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 or 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/

Abstract

The self may be construed as a cognitive structure representing a person's knowledge of him- or herself. This cognitive structure may take the form of a concept, image, or a node in an associative network of memories. Conscious states are not just represented in working memory (e.g., the "global workspace"), but must be linked to a mental representation of the self (as agent or patient, stimulus or experiencer), also represented in working memory. In unconscious mental life, as exemplified by automatic processing or explicit-implicit dissociations, this aspect of self-reference is missing, giving rise to "anoetic" effects. At the biophysical level of analysis, the self may be represented by a single "grandmother" neuron, a sparse network of neurons, or it may be widely distributed across the cortex. Viewed phylogenetically, ontogenetically, or culturally, the development of consciousness may be intimately tied to the development of the sense of self.

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 (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 aspect enables percepts, memories, thoughts, feelings, and desires to be represented in phenomenal awareness. It is through the controlling aspect that we exercise free will or agency. Conscious awareness is the prerequisite for conscious control: logically, we cannot 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, p. 221; italics and apostrophes original) :

> [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".

Based on his clinical observations of cases of hysteria, Pierre Janet (Janet, 1907, pp. 304-305) had a similar insight:

> 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 designates something very complicated. The question here is of the idea of personality, of my whole person…. 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.

And so did Claparede (Claparede, 1911/1951, p. 71), 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, music, car keys, a president, and a hamburger, 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 (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 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 concerning 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 qualitatively different from the mental representations of other entities which 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; 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 seem 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 broad class of situations; or it may be structured as a theory, which not only lists one's 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 knowledge structure stored in 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 . 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 (Tulving, 1972). In terms of a generic associative-network model of memory such as the various iterations of ACT (Anderson, 1976; Anderson et al., 2004; Anderson & Bower, 1973) or Mandler's (1980) activation-integration model of retrieval, then, the self can be thought of as a node in semantic memory representing oneself, linked to other nodes representing one's own characteristic physical and psychosocial features.

Autobiographical memories are nodes in episodic memory representing specific events and experiences; these are linked to other nodes representing the spatiotemporal context in which the event occurred, and 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 the further 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 (1868/1931), a Romantic philosopher who had argued in a best-selling treatise that The Unconscious (the initial capitals were his) pervaded the physical universe. A minor target was Helmholtz (1878/1968), who – reacting in part to Hartmann's excesses – had begun to hedge on

his notion of unconscious inference in perception. At the same time, however, James cited studies of hysteria and hypnosis by Janet (1889) and Binet (Binet & Fere, 1888) that seemed to provide clear evidence of unconscious mental life (e.g., p. 202-213). For example, he noted that, in cases of hysterical anesthesia patients who claimed to have no sense of touch would nonetheless adapt their hands to a grasped object. And in posthypnotic suggestion, subjects execute commands without knowing that they are doing so, or why. James resolved this apparent contradiction by concluding that consciousness could be divided: "*the total possible consciousness may be split into parts which coexist but mutually ignore each other,* and share the objects of knowledge between them" (p. 206). Unconscious mental life may have been a contradiction in terms for James, but he did not hesitate to write about "secondary or sub-conscious" (p. 210) selves, with its own stream of consciousness. Similarly, James's colleague Morton Prince (1905/1978, 1908) used the term *co-conscious* interchangeably with *sub-conscious* in his studies of Miss Beauchamp and other cases of multiple personality. Later, Hilgard (1977) revived James's approach in his neodissociation theory of divided consciousness (see also Kihlstrom, 1984, 1992). Hilgard proposed that, under some circumstances (like hypnotic analgesia; Hilgard, 1973), consciousness could be divided so that a subject would be unaware of a painful stimulus, even as that stimulus elicited undiminished physiological responses.

These subconscious streams of consciousness operated independently of each other, in the sense that information processed by one stream was not necessarily processed by the other(s). For all intents and purposes, these "sub-conscious" or "co-conscious" streams of mental activity are, for all intents and purposes, unconscious.

There is no contradiction, however, provided that we do not equate "consciousness" with "mental life" in the first place. Note, however, James's insistence that each stream of consciousness is associated with its own *self* – whether conscious or not.

The Unconscious and the Psychologist's Fallacy

When James wrote that "The distinction… *between the unconscious and the conscious being of the mental state*… is the sovereign means for believing what one likes in psychology, and of turning what might become a science into a tumbling-ground for whimsies" (James, 1890/1980, pp. 163-164), he was not rejecting the idea of unconscious (for him, *sub*-conscious) mental life. He was only asking for good standards for demonstrating the existence of unconscious percepts, memories, thoughts, and the like. Otherwise, we risk falling victim to a version of the psychologist's fallacy (James, 1890/1980, p. 196): imputing to subjects unconscious thoughts, memories, feelings, and desires which are not really theirs – but which they cannot deny, precisely because they are held to be unconscious.

Consider, as an illustration, the following image of psychodynamic psychotherapy from the days of the "recovered memory" epidemic:

> *Patient:* I'm afraid to be alone in the dark.
> *Therapist:* That's one of the symptoms of childhood sexual abuse.
> *Patient:* But I wasn't abused as a child.
> *Therapist:* Are you sure?
> *Patient:* Yes.
> *Therapist:* Let me tell you about the concept of repression.

This is a caricature, arguably, though some recovered-memory therapists did seem to behave that way (Kihlstrom, 1995c, 1996b, 1997b, 1998).

Freud himself made this mistake with his case of Dora, a young woman (then almost 18 years of age) who was sent by her father for treatment of a variety of neurotic and hysterical symptoms (Freud, 1905/1953). When Dora was 12, her father fell ill with tuberculosis, and was nursed by Frau K., a friend of the family (Dora's mother being busy running the household); the two eventually began an affair. Meanwhile, Dora became close to Frau K. and her children. Eventually, Frau K's husband, Herr K., began to make sexual advances on Dora. Dora complained to her father: Herr K. denied everything, and Dora's father sided with him – effectively trading her with him for Frau K.

All of this family pathology, fully accessible to Dora's consciousness, would be enough to drive any adolescent to distraction, but Freud had other ideas. Based mostly on his interpretation of Dora's dreams, Freud concluded that she was unconsciously in love with Herr K., feelings which were an unconscious substitute for her incestuous sexual desire for her father (he later concluded that Dora had also developed a homoerotic attraction to Frau K). All of this began, in Freud's view, when Dora witnessed the "primal scene" between her father and mother; she was already predisposed to hysteria by her father's syphilis (the late stages of which Freud had treated him two years previously); and her neurotic tendencies were exacerbated by a history of childhood masturbation. To the extent that Freud shared his interpretations with Dora, she rejected them. Nevertheless, he persisted, and she abruptly terminated treatment on New Year's Eve 1900, after 65 sessions spread out over three months. In his published reflections on the case, Freud reported that Dora had subsequently verified both her father's affair with Frau K. and Herr K.'s improper intentions toward

her. Still, Freud admitted only one error: he had failed to recognize the unconscious "transference" of Dora's affection for her father onto Freud himself (for comprehensive critical coverage of the "Dora" case, see Crews, 2017; Macmillan, 1991/1997).

