RT4053_C004.fm Page 77 Wednesday, July 21, 2004 5:39 PM

4

Conflict and Habit: A Social Cognitive Neuroscience Approach to the Self

MATTHEW D. LIEBERMAN AND NAOMI I. EISENBERGER

"But that's how biographies are. I mean, who's going to read about the peaceful life and times of a nobody employed at the Kawasaki Municipal Library"

—Haruki Murakami (1994), *Dance, Dance, Dance*

AU: Quote marks needed only if this is a character speaking within Murakami's novel.; pls. advise. Also, please add closing punctuation: question mark? period?

Novelist Haruki Murakami's claim is hardly contentious. We all prefer to read a biography full of unexpected events, tragic downfalls, and hard-won victories. They take us on a journey through which we hope to glean the character of the individual and perhaps some insight into human nature more generally. It is not that the simple life of the librarian has any fewer events filling the days or years but rather that those events follow an expected repetitive pattern with little variation over time. Indeed, there are no more hours in the day for a head of state than for Murakami's librarian, just more memorable ones.

Just as we prefer the miraculous to the mundane, the tragic to the trivial, and conflict over commonplace in choosing which biographies to read, we often rely on similar distinctions in understanding and defining ourselves. We look to those moments in our past when we were faced with obstables for which our daily routine, and the mental habits formed through this routine, could not guarantee safe passage. Should I be a lawyer or a doctor? Should I be a Democrat or Republican? Should I stand up for the student who is being picked on or keep quiet? If these situations have not been a part of one's routine, how is one to go forward? There seems to be no alternative but to "assert oneself" in these cases and use one's "free will," or at least that is how these episodes are often experienced in retrospect. These are the moments when the self seems to burst onto the scene,

0-415-93379-X/04/$0.00+$1.50

and thus these are the moments that we often take as self-defining in our own private autobiographies (Baumeister, 1986).

Why do these moments of conflict and the way we resolve them figure so prominently into our self-concepts? The main purpose of this chapter will be to suggest that a neural system (the *C-system*) specialized for controlled self-regulation processes can provide an explanation for the critical role of conflict and conscious choice in self-concept formation. That is to say that the computational properties of this neural system are biased toward encoding our mental and behavioral responses to conflicts rather than our habitual thoughts and behaviors. Because of the importance of facilitating timely adaptive responses to future episodes of conflict or choice points, we hypothesize that our solutions to conflicts will be more easily accessed for solving future conflicts and more strongly identified as part of our self-concept. Though we will focus a great deal on the neural system responsible for encoding postconflict thoughts and behaviors, there is also ample data to suggest this is not the only type of self-knowledge. We will also review the evidence for a second self-knowledge system, including data suggesting that there is a second neural system (the *X-system*) that supports this second kind of self-knowledge. This type of self-knowledge system does not rely on discrete episodes of conflict; rather, it is built up gradually over time through the integration of habitual thoughts, action patterns, and behavioral sequences.

In this chapter, we will first delve into the historical and functional accounts of the self in order to build the foundation for answering why moments of conflict or choice points constitute the most recognizable determinants of our self-concept. We will then expand on the neural structures underlying our explicit (*evidence-based*) self-knowledge system. Finally, we will review the less recognizable determinants of our self-concept as well as the neural correlates of this implicit (*intuitive*) self-knowledge system.

HISTORICAL CHANGES IN SELF-CONCEPT FORMATION

Baumeister (1986, 1987) proposed a radical hypothesis about the nature of the self-concept and how it has been transformed throughout history. He suggested that not only are there qualitatively distinct forms of self-definition processes that can shape the self-concept but that these processes have different effects on self-concept formation and have also been differentially present in Western civilization over the past millennium. As a result, self-concepts of people living in the past 2 centuries may be qualitatively different from those of people living in the middle ages. This is not just a matter of content, with medieval dwellers pondering their resilience to plagues and modern folks pondering over which character on television we are most like. Rather, Baumeister argued that only in the past few centuries has self-definition become problematic in such a way that the self, rather than being transparently equated with status and behavior in a rather simple fashion, is now something to be pondered and probed by all individuals—psychologists included. In essence, though people presumably have always had personalities and other defining characteristics, by this argument complex mental autobiographies are relatively modern psychological phenomena.[1]

AU: Pls provide year(s) (per APA style for first mention in a new paragraph). (Pls. also add years to subsequent citations within paragraphs only if needed to avoid ambiguity about which reference(s) you intend.)

Baumeister suggested that the distinct forms of self-definition processes, namely *given*, *achieved*, and *choice-driven* self-definition processes, differ in the difficulty they pose for the individual and the likelihood that the self-definition process will turn reflective such that the individual becomes aware of the process and its implications for self-definition. *Given* aspects of self-definition, including family lineage and gender, are present at birth and thus require no effort or decision making. *Achievement* aspects may be effortful or effortless processes (gaining wealth vs. becoming a parent) and typically only involve clear societal prescriptions. Until the 20th century, few people had internal conflicts over the prospect of achieving wealth or parenthood—it is simply what was done to the extent that one was capable. *Choice-driven* self-definition processes emerge when there are either no clear criteria or conflicting criteria for making a decision. For instance, how does one decide whether to be a professor or a doctor? Neither is objectively better, nor does society clearly value one more than the other. Each is better on some dimensions (doctors make more money and save lives, whereas professors choose their own avenue of study and advance human knowledge), but which dimensions are more important? Baumeister argued that when confronted with these conflicts, we look to our *self* to determine which is more important. It is unclear whether we find the answer in ourselves or construct an answer, which then becomes a defining part of our self. Either way, these choice points, for which behavior-guiding criteria are absent, are often in the highlights reel of our own *True Hollywood Story*.

An analysis of the changing social structure from the medieval period to our own reveals a shift in the landscape of self-definitional processes available. Baumeister showed that changes in the structure of society closely parallel the increasing frequency of people's reflecting on the nature of their identities. Medieval identity was simple and stable, defined primarily by givens such as social rank and gender. Many facets of identity that today are choice-driven or complex achievement processes were essentially givens in medieval life. One's occupation was most often determined by family lineage, and marriage was often arranged without any choice on the part of the betrothed. During this time period there is little evidence of self-reflection in existing cultural artifacts and almost no recovered autobiographies. In the centuries that followed, however, Protestantism provided people with religious alternatives, and later, industrialization and urbanization increasingly brought new opportunities for achievement and ultimately a variety of life choices that could only be made by assessing and asserting one's self. Accordingly, these centuries saw a boom in the number of artifacts indicating time devoted to self-reflection, such as personal diaries, autobiographies, and the development of an "inner life" in the characters of novels. It is of interest to note that during the same time period that Baumeister reviewed, mirrors changed from being rarely seen religious accoutrements to being the implicit enforcers of social norms and equipment used for self-discovery (Melchior-Bonnet, 1994).

AU: Pls. provide year(s).

FUNCTIONS OF THE SELF

Baumeister's historical account suggests that choices that produce the most internal anguish, as a result of lacking or conflicting criteria available for deciding, spur on the development of identity and become disproportionately salient in our resulting self-concepts. On the one hand, this seems obvious because experiences meeting these criteria are easy to bring to mind. Moreover, numerous theories of identity have suggested that these "nuclear episodes" that include "high points, low points, and turning points" and often focus on a sense of agency or the lack thereof (McAdams, 1993, p. 296) are important contributors to our identity as containing both continuity and change (Erikson, 1968; Harter, 1999; P. J. Miller, 1994; Prout & Prout, 1996; Thorne & McLean, 2002).

On the other hand, it is not clear from a mechanistic standpoint why these experiences should be more accessible than others. Eventually we will conclude that this is the case because conflicts engage the C-system, which produces robust episodic memories in order to facilitate the speedy dissolution of similar conflicts in the future. A discussion of one major function of the self, that of self-regulation, will provide a bridge between the phenomenology of choice conflicts and the neural bases of self-concept.

