Consciousness and Me-ness (Reprise)

John F. Kihlstrom
University of California, Berkeley

Article intended for a special section of Psychology of Consciousness on
“Hypotheses Regarding the Self as an Observer of as a ‘Subject of Consciousness’”

Running Head: Consciousness and the Self
Correspondence: John F. Kihlstrom
Department of Psychology
University of California, Berkeley
2121 Berkeley Way West
Berkeley, California 94720-1650

Email: jfkihlstrom@berkeley.edu
URL: https://www.ocf.berkeley.edu/~jfkihlstrom/
Consciousness and Me-Ness (Reprise)

“Think about it: There is no experience you’ve had that you were not at the absolute center of.”
-----David Foster Wallace (Wallace, 2005)

Consciousness has to do with two things: monitoring ourselves and our environment, so that we become aware of the world outside our minds (including the rest of the body) and our relation to it; and controlling ourselves and our environment, so that we voluntarily initiate and terminate various mental and behavioral activities. The monitoring function enables percepts, memories, thoughts, feelings, and desires to be represented in phenomenal awareness. It is through the controlling function that we exercise free will or agency. Conscious awareness is the prerequisite for conscious control: logically, it is difficult to see how we can consciously control things of which we are not consciously aware. It is by means of consciousness that we become aware of events, their meanings and implications, and plan and execute strategies for dealing with them.

All conscious experiences refer somehow to the self as the stimulus or experiencer of some mental state, or the agent or patient of some action (Kihlstrom, 1993a, 1997a). This insight dates back at least to William James who wrote, in his classic introspective analysis of the “Five Characters in Thought”, that “Thought tends to Personal Form” (James, 1890/1980):

[E]very thought is part of a personal consciousness…. In this room… there are a multitude of thoughts, yours and mine, some of which cohere mutually and some not…. Whether anywhere in the room there be a mere thought, which is nobody’s thought, we have no means of
ascertaining, for we have no experience of its like. The only states of consciousness that we naturally deal with are found in personal consciousnesses, minds, selves, concrete particular I’s and you’s….. It seems as if the elementary psychic fact were not thought or this thought or that thought, but my thought, every thought being owned…. On these terms the personal self rather than the thought might be treated as the immediate datum in psychology. The universal conscious fact is not “feelings and thoughts exist”, but “I think: and “I feel” (p. xxx; italics and apostrophes original).

Based on his clinical observations of cases of hysteria, Pierre Janet (Janet, 1907) had a similar insight regarding consciousness:

There are then in the “I feel”, two things in presence of each other: a small, new psychological fact, a little flame lighting up – “feel” – and an enormous mass of thoughts already constituted into a system – “I”. These two things mingle, combine; and to say “I feel” is to say that the already enormous personality has seized upon and absorbed that little, new sensation which has just been produced. The complete consciousness which is expressed by the words, “I see, I feel a movement”, is not completely represented by this little elementary phenomenon [i.e., of a sensation of vision or motion]. It contains a new term, the word “I”, which designated something very complicated. The question here is of the idea of personality, of my whole person….

And so did Claparede (Claparede, 1911/1951), based on one of the earliest observations of implicit memory and source amnesia in the amnesic syndrome (see also) (Claparede, 1911/1995; Kihlstrom, 1995a):

If one examines the behavior of such a patient, one finds that everything happens as though the various events of life, however well associated with each other in the mind, were incapable of integration with the me [i.e., the self] itself.

The point of these quotations is that consciousness and self are inextricably intertwined. All conscious mental states involve some form of self-reference: I see the painting, I hear the music; I remember where I put my car keys; I feel angry at the President, I want a hamburger. Without self-reference, there would be just a painting,
and music, and car keys, but no experience of sensing, perceiving, or remembering; without self-reference, the President and the hamburger would still be there, but the feeling and desire would not. As Searle (Searle, 1992) has rightly insisted, conscious mental states have a first-person ontology: they exist only insofar as someone experiences them. This is in contrast to other entities, such as molecules and mountains, which have a third-person ontology: they exist regardless of whether there is anyone (or anything) to observe them. For Searle, first-person subjectivity is an irreducible quality of consciousness – which means that reductionist explanations of consciousness cannot succeed because they leave its essence unexplained.