The Automatic and the Implicit

Fast-forward 100 years, and psychology has embraced not one but two somewhat different conceptions of unconscious mental life, each associated with operational definitions that can be used to distinguish between "the conscious and the unconscious being of the mental state" (Kihlstrom, 1987, 1995b).

Automaticity

Within scientific psychology, the revival of interest in unconscious mental life began in the mid-1970s with the introduction of a distinction between *controlled and automatic processing* (LaBerge, 1974; Posner & Snyder, 1975; Schneider & Shiffrin, 1977). Conventionally, automatic processes have four canonical features: inevitable evocation by a specific class of stimuli; incorrigible completion, meaning that they cannot be stopped once initiated; efficient execution, in that they consume little or nothing by way of attentional resources; and parallel processing, meaning that they create little or no interference with other, ongoing cognitive processes (additional features have also been proposed, but appreciable consensus has developed around these four). Automatic processes are unconscious in the strict sense of the term, in that they are executed outside of phenomenal awareness and conscious control (Shiffrin, 1997).

The distinction between automatic and controlled processes stimulated the development of a number of “dual-process” theories in cognitive and social psychology (Chaiken & Trope, 1999; Kahneman, 2011; Sherman, Gawronski, & Trope, 2014; Smith & DeCoster, 2000). According to these theories, many cognitive tasks can be performed either automatically or under cognitive control – although because automatic processes are characteristically fast, unconscious automaticity generally dominates conscious control. Automaticity has become so closely identified with unconscious processing that an anthology of papers exploring various aspects of automatic processing was titled *The New Unconscious* (Hassin, Uleman, & Bargh, 2005).

Even though the concept of automaticity, and its identification with unconscious processing, has gained wide acceptance, there remains a vigorous debate over how automatic processes should be defined (Moors, 2013, 2016). The four canonical features are not perfectly correlated with each other, raising the question of whether any of them are necessary, or any subset of features that are sufficient, to define a process as automatic. And while it is convenient to think of the canonical features, and thus automaticity itself, as qualitative in nature, either present or absent in an all-or-none fashion, it seems more likely that automaticity and control anchor a continuum. At the same time, more recent research has sometimes entailed a degradation in the definition of automaticity – so that, for example, a process is defined as automatic simply because it occurs in the absence of any specific instruction. In the final analysis, it seems best to consider the four canonical features as constituting a prototype of automaticity. The more features such a process displays, the more certain we can be that it operates automatically and unconsciously.

The differential contributions of automatic and controlled processing to task performance can be estimated by means of techniques such as the process-dissociation procedure (PDP), which employs a "Method of Opposition" to pit conscious and unconscious processes against each other (Jacoby, 1991; Yonelinas & Jacoby, 2012). Consider, for example, an experiment in which subjects are presented with a list of words containing the item *ashcan*, followed by a stem-completion test. In an "Inclusion" condition, subjects are instructed to complete each stem, if possible, with a word from the study list – in this case, completing *ash_* with *ashcan*. In an "Exclusion" condition, they are instructed to complete the stems with any word *but* one from the study list – perhaps *ashtray*, *ashamed*, *ashen*, *ashore*, or *ashram*. If the subjects do not consciously recognize *ash_* as related to a studied item, the relatively unlikely response of *ashcan* may nonetheless slip into their stem-completion performance by virtue of automatic priming effects. The PDP, thus, defines automatic, unconscious processing in terms of the lack of both awareness and control. As such, the PDP provides an operational definition of automatic processing to supplement the four prototypical features of inevitable evocation, incorrigible completion, efficient execution, and parallel processing. To the extent to which automatic processing dominates controlled processing, we can say that a process is being executed unconsciously.

Explicit and Implicit Cognition

The revival of interest in unconscious mental life received another boost from the discovery of dissociations between explicit and implicit memory. Specifically, neurological patients with the amnesic syndrome, who could not remember previously

studied pictures or words, nevertheless showed priming effects when asked to identify fragmented versions of those items (Warrington & Weiskrantz, 1968). Conceptually similar effects can also be observed in posthypnotic amnesia (Kihlstrom, 2020b), and in the functional amnesias of the dissociative disorders, such as dissociative identity disorder -- formerly known as multiple personality disorder (Eich, Macaulay, Lowenstein, & Dihle, 1997; Kihlstrom, 2005). They can also be observed in neurologically intact subjects with normal memory functions, such as savings in relearning (Nelson, 1978) and various priming effects (Jacoby & Dallas, 1981). Even if subjects remember seeing some of the primes on the study list, priming is relatively independent of level of processing and other factors which exert substantial effects on conscious recollection. Following Schacter (Schacter, 1987), we can define *implicit memory* as any effect of a past event on experience, thought, or action, in the absence of, or independent of, conscious recollection of that event.

Based on the model of implicit memory, the implicit-explicit distinction has been extended to other domains of mental life (Kihlstrom, 1987, 2012b). Thus, *implicit learning* can be defined as the effect of semantic or procedural knowledge on experience, thought, or action, in the absence of conscious awareness of that knowledge (Kihlstrom, 1996a). Reber (Reber, 1967; A. S. Reber, 1992), who initially coined the term, showed that subjects could learn to identify letter strings that that conformed to an artificial grammar, without being able to specify the grammatical rules themselves. The difference between implicit learning and implicit memory is that the former term applies to semantic and procedural knowledge acquired through experience, while implicit memory applies to episodic knowledge of specific events. In

implicit memory, the subject does not consciously remember a particular experience. In implicit learning, subjects may consciously remember the learning experience, but they are not consciously aware of what they have learned.

Similarly, *implicit perception* can be defined as the effect on experience, thought, and action of a stimulus in the present environment, in the absence of conscious perception of that stimulus (Kihlstrom, 1996a; Kihlstrom, Barnhardt, & Tataryn, 1992). Implicit perception includes the processing of "subliminal" stimuli which are presented at too low an intensity or for too short a duration to be consciously perceived (Dixon, 1971, 1981). But implicit perception is a broader term, because it also includes perception of stimuli that are not technically subliminal, but have been masked in such a manner as to prevent conscious perception (Cheesman & Merikle, 1984; Eich, 1984; Greenwald, Draine, & Abrams, 1996; Marcel, 1983). Implicit perception also covers examples of "preconscious" or "preattentive" processing where the stimuli in question are in no sense subliminal, such as parafoveal vision, dichotic listening, and inattentional blindness, and what might be called attentional blindness: repetition blindness, the attentional blink, and change blindness. Dissociations between explicit and implicit perception can also be seen in hypnotic blindness and deafness, as well as neuropsychological syndromes such as blindsight, prosopagnosia, and hemispatial neglect. Implicit perception differs from implicit memory in terms of the interval between priming stimulus event and the test (stimulus-onset asynchrony, or SOA). In implicit memory, the SOA is relatively long, and frequently includes some form of distraction; in implicit perception, the interval is typically short, with no distraction, such that both prime and test occur in what James (James, 1890/1980, p. 609) called "the specious present"

(see also White, 2020). In implicit memory, the subject was aware of the priming event at the time it occurred; in implicit perception, the subject was unaware of the prime. For this reason, priming for events occurring during general anesthesia (Kihlstrom & Cork, 2017; Kihlstrom, Schacter, Cork, Hurt, & Behr, 1990), which by virtue of the SOAs involved might be counted as an instance of implicit memory, better counts as implicit perception instead.

