CHAPTER 11

Consciousness, Introspection, and the Adaptive Unconscious

Wilhelm Hofmann and Timothy D. Wilson

> Consciousness is like the Trinity; if it is explained so that you understand it, it hasn't been explained correctly.
> —ROBERT J. JOYNT (1981, p. 108)

In the history of psychology, the view that a great deal of mental operations is carried out in an unconscious manner certainly has had its highs and lows. As early as the second half of the 19th century, British physicians and philosophers such as William Carpenter (1874) argued for the existence of nonconscious processes that carry out a large range of everyday mental operations from perception to behavior, a view also advocated by William James (1890). In contrast, Sigmund Freud (1915/1960) presented a more narrow perspective on the unconscious, according to which irrational, unacceptable motives, drives, and feelings are repressed into the unconscious. The psychodynamic point of view dominated and perhaps hindered the scientific exploration of unconscious processing during the first half of the last century. When the behaviorist revolution overtook psychology and the interest in mental processes plummeted altogether, the all-time low of research on the unconscious was reached.

Only with the advent of the cognitive revolution and more sophisticated methods to study mental processes has the tide turned again (e.g., Greenwald, 1992; Hassin, Uleman, & Bargh, 2005). Cognitive, personality, and brain researchers alike share a renewed interest in the exploration of mental processes outside of conscious awareness that nevertheless influence perception, judgments, feelings, and behavior. Together, these processes compose what we refer to as the *adaptive unconscious* (Wilson, 2002), adaptive in the sense that these processes are vital to everyday functioning. The adaptive unconscious is open to scientific investigations using experimental methods such as priming manipulations and indirect or implicit measures (see Section II, this volume), behavioral observations, neuropsychological data, and brain-imaging methods (see Ito, Chapter 5, this volume).

Just as the coastline of an island changes as the surrounding water rises or falls, so do changing views about the range of unconscious processing affect how much mental landscape is considered to be exclusively conscious territory. Faced with the progress in research on nonconscious social cognition in the past decades, one must admit that the sea level is rising fast. This development has sparked renewed interests in fundamental issues such as the functions of consciousness and the interplay between unconscious and conscious cognition. One particularly intriguing aspect of this interplay is whether, and to what extent, people may become consciously aware of the unconscious

underpinnings of their mental lives. On a more general level, this relates to the perennial question of how well we actually know ourselves (e.g., Wilson, 2002).

The present chapter is divided into two main parts. In the first, we begin with a selective review of the scope of unconscious processing. We then ask what types of processes may be reserved for conscious processing and how consciousness may achieve these functions. This leads us to introduce the global workspace approach of consciousness (e.g., Baars, 1997; Dehaene & Naccache, 2001). We believe that this approach, which has been emerging from the interplay of cognitive, neuroscientific, and philosophical investigations, may offer a useful conceptual framework for understanding the interplay between unconscious and conscious social cognition. In the second part, we consider in detail whether and how introspective insight into the adaptive unconscious may be possible. To do so, we largely focus on the relationship between implicit and explicit dispositions (i.e., attitudes, self-esteem, personality, self-concept). We propose a self-inference model that highlights the conditions under which more or less accurate explicit representations about nonconscious dispositions may be formed and organize the literature on implicit–explicit consistency along this model. Our main conclusion is that, even though self-insight into implicit dispositions is often poor, it is not impossible to obtain. Rather, self-insight into implicit dispositions will increase to the degree that valid mental or behavioral outcomes (such as gut feelings or nonverbal behaviors) are detected and used as a basis for self-inference.

Two conceptual issues need to be addressed at the outset. First, the distinction between unconscious versus conscious, which is in the focus of the present chapter, is typically part of dual-system or dual-process theories of the mind (Evans, 2008, for a review). In these models, the unconscious versus conscious distinction is generally viewed as one of several features associated with the broader distinction between automatic and controlled processing, alongside the features of unintentional versus intentional, effortless versus effortful, and uncontrollable versus controllable (Bargh, 1994). Whereas these features may often coincide, an all-or-none view of perfectly correlated features is clearly an oversimplification (see Moors & De Houwer, 2006, for a detailed analysis). Second, it should be noted that the term *unconscious* can refer to different features involved in a psychological process (e.g., Bargh, 1994; Gawronski, Hofmann, & Wilbur, 2006): (1) the conditions or stimuli that set a process in motion, (2) the process itself (i.e., the processing steps and algorithms involved), (3) the output of the process, and (4) the consequences of the output. In the present chapter, we take a conditional view of automaticity by arguing that the *output* of an automatic process may, under certain circumstances, become consciously available. If it becomes consciously available, a large range of follow-up processes are possible (e.g., transformations, corrections, self-inferences) that unconscious information cannot be subjected to. In other words, we view conscious availability as a highly consequential change of representational status that can (but often does not) happen with regard to the output of automatic processing, which would otherwise remain unconscious.

EVIDENCE FOR UNCONSCIOUS PROCESSING

Evidence for unconscious processing has accumulated in the domains of perceptual, affective, semantic, motor, and self-regulatory processes. A variety of methods have been used, including subliminal priming, in which normal participants are presented with stimulus material for such short durations that it cannot be consciously perceived, and supraliminal manipulations, in which participants are aware of the stimulus material (e.g., scrambled sentences, hidden rules) but unaware of how it affects them. Evidence for unconscious processing is obtained if such manipulations reliably affect perception, feelings, judgments, or behavior in spite of participants' reported unawareness. Further evidence for unconscious processing comes from studies on patients with brain lesion (Dietrich, 2007; Weisskrantz, 1997). Some of these lesions appear to wipe out aspects of conscious processing while leaving intact lower order unconscious processes of which patients are completely unaware.