Ramachandran (1995) publicly declared that humans do not have free will but suggested instead that we may have "free won't." This play on words harkens to the age-old discussion of the duality of the self as both the controller and the thing controlled (Baumeister, 1998; Lakoff & Johnson, 1999; Turner, 1976). When one says "I made myself keep studying," there seem to be two separate entities involved—one that wants to keep working and one that would prefer to bang on a drum all day. This can be partially resolved by considering the joint action of automatic and controlled cognitive processes. In this context, automatic processes are the habits and impulses that guide us through daily life with little effort or intention on our part (Bargh & Chartrand, 2001; Langer, 1989; Lieberman, 2000). These processes often run relatively autonomously, and because they have largely evolved or become conditioned to help us achieve our goals, they are often quite adaptive. When driving down the road, one hardly needs to think at all about all the various aspects of driving; with minimal attention, it just seems to happen. Such automaticities have their limitations, including spontaneous deployment at inappropriate times. For instance, it is adaptive to have our automatic driving habits guiding our behavior when we are driving up the street to the video store. Those same habits can be hazardous to one's health if they guide one's driving unchecked while in a foreign country with different driving laws (including driving on the "other" side of the road).

When our habits take us astray, we are then in need of "free won't," the capacity to stop our habits from running their course and possibly running us into oncoming traffic. Under optimal conditions, controlled processes are reasonably successful in correcting our behavior in light of the current context (Gilbert, 1989; Lieberman, 2003). Controlled processes typically involve effort, awareness, and intention—all the characteristics necessary to make these processes feel self-willed (Lieberman, Gaunt, Gilbert, & Trope, 2002) regardless of whether this phenomenological assessment is accurate (Wegner, 2002; Wilson, 2002) or even coherent (Dennett, 1984). Controlled processes are enormously flexible, limited only to the amount of information that can be processed at any one time, but not to the range of information that can be considered together or the originality of new representations generated. Because of this flexibility, it is easy to forget why controlled processes probably evolved: control.

Carver and Scheier (1981) developed an influential model of self-control guided by the insight that if controlled processes exist for the purpose of control, external or internal, then engineering models of control implemented in physical systems might shed light on human self-control. They drew on cybernetic models of self-regulation (Wiener, 1948) most simply exemplified in *test–operate–test–exit* (TOTE) units discussed by G. A. Miller, Galanter, and Pribram (1960). TOTE describes any computational mechanism with the capacity to (a) assess whether the current state of the world (limited to the world as detectable by the TOTE unit) deviates from the TOTE's standard of comparison or desired state of the world and (b) effect some change on the world until the current state matches the standard of comparison. Essentially the TOTE is a system that performs "tests" on the world, and when deviations from standards occur, the TOTE unit performs an "operation" on the world. The test–operate cycle continues until the test result indicates a match between the current state and the standard, at which point the TOTE unit "exits" until it is scheduled to begin new tests.

The beauty of the TOTE unit is that it is equally applicable to self-correcting systems as different as thermostats, individual humans, and complex governments. Within humans, there are many self-correcting systems for regulating bodily processes that could be described with TOTE units without any connection to consciousness or controlled processing. However, the TOTE units associated with our controlled processes are special because we are aware of their activity and experience TOTE functioning as coming from the self. Carver and Scheier (1981) and Duval and Wicklund (1973) have demonstrated in numerous experiments that state and trait self-awareness are intimately linked to the test function of controlled-processing TOTE units. Self-focused attention is typically either a response to a test indicating a mismatch from a standard or is involved in performing the test itself. The response to the mismatch can involve an assortment of reactions including self-evaluation (Higgins, 1987), generating reflected appraisals in which one infers the evaluations others are making of oneself (Lieberman & Rosenthal, 2001; Mead, 1934), and controlled processing operations to remove the mismatch. Each of these responses to the self-perceived mismatch is experienced as self-related.

If the TOTE units involved in controlled processing are typically experienced as generated "by the self and for the self," then a clear account can be given of why the increasing number of choices and conflicts presented in recent time periods would lead to greater reflection on the nature of the self than in earlier time periods. To the extent that goals, standards, and expectations are given at birth, TOTE units should be called upon less frequently. Under those conditions, the habits acquired while growing up would continue to be adaptive because they would remain in a relatively unchanging context. If the rules are set and constant, habits will perform exceptionally well. The modern world has opened up more and more aspects of life that involve rule changes, and the more those changes occur the more frequently TOTE units will be called upon in the service of overriding contextually inappropriate habits and keeping track of the self-assertions needed to guide behavior. To summarize, the more often habits conflict with current goals and expectations, the more often that TOTE units are called up for duty, and the more people should be cognizant of themselves as having an active self.

MULTIPLE MEMORY SYSTEMS

The preceding logic explains why the self should take up a greater part of the cultural consciousness as the presence of choice and conflict increases. This logic still does not explain why our mental and behavioral responses to these conflicts should be such salient aspects of our mental autobiographies. One can imagine that with a greater cultural emphasis on the self, people might be more likely to attend to and form more robust memories for their behavior in general without any special advantage for the kind of events that were catalysts for the greater emphasis.

The best explanation for this proposed memorial advantage for conflict-related events comes from the cognitive neuroscience of memory. We have known for almost half a century (Milner, Corkin, & Teuber, 1968)—since patient H.M. had most of his medial temporal lobes (MTLs), including the hippocampus, removed to treat his epilepsy—that there are multiple memory systems that are sensitive to different kinds of stimuli and have different operating characteristics. H.M., and many other patients with MTL damage, are dramatically impaired in their ability to form and retain new episodic memories. H.M. can meet new people several times, each time believing it to be the first time because he cannot retrieve a memory of the episode of the earlier meeting. This is because the MTL is critical to forming memories for particular episodes and for storing them, at least for several years (Squire, 1992). As bad as H.M.'s episodic memory is, he can form new habits, which comprise memory for procedures and conditioned associations. Since his surgery, H.M. has been trained to use a computer, but he does not know why he knows how to use it and he does not remember the learning episodes themselves. Conversely, patients with damage to the basal ganglia often have severe deficits in forming and using habits but are relatively spared in their capacity to form new episodic memories (Knowlton, Mangels, & Squire, 1996). Moreover, a number of studies suggest that the basal ganglia, critical for habit use, and the MTL, critical for episodic memory, may inhibit one another such that the activation of one system tends to deactivate the other (Lieberman, Chang, Chiao, Bookheimer, & Knowlton, in press; Packard, Hirsh, & White, 1989; Poldrack et al., 2001).

AU: If now published, please update.

The relation of these memory systems to one another suggests that as long as habits are successfully guiding our behavior, we are less likely to form strong robust episodic memories. This would account to some extent for the autobiographical salience of our reactions to choice and conflict, as habitual behavior will be relatively deemphasized in episodic memory.

NEURAL CORRELATES OF THE TOTE

In order to determine whether the salience of our memories for choice conflicts is due to the nonsalience of habits in episodic memory or because of something about choice reactions that actually increases the salience of these in episodic memory, we must determine the relation of TOTE-like self-regulation processes to episodic memory. Strictly speaking, there are no studies directly assessing this relationship. However, there are several findings suggestive of a special relationship between human TOTE-like processes and episodic memory.

First, it is well-established that successful encoding of episodic memories is related to depth of processing (Craik & Tulving, 1975). The more an individual mentally elaborates on the meaning of a stimulus, the more likely the individual will be able to recall the stimulus later. Paralleling these depth-of-processing effects, recent neuroimaging studies (Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Wagner et al., 1998) have shown that the extent to which the lateral prefrontal cortex (LPFC) is active during encoding significantly predicts retrieval success later.

The LPFC has been associated with linguistic (Bookheimer, 2002), working memory (Smith & Jonides, 1999), and causal processes (Satpute et al., 2003), among others. These processes all share the requirement of operating on and holding distinct multiple symbols and the capacity to flexibly and asymmetrically combine, compare, and sequence those symbols. For instance, "John loves Mary" is asymmetric because it does not imply that Mary loves John (although John might hope that it does). The fact that the LPFC possesses relatively sparsely coded representations (O'Reilly, Braver, & Cohen, 1999), using a relatively small number of neurons for each representation, may promote the ability to hold the representations separate from one another and thus keep track of the asymmetric relations between them (Holyoak & Hummel, 2000).[2]

This capacity for propositional representations that represent asymmetrical relations and implications between the different "objects" of a proposition could promote the capacity to hold context specific goals and rules in mind. These context-specific rules could temporarily bind symbols that ordinarily are not associated with one another. Instead of merely being able to represent that *A* goes with *B*, this capacity allows us to represent that *A* goes with *B*, but only right now in context *C*. This flexibility would allow the LPFC to "think outside the box," overcoming automatic habits and associations by incorporating contextually relevant information into goals and action plans. In other words, the LPFC could guide behavior toward current standards. We have suggested elsewhere (Lieberman, Jarcho, & Satpute, 2003; Lieberman et al., 2002) that the LPFC along with the anterior cingulate cortex (ACC), posterior parietal cortex (PPC), and MTL together perform the TOTE functions of human controlled processes. We call this group of four structures the C-system (for the *C* in *reflective consciousness*).