Continuing the analogies, *implicit thought* can be defined as the effect of an internally generated idea, or image, on experience, thought, and action in the absence of conscious awareness of that idea or image (Dorfman, Shames, & Kihlstrom, 1996; Kihlstrom, Shames, & Dorfman, 1996). Unconscious processing has played a role in theories of thinking and problem-solving at least since Wallas (1926) identified five stages of thought: preparation, incubation, intimation (intuition), illumination (insight), and verification. The role of unconscious processing in incubation remains a controversial subject (Sio & Ormerod, 2009), while intuition has come to be identified with automatized, and somewhat unreliable, processes in judgment and decision-making (Kahneman, 2011; Ross, 1977; Tversky & Kahneman, 1974). Intuition began to come back into psychology's good graces, however, with research by Bowers and his colleagues (Bowers, Regehr, Balthazard, & Parker, 1990) showing that subjects could identify which of two difficult verbal or pictorial problems was soluble, without also being able to identify what the solutions were (see also Dijksterhuis & Strick, 2016; Fleck, Beeman, & Kounios, 2013; Topolinski, 2018). These intuitions seem to be mediated by priming, similar to that encountered in implicit memory, implicit perception, and the "feeling of knowing" (Hart, 1965; Nelson & Narens, 1990). The difference is that what is

primed is neither a memory – a representation of some past event – nor a percept – a representation of some object in the current stimulus environment. Rather, what is primed is pre-existing knowledge, in either verbal or imagistic form.

Explicit and Implicit Emotion and Motivation

Beyond cognition, the explicit-implicit distinction can be applied in the domains of emotion and motivation. Following the example of implicit memory, implicit emotion can be defined as the influence on experience, thought, or action of a positive or negative affective state, in the absence of conscious awareness of that state (Kihlstrom, Mulvaney, Tobias, & Tobis, 2000). A dissociation between explicit and implicit emotion was anticipated in Lang's (Lang, 1968) multicomponent theory of emotion, and Rachman and Hodgson's (Rachman & Hodgson, 1974) concept of desynchrony among subjective, behavioral, and psychophysiological measures of fear. Although actual evidence of unconscious emotion remains somewhat sparse (e.g., Feldman_Barrett, Niedenthal, & Winkielman, 2005; Winkielman & Berridge, 2004), LeDoux (1995, 2000; LeDoux & Brown, 2017) has offered a model of the brain circuits involved in fear that allows for the predicted dissociation between the subjective experience of emotion and the physiological and behavioral expressions of emotional processing. Implicit emotion, in the form of unconscious attitudes for or against some class of objects, may be tapped by the Implicit Attitudes Test (IAT; Greenwald & Banaji, 2017; Greenwald & Lai, 2020; Greenwald, McGhee, & Schwartz, 1998)

Similarly, we can define implicit motivation as the influence of a motive on experience, thought, and action in the absence of conscious awareness (Kihlstrom,

2019; McClelland, Koestner, & Weinberger, 1989). Implicit motivation is commonly measured with the Picture-Story Exercise, a variant on the Thematic Apperception Test (Murray, 1943), in which subjects write stories in response to pictorial stimuli (Schultheiss & Brunstein, 2010). As with implicit emotion, the evidence for implicit motivation is somewhat controversial, but given the evidence for various aspects of implicit cognition, the hypothesis that feelings and desires can be unconscious, as well as percepts, memories, and thoughts, cannot be dismissed out of hand.

The Paradox of H.M.

In principle at least, automatic, unconscious processing can be identified by applying aspects of the fourfold operational definition of automaticity described earlier. For example, to the extent that a process consumes little or no cognitive capacity, or creates little or no interference with other, ongoing cognitive operations, we can consider it to be automatic and unconscious. Alternatively, Jacoby's (Jacoby, 1991) PDP can be used to demonstrate that automatic processing dominates performance of a particular task.

Unfortunately, dissociations between explicit and implicit cognition, emotion, and motivation offer no such easy identification, because they all depend on self-report. For example, we construe priming as an expression of implicit memory when it occurs in the absence of conscious recall or recognition of the prime. The problem is that, as an aspect of psychology's "conscious shyness" (Flanagan, 1992) psychologists have long distrusted self-reports, and introspection in general, as a window onto the mind (Baumeister, Vohs, & Funder, 2007; Nisbett & Wilson, 1977). Jacoby's PDP is of some

help in this respect, because the inability to exclude target items depends on the lack of conscious awareness: the priming effects that constitute the bulk of evidence for implicit cognition, emotion, and motivation are themselves the products of automatic processing. But even in this case successful performance on the Exclusion task depends on self-report. The problem was resolved to the satisfaction of some theorists by studying explicit-implicit dissociations in neurological patients with the amnesic syndrome, blindsight, and the like. But that tack only raised what I think of as *the paradox of H.M.*: some psychologists only believe the self-reports of subjects who are brain-damaged. There is another way to solve the problem of self-reports, though, and that is to create conditions in which subjects will feel comfortable reporting accurately on what they perceive, remember, know, think, feel, and desire (Harris, 1988; Kihlstrom, 2002, 2020a; Orne, 1962, 1973).

The Automatic, the Implicit, and the Self

What does this have to do with the self? It appears that, in both categories of unconscious mental life (automaticity and explicit-implicit dissociations), processing does not make contact with the mental representation of the self currently activated in working memory. And because conscious awareness requires contact with some mental representation of the self, both automatic processing and expressions of implicit memory (perception, thought, emotion, motivation) occur outside conscious awareness.

In automatic processing, nodes representing the goal and condition of an action are activated in working memory, where they can make contact with the activated mental representation of the self. But, at least in Anderson's ACT model, the action

itself is executed outside working memory (Anderson, 1992). Thus, skilled drivers of standard-shift automobiles will be aware of where they are (in a car with the shift in first gear) and what they want to do (move the car forward), but they will only be dimly aware, if at all, of exactly what they are doing to accomplish this goal (ease up on the clutch slowly while pressing down gently on the accelerator). Novices consciously follow a set of steps to get the car moving, and they can tell you what they're doing while they're doing it; experts just do it automatically, and they find it hard to describe what they are doing. For a skilled driver, manual shifting is experienced as automatic, not just because it is overlearned, but because the action itself is not represented in working memory. And because it is not represented in working memory, it cannot contact the mental representation of the self which *is* active in working memory. As a result, the person is consciously aware of where s/he is, and what s/he wants to do, but is not consciously aware of what he or she is doing.

The self plays a somewhat different role in explicit-implicit dissociations. Consider first the example of memory: subjects study a list containing a word like *ashtray*, and later complete various tests of explicit memory. The encoding process will activate a representation of that word in semantic memory, which we can call the item-node, and link it to nodes representing the self (e.g., "I studied this item") and the spatiotemporal context in which the experience occurred (e.g., "This item was on the second list"). In a standard test of free recall, subjects are simply asked what items they remember from the list they just studied. Such a query will activate nodes representing the self and the spatiotemporal context in which the study phase occurred. If these

nodes are linked to the item-node for *ashcan* then the subject will recall having studied that word.