First and most importantly for the present purpose, unconscious processes have also been identified in affective processing (e.g., Murphy & Zajonc, 1993). This point has been made most prominently by the pioneering work of LeDoux (1996) on the brain's fear circuit. LeDoux argued that the limbic system, and in particular the amygdala, is part of an automatic danger-detection system. It has privileged access to incoming sensory information at a relatively crude level of perceptual analysis—before the results of more accurate, but also more time-consuming, high-level perceptual analyses can enter conscious awareness. The amygdala quickly

scans early perceptual processing outputs for signs of danger and can automatically trigger a fear response if it detects such signs. Because the analysis is fast and crude, however, errors can happen (such as when mistaking a tree trunk for a crocodile). Amygdala activation has been demonstrated in response to subliminally presented emotional stimuli, indicating that these affective responses do not require conscious awareness, identification, or additional cognitive processing (Whalen et al., 1998). In the domain of prejudice, Phelps and colleagues (2000) found that amygdala activation in response to outgroup members correlated with implicit attitude scores on the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998). Broadly, it has been argued that all kinds of simple affective processes involve a potentially unconscious contribution from subcortical processes (Berridge, 2003). This core affect (Russell, 2003) or, more colloquially, gut feeling may not become fully conscious (in the sense of access consciousness defined later) but may nevertheless influence behavior and decision making in ways that people do not consciously recognize (Bechara, Damasio, Tranel, & Damasio, 1997; Winkielman, Berridge, & Wilbarger, 2005). Recently, researchers argued that even specific emotions such as guilt can be primed nonconsciously to influence later behavior (Zemack-Rugar, Bettman, & Fitzsimons, 2007).

Evidence for unconscious processing has also been obtained in the domains of perceptual processing (e.g., studies on blindsight; Weisskrantz, 1997), semantic processing (Devine, 1989; Marcel, 1983; Merikle, Joordens, & Stolz, 1995), implicit memory (e.g., Schacter, 1987), and with regard to the execution of overt behavior (e.g., the perception–behavior link; Bargh, Chen, & Burrows, 1996; Chartrand & Bargh, 1999). Finally, even self-regulatory processes, which have typically been attributed to be under the sole domain of conscious operations, may be carried out nonconsciously. Specifically, Bargh, Gollwitzer, Lee-Chai, Barndollar, and Troetschel (2001) argued that goals can be primed by external triggers and then guide self-regulatory behavior without the individual consciously intending to do so (see Ferguson & Porter, Chapter 17, this volume).

Taken together, an entire array of perceptual, affective, semantic, motor, and self-regulatory processes has been shown to occur outside of conscious awareness. These findings make a strong case for views stressing the modularity of mind (e.g., Fodor, 1983), where the human brain is seen as a massive parallel processing system in which special subsystems, or "modules," are dedicated to specific computational purposes. As a result of the modular architecture of the mind, many mental operations and even sequences of interconnected operations (such as the perception–behavior or the perception–affect link) can proceed unconsciously. Clearly, the efficiency gained by delegating a great deal of brainwork to specialized mental compartments or subsystems that need no conscious awareness cannot be overestimated (Bargh, 2005; James, 1890). However, given that such a considerable amount of mental processing appears to occur nonconsciously, one is led to wonder what consciousness, the "cream on the cake of mentality" (Armstrong, 1980), is good for after all? Framed differently, what are the computational and evolutionary advantages enabled by conscious processing?

THE CASE FOR CONSCIOUS PROCESSING

The question about the functional utility of consciousness has been raised by neuroscientists, philosophers, and cognitive psychologists alike. Not surprisingly, quite divergent answers have been proposed. Still, asking the functional question is probably among the most fruitful ways to approach the thorny issue of consciousness (Dennett, 2001). From our reading of the literature, there seems to be some convergence in at least four classes of interrelated functions requiring consciousness (e.g., Baars, 1997; Dietrich, 2007; for an excellent overview, see Dehaene & Naccache, 2001): active information maintenance, flexible combination of information (including rule-based reasoning), the generation of intentional behavior, and the creation of a sense of "self."[1] These functions appear to build on each other in order to enable mental achievements of increasing complexity.

Active Information Maintenance

First, many automatic processing modules appear to have their own domain-specific memory buffer (e.g., iconic memory in the visual system). However, information in these buffers decays very quickly (Sperling, 1960). One primary purpose of consciousness, therefore, may be its capacity to maintain (selected) information in an active state so that it can be used for mental manipulations of all kinds (Dehaene & Naccache, 2001). The ability to bridge temporal gaps by maintaining active internal representations of objects, persons, and so on has been linked to working memory, especially

to the episodic buffer as a common storage system (Baddeley, 2007). One important implication of a common temporary store is that information may be represented in a shared representational format that is closely linked to thought and language (Dehaene & Naccache, 2001). Note that active information maintenance is distinct from long-term memory. The latter is probably best viewed as a modular subsystem contributing information to consciousness (Dehaene & Naccache, 2001). For instance, when asked "What is 'Tiger' Woods's actual first name?" you have to hold the question temporarily in mind (active information maintenance) and then search your long-term memory in the hope for an answer.

Flexible Combination of Information

A second type of conscious mental activity is the ability to combine information in a highly flexible way. This ability forms the basis of deliberate, rule-based reasoning, which requires the flexible selection, manipulation, and combination (e.g., weighting) of information. This idea lies at the heart of influential conceptualizations of working memory as involving not only an episodic buffer used for active information maintenance (see prior discussion) but also the ability to perform a large range of mental operations "on top of" the represented information (Baddeley, 2007; Kane, Bleckley, Conway, & Engle, 2001). Furthermore, the flexible manipulation of information is associated with a subjective feeling of mental effort. Take, for example, mental arithmetic (e.g., solve 24×13) or anagrams (e.g., find an anagram for "scones cousins").[2] In fact, abundant research attests to the limited capacity and resource dependence of executive operations such as switching between tasks, mental transformations, negations, or response inhibition (Deutsch, Gawronski, & Strack, 2006; Schmeichel, Vohs, & Baumeister, 2003). This limited capacity and resource dependence set considerable constraints on the scope and influence of conscious operations.

