what its implications were. How we divide our lives into epochs—pre- and post- tenure, first marriage and second one, before and after children and grandchildren (Skowronski, et al., 2007). At it includes causal attributions: how we explain the course our lives took, may say as much about us as the events themselves.

Recollective experience may also be an important feature of the autobiographical self (Gardiner, 1988; Tulving, 1985). Some of our personal memories are actually remembered, accompanied by concrete awareness of oneself in the past; others are more abstract, closer to semantic memories—perhaps because they were acquired second-hand, from relatives and scrapbooks; still others are based only on an intuitive feeling that something happened; still others may reflect nothing more than beliefs about the past, in the absence of any personal recollection or independent evidence. Whatever their epistemological status, all autobiographical memories are mental representation of the past, which means that they may diverge in important ways from the historical fact of the matter. The events that we remember may not have happened the way we remember them (Neisser & Harsch, 1993); they may not have happened at all (Loftus, 2003). True or false, they are still our memories; they are still part of ourselves.

THE LIGHT'S BETTER OVER HERE

Beginning with PET and fMRI, rapid advances in neuroimaging offer the possibility that social neuroscientists will, someday soon, be able to identify the brain structures that are involved in representing and processing self-knowledge. But the success of this (or any similar) project will not depend on the strength of the magnet, the power of the image-processing algorithm, or the luck of selecting the appropriate region of interest. Rather, success will depend, first and foremost, on how we think about the self. There are likely to be different selves, depending on whether we are looking at the self-concept, the self-image, or the autobiographical self. And within each of these categories, there may be many different self-concepts and self-images, and self-narratives. And, as demonstrated by studies employing the self-reference effect, the outcome of the search will depend critically on the control task used to distinguish between self and others. The search, after all, will be for whatever in the brain is distinctive to the self.

On the other hand, we may need to be prepared for the possibility that there is no distinctive neural structure that corresponds to the self (in any of its incarnations). It may be, for example, that the self is just a person, like any other, embedded in a dense network of representations of other people. Or it may be that we cannot help thinking about others when we think about ourselves, and vice-versa—a situation that would make it impossible to isolate a module or node distinctive to the self. If there is a self node in the brain, whether localized in a single unit or sparsely distributed, it will be found only through the sheer luck of electrode placement; and it is likely to be in a different place for each individual subject. If the self is densely distributed, we will be out of luck entirely. Similarly, there may not be a distinct module that mediates self-referential processing. Thinking about the self may just be a special case of thinking about people in general; and social information-processing may not be distinct from nonsocial processing. Given all the functions that neuroimaging studies have assigned to the medial prefrontal cortex, this seems very likely. The self exists in the mind, to be sure. As James noted in the same passage as began this paper, "No psychology… can question the *existence* of personal selves" (1890/1980, p. 226). But the self may not exist in the brain in a way that permits us to find it.

REFERENCES

Allport, G. W. (1961). *Pattern and growth in personality*. New York: Holt, Rinehart, & Winston.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*, 1–171.

Anderson, J. R. (2000). *Cognitive psychology and its implications* (5th Ed.). New York: Worth Publishers, Inc.

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *Journal of Physiology*, *119*, 69–88.

Barlow, H. B. (1995). The neuron in perception. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 415–434). Cambridge, MA: MIT Press.

Bertenthal, B. I., & Fischer, K. W. (1978). Development of self-recognition in the infant. *Developmental Psychology*, *14*, 44–50.

Bower, G. H., & Gilligan, S. G. (1979). Remembering information related to one's self. *Journal of Research in Personality*, *13*, 420–432.

Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, *116*, 220–251.

Burke, P. A., Kraut, R. E., & Dworkin, R. H. (1984). Traits, consistency, and self-schemata: What do our methods measure? *Journal of Personality & Social Psychology*, *47*, 568–579.

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12, pp. 3–52). New York: Academic Press.

Cooley, C. H. (1902). *Human nature and the social order*. New York: Scribner.