The process is somewhat reversed for recognition: presentation of the cue *ashcan* will activate the corresponding item-node in memory; if it is linked to the self and context nodes, the word will be recognized. If for any reason the link to the self-node is absent or degraded, perhaps because the word was poorly encoded to begin with, conscious recollection will fail. Nevertheless, recognition may not fail entirely. If the retrieval process contacts just the activated item-node, successful recognition may be mediated by a priming-based feeling of familiarity (Yonelinas, 2002). If it also contacts the context-node, recognition may be mediated by the abstract knowledge that the item was on the list (Kihlstrom, 2020c). But without the link between the item-node and the self-node, successful recognition will not be accompanied by conscious recollection of the study episode. It will be "anoetic", or at most "noetic", in Tulving's (1985) sense.

While explicit episodic memory – what Tulving (1985) would call "autonoetic" conscious recollection – requires a link between the representation of an event and a representation of the self as the agent, patient, stimulus, or experiencer of that event, the usual tests of implicit memory require no contact with the self-node at all. In a stem-completion, the subject may be asked to complete the stem *ash_* with a legal English word. A search of semantic memory will find nodes representing candidate words, including *ashtray* and *ashram*, but the residual activation attached to *ashcan* will facilitate that word as a response, yielding a repetition priming effect. Similarly for a fragment-completion task, in which the subject is presented with a partial word such as *a_h_a_*; or a perceptual identification test, in which the whole word ~~*ashcan*~~ is presented

in degraded form. The subjects do not have to say: "I remember that I saw the word *ashcan*". All they have to say is "That word looks like *ashcan*". In a semantic priming test, a subject might be asked to name an object found in the back yard, or even an American artistic movement: "The answer is ***ashcan"*** – again, a response worded in the third rather than first person (and in neither case the response most likely to come to mind first). The subject may have a conscious experience ("It looks like it could be *ashcan*'); but it is not a conscious recollection of the past. On a test of savings in relearning (Nelson, 1978), subjects only have to remember the *second* series of acquisition trials; they do not have to remember anything about the first series. In Claparede's (1911/1951) famous case of source amnesia, the patient didn't say "I remember that you pricked my hand" – based on conscious recollection of a specific episode; instead, she said "Sometimes people hide pins in their hands", based on generic knowledge acquired from past experience (see also) (Kihlstrom, 1995a).

Succeeding on tests of explicit memory requires that subjects answer in the first person, about their own personal pasts; succeeding on tests of implicit memory does not. The same comparison holds true for other dissociations between explicit and implicit cognition, emotion, and motivation. On a lexical decision test of implicit perception (Marcel, 1983), subjects do not say "I saw the word *doctor*; instead, they say that "*Nurse* is a legal English word" more quickly than they would in the absence of the prime. In implicit learning (Reber, 1967), they do not say "I know that if the first letter in the string is an *S*, it must be followed by a *T*; instead, they say "That string is grammatical" or "This one looks right, but the other one does not". When presented with a Dyads of Triads problem (Bowers et al., 1990), subjects say, "I don't know what

the solution is, but *this* problem is soluble while the other one is not". In the particular form of desynchrony that dissociates explicit and implicit emotion (Hodgson & Rachman, 1974), a phobic patient successfully treated with systematic desensitization can say "I'm not afraid of snakes", but when he sees one his heart rate still increases and he only begrudgingly takes his children to the reptile house at the zoo. In implicit motivation, the subject does not report, on a questionnaire, that she is highly motivated to achieve; instead, she tells a story about *someone else* who accomplishes a goal or surmounts some challenge with a combination of skill and grit. In all cases, the answer posed by the explicit test must be answered in the first person, because it requires a report of conscious experience; the response on the implicit test is formulated in the third person, because the basis for the response is unconscious. The subject is aware of telling a story, but not aware of projecting her own motives on the protagonist.

Neural Correlates of Consciousness and Selfhood

To the extent that early cognitive psychologists thought about consciousness at all (Baars & Banks, 1992), they identified it with attention and short-term memory. In the classic multistore model of memory (Atkinson & Shiffrin, 1968; Malmberg, Raaijmakers, & Shiffrin, 2019; Shiffrin, 1999), attention was the gateway to short-term memory, and pre-attentive processing was synonymous with pre-conscious processing (Pashler, 1998). Short-term memory was eventually replaced by working memory, but the identification still held (Baars & Franklin, 2003; Baddeley, 1992; Velichkovsky, 2017). Baars's (Baars, 1988) Global Workspace Theory (GWT) and its neural cognate, Global Neuronal Workspace Theory (GNWT; Dehaene, 2014), identifies consciousness

with a “global workspace”, similar to working memory, which makes the individual contents of various subordinate information-processing modules available to each other – enabling us, for example, to think about what we are perceiving. However, the viewpoint presented here is that residence in working memory, while necessary, is not sufficient for consciousness. In addition, some mental content must be linked to an activated mental representation of the self, also resident in working memory. This link provides the cognitive basis for James’s “universal conscious fact”: “*I* think, *I* feel”.

If consciousness and the self are inextricably linked, the ongoing search for the neural correlates of consciousness (Koch, Massimini, Boly, & Tononi, 2016) raises the question of the neural representation of the self (Kihlstrom, 2012a). Cognitive neuroscience has made great progress in identifying the neural substrates of working memory (D'Esposito & Postle, 2015), but has generally failed to identify any region of the brain uniquely associated with self-referential processing (Gillihan & Farah, 2005). But perhaps this is not the right way to look. If there are “grandmother”-like cells in the medial temporal lobe associated with popular-culture figures like Jennifer Aniston and Halle Berry (Bowers, 2009; Quian Quiroga, Kreiman, Koch, & Fried, 2008), surely there are similar cells associated our mental representations of ourselves, which can be identified with single-unit recording -- and lucky electrode placement. On the other hand, parallel distributed processing theories (McClelland & Rumelhart, 1986) suggest that the neural representations of particular people, like other concepts, are distributed widely across the cortex. If so, then other, computationally intensive, technologies may be needed to discern the neural representation of self; and that neural representation is likely to be different for each individual (Kay, Naselaris, Prenger, & Gallant, 2008).

Origins of Consciousness and Selfhood

The centrality of the self in consciousness suggests that a person is not fully conscious unless and until he or she has developed an internal mental representation of him- or herself – a cognitive structure that can be linked to the mental representation of ongoing experience, thought, and action. Human infants pass Gallup's (1970) mirror self-recognition test by the time they are 24 months of age (Nielsen & Dissanayake, 2004), indicating that a clear mental representation of self emerges during the second year of life. Similarly, children pass the false-belief test of a theory of mind by the time they are four or five years old (Wellman, Cross, & Watson, 2001), indicating that, at least by this time, they recognize mental states as such, and appreciate that others' mental states may be different from their own. Full-blown adult consciousness has arrived.

This does not mean that children younger than five are not conscious. In fact, Gopnik (2009) has argued that infants may be *more* conscious than grownups – partly because so much of adult thought and action is automatized, and partly because the development of attention enables adults to filter out so much of what goes on in the world around them. Moreover, infants as young as 15 months pass nonverbal versions of the false-belief test (Onishi & Baillargeon, 2005) – an age that aligns much more closely with the results from mirror self-recognition. Nor does it mean that non-human animals, who generally do not pass either mirror self-recognition (Gallup & Anderson, 2020) or the false-belief test (Call & Tomasello, 2008), lack consciousness. As de Waal (2016) has argued, the subjective world of other animals may be very different from our

own. In the final analysis, the Darwinian principle of evolutionary continuity argues against denying consciousness to nonhuman animals – at least those who have a cerebral cortex. If we have it (and we do) most of them must have something like it, too.