We do not mean to suggest that consciousness is "smart" and unconscious processing is "dumb." Recent research suggests that distracting people from consciously thinking about the information before letting them choose leads to higher quality decisions (Dijksterhuis, Bos, Nordgren, & van Baaren, 2006; but see Acker, 2008), at least for complex choice tasks involving large amounts of information (Payne, Samper, Bettman, & Luce, 2008). This may be the case because distraction may prevent people from consciously weighing the presented information in an overly selective manner. Most researchers agree, however, that conscious thought is required to follow set rules when information has to be combined in a precise and selective manner (Dijksterhuis & Nordgren, 2006).

Generation of Intentional Behavior

Third, consciousness seems to be strongly associated with the generation of willful, intentional behavior (Baars, 1997; Dehaene & Naccache, 2001). What we refer to here is the set of complex processes involved when people plan, initiate, and correct goal-directed behavior, particularly those that involve long-term planning (e.g., Gilbert & Wilson, 2007; Wilson, 2002). These processes include (1) conscious deliberation about and simulation of the costs and benefits of future action, (2) the commitment to a particular action, (3) the ability to assemble action plans by combining subgoals into a goal hierarchy or sequence, ranging from concrete to abstract, and (4) the ability to adapt to discrepancies between actual and ideal states by inhibiting or overriding interfering behaviors during goal pursuit.

Note that the research on nonconscious goal pursuit cited previously (Bargh et al., 2001) seems to call into question whether consciousness is necessary for intentional behavior to occur. There is more and more evidence that each of these stages can also be carried out nonconsciously (Hassin, Aarts, Eitam, Custers, & Kleiman, 2009). We believe that the essential difference between conscious and nonconscious goal pursuit lies in the flexibility enabled by conscious processing. Consciousness may not be needed for—and may even hamper—the initiation and performance of context-appropriate action plans that, because of their frequent occurrence, have become consolidated into automatized routines (Baumeister, 1984; Cleeremans & Jiménez, 2002). However, consciousness may be needed to provide the "workplace" where action plans can be generated anew or significantly modified and corrected in response to an ever-changing environment (Bargh, 2005; Bongers, Dijksterhuis, & Spears, 2008; Dehaene & Naccache, 2001; Tononi & Edelman, 1998). Somewhat ironically then, "one of the primary objectives of conscious processing may be to eliminate the need for itself in the future by making learned skills as automatic as possible" (Bargh, 2005, p. 53)—but to be back on the alert in case things go wrong. For an illustrative field experiment, simply exchange the coffee machine in your

department by a completely different type and observe how people have to suppress their (now inappropriate) routine actions, use all their conscious attention to figure out the new operating rules, and after only a few days fetch their coffee whistling as absentmindedly as usual.

Typically, our intentional behaviors are accompanied by a subjective feeling of conscious will or a sense of "agency." Yet there are quite diverging views about whether this co-occurrence should be interpreted in causal terms (i.e., conscious operations actually causing intentional behavior) or rather as a spurious correlation (i.e., conscious will and intentional behavior both caused by unconscious mechanisms, reducing conscious will to an epiphenomenon; Libet, Gleason, Wright, & Perl, 1983; Wegner, 2002). Some authors (e.g., Libet, 1999; Wilson, 2002) have adopted something like a middle position, assigning consciousness the role neither of an all-controlling agent nor of a totally passive and utterly inconsequential recipient of fait accompli. A metaphor to describe this midline position is that of a chief executive presiding over a number of independent departments in a well-functioning company. The department officers set their own agendas and do not typically inform the executive about their every move. However, the executive still has a certain picture about what is and should be going on in the company. During times of crisis, he or she can use his or her power to veto (Libet, 1999) or modify important decisions under way if they do not conform to his or her vision of the company's future.

Sense of Self

Consciousness surely contributes to a sense of self. Each conscious state is typically experienced as a unity (Tononi & Edelman, 1998). The integrated, unitary nature of consciousness may provide the experiential basis for a feeling of selfhood. On a larger time scale, conscious experiences are condensed into an autobiographical self through the temporal integration of events into a coherent, personal narrative (Dietrich, 2007; Wilson, 2002). Thus, as a fourth function, consciousness may allow for the development of relatively stable explicit self-views, self-concepts, beliefs, and attitudes. These explicit representations involving the self "bind" together what would otherwise be relatively meaningless reactions on the spur of the moment. Clearly, having a sense of self or "self model" (Vogeley, Kurthen, Falkai, & Maier, 1999) is a great organizing principle. It endows humans with the capacity to coordinate their activities across long time spans (e.g., pursuing a university degree) and in accordance with basic orientations that have proven functional in the past.

CONTEMPORARY APPROACHES TO THE NATURE OF CONSCIOUSNESS: CONSCIOUSNESS AS GLOBAL WORKSPACE

Having sketched a short list of functions that may be associated with consciousness, the next issue to address is obviously how these functions are achieved. That's the million-dollar question in consciousness research. It would be either naive or arrogant to claim that psychology has even come close to solving this riddle. However, the last decade or so has seen great progress in conceptualizing possible underlying mechanisms. This progress has been facilitated by the tools of cognitive neuroscience, which have made it possible to explore the neural architecture that supports consciousness. An extensive review of different approaches to the nature of consciousness is beyond the scope of this chapter (for a review, see Atkinson, Thomas, & Cleeremans, 2000). To sacrifice breadth for detail, we discuss only one basic idea toward which a number of scholars have been converging from quite different fields (Dennett, 2001): the idea of consciousness as a global workspace (Baars, 1997; Dehaene & Naccache, 2001). We then use the global workspace model as a background and elaborate a more specific model about the relation between implicit and explicit cognition and about possible ways by which accurate self-inferences about the adaptive unconscious may be possible.