**The Self as a Knowledge Structure in Declarative Memory**

If consciousness entails self-reference, then the next question is: what is the self? Much ink has been spilled over this question, such as the distinction between self-as-object and self-as-subject (Allport, 1961; Mead, 1934), whether the self is an illusion (Kunzendorf, 1988, 2022), whether an individual has a core self as opposed to a multiplicity of selves (Markus & Nurius, 1986), and even the question of whether the self can be understood using current scientific methodologies (Klein, 2012). From a cognitive point of view, however, we can simply define the self as one’s mental representation of oneself – recording a person’s fund of knowledge of him- or herself (for comprehensive overviews and relevant references, see) (Kihlstrom, 1993b, 2012a; Kihlstrom & Cantor, 1984; Kihlstrom et al., 1988; Kihlstrom & Cunningham, 1991; Kihlstrom, Marchese-Foster, & Klein, 1997).
As such, the self is not different in kind from the mental representations of other entities that we carry around in our heads. Cognitive psychologists (e.g.,) (Anderson, 1995) generally identify two broad forms of mental representation. *Perception-based* representations which provide information about the physical appearance of an object or event; these analog representations generally take the form of mental images. By contrast, *meaning-based* representations are abstracted from perceptual details, and provide information about the meaning of an object or event; they generally take the form of sentence-like propositions.

From such a starting position, we can apply the theoretical apparatus of cognitive psychology to explicate such folk-psychological concepts as *self-concept* and *self-image*. Given what we know about other natural categories (Murphy, 2002; E. E. Smith & Medin, 1981), it seems unlikely that the self-concept is structured as a proper set of features that are singly necessary and jointly sufficient to distinguish oneself from all others. However, there are other models of conceptual structure that are more viable. The self may be structured as a prototype, whose features are only imperfectly correlated with the self; or it may be structured as a collection of exemplars, each relevant to a class of situations; or it may be structured as a theory, which not only lists characteristic features but explains (at least to the individual him- or herself) how he or she got that way. In similar fashion, far from being identified solely with self-esteem, the self-image is better construed as a perception-based mental representation of oneself – how one’s face and the rest of one’s body appears to oneself (Schilder, 1938). Mostly, the self-image is a reflection of what we see in a mirror (Mita, Dermer, & Knight, 1977),
but our mental image of ourselves can be markedly different from what we (or others) actually see (Fallon & Rozin, 1985).

For present purposes, however, it is most convenient to think of the self as a declarative knowledge structure stored in semantic memory. Cognitive psychology commonly distinguishes between declarative knowledge, which consists of factual statements about the world represented as sentence-like propositions, and procedural knowledge, which consists of motor and mental skills represented by productions consisting of a goal, a condition, and an action which will achieve the goal under the condition stated (Anderson, 1995; Winograd, 1972). In theory, productions are executed automatically provided that the relevant goals and conditions are represented in working memory (Anderson, 1992). Declarative knowledge, in turn, comes in two forms: episodic memory for events and experiences associated with a particular spatiotemporal context, and semantic memory for abstract, context-free knowledge. In terms of a generic associative-network model of memory (e.g.,) (Anderson, 1983), then, the self can be thought of as a node in semantic memory representing oneself, linked to other nodes representing characteristic physical and psychosocial features of the person.

Autobiographical memories are nodes in episodic memory representing specific events and experiences; these are linked to each other by associations representing temporal and other relations (Kihlstrom, 2009); each event-node is also linked to the self-node by associations representing the self in one of the semantic roles that a person can take: as the agent or patient of some action, or the stimulus or experiencer of some state (Brown & Fish, 1983; Fillmore, 1968). In a conscious individual, the self-
node is activated in working memory, where it can become linked to other activated nodes representing ongoing experiences, thoughts, and actions. Under various circumstances, such as repetition, the links to the self and context are weakened or dissolved, such that what once was a specific episodic memory becomes a more generic semantic one. But before that happens, any representation activated in working memory alongside the mental representation of the self will be represented in phenomenal awareness. Note that it is not activation in working memory that makes some mental state conscious; what is required is activation of a link to the self.

James on Unconscious Mental Life

It is common for people to identify consciousness with all mental life. Even William James did it, as when his original definition of psychology as “the science of mental life” (James, 1890/1980) (, p. 1) changed to “the description and explanation of states of consciousness as such (James, 1892/1980) (, p. 1). Because he identified consciousness with thought (by which he meant all forms of mental life), the notion of unconscious mental life (as opposed to unconscious brain processes) struck him as a contradiction in terms. Adopting the doctrine of esse est sentiri (to be is to be felt), he argued that the essence of consciousness (its “to be”) is to be sensed (p. 172). Mental states are felt; therefore, they cannot be unconscious. In Chapter 6 of the Principles, James went on to consider and rebut ten “proofs” of the existence of unconscious mental states (James, 1890/1980) (, pp. 162-176).

A major target of James’s critique was Hartmann (Hartmann, 1868/1931), a Romantic philosopher who had argued in a best-selling treatise that The Unconscious