Perhaps human consciousness has a developmental trajectory which can be traced from both an ontogenetic and phylogenetic point of view (Gould, 1977). Something like this has been proposed by Damasio (1999, 2010), who distinguishes between unconscious neural activity and three evolutionary stages in the development of consciousness: a preconscious "protoself" which collects information about the organism's internal state, presumably including Head's "body schema" (Haggard & Wolpert, 2005; Head & Holmes, 1911; Reed & Farah, 1995), but does not necessarily give rise to any phenomenal experience; a "core consciousness" which generates qualitative feeling states such as pain; and an "extended consciousness" which depends on memory to go beyond the "here and now". Perhaps the full consciousness characteristic of neurologically intact humans (at least) arises when the person achieves the cognitive capacity to connect qualitative feeling states to the mental representation of the self, as well as to the spatiotemporal context in which the experience occurred: "*I* am seeing something now"; "*I* heard something then".

The origins of consciousness may be viewed culturally as well as phylogenetically and ontogenetically. Based largely on literary evidence, Jaynes (Jaynes, 1976) provocatively suggested that consciousness as we experience it arose in historical time – roughly speaking, late in the 2nd millennium BCE, laying the foundation for the "Axial Age" of Socrates, Gautama Buddha, and Confucius (Armstrong, 2006; Jaspers, 1949/1953). Before that time, Jaynes argues, humans

attributed their mental states to the gods; afterwards, in the wake of certain cultural changes (like the rise of cities), they understood that mental states were the product of their own minds (see also Kuijsten, 2006; for a critique, see Block, 1981). Put bluntly, the Mycenean Greeks of the *Iliad* do not think much for themselves – the gods mostly tell them what to do (Achilles is a partial exception); but Odysseus, Homer's "man of twists and turns", is always planning, always trying to deceive and outthink other people (the Trojan Horse was his idea).

In part, Jaynes's theory of the bicameral mind was inspired by then-new research on split-brain patients (Sperry, 1969; see also Gazzaniga, 1978): it is as if "the speech of the gods" (p. 105) and other hallucinations arose in the right hemisphere, and were responded to by the left hemisphere. But Jaynes never suggested that the corpus callosum evolved in historical time to connect the two hemispheres with each other – and, indeed, there does not seem to have been much structural change in the human brain since the Upper Paleolithic, 35,000 years ago (Neubauer, Hublin, & Gunz, 2019). If Jaynes had been writing just a little bit later, he might have found alternative inspiration in the developmental *theory of mind* (Lillard, 1998; Perner & Dienes, 2003; Premack & Woodruff, 1978). Perhaps the ancients had conscious experiences but did not recognize them as the product of their own minds. Perhaps at some point in historical time, again in response to cultural change, people recognized that they (and others) had minds of their own and could think for themselves. By taking ownership of their own mental states, they made the transition from "feelings and thoughts exist" to "*I* think" and "*I* feel".

References

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

Anderson, J. R. (1976). *Language, Memory, and Thought*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology, 105*(2), 165-180. doi: http://dx.doi.org/10.2307/1423026

Anderson, J. R. (1995). *Cognitive psychology and its implications* (4th ed.). New York, NY, USA: W. H. Freeman & Co, Publishers.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review, 111*(4), 1036-1060. doi: http://dx.doi.org/10.1037/0033-295X.111.4.1036

Anderson, J. R., & Bower, G. H. (1973). *Human Associative Memory*. Washington, D.C.: V. H. Winston & Sons.

Armstrong, K. (2006). *The great transformation: The beginning of our religious traditions* New York: Knopf.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-105). New York: Academic Press.

Baars, B. J. (1988). *A cognitive theory of consciousness*. New York: Cambridge University Press.

Baars, B. J., & Banks, W. P. (1992). On returning to consciousness. *Consciousness & Cognition, 1*(1), 1-2. doi: https://doi.org/10.1016/1053-8100(92)90036-A

Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Sciences, 7*(4), 166-172. doi: http://dx.doi.org/10.1016/S1364-6613(03)00056-1

Baddeley, A. D. (1992). Consciousness and working memory. *Consciousness & Cognition, 1*(1), 3-6. doi: https://doi.org/10.1016/1053-8100(92)90037-B

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? *Perspectives on Psychological Science, 2*(4), 396-403(398). doi: http://dx.doi.org/10.1111/j.1745-6916.2007.00051.x

Binet, A., & Fere, C. (1888). *Animal Magnetism*. New York: Appleton-Century.

Block, N. (1981). Review of Julian Jaynes's *The origins of consciousness in the beakdown of the bicameral mind*. *Cognition and Brain Theory, 4*(1), 81-83.

Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review, 116*(1), 220-251. doi: http://dx.doi.org/10.1037/a0014462

Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology, 22*(1), 72-110. doi: http://dx.doi.org/10.1016/0010-0285(90)90004-N

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition, 14*(3), 237-273. doi: http://dx.doi.org/10.1016/0010-0277(83)90006-9

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*(5), 187-192. doi: http://dx.doi.org/10.1016/j.tics.2008.02.010
Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York: Guilford.
Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception & Psychophysics, 36*, 387-395. doi: http://dx.doi.org/10.3758/BF03202793
Claparede, E. (1911/1951). [Recognition and me-ness]. In D. Rapaport (Ed.), *Organization and pathology of thought: Selected sources* (pp. 58-75). New York: Columbia University Press.
Claparede, E. (1911/1995). Recognition and selfhood. *Consciousness & Cognition, 4*(4), 371-378. doi: https://doi.org/10.1006/ccog.1995.1044
Crews, F. C. (2017). *Freud: The making of an illusion* (Vol. Metropolitan): New York.
D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology, 66*, 115-142. doi: http://dx.doi.org/10.1146/annurev-psych-010814-015031
Damasio, A. (2010). *Self comes to mind: Constructing the conscious brain*. New York: Pantheon.
Damasio, A. R. (1999). The Feeling of What Happens: Body and Emotion in the Making of Consciousness.
de Waal, F. (2016). *Are we smart enough to know how smart animals are?* New York: Norton.
Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. New York: Viking.
Dijksterhuis, A., & Strick, M. (2016). A Case for Thinking Without Consciousness. *Perspectives on Psychological Science, 11*(1), 117-132. doi: https://doi.org/10.1177/1745691615615317
Dixon, N. F. (1971). *Subliminal perception: The nature of a controversy*. London: McGraw-Hill.
Dixon, N. F. (1981). *Preconscious processing*. Chichester, U.K.: Wiley.
Dorfman, J., Shames, V. A., & Kihlstrom, J. F. (1996). Intuition, incubation, and insight: Implicit cognition in problem solving. In G. Underwood (Ed.), *Implicit cognition.* (pp. 257-296). Oxford: Oxford University Press.
Eich, E. (1984). Memory for unattended events: Remembering with and without awareness. *Memory and Cognition, 12*(2), 105-111. doi: http://dx.doi.org/10.3758/BF03198423
Eich, E., Macaulay, D., Lowenstein, R. J., & Dihle, P. H. (1997). Memory, amnesia, and dissociative identity disorder. *Psychological Science, 8*(6), 417-422. doi: http://dx.doi.org/10.1111/j.1467-9280.1997.tb00454.x
Fallon, A. E., & Rozin, P. (1985). Sex differences in perception of desirable body shape. *Journal of Abnormal Psychology, 94*(1), 102-105. doi: http://dx.doi.org/10.1037/0021-843X.94.1.102
Feldman_Barrett, L., Niedenthal, P. M., & Winkielman, P. (Eds.). (2005). *Emotion and consciousness*. New York: Guilford.
Fillmore, C. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart and Winston.