The Global Workspace as a Communication Platform

Many cognitive theories share the assumption that conscious processing goes beyond the modularity of dedicated subsystems that operate in parallel along established neural pathways. Instead, it is supposed to be supported by a functional architecture that allows for a highly flexible exchange of information across participating processing units (e.g., Baddeley, 1986; Posner, 1994; Shallice, 1988). One idea that has received a great deal of attention in consciousness research (e.g., Baars, 1997) is that of a global workspace "with long-distance connectivity that can potentially interconnect multiple specialized brain areas in a coordinated though variable manner" (Dehaene & Naccache,

2001, p. 13). The computational benefit enabled by such a global workspace is that modular systems that are not directly interconnected to each other nevertheless receive access to each other's content, much like through a shared theater stage (Baars, 1997) or, to use our preferred metaphor, an internal news program.[3] In other words, the global workspace provides a common communication platform onto which important "headlines" from a potentially great range of processing modules are broadcast. This global workspace seems to be highly interconnected with thought and language. Mental contents that enter the global workspace can be readily represented in the currency of "thought" (e.g., "The answer is 'Eldrick'!") or, to use a term applied by Strack and Deutsch (2004; see also Deutsch & Strack, Chapter 4, this volume), in a *propositional* format. These propositional representations can then be communicated via language by drawing on serial speech production centers (Dehaene & Naccache, 2001). Of course, people may sometimes lack the precise words for describing their internal states or be unwilling to report these states explicitly. Nevertheless, the connection between consciousness and language seems to be so close that many authors have made the verbal reportability a central defining feature of consciousness (e.g., Weisskrantz, 1997).

Access to the Global Workspace

The next question to address, of course, is, which particular subsystems share a subscription to the common news channel and which do not? The assumption here is that brain modules not interconnected to the global workspace are permanently cut off from it and can, therefore, never participate in the internal broadcasting of conscious content (for instance, brain stem systems for controlling body functions).

Dehaene and Naccache (2001) suggest that several classes of neural subsystems appear to participate in the workspace: (1) a working memory system used for active representation and manipulation of information, (2) perceptual systems that provide sensual information about the present state of the environment and the body, (3) long-term memory networks that retrieve condensed past workspace outputs (e.g., knowledge laid down in semantic memory) or reinstate past workspace states (e.g., episodic memory of past experiences), (4) evaluation circuits that provide valence in relation to previous experience, (5) motor circuits that are concerned with the preparation and execution of actions, and (6) special processing units dedicated to extract and interpret information about the self (see also Damasio, 1999), allowing for the long-term temporal integration of conscious experiences into a personal narrative or autobiographical self. These systems can all contribute and exchange a wide range of potential contents within the global workspace through their mutual interconnections. This may account for the amazingly rich and differentiated spectrum of conscious states (Tononi & Edelman, 1998).

However, only a tiny fraction of the information from subsystems that are potentially interconnected to the global workspace gains access at a given point in time (to determine each global workspace state). Whether the output of a given process is actually recruited into consciousness may depend on at least two dynamic parameters. First, the information stemming from nonconscious processing has to have a minimal amount of (bottom-up) activation that exceeds a certain threshold. Some processes may simply be too weak to yield sufficient degrees of ongoing activation in order for the represented information to be recruited into the global workspace (see Cleeremans & Jiménez, 2002, for a more differentiated view). Yet the same processes may still be strong enough to trigger further nonconscious processing and may eventually even produce significant behavioral output. This is presumably the case in the subliminal priming studies reviewed previously, where the neural activation triggered by the subliminal stimulus is assumed to fall in between a minimal threshold of information processing and a consciousness threshold, above which information can be recruited into the global workspace.

Second, whether information becomes part of the current workspace is assumed to depend on a mechanism of top-down attentional amplification (Dehaene & Naccache, 2001; Lamme, 2003; Posner, 1994). Most theorists share the postulate that a selective attention system is responsible for gating access to and residence time in consciousness. Specifically, the orientation of attention, similar to a spotlight, is the "mechanism by which modular processes can be temporarily mobilized and made available to the global workspace, and therefore to consciousness" (Dehaene & Naccache, 2001, p. 14). In other words, ongoing activity is not sufficient for information to enter consciousness. It also has to be amplified and maintained over a sufficient amount of time in order to become available for other processes participating in the workspace.[4]

The capacity for what is represented in consciousness at a given point in time is severely

limited. Therefore, at any given time, multiple modular processing outputs compete for access to the global workspace, or "fame in the brain" as Dennett (2001) has put it. Because access is limited and competitive, the contours of the global workspace are not constant (Dehaene & Naccache, 2001; Tononi & Edelman, 1998). Rather, global workspace states change dynamically as, depending on the context, different information gains temporary access and fades out again, giving rise to what James (1890) has called the "stream of consciousness."

Three "Zones" of Consciousness

In sum, the adopted framework of a global workspace leads to three different zones of consciousness, as illustrated in Figure 11.1. The first zone includes the class of information that is forever impervious to consciousness because it is part of lower order subroutines that cannot per se share their contents in the common communication protocol because of a lack of neural connectivity with the workspace (Dehaene & Naccache, 2001). Take, for instance, the "low-level" computations performed by separate dedicated subsystems in your visual cortex. No matter how hard you try (without using a scalpel), you will never be able to perceive separately the intermediate results from the dorsal stream (the "where" pathway) and the ventral stream (the "what" pathway) *before* they are actually integrated on a higher order level of processing. The first zone may be referred to as the realm of unconscious processing in the strict sense. Even though impervious to consciousness, unconscious modular processing can be the source both of processing outputs from the following zone (phenomenal consciousness) and of behavioral output (e.g., Bargh et al., 1996).