Craik, F.I.M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671–684.

Craik, F.I.M., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., et al. (1999). In search of the self: A positron emission tomography study. *Psychological Science*, *10*, 26–34.

Crovitz, H. F. (1970). *Galton's walk; Methods for the analysis of thinking, intelligence and creativity*. New York: Harper & Row.

Fallon, A. E., & Rozin, P. (1985). Sex differences in perception of desirable body shape. *Journal of Abnormal Psychology*, *94*, 102–105.

Gallagher, S. (2005). *How the body shapes the mind*. New York: Oxford University Press.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313.

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin*, *131*, 76–97.

Goldman-Rakic, P. S., Cools, A. R., & Srivastava, K. (1996). The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society B*, *351*, 1445–1453.

Gross, C. G. (2002). Genealogy of the "grandmother cell." *The Neuroscientist*, *8*, 512–518.

Haggard, P., & Wolpert, D. (2005). Disorders of body scheme. In H.-J. Freund, M. Jeannerod, M. Hallett, & R. Leiguarda (Eds.), *Higher-order motor disorders: From neuroanatomy and neurobiology to clinical neurology*. Oxford: Oxford University Press.

Hatfield, G. (2000). The brain's "new" science: Psychology, neurophysiology, and constraint. *Philosophy of Science*, *67*, S388–S403.

Head, H., & Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain*, *34*, 102.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hilts, P. J. (1995). *Memory's ghost: The strange tale of Mr. M and the nature of memory*. New York: Simon & Schuster.

Holzman, P. S., & Rousey, C. (1966). The voice as a percept. *Journal of Personality & Social Psychology*, *4*, 79–86.

Holzman, P. S., Rousey, C., & Snyder, C. R. (1966). On listening to one's own voice: Effects on psychophysiological responses and free associations. *Journal of Personality & Social Psychology*, *4*, 432–441.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, *148*, 574–591.

Hughes, S. M., & Nicholson, S. E. (2010). The processing of auditory and visual recognition of self-stimuli. *Consciousness & Cognition*, *19*, 1124–1134.

Hume, D. (1739–1740/1978). *A treatise of human nature*. Oxford: Clarendon.

James, W. (1890/1980). *Principles of psychology*. Cambridge, MA: Harvard University Press.

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis* (2nd Ed.). New York: Wiley.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*, 352–355.

Keenan, J. M., & Baillet, S. D. (1980). Memory for personally and socially significant events. In R. S. Nickerson (Ed.), *Attention and performance* (Vol. 8, pp. 651–659). Hillsdale, NJ: Erlbaum.

Kelly, G. (1955). *The psychology of personal constructs*. New York: Norton.

Kihlstrom, J. F. (1993). What does the self look like? In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 5, pp. 79–90). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kihlstrom, J. F. (2009). "So that we might have roses in December": The functions of autobiographical memory. *Applied Cognitive Psychology*, *23*, 1179–1192.

Kihlstrom, J. F. (2010). Social neuroscience: The footprints of Phineas Gage. *Social Cognition*, *28*, 757–783.

Kihlstrom, J. F., Beer, J. S., & Klein, S. B. (2002). Self and identity as memory. In M. R. Leary & J. Tangney (Eds.), *Handbook of self and identity* (pp. 68–90). New York: Guilford Press.

Kihlstrom, J. F., & Cantor, N. (1984). Mental representations of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 1–47). New York: Academic.

Kihlstrom, J. F., Cantor, N., Albright, J. S., Chew, B. R., Klein, S. B., & Niedenthal, P. M. (1988). Information processing and the study of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 145–178). San Diego, CA: Academic Press, Inc.

Kihlstrom, J. F., & Canter Kihlstrom, L. (1999). Self, sickness, somatization, and systems of care. In R. J. Contrada & R. D. Ashmore (Eds.), *Self, social identity, and physical health: Interdisciplinary explorations* (Vol. 2, pp. 23-42). New York: Oxford University Press.