Flanagan, O. (1992). *Consciousness reconsidered*. Cambridge, Ma.: MIT Press.
Fleck, J. I., Beeman, M., & Kounios, J. (2013). Insight. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 779-794). New York: Oxford University Press.
Freud, S. (1905/1953). *Fragment of an analysis of a case of hysteria* (Vol. 7). London: Hogarth Press and the Institute of Psycho-analysis.
Gallup, G. G. (1970). Chimpanzees: Self-recognition. *Science, 167*(3947), 86-87. doi: http://dx.doi.org/10.1126/science.167.3914.86
Gallup, G. G., & Anderson, J. R. (2020). Self-recognition in animals: Where do we stand 50 years later? Lessons from cleaner wrasse and other species. *Psychology of Consciousness: Theory, Research, & Practice, 7*(1), 46-58. doi: http://dx.doi.org/10.1037/cns0000206
Gazzaniga, M. L., Joseph (1978). The Integrated Mind. (1978). *The integrated mind*. New York: Springer.
Gillihan, S. J., & Farah, M. J. (2005). Is Self Special? A Critical Review of Evidence From Experimental Psychology and Cognitive Neuroscience. *Psychological Bulletin, 131*(1), 76-97. doi: http://dx.doi.org/10.1037/0033-2909.131.1.76
Gopnik, A. (2009). *The philosophical baby: What childlren's minds tell us about truth, love, and the meaning of life*. New York: Farrar, Straus, & Giroux.
Gould, S. J. (1977). *Ontogeny and phylogeny*. Cambridge, Ma.: Harvard University Press.
Greenwald , A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist, 72*(9), 861 -871. doi: http://dx.doi.org/10.1037/amp0000238
Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science, 273*, 1699-1702. doi: http://dx.doi.org/10.1126/science.273.5282.1699
Greenwald, A. G., & Lai, C. K. (2020). Implicit Social Cognition. *Annual Review of Psychology, 71*, 419-445. doi: http://dx.doi.org/10.1146/annurev-psych-010419-050837
Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality & Social Psychology, 74*, 1464-1480. doi: http://dx.doi.org/10.1037/0022-3514.74.6.1464
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* (pp. 261–271). Oxford: Oxford University Press.
Harris, B. (1988). Key words: A history of debriefing in social psychology. In J. G. Morawski (Ed.), *The rise of experimentation in american psychology* (pp. 188-212). New Haven, Ct.: Yale University Press.
Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*, 208-216. doi: http://dx.doi.org/10.1037/h0022263
Hartmann, E. v. (1868/1931). *Philosophy of the unconscious: Speculative results according to the inductive method of physical science*. London: Routledge and Kegan Paul.

Hassin, R. R., Uleman, J. S., & Bargh, J. A. (Eds.). (2005). *The new unconscious*. New York: Oxford University Press.
Head, H., & Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain, 34*(2-3), 102-254. doi: http://dx.doi.org/10.1093/brain/34.2-3.102
Helmholtz, H. v. (1878/1968). The facts of perception (Originally published 1878). In R. M. Warren & R. P. Warren (Eds.), *Helmhotz on perception: Its physiology and development* (pp. 205-231). New York: Wiley.
Hilgard, E. R. (1973). A neodissociation interpretation of pain reduction in hypnosis. *Psychological Review, 80*(5), 396-411. doi: http://dx.doi.org/10.1037/h0020073
Hilgard, E. R. (1977). *Divided consciousness: Multiple controls in human thought and action*. New York: Wiley-Interscience.
Hodgson, R. E., & Rachman, S. (1974). Desynchrony in measures of fear. *Behavior Research & Therapy, 12*(4), 319-326. doi: http://dx.doi.org/10.1016/0005-7967(74)90006-0
Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language, 30*(5), 513-541. doi: http://dx.doi.org/10.1016/0749-596X(91)90025-F
Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*(3), 306-340. doi: http://dx.doi.org/10.1037/0096-3445.110.3.306
James, W. (1890/1980). *Principles of Psychology*. Cambridge, Ma.: Harvard University Press.
James, W. (1892/1980). *Psychology: Briefer course*. Cambridge, Ma.: Harvard University Press.
Janet, P. (1889). *L'automatisme psychologique [Psychological automatisms]*. Paris: Alcan.
Janet, P. (1907). *The major symptoms of hysteria*. New York: Macmillan.
Jaspers, K. (1949/1953). *The Origin and the Goal of History* (M. Bullock, Trans.). New Haven, Ct.: Yale University Press.
Jaynes, J. (1976). *The origins of consciousness in the breakdown of the bicameral mind*. Boston: Houghton Mifflin.
Kahneman, D. (2011). *Thinking fast and slow*. New York: Farrar, Straus & Giroux.
Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452*(7185), 352-355. doi: http://dx.doi.org/10.1038/nature06713
Kihlstrom, J. F. (1984). Conscious, subconscious, unconscious: A cognitive perspective. In K. S. Bowers & D. Meichenbaum (Eds.), *The unconscious reconsidered* (pp. 149-211). New York: Wiley.
Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237*(4821), 1445-1452. doi: http://dx.doi.org/10.1126/science.3629249
Kihlstrom, J. F. (1992). Dissociation and dissociations: A comment on consciousness and cognition. *Consciousness & Cognition, 1*(1), 47-53. doi: http://dx.doi.org/10.1016/1053-8100(92)90044-B
Kihlstrom, J. F. (1993a). The psychological unconscious and the self *Experimental and theoretical studies of consciousness.* (pp. 147-167). Chichester, England UK: John Wiley & Sons.