The second zone is given by the subclass of higher order processing outputs from modular subsystems that has the *potential* to be recruited into the workspace but does not (currently) gain access to it, either because it is too weakly represented or because it does not receive top-down attentional amplification. This zone refers to what has under different names and emphases been called fringe consciousness (James, 1890), preconscious (Freud, 1924/1968), experiential awareness (Strack & Deutsch, 2004), primary experience

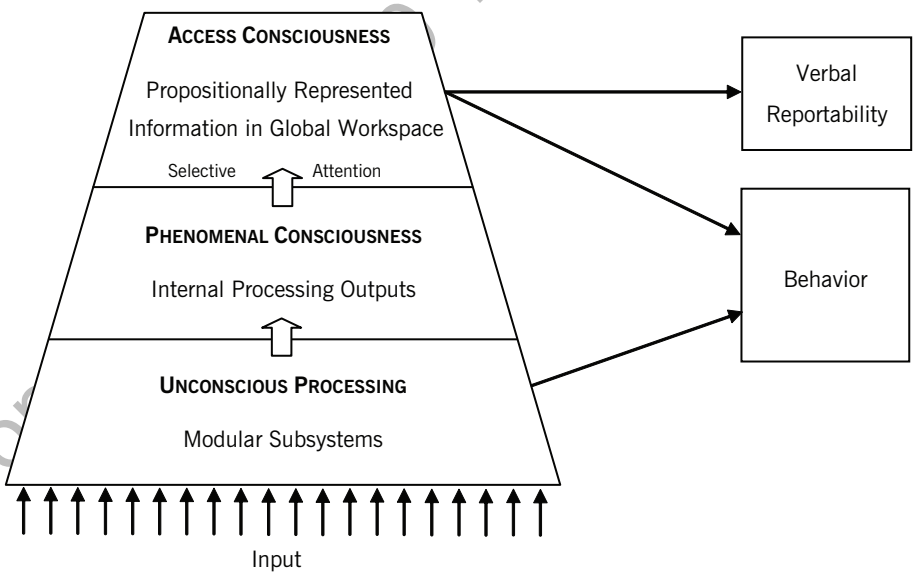

FIGURE 11.1. Three "zones" of consciousness. At the bottom of the information-processing pyramid is the realm of unconscious processing delegated to modular subsystems. A portion of processing outputs from these modular subsystems may gain the status of phenomenal consciousness. Only a small subset of the rich spectrum of phenomenological experience enters the global workspace through mechanisms of selective attention, thereby gaining the status of access consciousness. Access-conscious information is assumed to be represented in a propositional format and is therefore, in principle, verbally reportable. Behavior can be generated via both unconscious processing and consciously formed action plans.

(Farthing, 1992), first-order experience (Lambie & Marcel, 2002), and, the term adopted here, *phenomenal consciousness* (Block, 1995). Phenomenal consciousness may encompass a vast range of sensations and experiences, including perceptions, feelings, and other bodily sensations such as pain that people are having without being aware (in the sense of knowing) that they are having them (Frijda, 2005; Lambie & Marcel, 2002). Importantly, information represented in this way may, under certain conditions (i.e., attentional amplification), become fully consciously accessible and gain the status of the third zone (Dehaene & Naccache, 2001; Lamme, 2003).

The third zone is reserved for the elite information that has passed the gates of selective attention and, therefore, has become recruited into the global workspace. As part of the global workspace, it becomes accessible to a large range of different processing modules, including long-term memory, self-processing units, and speech production centers that translate the information into a propositional format (see prior discussion). This zone represents what has been termed noetic awareness (Strack & Deutsch, 2004), reflective experience (Farthing, 1992), second-order experience (Lambie & Marcel, 2002), and *access consciousness* (Block, 1995). This transition to access consciousness is where an informational "quantum leap" takes place: Suddenly, experience is no more just experience. It has been tagged as such and becomes the object of higher order thoughts directed at it (Rosenthal, 1993; Schooler, 2002), such as when you eventually realize during an interesting sofa conversation that your foot has gone asleep. Access consciousness, therefore, refers to informational contents that people are aware of in the sense of *knowing*. These contents can be communicated, at least approximately, in the form of propositional statements about oneself and the world.

GETTING TO KNOW THE ADAPTIVE UNCONSCIOUS

The discovery that many processes take place outside of conscious awareness has raised the question about what people can actually know about their mental life (Wilson, 2002). In other words, to what degree can we gain conscious self-insight into our adaptive unconscious? In the following, we consider two approaches to this question. A first line of research has demonstrated the limited nature of people's introspective ability. Because this research has been summarized in great detail elsewhere (Nisbett & Wilson, 1977b), we only provide a short overview of the major conclusions from this work. The second line of research, which we focus on in the present chapter, deals with implicit–explicit consistency, that is, the match or mismatch between people's explicit self-reports about their attitudes, self-esteem, self-concept, or motives on the one hand and their scores in thematically corresponding implicit measures on the other.

Completing a questionnaire about oneself (e.g., "Do I like green eggs and ham?") presupposes some kind of introspective activity. Introspection is a very broad term and can involve quite different activities ranging from meditation to psychoanalytic therapy (Wilson, 2002). For the following analysis, we narrow the term down to those cases in which people reflect about their own dispositions or behaviors and define introspection as a conscious mental activity by which attention is directed toward one's own phenomenal sensations and experiences in an attempt to form a self-referential proposition about these experiences with the use of inferential rules. Introspection thus defined is a complex inference process by which mental experiences may become access conscious and become integrated into a coherent propositional judgment involving the self (e.g., "I do not like green eggs and ham; the mere sight of them makes me sick").

Introspective Limits into the Causes of Behavior

A first wave of interest into people's introspective abilities has been sparked by the general finding that people are often unaware of the effects of a given stimulus on their own behavior (for a review, see Nisbett & Wilson, 1977b). In a series of experiments, Nisbett and Wilson showed that participants' verbal reports about the causes of their behavior often stand in stark contrast to the true state of affairs. For instance, in one study Nisbett and Wilson (1977a) manipulated the warmth or coldness of a foreign college teacher presented in a video interview. One group watched the lecturer answer a series of questions in an extremely warm and friendly manner. The second group saw the same person answer exactly the same questions in a cold and distant manner. They then rated the teacher on three attributes that were kept constant across conditions: physical attractiveness, mannerisms, and accent. Consistent with the so-called halo effect, students who saw the warm version of the teacher rated his physical attractiveness, mannerisms, and accent as more appealing than

did students who saw the cold version. However, when asked whether their liking for the teacher had influenced their ratings of the three attributes, participants in both groups denied any effect this might have had. Even more striking, those in the cold condition claimed that their impressions of the three attributes had influenced their overall liking for him. In other words, they had inverted the true causal relationship between the variables in this study.