AU: Full term was needed instead of ERP—edit correct for your meaning?

There are numerous functional magnetic resonance imaging (fMRI) and event-related potential studies suggesting that the ACC is sensitive to discrepancies between perceptions and impulses, on the one hand, and current expectations and goals on the other hand (Braver, Barch, Gray, Molfese, & Snyder, 2001; Kiehl, Liddle, & Hopfinger, 2000). It is sensitive to conflicts as minor as the automatic impulse to read a color word (*r-e-d*) during the Stroop task when the goal is to say the color of the ink that the word is written in (*blue*). It is also sensitive to major conflicts such as physical pain (Lieberman et al., 2003; Rainville, Duncan, Price, Carrier, & Bushnell, 1997) and social exclusion (Eisenberger, Lieberman, & Williams, 2003). In a series of elegant studies, Botvinick, Braver, Barch, Carter, and Cohen (2000) and Carter et al. (2000) have shown that the ACC acts as a conflict monitor, performing the test component of the TOTE unit. Rather than performing the subsequent TOTE operations itself, the ACC acts as an alarm that signals the LPFC to begin performing operations (see Hunter et al., 2003, for research demonstrating the temporal sequencing of ACC and LPFC operations).

If the ACC performs the test and the LPFC handles the operations, what role is left for the MTL? A comparison between the TOTE units implemented in thermostats and human controlled processing will suggest an answer that will also address the larger question of the overrepresentation of choice conflicts in our mental autobiographies. Thermostats have a single goal or standard to test for—the temperature level set by the occupant of the room. In addition, in any given season there is typically only one way the temperature can deviate from the standard: it can be too cold in winter and too warm in summer. In the summer, then, each and every time a mismatch is detected, the thermostat automatically triggers the air conditioning to come on. For humans, things are not so simple—not even close. At any one time, there are virtually an infinite number of standards that might not be met. Everything from uncomfortable clothes, aches and pains, hunger, negative nonverbal feedback from friends, and subpar performance at work or on a test can all grab the attention of the ACC.

Leaving aside the issue of how the standards are formed and maintained (Higgins, 1997; Mead, 1934), it is no simple matter for the LPFC to perform the appropriate operation to fix the problematic situation. This is especially unfortunate because one of the defining and unique features of the LPFC is that its functions are severely limited by processing constraints. LPFC computations, characterized by the constraints of working memory (Smith & Jonides, 1999), seem to operate on symbolic representations in serial fashion with only seven, plus or minus two, bits of information in use at a time in the service of a single thought at a time (James, 1890/1950; G. A. Miller, 1956). Attempts to handle more information simultaneously lead to a degradation in performance as evidenced by dual-task and cognitive-load studies (Gilbert, 1989). Moreover, sustained use of working memory, even within its constraints, can deplete working memory effectiveness for short periods (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Vohs & Heatherton, 2000). Given the limitations of the LPFC, the less work it performs in general the better able it will be to perform when it is really needed.

AU: There is no Higgins '97 in reference list. Should this be Higgins '87?

The fragility of LPFC processing helps explain why the ACC performs the TOTE tests. The LPFC is able to effectively rest until it is called upon by a mismatch detected in the ACC. The MTL also serves to compensate for LPFC fragility by preserving a record of how previous conflicts were resolved by the LPFC. When the same situation arises in the future, automatic habits are likely to be little changed. For instance, driving for 5 min in a country where driving on the left side of the road is the norm will have little effect on one's preexisting habit to drive on the right side of the road. Indeed, what good would our habits be if they were upended so easily (McClelland, McNaughton, & O'Reilly, 1995)? Instead, habits are decontextualized representations that are insensitive to the constraints of the particular situation and change only with numerous instances of a new behavior, perception, or contingency. Because of the habit system's intransigence, renewed exposure to the situation that activated the ACC before is likely to activate it again.

Recall that episodic memories in the MTL are better encoded to the extent that there is deeper processing associated with LPFC operations. In the context of TOTE functions this means that as the number and complexity of LPFC operations performed in response to an ACC mismatch increases, so too does the strength of the episodic memory laid down in the MTL. In other words, we have good episodic memories for big problems that were difficult to solve. Why? Because those who cannot remember—recall how they solved a problem in—the past are doomed to repeat it (and thus must figure out the solution again). If the LPFC can retrieve a solution from the MTL's records of past responses to conflicts, then it can focus on implementing the solution rather than on rediscovering it. Thermostats only need to "remember" a single solution and thus would not benefit from a memory bank. A thermostat's memory bank would have line after line of "At 3:42 on a Wednesday afternoon, turned on air conditioning. At 6:07 on a Tuesday evening, turned on air conditioning. At 10:15 on a Saturday night, turned on air conditioning." For humans, however, this database of solutions to past problems (which have not occurred frequently enough to change our habits) is invaluable. Tommy may not mind putting in the effort to figure out the answer to a math problem once (288,499 × 25 + 1,462,834 = 8,675,309), but it would be nice to have an episodic memory of the answer to turn to if asked again seconds later.

This brings us to the solution of the major question of this chapter. Why are our mental autobiographies filled more with memories of our responses to difficult choices and conflicts than with memories of banal everyday activity? It seems that there are at least two complementary reasons for this. In part, this occurs because the successful deployment of our automatic habits may directly interfere with the formation of new episodic memories given the competitive relationship between the basal ganglia and MTL (Poldrack & Packard, 2003). More important, however, is that episodic memory may be an integral part of the TOTE functions of the C-system, forming new episodic representations to the extent that LPFC operations occur, and serving as a shortcut to the previous solution when the situation arises again. Thus, if the computations of the C-system involve looking to the self and constructing new solutions to conflicts for which habits of mind and behavior are ineffectual, these self-infused solutions will be recorded in episodic memory.

EXPERIENCING THE SELF IN THE C-SYSTEM

Now that we have addressed the question of why our reactions to conflicts should be overrepresented in our autobiographical memory, we would like to backtrack and address one of our earlier assumptions in greater detail. Earlier, we noted that one way the TOTE-like processes of humans and thermostats differ is that we experience our TOTE processes as coming from the self and (certain Buddhist doctrines notwithstanding) thermostats do not. This is a critical assumption because otherwise the solutions to our choice conflicts might be recorded into episodic memory without their being linked to the self. We might have memories of this and that having happened without having a sense that we were the agent at the center of the action. Recent research in cognitive neuroscience has begun to shine a light on the link between experienced self-processes and the structures of the C-system in terms of self-awareness, self-control, and self-knowledge. We address each of these topics in turn.

Self-awareness. Self-awareness refers to the ability to turn one's attention and thoughts to oneself. A nuanced understanding of self-awareness would fill volumes and even then would most likely leave us feeling that something basic about self-awareness was still not addressed. In the meantime, a number of neuroimaging studies have implicated the ACC in self-awareness across a variety of domains. When individuals are asked to reflect on their emotions (Lane, Fink, Chua, & Dolan, 1997) or their actions (Jueptner et al., 1997), rather than merely experiencing them, there is greater activity in the ACC. Additionally, when individuals are asked to consider a scenario and reflect on how they would feel and act, the ACC is again more active (Vogeley et al., 2001). Finally, ACC activity is found when individuals are asked to reflect on their physical traits (Kjaer, Nowak, & Lou, 2002).

AU: If now published, please update.

AU: Please give year of draft here.