Kihlstrom, J. F. (1993b). 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, USA: Lawrence Erlbaum Associates, Inc.
Kihlstrom, J. F. (1995a). Memory and consciousness: An appreciation of Claparede and *Recognition et Moiite*. *Consciousness & Cognition: An International Journal, 4*(4), 379-386. doi: http://dx.doi.org/10.1006/ccog.1995.1045
Kihlstrom, J. F. (1995b). The rediscovery of the unconscious. In H. Morowitz & J. L. Singer (Eds.), *The mind, the brain, and complex adaptive systems.* (pp. 123-143). Reading, MA, USA: Addison-Wesley Publishing Co, Inc.
Kihlstrom, J. F. (1995c). The trauma-memory argument. *Consciousness & Cognition: An International Journal, 4*(1), 63-67. doi: http://dx.doi.org/10.1006/ccog.1995.1004
Kihlstrom, J. F. (1996a). Perception without awareness of what is perceived, learning without awareness of what is learned. In M. Velmans (Ed.), *The science of consciousness: Psychological, neuropsychological and clinical reviews.* (pp. 23-46). London, England UK: Routledge.
Kihlstrom, J. F. (1996b). The trauma-memory argument and recovered memory therapy. In K. Pezdek & W. P. Banks (Eds.), *The recovered memory/false memory debate.* (pp. 297-311). San Diego, CA, USA: Academic Press, Inc.
Kihlstrom, J. F. (1997a). Consciousness and me-ness. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 451-468). Mahwah, N.J.: Erlbaum.
Kihlstrom, J. F. (1997b). Suffering from reminiscences: Exhumed memory, implicit memory, and the return of the repressed. In M. A. Conway (Ed.), *Recovered memories and false memories.* (pp. 100-117). Oxford, England UK: Oxford University Press.
Kihlstrom, J. F. (1998). Exhumed memory. In S. J. Lynn & K. M. McConkey (Eds.), *Truth in memory* (pp. 3-31). New York: Guilford Press.
Kihlstrom, J. F. (2002). Demand characteristics in the laboratory and the clinic: Conversations and collaborations with subjects and patients. *Prevention & Treatment [Special issue honoring Martin T. Orne], 5*(1), Article_36c. doi: http://dx.doi.org/10.1037/1522-3736.5.1.536c
Kihlstrom, J. F. (2005). Dissociative disorders. *Annual Review of Clinical Psychology, 1*, 227-253. doi: http://dx.doi.org/10.1146/annurev.clinpsy.1.102803.143925
Kihlstrom, J. F. (2009). "So that we might have roses in December": The functions of autobiographical memory. *Applied Cognitive Psychology, 23*, 1179-1192. doi: http://dx.doi.org/10.1002/acp.1618
Kihlstrom, J. F. (2012a). Searching for self in mind and brain. *Social Cognition, 30*(4), 367-379. doi: http://dx.doi.org/10.1521/soco.2012.30.4.367
Kihlstrom, J. F. (2012b). Unconscious processes. In D. Reisberg (Ed.), *Oxford Handbook of Cognitive Psychology* (pp. 176-186). Oxford: Oxford University Press.
Kihlstrom, J. F. (2019). The motivational unconscious. *Personality & Social Psychology Compass, 13*(5), e12466. doi: http://dx.doi.org/10.1111/spc3.12466
Kihlstrom, J. F. (2020a). Ecological validity and "ecological validity". *Perspectives on Psychological Science*, in press.

Kihlstrom, J. F. (2020b). Posthypnotic amnesia: Using hypnosis to induce forgetting. In D. Groome & M. Eysenck (Eds.), *Forgetting: Explaining memory failure*. Thousand Oaks, Ca.: SAGE.

Kihlstrom, J. F. (2020c). Varieties of recollective experience. *Neuropsychologia*, in press. doi: https://doi.org/10.1016/j.neuropsychologia.2019.107295

Kihlstrom, J. F., Barnhardt, T. M., & Tataryn, D. J. (1992). Implicit perception. In R. F. Bornstein & T. S. Pittman (Eds.), *Perception without awareness: Cognitive, clinical, and social perspectives.* (pp. 17-54). New York, NY, USA: The 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: Social psychological studies of the self: Perspectives and programs., pp. 145-178). San Diego, CA, USA: Academic Press, Inc.

Kihlstrom, J. F., & Cork, R. C. (2017). Anesthesia. In S. Schneider & M. Velmans (Eds.), *Blackwell companion to consciousness* (2nd ed., pp. 682-694). Oxford: Blackwell.

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

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, NY, USA: Cambridge University Press.

Kihlstrom, J. F., Mulvaney, S., Tobias, B. A., & Tobis, I. P. (2000). The emotional unconscious. In E. Eich, J. F. Kihlstrom, G. H. Bower, J. P. Forgas & P. M. Niedenthal (Eds.), *Cognition and emotion* (pp. 30-86). New York: Oxford University Press.

Kihlstrom, J. F., Schacter, D. L., Cork, R. C., Hurt, C. A., & Behr, S. E. (1990). Implicit and explicit memory following surgical anesthesia. *Psychological Science, 1*(5), 303-306. doi: http://dx.doi.org/10.1111/j.1467-9280.1990.tb00222.x

Kihlstrom, J. F., Shames, V. A., & Dorfman, J. (1996). Intimations of memory and thought. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 1-23). Mahwah, N.J.: Erlbaum.

Klein, S. B. (2012). The self and science: Is it time for a new approach to the study of human experience? *Current Directions in Psychological Science, 21*(4), 253-257. doi: http://dx.doi.org/10.1177/0963721412447623

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). The neural correlates of consciousness: Progress and problems. *Nature Reviews: Neuroscience, 17*(5), 307-321. doi: https://doi.org/10.1038/nrn.2016.22

Kuijsten, M. (Ed.). (2006). *Reflections on the dawn of consciousness: Julian Janes's bicameral mind theory revisited*. Henderson, Nv.: Julian Jaynes Society.

Kunzendorf, R. G. (1988). Self-consciousness as the monitoring of cognitive states: A theoretical perspective. *Imagination, Cognition, and Personality, 7*(1), 3-22. doi: http://dx.doi.org/10.2190/TW3U-0886-Q5TA-XHM4
Kunzendorf, R. G. (2022). Source Monitoring as an Explanation for the Illusion of 'Self as Subject'. *Psychology of Consciousness*, in press.
LaBerge, D., & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293-323. doi: http://dx.doi.org/10.1016/0010-0285(74)90015-2
Lang, P. J. (1968). Fear reduction and fear behavior: Problems in treating a construct. In J. M. Schlein (Ed.), *Research in psychotherapy* (Vol. 3, pp. 90-103). Washington, D.C.: American Psychological Association.
LeDoux, J. E. (1995). Emotion: Clues from the brain. *Annual Review of Psychology, 46*, 209-235. doi: http://dx.doi.org/10.1146/annurev.ps.46.020195.001233
LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23*, 155-184. doi: http://dx.doi.org/10.1146/annurev.neuro.23.1.155
LeDoux, J. E., & Brown, R. (2017). Emotions as higher-order states of consciousness. *Proceedings of the National Academy of Sciences, 114*(10), E2016-E2025. doi: http://dx.doi.org/10.1073/pnas.1619316114
Lillard, A. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin, 123*(1), 3-32. doi: https://doi.org/10.1038/217972a0
Macmillan, M. (1991/1997). *Freud evaluated: The completed arc.* Cambridge, Ma.: MIT Press.
Malmberg, K. J., Raaijmakers, J. G. W., & Shiffrin, R. M. (2019). 50 years of research sparked by Atkinson and Shiffrin (1968). *Memory & Cognition, 47*(4), 561-574. doi: https://doi.org/10.3758/s13421-019-00896-7
Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*(3), 252-271. doi: http://dx.doi.org/10.1037/0033-295X.87.3.252
Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15*(2), 197-237. doi: http://dx.doi.org/10.1016/0010-0285(83)90009-9
Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist, 41*(9), 954-969. doi: http://dx.doi.org/10.1037/0003-066X.41.9.954
McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96*(4), 690-702. doi: http://dx.doi.org/10.1037/0033-295X.96.4.690
McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, Ma.: MIT Press.
Mead, G. H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.
Mita, T. H., Dermer, M., & Knight, J. (1977). Reversed facial images and the mere-exposure hypothesis. *Journal of Personality & Social Psychology, 13*(8), 89-111. doi: http://dx.doi.org/10.1037/0022-3514.35.8.597
Moors, A. (2013). Automaticity. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 163-175). New York: Oxford University Press.