These and similar results across a wide range of domains (Nisbett & Wilson, 1977b; see also Eastwick & Finkel, 2008, for a recent extension to romantic partner preferences) support the idea that introspecting about the reasons for their behavior is a constructive process in which people have only limited conscious access to the true underlying processes. The true underlying processes, however, may often be produced by the adaptive unconscious (such as, for instance, the halo effect). When asked to provide causal explanations for their behavior, people may retrieve lay theories about themselves (i.e., chronic self-views or personal narratives) or about how certain events are caused and generate explanations that are consistent with these lay theories. However, people may often fail to recognize that the explanations thus generated are incomplete and inaccurate. As Nisbett and Wilson (1977b) implied, introspective reports should become more accurate to the extent that influential stimuli aspects are salient and plausible causes of behavior so that they are attended to and used as a basis for causal inference.

The Implicit–Explicit Relationship: A Self-Inference Model

The research program just described has provided evidence for limited introspective access into the nonconscious underpinnings of mental life by demonstrating discrepancies between people's introspective reports and influential aspects in their stimulus environment that have been varied experimentally. However, the advances in the assessment of implicit representations such as implicit attitudes, self-esteem, and personality self-concepts—without which this Handbook would not have been written—have opened up a new way of studying the issue of introspective access to the adaptive unconscious. Specifically, by investigating the degree of correspondence (or consistency) between explicit and implicit measures of the same construct, new insights about the relationship between implicit and explicit cognition and the factors that moderate it may be gained.

Much of the early debate on implicit social cognition was sparked from empirical findings showing a dissociation between implicit and explicit measures of attitudes, self-esteem, motives, and personality (Greenwald & Banaji, 1995; for a review, see Wilson, Lindsey, & Schooler, 2000). This lack of correspondence between implicit and explicit measures was often taken as evidence for the existence of implicit representations (with regard to objects, people, and the self) that were "introspectively unidentified (or inaccurately) identified" (e.g., Greenwald & Banaji, 1995, p. 5). However, the claim of complete independence may be too strong. That is, the relationship between implicit representations and explicit representations may be *conditional* rather than fixed. Such a view is theoretically consistent with the prior idea that access to consciousness is the result of a competition among many modular processes depending on features of the processes themselves (strength of activation) as well as on the surrounding context (attentional amplification). Hence, the result of unconscious processing may, at different times, become access-conscious or not. Often, dissociations between implicit and explicit representations may emerge if implicit processing outputs are too weak, not attended to, misinterpreted, or suppressed for motivational reasons, and these dissociations may have important consequences for further information processing and well-being (Wilson & Dunn, 2004; Wilson et al., 2000). Given the right conditions, however, implicit processing outputs may enter access consciousness and factor into people's explicit representations about themselves and the world. In these cases, at least gradual self-insight into workings of the adaptive unconscious should be possible.

The conditional view to implicit–explicit consistency suggests a *moderator* approach by which changes in the correlation between implicit and explicit measures are investigated as a function of cognitive or motivational factors. To organize the empirical research that has adopted such an approach, we suggest a self-inference model for implicit–explicit consistency (Figure 11.2) that allows putting the observed correlation between explicit self-reports and implicit measures (arrow i in Figure 11.2) into a broader theoretical perspective (see also Hofmann, Gschwendner, Nosek, & Schmitt, 2005).[5]

As a starting point, we adopt a dual-representation view according to which implicit and explicit dispositions have distinct mental representations (e.g., Strack & Deutsch, 2004; Wilson et al., 2000). More specifically, most models

of implicit cognition propose that implicit dispositions are represented in an associative format (Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). For instance, a negative implicit attitude toward obese people is understood as an associative link between the concept of *obese* and a negative evaluative node. Conscious, explicit dispositions are represented in a propositional format, that is, through a statement about concepts and their interrelations (e.g., "I do not like obese people"). In contrast to associative links, propositional statements carry a truth value and can, therefore, be accepted as true or rejected as false (e.g., Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004).

Viewed this way, the issue of introspective access becomes the question of how information that is represented in an associative format can be translated into a propositional format. Clearly, distinct formats (i.e., associative vs. propositional) preclude direct introspective access in the sense of a direct explicit oversight of implicit representational structures (e.g., Nisbett & Wilson, 1977b). We cannot spot our associations directly. Any approach concerned with the transition from implicit to explicit representations must, therefore, specify *how* such a translation process may be brought about. We believe that much conceptual clarity can be gained by viewing this translation as a *self-inference* process and by drawing on the logic of prominent models of social inference (e.g., Brunswik, 1956; Funder, 1999). That is, even though people cannot directly "take a look" at their implicit associations, they may be able to infer valid self-insights about these properties by registering the perceivable processing outputs that these associations produce.[6] Specifically, the activation of implicit representations may generate (1) inner phenomenal experiences or (2) behavioral cues, which may assist in drawing inferences about the nature of the underlying implicit representation. The accuracy of the self-inference process will depend on the quality of three consecutive steps: First, implicit representations have to be reliably associated with phenomenal or behavioral cues in the first place (cue validity). Second, these valid cues have to be noticed (cue detection). Third, the information contained in these cues has to be integrated into the explicit self-inference (cue utilization). As we will see, obstacles to accurate self-insights about implicit representations may lurk at each stage of this process.

In the following, we first consider cue validity and cue detection for the two plausible routes to self-knowledge about implicit representations, that is, via phenomenal cues (path a × path b in Figure 11.2) or via the self-observation of behavioral cues (path c × path e). If successfully taken, information contained in valid cues will become consciously available for self-inference. We then discuss factors such as additional information retrieval and invalidation processes that may prevent consciously accessible information from having a traceable impact on the final explicit representation. These factors are represented by a feedback loop (arrow f in Figure 11.2) to indicate that conscious content is iteratively transformed through the additional processes involved. Finally, we highlight some issues of measurement (involving paths g and h in Figure 11.2) one should be aware of when drawing conclusions about self-insight based on the correlation between implicit and explicit measures.