Though brain localization is an important first step in understanding the neural bases of any mental process (Lieberman & Pfeifer, in press), it is unsatisfactory as an end in itself (Willingham & Dunn, in press). After determining that the ACC is related to self-awareness, the next obvious question to ask is why. Which aspects of ACC computations are critical for self-awareness? We recently addressed this question in an fMRI study (Eisenberger, Lieberman, & Satpute, XXXX). Given that the ACC's mismatch detection function is a good candidate for the TOTE test function in human controlled processing and given that self-awareness has been theoretically linked to TOTE processes, we hypothesized that the reactivity of the ACC to mismatches would predict the frequency and accuracy of self-awareness processes outside the scanner. We reasoned that more reactive ACC's should produce TOTE tests that are more sensitive to mismatches leading to more frequent episodes of self-awareness as well as more sensitive, or accurate, self-awareness.

AU: If now published, please update.

In this study, participants were first scanned while performing an "oddball" task during which they were presented with a sequence of letters on the screen. Eighty percent of the letters were the letter *X*, but participants were instructed only to press a button whenever they saw a letter other than *X*. Because the base-rate expectation of seeing an *X* is 80%, seeing other letters violates this expectation and leads to activation of the ACC (Braver et al., 2001; Menon, Adleman, White, Glover, & Reiss, 2001; Weissman, Giesbrecht, Song, Mangun, & Woldorff, 2003). A week or more after the scanning session, participants returned to the behavioral lab, where frequency and accuracy of self-awareness were assessed. First, participants filled out questionnaires including a self-consciousness scale that measures, among other things, frequency of self-awareness (Fenigstein, Scheier & Buss, 1975). Participants then exercised vigorously for 1 min and then reported reported on how physiologically aroused they thought they were, from 0% (perceived arousal before exercising) to 100% (perceived arousal after exercising) every 2 min until 10 min postexercise. We measured actual physiological arousal at the same time points in terms of a gender-neutral measure of arousal, rate pressure product (Pham, Taylor, & Seeman, 2001), which combines heart rate and systolic blood pressure. We found that ACC reactivity to the oddball trials, relative to nonoddball trials, predicted the accuracy of arousal self-awareness extremely well ($r^2 = .50$) even after covarying out individual differences in arousal curves. Additionally, ACC reactivity correlated highly with the self-report measure of self-awareness ($r = .76$). Interestingly, ACC reactivity predicted arousal self-awareness better than self-reported self-awareness predicted this behavioral measure. Finally, ACC reactivity also correlated highly with neuroticism ($r = .69$), which can also be seen as related to dispositional self-awareness.

This study, like several before it, demonstrates a link between self-awareness and the ACC. Unlike previous studies, it helps explain why this link exists by connecting self-awareness to a particular neurocognitive process in the ACC—namely, reactivity to mismatches. Additionally this study provides some of the best evidence to date supporting the contention that self-awareness is linked to the TOTE unit's test function.

Discrepancy detection as a trigger for self-awareness and subsequent self-control is, generally speaking, an adaptive mechanism that goes well beyond the limitations of simple habits. We would not want to leave the reader with the impression that self-focused attention in all forms is always a good thing. In fact, several psychological disorders, including clinical depression and anxiety, are associated with elevated levels of self-focus (Ingram, 1990). Consistent with our account of self-focus and ACC activity, these self-focus-related disorders typically involve abnormal ACC functioning (Benkelfat et al., 1995; Davidson, Pizzagalli, Nitschke, & Putnam, 2002; Kimbrell et al., 1999; Pizzagalli et al., 2001; Ursu, Stenger, Shear, Jones, & Carter, 2003). Thus, although self-focused attention may be useful under various conditions, it can be problematic in its extreme forms.

Self-control. It is commonly believed that the LPFC is central to working memory processes most clearly aligned with effortful top–down processes that regulate behavior (E. K. Miller & Cohen, 2001; Smith & Jonides, 1999). The LPFC is believed to perform at least three types of processes that would greatly facilitate self-control. First, the ventral LPFC is involved in the suppression or disruption of unwanted cognitive, affective, or behavioral responses (Aron, Fletcher, Bullmore, Sahakian, & Robbins, 2003; Eisenberger, Lieberman,

& Williams, 2003; Iversen & Mishkin, 1970; Monchi, Petrides, Petre, Worsley, & Dagher, 2001; Ochsner, Bunge, Gross, & Gabrieli, 2002; Preibisch et al., 2003; Small, Zatorre, Dagher, Evans, & Jones-Gotman, 2001). Second, the dorsal LPFC is involved in boosting the strength of weaker, but contextually appropriate, representations and action plans (Kosslyn, Thomson, & Alpert, 1997; E. K. Miller & Cohen, 2001; Tomita, Obayashi, Nakahara, Hasegawa, & Miyashita, 1999). Third, the LPFC along with the frontopolar region of the prefrontal cortex (PFC) can flexibly combine symbolic representations using propositional rules to consider novel courses of action and ultimately set one in motion (Kroger et al., 2002; Waltz et al., 1999; Zysset, Huber, Ferstl, & von Cramon, 2002).

Though each of these processes contribute to self-control, it is the phenomenological experience of being the author of these acts of self-control, the feeling that "I am planning" or "I am suppressing an impulse" that links them to self-concept. Despite the fact that all experience is produced by our own neural activity, the great majority of these experiences are attributed to something external to oneself. When faced with an American flag, only patriots, poets, and philosophers would be expected to say that a bit of their consciousness is red, white, and blue. When we see John shove Michael (because Mary loves Michael, not John), we believe the aggressiveness of the act is out there in the world, not an aspect of our conscious experience dependent on our goals, beliefs, and values (Griffin & Ross, 1991). When we engage in acts of self-control, be it holding our breath under water, fasting for a religious holiday, or rehearsing a nine-digit number, we almost always feel a sense of authorship for the act. It feels like no mere accident happening to us but instead feels intentional—intended by us. Indeed, it is hard to imagine ever finding oneself accidentally rehearsing a nine-digit number, for as soon as we stop intending to do so, active rehearsal stops (though a trace may have been laid down in long-term memory).

A number of studies have implicated the PPC in assessing whether oneself or another was responsible for an action (Chaminade & Decety, 2002; Farrer & Frith, 2002; Ruby & Decety, 2001; Taylor, 2001), although its exact role is unclear. In other words, it has not been determined whether the PPC participates in all judgments of authorship or just those involving the perception of actions. For example, Gusnard, Akbudak, Shulman, and Raichle (2001) found that when individuals were differentiating their emotional reaction from the emotional reaction of others, the dorsomedial PFC (adjacent to the ACC) rather than the PPC was involved. Apart from neuroimaging studies, a number of neuropsychological investigations also implicate the PPC in the experience of authorship for one's body and its actions. *Anosognosia* refers to a condition in which patients have some kind of impairment but do not recognize that they have it (Galin, 1992). When patients have had a stroke that has paralyzed one side of their body as well as damaged inferior parietal cortex they will sometimes, and often only temporarily, become anosognosic. Ramachandran (1995) provided a vivid case history of such a patient who believes that her paralyzed left arm is just as able as her right arm. Despite incontrovertable evidence that she cannot control this arm in any way, she continues in her belief. This suggests that the PPC plays an important role in the experience of self-efficacy, control, and authorship. Similarly, when individuals have alien hand syndrome and experience their arm movements as controlled by an external force, there is increased PPC activity (Spence et al., 1997), which drops off with symptom reductions.

Self-knowledge. In the first neuroimaging study of self-knowledge, Craik et al. (1996) found that self-knowledge judgments activated the right PFC. This is consistent with a number of neuropsychological investigations suggesting a link between this area and self-knowledge (B. L. Miller et al., 2001; Stuss, Picton, & Alexander, 2001). To date, there have now been at least seven neuroimaging studies, including three positron emission tomography (PET) studies (Craik et al., 1999; Fink et al., 1996; Kjaer, Nowak, & Lou, 2002) and four fMRI studies (Johnson et al., 2002; Kelley et al., 2002; Kircher et al., 2000; Lieberman, Jarcho, & Satpute, 2003), each using relatively similar paradigms in which participants had to judge whether words were self-descriptive. As seen in Table 4.1, six of the seven studies found activation in the medial aspect of PPC, called the precuneus, and adjacent posterior cingulate. The precuneus, along with the MTL, is associated with successful episodic recall (Cabeza & Nyberg, 2000). In one rodent study (Izquierda et al., 1997), the precuneus was the only structure examined that if ablated any time after learning would prevent successful recall. MTL structures were critical for the first month after encoding, but eventually retrieval could function without the MTL. This supports the basic assumption that the self-concept is dependent on memory for autobiographical episodes. TOTE test functions associated with the ACC bring self-awareness online when conflicts occur that our habits cannot handle. Self-control is then exerted in its varied forms, implemented by the LPFC and labeled as self-authored by the PPC and perhaps medial PFC. To the extent that the LPFC is engaged in the conflict resolution, these operations should be encoded more robustly in the MTL and later retrieved during self-knowledge judgments by the precuneus and the PFC.