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology, 67*, 263-287. doi: http://dx.doi.org/10.1146/annurev-psych-122414-033550
Murphy, G. L. (2002). *The big book of concepts*. Cambridge, Ma.: MIT Press.
Murray, H. A. (1943). *Thematic Apperception Test*. Cambridge, Ma.: Harvard University Press.
Nelson, T. O. (1978). Detecting small amounts of information in memory: Savings for nonrecognized items. *Journal of Experimental Psychology: Human Learning & Memory, 4*(5), 453-468. doi: http://dx.doi.org/10.1037/0278-7393.4.5.453
Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The Psychology of learning and motivation* (Vol. 26, pp. 125-173). New York: Academic.
Neubauer, S., Hublin, J.-J., & Gunz, P. (2019). The evolution of modern human brain shape. *Science Advances, 4*(1), eaao5961. doi: http://dx.doi.org/10.1126/sciadv.aao5961
Nielsen, M., & Dissanayake, C. (2004). Pretend play, mirror self-recognition and imitation: A longitudinal investigation through the second year. *Infant Behavior & Development, 27*(3), 342-365. doi: http://dx.doi.org/10.1016/j.infbeh.2003.12.006
Nisbett, R. E., & Wilson, D. S. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-253. doi: http://dx.doi.org/10.1037/0033-295X.84.3.231
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255-258. doi: http://ds.doi.org/10.1126/science.1107621
Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776-783. doi: http://dx.doi.org/10.1037/h0043424
Orne, M. T. (1973). Communication by the total experimental situation: Why it is important, how it is evaluated, and its significance for the ecological validity of findings. In P. Pliner, L. Krames & T. Alloway (Eds.), *Communication and affect* (pp. 157-191). New York: Academic.
Pashler, H. (Ed.). (1998). *Attention*. Philadelphia: Taylor & Francis.
Perner, J., & Dienes, Z. (2003). Developmental aspects of consciousness: How much theory of mind do you need to be consciously aware? *Consciousness and Cognition, 12*(1), 63-82. doi: http://dx.doi.org/10.1016/S1053-8100(02)00010-7
Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55-85). New York: Wiley.
Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral & Brain Sciences, 4*(4), 515-526. doi: http://dx.doi.org/10.1017/S0140525X00076512
Prince, M. (1905/1978). *The dissociation of a personality*. New York: Oxford University Press.
Prince, M. (1908). The Unconscious. *Journal of Abnormal Psychology, 3*(4), 261-297. doi: http://dx.doi.org/10.1037/h0073698

Quian Quiroga, R., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not `Grandmother-cell' coding in the medial temporal lobe. *Trends in Cognitive Sciences, 12*(3), 87-91. doi: http://dx.doi.org/10.1016/j.tics.2007.12.003

Rachman, S., & Hodgson, R. E. (1974). Synchrony and desynchrony in measures of fear. *Behaviour Research & Therapy, 12*(4), 311-318. doi: http://dx.doi.org/10.1016/0005-7967(74)90005-9

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior, 6*(6), 855-863. doi: http://dx.doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1992). The cognitive unconscious: An evolutionary perspective. *Consciousness & Cognition, 1*(2), 93-133. doi: http://dx.doi.org/10.1016/1053-8100(92)90051-B

Reed, C. L., & Farah, M. J. (1995). The psychological reality of the body schema: A test with normal participants. *Journal of Experimental Psychology: Human Perception & Performance, 21*, 334-343. doi: http://dx.doi.org/10.1037/0096-1523.21.2.334

Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173-220). San Diego: Academic Press.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(3), 501-518. doi: http://dx.doi.org/10.1037/0278-7393.13.3.501

Schilder, P. (1938). *Image and appearance of the human body.* London: Kegan, Paul, Trench, & Tribner.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66. doi: http://dx.doi.org/10.1037/0033-295X.84.1.1

Schultheiss, O. C., & Brunstein, J. C. (Eds.). (2010). *Implicit motives*. Oxford: Oxford University Press.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, Mass.: MIT Press.

Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.). (2014). *Dual-process theories of the social mind*. New York: Guilford.

Shiffrin, R. M. (1997). Attention, automatism, and consciousness *Scientific approaches to consciousness.* (pp. 49-64). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Shiffrin, R. M. (1999). 30 years of memory *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson-Shiffrin model.* (pp. 17-33). Mahwah, NJ, US: Lawrence Erlbaum Associates, Inc., Publishers.

Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin, 135*(1), 94-120. doi: http://dx.doi.org/10.1037/a0014212

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

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*(2), 108-131. doi: http://dx.doi.org/10.1207/S15327957PSPR0402_01

Sperry, R. W. (1969). A modified concept of consciousness. *Psychological Review, 76*(6), 532-536. doi: http://dx.doi.org/10.1037/h0028156
Topolinski, S. (2018). The sense of coherence: How intuition guides reasoning and thinking. In L. J. Ball & V. A. Thompson (Eds.), *Routledge international handbook of thinking and reasoning* (pp. 559-574). New York: Taylor& Francis.
Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press.
Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26*, 1-12. doi: http://dx.doi.org/10.1037/h0080017
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124-1131. doi: http://dx.doi.org/10.1126/science.185.4157.1124
Velichkovsky, B. B. (2017). Consciousness and working memory: Current trends and research perspectives. *Consciousness and Cognition, 55*(1), 35-45. doi: http://dx.doi.org/10.1016/j.concog.2017.07.005
Wallace, D. F. (2005). *This is Water*. Paper presented at the Commencement Address at Kenyon College. https://fs.blog/2012/04/david-foster-wallace-this-is-water/
Wallas, G. (1926). *The art of thought*. New York: Harcourt Brace.
Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature, 217*, 972-974. doi: https://doi.org/10.1038/217972a0
Wellman, H. M., Cross, D., & Watson, J. B. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655-684. doi: http://dx.doi.org/10.1111/1467-8624.00304
White, P. A. (2020). The perceived present: What is it, and what is it there for? *Psychonomic Buletin & Review, 27*(4), 583-601. doi: . Psychon Bull Rev 27, 583–601 (2020). https://doi.org/10.3758/s13423-020-01726-7
Winkielman, P., & Berridge, K. C. (2004). Unconscious Emotion. *Current Directions in Psychological Science, 13*(3), 120-123(124). doi: http://dx.doi.org/10.1111/j.0963-7214.2004.00288.x
Winograd, T. (1972). *Understanding Natural Language*. New York, N.Y.: Academic Press.
Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language, 46*(3), 441-517. doi: http://dx.doi.org/10.1006/jmla.2002.2864
Yonelinas, A. P., & Jacoby, L. L. (2012). The process-dissociation approach two decades later: Convergence, boundary conditions, and new directions. *Memory & Cognition, 40*, 663-680. doi: http://dx.doi.org/10.3758/s13421-012-0205-5