Kihlstrom, J. F., & Cunningham, R. L. (1991). Mapping interpersonal space. In M. Horowitz (Ed.), *Person schemas and maladaptive interpersonal patterns* (pp. 311–336). Chicago, IL: The University of Chicago Press.

Kihlstrom, J. F., & Klein, S. B. (1994). The self as a knowledge structure. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes* (Vol. 1, pp. 153–208). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kihlstrom, J. F., & Klein, S. B. (1997). Self-knowledge and self-awareness. In J. G. Snodgrass & R. L. Thompson (Eds.), *The self across psychology: Self-recognition, self-awareness, and the self concept* (pp. 5–17). New York: New York Academy of Sciences.

Kihlstrom, J. F., Marchese-Foster, L. A., & Klein, S. B. (1997). Situating the self in interpersonal space. In U. Neisser & D. A. Jopling (Eds.), *The conceptual self in context: Culture, experience, self-understanding* (pp. 154–175). New York: Cambridge University Press.

Kircher, T.T.J., & David, A. (Eds.). (2003). *The self in neuroscience and psychiatry*. Cambridge: Cambridge University Press.

Klein, S. B., & Kihlstrom, J. F. (1986). Elaboration, organization, and the self-reference effect in memory. *Journal of Experimental Psychology: General*, *115*, 26–38.

Klein, S. B., & Loftus, J. (1993). The mental representation of trait and autobiographical knowledge about the self. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 5, pp. 1–49). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Klein, S. B., Loftus, J., & Kihlstrom, J. F. (1996). Self-knowledge of an amnesic patient: Toward a neuropsychology of personality and social psychology. *Journal of Experimental Psychology: General*, *125*, 250–260.

Klein, S. B., Robertson, T. E., Gangi, C. E., & Loftus, J. (2008). The functional independence of trait self-knowledge: Commentary on Sakaki (2007). *Memory*, *16*, 556–565.

Konorski, J. (1967). *Integrative activity of the brain: An interdisciplinary approach*. Chicago: University of Chicago Press.

Lashley, K. S. (1950). In search of the engram. *Symposia of the Society for Experimental Biology*, *4*, 454–482.

Locke, J. (1690/1959). *An essay concerning human understanding*. New York: Dover.

Loftus, E. (2003). Our changeable memories: Legal and practical implications. *Nature Reviews Neuroscience*, *4*, 231–233.

Maher, B. A. (Ed.). (1968). *Clinical psychology and personality: The selected papers of George Kelly*. New York: Wiley.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personal and Social Psychology*, *35*, 63–78.

McAdams, D. P. (1993). *The stories we live by: Personal myths and the making of the self*. New York: William Morrow & Co, Inc.

McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing: Ex-*

plorations in the microstructure of cognition (Vol. 1). Cambridge, MA: MIT Press.

McGuire, W. J. (1984). Search for the self: Going beyond self-esteem and the reactive self. In R. A. Zucker, J. Aronoff, & A. I. Rabin (Eds.), *Personality and the prediction of behavior* (pp. 73–120). New York: Academic.

McGuire, W. J., & McGuire, C. V. (1981). The spontaneous self-concept as affected by personal distinctiveness. In M. D. Lynch, A. A. Norem-Hebeisen, & K. Gergen (Eds.), *Self-concept: Advances in theory and research* (pp. 147–171). New York: Ballinger.

McGuire, W. J., & Padawer-Singer, A. (1976). Trait salience in the spontaneous self-concept. *Journal of Personality and Social Psychology*, *33*, 743–754.

Mead, G. H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.

Miller, E., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Mita, T. H., Dermer, M., & Knight, J. (1977). Reversed facial images and the mere-exposure hypothesis. *Journal of Personality & Social Psychology*, *13*, 89–111.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Neisser, U., & Fivush, R. (Eds.). (1994). *The remembering self: Construction and accuracy in the self-narrative*. New York: Cambridge University Press.