Route 1: Introspective Access to Mental Experiences

According to the model, translation can work only if there are valid internal cues that are associated with implicit representations in the first place (path a in Figure 11.2). In principle, many different kinds of phenomenal experiences may be triggered by implicit representations and serve as valid cues, among them gut feelings or core affect (Russell, 2003) as well as experiences of fluency, familiarity, or confidence (e.g., Petty, Briñol, & Tormala, 2002; Reber & Schwarz, 2001; Schwarz & Clore, 1996; Topolinski & Strack, 2009). To illustrate our point, we focus on gut feelings as particularly relevant and likely cues for such self-inferences, especially in the domains of attitudes and self-esteem (Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Gawronski & Bodenhausen, 2006; Phelps et al., 2000). For other domains involving semantic processing, such as implicit stereotyping or the implicit self-concept, knowledge-related mental cues such as fluency, familiarity, or confidence may prove to be more relevant.

In accordance with the prior approach to consciousness, such mental cues do not necessarily enjoy a conscious status. They reflect, in essence, an experiential, preconscious, or phenomenal state of awareness (Strack & Deutsch, 2004). These signals can be weak, ambiguous, or simply unattended to because attention is directed elsewhere. Often, therefore, these cues may fade into oblivion before ever becoming fully consciously available for self-inferences. At the same time, they may still be strong enough to bias judgments and behavior in an unconscious manner (e.g., Winkielman et al., 2005). Following the global workspace

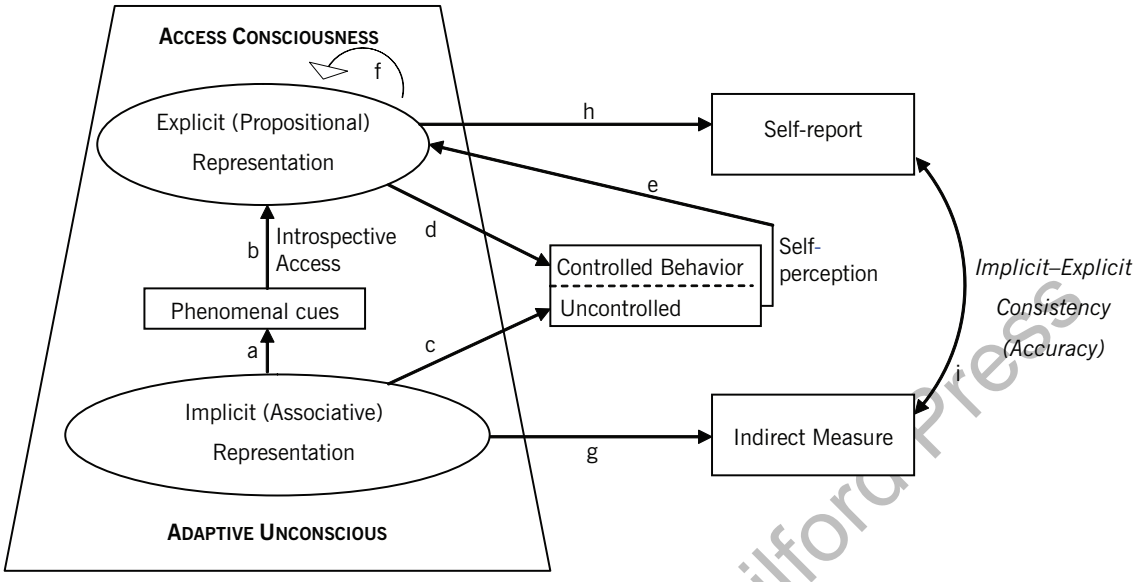

FIGURE 11.2. A self-inference model for implicit–explicit consistency. According to the model, accurate explicit representations of implicit representations depend on whether valid phenomenal or behavioral cues are available (path a or path c, respectively) and whether these cues are detected (paths b and e, respectively). Furthermore, accuracy as assessed by implicit–explicit consistency (i) can be impaired by additional information integration or invalidation processes (arrow f) and by poor measurement (paths g and h).

approach, whether valid cues are introspectively accessible (path b) may hinge on at least two parameters: their strength of activation and whether they receive attentional top-down amplification (Dehaene & Naccache, 2001).

Activation Strength

Given that access to consciousness is the result of a dynamic competition among processing outputs for "fame in the brain," strong representations may yield outputs that have a higher potential for becoming available to access consciousness than weak representations. As a consequence, implicit and explicit representations may correspond more closely for strong representations. Nosek (2005) tested this hypothesis by using a combined strength factor of attitude importance, thought frequency, and familiarity as a moderator of implicit–explicit consistency. Across 57 attitude domains, a multilevel analysis revealed that stronger representations were associated with greater consistency between implicit and self-reported evaluations compared with weaker attitudes. Comparable effects have been observed for attitude *importance* as a proxy for representational strength (Hofmann, Gschwendner, & Schmitt, 2005; Karpinski, Steinman, & Hilton, 2005). Although it is difficult to pinpoint the exact mechanism responsible for these findings, these data generally support the idea that strong implicit representations yield mental experiences that are more easily detectable than those stemming from weak representations.

Is activation strength linearly related to better access into consciousness? Some authors have argued that strength (in terms of frequency of operation) may, rather, have an inverse U-shaped relationship with availability to access consciousness (Cleeremans & Jiménez, 2002). Novel representations may yet be too weak to become noticeable, whereas highly overlearned representations no longer attract attention (even though their representations may be very strong). This is clearly an area for future research and points to the interplay between representational properties and the modulating role of attention that are considered next.