INTUITION-BASED SELF-KNOWLEDGE

If things were so simple as the previous summary suggests, we would be on to the reference section by now. For better or for worse, the story of self-concepts has a second act. Though the account of self-concepts as drawing on episodic memories of our reactions to important choice points fits very well with our folk theory of self-concepts and is consistent with much of the existing imaging data, there is a growing body of work suggesting that this account of self-knowledge is incomplete in important ways.

The problem, in a nutshell, is that episodic memory is not critical for many kinds of self-judgments. Imagine Jerry Seinfeld being asked to judge whether or not he is funny. At this point in his career, with all the success and laughter his comedy has produced, it seems possible that he would just know this without having to reflect on and evaluate memories of discrete comedic performances from the past. True, barring a nasty bump on the head, he could engage C-system processes to consult all those episodic memories, evaluate them, and construct

TABLE 4.1. Neural Correlates of Self-Knowledge Retrieval From Seven Studies

Brain region	Brodmann area(s)	Side	Kircher et al. (2000)	Johnson et al. (2002)	Craik et al. (1999)	Kelley et al. (2002)	Fink et al. (1996)	Lieberman et al. (2003)	Kjaer et al. (2002)	Total
Precuneus and posterior cingulate	7/31	Both	•	•		•	•	•	•	6/7
Medial and ventromedial prefrontal	9/10/11	Both		•	•			•	•	4/7
Inferotemporal	21/38	Right		•			•	•		3/7
Inferior parietal	40	Both	•					•	•	3/7
Ventrolateral prefrontal	44/45/47	Right			•			•		2/7
Basal ganglia		Left	•					•		2/7
Insula		Both	•				•			2/7

Note. Check marks indicate areas of activation.

AU: Note was added; could you pls. check wording ("activation" correct for your meaning?) and revise if needed.

what we have called *evidence-based* self-knowledge (Lieberman, Jarcho, & Satpute, 2003). But does he need to?

In a series of behavioral and neuropsychological studies, Klein and his colleagues have shown that evidence-based self-knowledge does not appear to be necessary for various self-knowledge judgments. In one series of studies (Klein, Loftus, Trafton, & Fuhrman, 1992), participants showed no reaction time advantage when making self-knowledge judgments immediately after the activation of relevant autobiographical memories, relative to when no autobiographical memories were preactivated. If episodic retrieval is used in making self-knowledge judgments, one would expect that making the relevant memories more accessible would facilitate those judgments, but here it did not. The activation of autobiographical memories only improved reaction times when participants were making judgments about themselves in a domain that was relatively new to them. This suggests that early in the development of any area of self-knowledge, particular episodes are important elements of the self-concept in that domain. With growing experience, however, Klein's data suggests that the knowledge is recompiled in such a manner as to render the link to the particular episodes unnecessary.

From behavioral data alone, it is unclear whether a single representation undergoes a transformation from being evidence-based to being something else or whether there are multiple distinct self-knowledge representations forming in parallel. Klein and colleagues used neuropsychological case studies to shed light on the issue of single versus multiple self-knowledge systems (for a review of all of these case studies, see Klein, Rozendal, & Cosmides, 2002). A series of patients with congenital or acquired deficits in episodic memory have proved able to produce self-knowledge judgments as accurate as those of healthy controls. In the best known of these cases, patient W.J. suffered a traumatic head injury that temporarily rendered her incapable of retrieving memories of events that had occurred in the previous 12 months. Despite this impairment in episodic memory, W.J. was able to produce personality ratings for herself that were highly correlated with the ratings she produced after she regained access to her episodic memories (Klein, Loftus, & Kihlstrom, 1996).

These studies make a compelling case for the multiple self-knowledge systems position. However, they primarily shed light on what the second self-knowledge system is not rather than illuminating what it is. We know that representations from the second system do not depend on evidence from the autobiographical record generated in the C-system, at least once these representations have fully matured. However, we do not know what type of self-related information this second self-knowledge system is dependent upon.

Because of the independence from autobiographical evidence, we have characterized the second system as an *intuition-based* self-knowledge system. In other unrelated research on judgment and decision making, attribution, and prejudice (Lieberman, 2000; Lieberman et al., in press; Lieberman, Eisenberger, & Crockett, 2003 ; Lieberman et al., 2002; Lieberman, Hariri, Jarcho, & Bookheimer, 2003), we have found evidence of a second neurocognitive system called the X-system (for the *x* in *reflexive*; it includes the basal ganglia, ventromedial PFC, amygdala, and lateral temporal cortex), which is typically the automatic social-

AU: If now published, please update.

cognitive counterpart to the C-system's controlled processes.[3] We hypothesized that the same relation would hold with regard to self-knowledge such that intuition-based self-knowledge would be subserved by the X-system. If supported, this is a case where merely finding where in the brain a process occurs can yield theoretical fruit, because there is already a reasonable understanding of the characteristics of the X-system (e.g., associative learning, parallel processing). Given that we mostly know what intuition-based self-knowledge is not, this link would suggest some answers to what it is.

AU: Pls. note edits were made throughout for APA style to avoid ital. for emphasis. If any changes are needed to chapter as edited, please provide specific changes you'd like.

Lieberman, Jarcho, and Satpute (2003) tested the hypotheses that there were two distinct self-knowledge systems, evidence-based and intuition-based, and that these depended on two neurocognitive systems, the C-system and X-system, respectively. Each participant was an experienced athlete or actor. The athletes and actors were asked to make self-knowledge judgments regarding the applicability of traits words relevant to each domain (athleticism and acting). Thus, participants made judgments in both a high-experience domain and a low-experience domain. When the neural activity was compared across these different judgments, all but one of the regions more active for high-experience judgments than low-experience judgments were X-system regions. In this comparison all regions of the X-system were more active, including the basal ganglia, ventromedial PFC, amygdala, and lateral temporal cortex. The only C-system region active in this comparison was the PPC. The dorsolateral PFC was the only region of the brain that was significantly more active for low-experience judgments, although the right hippocampus in the MTL was also significant once reaction times were controlled for.

We believe this study clearly shows two self-knowledge systems at work in distinct neural systems. The C-system produced greater activation when making low-experience domain self-judgments presumed to rely on evidence-based self-knowledge, whereas the X-system produced greater activation when making high-experience domain self-judgments presumed to rely on intuition-based self-knowledge. So what do these results buy us? They help us to make inroads into the operating principles of intuition-based self-knowledge based on what is already known about the characteristics of the X-system.

The X-system (again, the amygdala, basal ganglia, ventromedial PFC, and lateral temporal cortex) is hypothesized to automatically generate the affective and social components of the stream of consciousness and produce a great deal of the habits and impulses that guide our daily activity (for full reviews, see Lieberman et al., 2002; Lieberman, Hariri, et al., 2003; Lieberman, Jarcho, & Satpute, 2003). The basal ganglia and ventromedial PFC have both been identified as playing a role in learning abstract relationships between features of the environment and the affective significance of these feature without conscious awareness or intention (Bechara, Damasio, Tranel, & Damasio, 1997; Cromwell & Schultz, 2003; Knowlton, Mangels, & Squire, 1996; Lieberman, 2000). The amygdala is also strongly identified with automatic affective responses, in particular responding to the threat value of environmental stimuli (LeDoux, 1996), even without the conscious perception of these stimuli (Whalen et al., 1998). The lateral temporal cortex has been more frequently associated with semantic associations than affect (Burton, Diamond, & McDermott, 2003; Copland et al., 2003; Mummery, Shallice, & Price, 1999), but it may store various social-cognitive associations that would modulate the automatic affective responses in other parts of the X-system. This suggests that intuition-based self-knowledge is more affect-based than evidence-based self-knowledge. C-system representations may be about affect, but they are still likely to be propositions. X-system representations are much closer to the primitives of affective experience. This is a new positive piece of information about intuition-based self-knowledge that can be inferred from simply knowing what structures in the brain are responsible for this type of self-knowledge.