Neisser, U., & Harsch, N. (1993). Phantom flashbulbs: False recollections of hearing the news about Challenger. In U. Neisser & N. Harsch (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp. 9–31). New York: Cambridge University Press.

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J.D.E., Kihlstrom, J. F., et al. (2005). The neural correlates of direct and reflected self-knowledge. *NeuroImage*, *28*, 797–814.

Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, *14*, 481–487.

Palmer, S. E. (1999). *Vision science: photons to phenomenology*. Cambridge, MA: MIT Press.

Penfield, W. (1955). The permanent record of the stream of consciousness. *Acta Psychologica*, *11*, 47–69.

Phillips, K. A. (1991). Body dysmorphic disorder: The distress of imagined ugliness. *American Journal of Psychiatry*, *148*, 1138–1149.

Platek, S. M., Thomson, J. W., & Gallup, G. G. (2004). Cross-modal self-recognition: The role of visual, auditory, and olfactory primes. *Consciousness & Cognition*, *13*, 197–210.

Quiroga, R. Q., & Kreiman, G. (2010a). Measuring sparseness in the brain: Comment on Bowers (2009). *Psychological Review*, *117*, 291–299.

Quiroga, R. Q., & Kreiman, G. (2010b). Postscript: About grandmother cells and Jennifer Aniston neurons. *Psychological Review*, *117*, 297–299.

Quiroga, R. Q., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not 'Grandmother-cell' coding in the medial temporal lobe. *Trends in Cognitive Sciences*, *12*, 87–91.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*, 1102–1107.

Robinson, J. A. (1976). Sampling autobiographical memory. *Cognitive Psychology*, *8*, 578–595.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self reference and the encoding of personal information. *Journal of Personality & Social Psychology*, *35*, 677–688.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–233.

Rosenberg, S., & Gara, M. A. (1985). The multiplicity of personal identity. *Review of Personality and Social Psychology*, *6*, 87–113.

Rozin, P., & Fallon, A. (1988). Body image, attitudes to weight, and misperceptions of figure preferences of the opposite sex: A comparison of men and women in two generations. *Journal of Abnormal Psychology*, *97*, 342–345.

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge, MA: MIT Press.

Sarbin, T. R. (1952). A preface to a psychological analysis of the self. *Psychological Review*, *59*, 11–22.

Sherrington, C. (1940). *Man on his nature*. Oxford: Oxford University Press.

Singer, J. A., & Salovey, P. (1993). *The remembered self: Emotion and memory in personality*. New York: Free Press.

Skowronski, J. J., Ritchie, T. D., Walker, W. R., Betz, A. L., Sedikides, C., Bethencourt, L. A. & Martin, A. L. (2007). Ordering our world: The quest for traces of temporal organization in autobiographical memory. *Journal of Experimental Social Psychology*, *43*, 850–856.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: MIT Press.

Stuss, D. T. (1991a). Disturbance of self-awareness after frontal-system damage. In G. Prigatano & D. L. Schacter (Eds.), *Awareness of deficit after brain injury: Clinical and theoretical issues* (pp. 63–83). New York: Oxford University Press.

Stuss, D. T. (1991b). Self, awareness, and the frontal lobes: A neuropsychological perspective. In J. Strauss & G. R. Goethals (Eds.), *The self: interdisciplinary approaches* (pp. 255–278). New York: Springer-Verlag.

Thompson, C. P., Skowronski, J. J., Larsen, S. F., & Betz, A. (1996). *Autobiographical memory: Remembering what and remembering when*. Hillsdale, NJ: Erlbaum.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, *26*, 1–12.

Tulving, E. (1993). Self-knowledge of an amnesic individual is represented abstractly. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 5, pp. 147–156). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Walker, B. M., & Winter, D. A. (2007). The elaboration of personal construct psychology. *Annual Review of Psychology*, *58*, 453–477.