Another critical feature of all the X-system structures is that the formation of new representations is typically slow and incremental (Damasio, 1994; McClelland, McNaughton, & O'Reilly, 1995), whereas C-system structures typically form complete representations quickly based on single trials.[4] This has two major implications for our understanding of evidence-based and intuition-based self-knowledge. First, it is likely to require numerous repetitions in a domain before intuition-based self-knowledge will mature enough to dominate cognition and behavior when it is needed. Though the work by Klein et al. (1992) demonstrated that evidence-based self-knowledge was guiding behavior most when individuals made low-experience domain self-knowledge judgments, it was not clear why this should be the case. The fact that intuition-based self-knowledge is implemented in a neural system that is well-documented as having a slow incremental learning algorithm helps to explain Klein's findings. The second implication that this finding has is that it suggests that most intuition-based self-knowledge cannot be updated quickly. Convincing individuals with longstanding low self-esteem that they are deserving of greater self-esteem may lead to the modification of some linguistic propositions in the C-system, but it will probably have little effect on the X-system. The X-system seems to be less sensitive to linguistic input, whether it be a friend's, a therapist's, or even one's own interior thoughts, and more sensitive to repeated exposure to an environment with a stable set of underlying relationships between stimuli.

INTUITION-BASED SELF-KNOWLEDGE AND IMPLICIT SELF-PROCESSES

AU: There was no Spencer (single author) in this vol.--edit ok for your meaning?

There is a natural desire to identify intuition-based self-knowledge with implicit self-processes (see Spencer, Jordan, Logel, & Zanna, this volume), and we suspect that there is some overlap between the two. We do, however, hesitate to suggest they are synonymous. Implicit representations are typically those that cannot be brought to mind explicitly and are instead revealed through various other responses that imply that a representation must be present and guiding cognition even though it is not consciously accessed (Schacter, 1992).

In all of Klein's work, as well as our own imaging work, subjects explicitly and successfully answered questions about their own self-concepts. Clearly, this knowledge does not pass the litmus test for being implicit. What may appear to be implicit is the evidentiary basis for intuitive self-knowledge; however, we believe this is not the case. There is no reason to believe that there are links from intuition-based self-knowledge to implicit representations of the evidence supporting that knowledge (Lieberman et al., 2002). Rather, as long as each episode that provides evidence incrementally alters intuition-based self-knowledge as it happens, it would have its effect without leaving a representational trace of itself in the X-system. Every time a hammer hits a nail, it will be embedded further into a piece of wood. In so doing, the hammer incrementally changes the status of the nail but the nail does not require a memory of the hammer in order to maintain its new status.

MEDIAL PREFRONTAL CORTEX AND SELF-PROCESSES

A number of studies suggest that the medial prefrontal cortex (mPFC) also plays a role in self-processes (see Table 4.1). It is unclear at this time for multiple reasons what this role is and where mPFC would fit with respect to the X- and C-systems. Although future reviews may well include this region of the brain as a major component of self-processing, it is too soon to make such a claim. The mPFC is a very large area of cortex comprising no less than three Brodmann's areas (9, 10, 11). It is likely that different areas are involved in different kinds of computations, but for now there is no agreed upon nomenclature for dividing the mPFC into its constituent parts. The upshot of this is that self studies reporting activity in mPFC appear to be talking about the same area of the brain as one another when in fact these studies are reporting activations that are quite distinct.

A second issue is that the mPFC has been identified with social cognition more generally and not just self-processes. It is more active when we are trying to understand the intentions of others (Gallagher & Frith, 2002) or even just imputing intention to moving cartoon objects (Schultz et al., in press). It is also more active when processing information related to a person than an object (Mitchell, Heatherton, & Macrae, 2002). Thus, it is difficult to draw any conclusions about whether the mPFC is playing a specific role in self-processing.

AU: If now published, please update.

Finally, the mPFC has an unusal property that makes drawing inferences about its role even more difficult. A review of dozens of neuroimaging studies (Raichle et al., 2001) indicates that the mPFC is more active at rest than during almost any kind of mental activity a person engages in. In other words, engaging in almost any kind of mental activity seems to interfere with whatever it is the mPFC does when the rest of the brain is at rest. Most previous studies ostensibly showing mPFC increases are really only showing smaller decreases during self- or social cognition than during some control task (Kelly et al., 2002; Mitchell, Heatherton, & Macrae, 2002). We have recently shown true mPFC increases relative to a resting baseline when participants were watching realistic social interactions between two people (Iacoboni et al., in press); however, it is unclear at this point exactly which features of the social interaction led to this increase. All these issues taken together suggest that it would be premature to make any claims about the specific role of the mPFC in self-processes.

AU: If now published, please update.

CONCLUSION

To be sure, this was meant to be a review of research on self-processes, specifically those involved in defining our "selves." We hope that we have provided some insight into why conflict and our response to it plays such a prominent role in our autobiographical stories. Moreover, we hope we have provided a framework for understanding the multiple self-knowledge systems and their neurocognitive bases.

In essence, we have reviewed two self-knowledge systems with two separate underlying neural subcomponents. The evidence-based self-knowledge system, which contributes to the importance of conflict in our self-definition, is composed of C-system structures (the LPFC, ACC, PPC, and MTL), involved in the controlled regulation of behavior when something goes wrong or no clear behavioral response is available. The intuition-based self-knowledge system, which is built up gradually over time through repeated habits and behaviors, is comprised of X-system structures (the basal ganglia, ventromedial PFC, amygdala, and lateral temporal cortex), involved in the automatic enactment of behavioral responses. An important purpose of the C-system is to exert control when no clear habitual response exists and to record these new behavioral responses should a similar situation arise again. Thus, the conflicts that we face (deciding whether to become a doctor or lawyer) become the basis for our evidence-based self-knowledge, whereas our habitual behavioral responses (how we respond to our patients or clients) become the basis for our intuition-based self-knowledge.

The subtext of this chapter, however, was meant to demonstrate the value of cognitive neuroscience research and neuroimaging tools in advancing social psychological theories of the self. By understanding the neural components involved in automatic and controlled processes, we can begin to disentangle the complexity of self-processes such as those involved in self-knowledge, self-esteem, self-enhancement, or self-regulation. Neural activity in specific structures provides us with clues about the type of cognition that occurs there, and the interactions between structures tells us about how these types of processes support or interfere with each other. As neural data continue to inform us of the different types of processes involved in self-knowledge, so too should social psychological theories of the self be updated to accommodate the implications of cognitive neuroscience data. If, through a better understanding of the neural processes involved in self-knowledge, readers have had to rethink their theories of the self, then we have done our job.

NOTES

1. Baumeister did not suggest that the nature of self-concept processes is the same for everyone at a certain point in history. Rather, he suggested that generalizations can be made about the typical experience from the time periods.

2. Though we have not yet reached the discussion of posterior parietal cortex, it is worth noting here that it works in concert with the LPFC for working memory tasks and also has very sparse representations (Gottlieb, Kusonoki, & Goldberg, 1998).

AU: Please add an entry for this to your reference list—it's missing.

3. We do not mean to imply these structures are all necessarily coordinated with each other neurally. Rather, we believe they serve a common set of functions which bind them even if "one hand doesn't know what the other is doing." Though there is some evidence of neural connectivity for some of the structures within each system and especially the C-system, the full extent of this connectivity is beyond the scope of this chapter and is still largely unknown.
4. Though the amygdala often forms representations of threat cues incrementally, it is also capable, at least in rodents, of single-trial learning. This makes sense in light of the differential need to learn threat-versus-reward cues quickly; however, it does make the amygdala somewhat anomolous within the X-system.

AU: Was "beyond the scope of this beyond" in orig. ms.—edit correct for your meaning?

REFERENCES

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience, 6*, 115–116.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*, 462–479.

Baumeister, R. F. (1986). *Identity.* New York: Oxford University Press.

Baumeister, R. F. (1987). How the self became a problem: A psychological review of historical research. *Journal of Personality and Social Psychology, 52*, 163–176.

Baumeister, R. F. (1998). The self. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 680–740). Boston: McGraw-Hill.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 65*, 317–338.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275,* 1293–1295.

AU: If issue date (month and day) are readily available to you, pls. provide (as Science is formatted as a magazine ref. in the APA manual).

Benkelfat, C., Bradwejn, J., Meyer, E., Ellenbogen, M., Milot, S., Gjedde, A., & Evans, A. (1995). Functional neuroanatomy of CCK4-induced anxiety in normal healthy volunteers. *American Journal of Psychiatry, 152*, 1180–1184.

Bookheimer, S. Y. (2002). Functional MRI of language: New approaches to understanding cortical organization of semantic processing. *Annual Review of Neuroscience, 25*, 151–188.

Botvinick, M. M., Braver, T. D., Barch, D. M., Carter, C. S., & Cohen, J. D. (2000). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624–652.

Braver, T. D., Barch, D. M., Gray, J. R., Molfese, D. L., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of frequency, inhibition, and errors. *Cerebral Cortex, 11*, 825–836.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science, 281*, 1185–1187.

AU: Issue cover date avail.?

Burton, H., Diamond, J. B., & McDermott, K. B. (2003). Dissociating cortical regions activated by semantic and phonological tasks: An fMRI study of blind and sighted people. *Journal of Neurophysiology, 90*, 1965–1982.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition: II. An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 12*, 1–47.

Carter, C. S., MacDonald, A. W., Botvinick, M. M., Ross, L. L., Stenger, V. A., Noll, D., et al. (2000). Parsing executive processes: Strategic vs. evaluative functions of the anterior cingulate cortex. *Proceedings of National Academy of Sciences, 97*, 1944–1948.

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control theory approach to human behavior.* New York: Springer-Verlag.

Chaminade, T., & Decety, J. (2002). Leader or follower? Involvement of the inferior parietal lobule in agency. *NeuroReport, 13*, 1975–1978.

Copland, D. A., de Zubicaray, G. I., McMahan, K., Wilson, S. J., Eastburn, M., & Chenery, H. J. (2003). Brain activity during automatic semantic priming revealed by event-related functional magnetic resonance imaging. *NeuroImage, 20*, 302–310.

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.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

Cromwell, H. C., & Schultz, W. (2003). Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *Journal of Neurophysiology, 89*, 2823–2838.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

Davidson, R. J., Pizzagalli, D., Nitschke, J. B., & Putnam, K. (2002). Depression: Perspectives from affective neuroscience. *Annual Review of Psychology, 53*, 545–574.

Dennett, D. C. (1984). *Elbow room: The varieties of free will worth wanting*. Cambridge, MA: MIT Press.

Duval, S., & Wicklund, R. A. (1973). Effects of objective self-awareness on attribution of causality. *Journal of Experimental Social Psychology, 9*, 17–31.

Eisenberger, N. I., Lieberman, M. D., & Satpute, A. B. (**XXXX**). *The neural correlates of self-awareness: An fMRI study of anterior cingulate cortex reactivity.* Manuscript in preparation.

AU: Please give year of draft here. Or, update all publishing info?

AU: Issue cover date avail.?

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An fMRI study of social exclusion. *Science, 302*, 290–292.

Erikson, E. H. (1968). *Identity, youth, and crisis*. New York: Norton.

Farrer, C., & Frith, C. D. (2002). Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *NeuroImage, 15*, 596–603.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43*, 522–527.

Fink, G. R., Markowitsch, H. J., Reinkemeier, M., Bruckbauer, T., Kessler, J., & Heiss, W. (1996). Cerebral representation of one's own past: Neural networks involved in autobiographical memory. *Journal of Neuroscience, 16*, 4275–4282.

Galin, D. (1992). Theoretical reflections on awareness, monitoring, and self in relation to anosognosia. *Consciousness and Cognition, 1*, 152–162.

Gallagher, H. L., & Frith, C. D. (2002). Functional imaging of "theory of mind." *Trends in Cognitive Sciences, 7*, 77–83.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.

Griffin, D. W., & Ross, L. (1991). Subject construal, social inference, and human misunderstanding. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 319–359). San Diego, CA: Academic Press.

Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proceedings of the National Academy of Science, USA, 98*, 4259–4264.

Harter, S. (1999). *The construction of the self: A developmental perspective.* New York: Guilford Press.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*, 319–340.

Holyoak, K. J., & Hummel, J. E. (2000). The proper treatment of symbols in a connectionist architecture. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines.* Mahwah, NJ: Erlbaum.

AU: Pls. provide page range.

Hunter, M. D., Farrow, T. F., Papadakis, N. G., Wilkinson, I. D., Woodruff, P. W., & Spence, S. A. (2003). Approaching an ecologically valid functional anatomy of spontaneous "willed" action. *NeuroImage, 20*, 1264–1269.

Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (in press). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *NeuroImage*.

AU: If now published, please update.

Ingram, R.E. (1990). Self-focused attention in clinical disorders: Review and a conceptual model. *Psychological Bulletin, 107*, 156–176.

Iversen, S. D., & Mishkin, M. (1970). Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. *Experimental Brain Research, 11*, 376–386.

Izquierda, I., Quillfield, J. A., Zanatta, M. S., Quevedo, J., Schaeffer, E., Schmitz, P. K., et al. (1997). Sequential role of hippocampus and amygdala, entorhinal cortex and parietal cortex in formation and retrieval of memory for inhibitory avoidance in rats. *European Journal of Neuroscience, 9*, 786–793.

James, W. (1890/1950). *The principles of psychology.* New York: Dover.

Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., & Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain, 125*, 1808–1814.

Jueptner, M., Stephan, K. M., Frith, C. D., Brooks, D. J., Frackowiak, R. S., & Passingham, R. E. (1997). Anatomy of motor learning: I. Frontal cortex and attention to action. *Journal of Neurophysiology, 77*, 1313–1324.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience, 14,* 785–794.

Kiehl, K. A., Liddle, P. F., & Hopfinger, J. B. (2000). Error processing and the rostral anterior cingulate: An event-related fMRI study. *Psychophysiology, 37*, 216–223.

Kimbrell, T. A., George, M. S., Parekh, P. I., Ketter, T. A., Podell, D., M., Danielson, A. L., et al. (1999). Regional brain activity during transient self-induced anxiety and anger in healthy adults. *Biological Psychiatry, 46*, 454–465.

AU: Is this journal paginated by issue rather than volume? If not, pls. delete issue no.

Kircher, T. T. J., Senior, C., Phillips, M. L., Benson, P. J., Bullmore, E. T., Brammer, M., et al. (2000). Towards a functional neuroanatomy of self-processing: Effects of faces and words. *Cognitive Brain Research, 10*(1–2), 133–144.

Kjaer, K. W., Nowak, M., & Lou, H. C. (2002). Reflective self-awareness and conscious states: PET evidence for a common midline parietofrontal core. *NeuroImage, 17*, 1080–1086.

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., Loftus, J., Trafton, J. G., & Fuhrman, R. W. (1992). Use of exemplars and abstractions in trait judgments: A model of trait knowledge about the self and others. *Journal of Personality and Social Psychology, 63*, 739–753.

Klein, S. B., Rozendal, K., & Cosmides, L. (2002). A social-cognitive neuroscience analysis of the self. *Social Cognition, 20*(2), 105–135.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273*.

AU: Please provide page range If issue cover date is readily available, pls. also provide.

Kosslyn, S. M., Thompson, W. L., & Alpert, N. M. (1997). Neural systems shared by visual imagery and visual perception: A positron emission tomography study. *NeuroImage, 6*, 320–334.

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. *Cerebral Cortex, 12*, 477–485.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.

Lane, R. D., Fink, G. R., Chau, P. M.-L., & Dolan, R. J. (1997). Neural activation during selective attention to subjective emotional responses. *NeuroReport, 8*, 3969–3972.

Langer, E. J. (1989). *Mindfulness*. Reading, MA: Addison-Wesley.

LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York: Simon & Schuster.

Lieberman, M. D. (2000). Intuition: A social cognitive neuroscience approach. *Psychological Bulletin, 126*, 109–137.

Lieberman, M. D. (2003). Reflective and reflexive judgment processes: A social cognitive neuroscience approach. In J. P. Forgas, K. Williams, & W. V. Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 44–67). New York: Cambridge University Press.

AU: Was a lowercase "v."—edit correct? (If this is supposed to be "von" please spell out in full.)

Lieberman, M. D., Chang, G. Y., Chiao, J. Y., Bookheimer, S. Y., & Knowlton, B. J. (in press). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience*.

AU: If now published, please update.

Lieberman, M. D., Eisenberger, N. I., & Crockett, M. (2003). *An fMRI study of automatic behavior: A disruption account*. Manuscript in preparation.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology, 34*, 199–249.

Lieberman, M. D., Hariri, A., Jarcho, J. J., & Bookheimer, S. Y. (2003). *Two captains, one ship: A social cognitive neuroscience approach to disrupting automatic affective processes*. Unpublished manuscript.

Lieberman, M. D., Jarcho, J. M., Berman, S., Naliboff, B., Suyenobu, B. Y., Mandelkern, M., et al. (**XXXX**). *The neural correlates of placebo effects: A disruption account*. Manuscript submitted for publication.

AU: Please give year of draft here.

AU: If now published, please update.

Lieberman, M. D., Jarcho, J. M., & Satpute, A. B. (2003). *Dissociating evidence-based and intuition-based self-knowledge: An fMRI study*. Manuscript submitted for publication.

Lieberman, M. D., & Pfeifer, J. H. (in press). The self and social perception: Three kinds of questions in social cognitive neuroscience. In A. Easton & N. Emery (Eds.), *Cognitive neuroscience of emotional and social behavior*. Philadelphia: Psychology Press.

AU: If now published, pls. update with year and page range

Lieberman, M. D., & Rosenthal, R. (2001). Why introverts can't always tell who likes them: Multitasking and nonverbal decoding. *Journal of Personality and Social Psychology, 80*, 294–310.

McAdams, D. P. (1993). *The stories we live by: Personal myths and the making of the self.* New York: Guilford Press.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419–457.

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

Melchior-Bonnet, S. (1994). *The mirror.* New York: Routledge.

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a go/no go response inhibition task. *Human Brain Mapping, 12*, 131–143.

Miller, B. L., Seeley, W. W., Mychack, P., Rosen, H. J., Mena, I., & Boone, K. (2001). Neuroanatomy of the self: Evidence from patients with frontotemporal dementia. *Neurology, 57*, 817–821.

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

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston.

Miller, P. J. (1994). Narrative practices: Their role in socialization and self-construction. In U. Neisser & R. Fivush (Eds.), *The remembering self* (pp. 158–179). New York: Cambridge University Press.

Milner, B., Corkin, S., & Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: Fourteen year follow-up study of H.M. *Neuropsychologia, 6*, 215–234.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N., (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences, 99*, 15238–15243.

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin card sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience, 21*, 7733–7741.

Murakami, H. (1994). *Dance, dance, dance: A novel*. New York: Vintage Books.

Mummery, C. J., Shallice, T., & Price, C. J. (1999). Dual-process model in semantic priming: A functional imaging perspective. *NeuroImage, 9*, 516–525.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375–411). New York: Cambridge University Press.

Ochsner, K. N., Bunge, S. A., Gross, J. J., & Gabrieli, J. D. (2002). Rethinking feelings: An FMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience, 14*, 1215–1229.

Packard, M. G., Hirsh, R., & White, N. M. (1989). Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: Evidence for multiple memory systems. *Journal of Neuroscience, 9*, 1465–1472.

Pham, L. B., Taylor, S. E., & Seeman, T. E. (2001). Effects of environmental predictability and personal mastery on self-regulatory and physiological processes. *Personality and Social Psychology Bulletin, 27*, 611–620.

Pizzagalli, D., Pascual-Marqui, R. D., Nitschke, J. B., Oakes, T. R., Larson, C. L., Abercrombie, H. C., et al. (2001). Anterior cingulate activity as a predictor of degree of treatment response in major depression: Evidence from brain electrical tomography analysis. *American Journal of Psychiatry, 158*, 405–415.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., et al. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546–550.

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia, 41*, 245–251.

Preibisch, C., Neumann, K., Raab, P., Euler, H. A., Gudenberg, A. W. von, Lanfermann, H., et al. (2003). Evidence for compensation for stuttering by the right frontal operculum. *NeuroImage, 20*, 1356–1364.

Prout, H. T., & Prout, S. M. (1996). Global self-concept and its relationship to stressful life conditions. In B. A. Bracken (Ed.), *Handbook of self-concept: Developmental, social and clinical considerations* (pp. 259–286). New York: Wiley.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98*, 676–682.

AU: Issue cover date avail.?

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science, 277*, 968–971.

Ramachandran, V. S. (1995). Anosognosia in parietal lobe syndrome. *Consciousness and Cognition, 4*, 22–51.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience, 4*, 546–550.

Satpute, A. B., Sellner, D., Waldman, M. D., Tabibnia, G., Holyoak, K. J., & Lieberman, M. D. (XXXX). *An fMRI study of causal versus associative judgments*. Manuscript in preparation.

AU: Please give year of draft here.Or, if now published, please update.

Schacter, D. L. (1992). Understanding implicit memory: A cognitive neuroscience approach. *American Psychologist, 47*, 559–569.

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., et al. (in press). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philosophical Transactions of the Royal Society of London (Series B)*.

Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C., & Jones-Gotman, M. (2001). Changes in brain activity related to eating chocolate: From pleasure to aversion. *Brain, 124*, 1720–1733.

AU: Issue cover date avail.?

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*, 1657–1661.

Spence, S. A., Brooks, D. J., Hirsh, S. B., Liddle, P. F., Meehan, J., & Grasby, P. M. (1997). A PET study of voluntary moevement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain, 120*, 1997–2011.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99*, 195–231.

Stuss, D. T., Picton, T. W., & Alexander, M. P. (2001). Consciousness, self-awareness, and the frontal lobes. In S. P. Salloway & P. F. Malloy (Eds.), *The frontal lobes and neuropsychiatric illness* (pp. 101–109). Washington, DC: American Psychiatric Publishing.

Taylor, J. G. (2001). The central role of the parietal lobes in consciousness. *Consciousness and Cognition, 10*, 379–417.

Thorne, A., & McLean, K. C. (2002). Gendered reminiscence practices and self-definition in late adolescence. *Sex Roles, 46*, 267–277.

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top–down signal from prefrontal cortex in executive control of memory retrieval. *Nature, 401*, 699–701.

Turner, R. H. (1976). The real self: From institution to impulse. *American Journal of Sociology, 81*, 989–1016.

Ursu, S., Stenger, V. A., Shear, M. K., Jones, M. R., & Carter, C. S. (2003). Overactive action monitoring in obsessive–compulsive disorder: Evidence from functional magnetic resonance imaging. *Psychological Science, 14*, 347–353.

Vogeley, K., Bussfeld, P., Newen, A., Hermann, S., Happe, F., Falkai, P., et al. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *NeuroImage, 14*, 170–181.

Vohs, K. D., & Heatherton, T. F. (2000). Self-regulatory failure: A resource-depletion approach. *Psychological Science, 11*, 249–254.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., et al. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brian activity. *Science, 281*, 1188–1191.

AU: Issue cover date avail.?

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*, 119–125.

Wegner, D. T. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Weissman, D. H., Giesbrecht, B., Song, A. W., Mangun, G. R., & Woldorff, M. G. (2003). Conflict monitoring in the human anterior cingulate cortex during selective attention to global and local object features. *NeuroImage, 19*, 1361–1368.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience, 18*, 411–418.

Wiener, N. (1948). *Cybernetics: Control and communication in the animal and the machine*. Cambridge, MA: MIT Press.

Willingham, D. B., & Dunn, E. W. (in press). What neuroimaging and brain localization can do, cannot do, and should not do for social psychology. *Journal of Personality and Social Psychology.*

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard University Press.

Zysset, S., Huber, O., Ferstl, E., & von Cramon, D. Y. (2002). The anterior frontomedian cortex and evaluative judgment: An fMRI study. *NeuroImage, 15*, 983–991.