Focus of Attention

Whether attention is focused on valid diagnostic cues or directed somewhere else should have a decisive influence on the detection of these cues. For instance, Hofmann, Gschwendner, and Schmitt (2005) reasoned that, all else being equal, people

high in private self-consciousness (Fenigstein, Scheier, & Buss, 1975) should direct their attention inward more often than people low in private self-consciousness and become more sensitive with regard to inner mental experiences (especially affective and visceral signals). However, in an initial set of studies, private self-consciousness did not by itself yield increased implicit–explicit correspondence (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Hofmann, Gschwendner, & Schmitt, 2005).

These first findings are consistent with the idea that introspection does not necessarily lead to greater accuracy. Rather, the focus at which attention is directed seems to be crucial. Often introspection may involve an extensive information search about *why* one holds a particular attitude or trait (Wilson, Dunn, Kraft, & Lisle, 1989). This may direct the spotlight of attention away from relevant phenomenal cues such as immediately experienced feelings (Gawronski & LeBel, 2008; Hixon & Swann, 1993; Storbeck & Clore, 2008). In a stringent test of this assumption, Gawronski and LeBel (2008) manipulated the focus of attention during introspection about their attitudes. Consistent with the assumed role of attentional focus, implicit–explicit correspondence was close to zero when participants were instructed to focus on the reasons for their preferences; however, substantial implicit–explicit correspondence emerged when participants were instructed to focus on their affective reactions instead. Similarly, Smith and Nosek (2007) showed that implicit and explicit attitudes formed a single factor when participants completed both measures under an affective focus manipulation. In contrast, implicit and explicit attitudes diverged when a cognitive focus was induced. These results are consistent with a meta-analysis across many content domains in which IAT measures correlated more highly with self-reports that implied an affective compared with a cognitive focus (Hofmann, Gawronski, et al., 2005).

Route 2: Self-Observation of Diagnostic Behavior

Rather than "looking inward," a person may take a "look-outward" approach to infer something valid about his or her adaptive unconscious. In accordance with self-perception theory (Bem, 1972), it has been suggested that individuals may gain conscious insights about implicit dispositions by self-observing diagnostic behaviors (path e in Figure 11.2) that are caused by implicit dispositions (path c) (e.g., Gawronski et al., 2006; Nosek, 2005; Wilson, 2002). Via this behavioral detour, individuals may generate accurate self-inferences about their implicit representations.

What kinds of behavioral cues are potentially diagnostic with regard to underlying implicit dispositions (path c)? Accumulating research now shows that implicit dispositions related to social interactions, such as implicit prejudice, implicit anxiety, or implicit shyness, are particularly strongly reflected in nonverbal, relatively uncontrolled modes of behavior (e.g., Asendorpf, Banse, & Mücke, 2002; Dovidio, Kawakami, & Gaertner, 2002; Egloff & Schmukle, 2002). In contrast, explicit dispositions have been shown to primarily predict controlled behaviors (e.g., Asendorpf et al., 2002; Dovidio et al., 2002; indicated by path d in the model). During social interaction, people typically pay relatively little attention to the uncontrolled (e.g., nonverbal) aspects of their behavior, quite in contrast to what they say or what they intentionally do. Hence, asking individuals to adopt the visual perspective of an observer and to focus on these normally unattended behavioral cues may increase self-perceivers' accuracy about their underlying implicit dispositions.

Plausible as the self-perception route seems to be, the available evidence suggests that it is more like a bumpy road than a highway to the adaptive unconscious. Hofmann, Gschwendner, and Schmitt (2009) investigated whether self-perceivers can detect their own implicit dispositions such as implicit extraversion or implicit anxiety from nonverbal behavioral cues contained in video feedback. Across three studies, near-zero correlations between participants' implicit dispositions and their behavioral self-assessment on the respective nonverbal dimensions were consistently obtained. One may object that perhaps the behaviors under investigation were not diagnostic of implicit dispositions. This alternative interpretation can be ruled out by the fact that neutral observers reliably detected participants' implicit dispositions from the videotapes (Hofmann et al., 2009). This suggests that the videos actually did contain valid information about implicit dispositions, which, however, remained undetected by self-perceivers (see also Dovidio et al., 2002).

The observed discrepancy between self and other perception suggests that people have a blind spot when it comes to the detection of their own implicit dispositions from behavioral information. Somewhat paradoxically then, observers making accurate inferences from behavioral cues may sometimes know a target person's implicit dispositions better than the target him- or herself, at least when he or she has no privileged introspective access via route 1. A number of cognitive and

motivational factors may account for such a blind spot (for a more detailed discussion, see Hofmann et al., 2009). For instance, individuals may harbor chronic self-schemas (Markus, 1977), which may bias the perception and categorization of one's own behavior, leading to less accurate self-observations than observations made by unbiased observers. Such perceptual barriers may be overcome if accuracy takes precedence over consistency motivation, for instance when high incentives are provided or when people learn to view themselves through the eyes of others (Wilson, 2002; Wilson & Dunn, 2004).

Limits to Cue Utilization: Information Retrieval and Invalidation Processes

So far we have argued that, at least under certain conditions, potentially diagnostic information about underlying implicit dispositions may become consciously available in a propositional format, but this is only half the story. For the self-inference process to be completed, the potentially diagnostic information has to be integrated into the explicit representation (cue utilization). Consistent with the global workspace approach, conscious information may become subject to additional processing (arrow f in Figure 11.2) that leads to a flexible manipulation of the contents of consciousness in relation to other propositionally available information. Because of the large number of possible operations, these additional transformations of conscious content can take many forms. Here, we pinpoint two broad classes of operations that we believe are key when people form an explicit representation about their attitudes, self-esteem, or self-concept: further information retrieval and invalidation processes.

Further Information Retrieval and Integration

When forming an explicit representation of their dispositions, people are likely to retrieve information from long-term memory. Such information may include past behaviors, knowledge about the world, and knowledge about the self. Consistent with the basic tenets of information integration theory (Anderson, 1981), such additional information will dilute the residual weight or impact of the diagnostic cue for the resulting explicit representation. For instance, when introspecting about whether to buy a given car, a person's explicit representation may be based on more than just the propositional translation of his or her gut feelings about the car, incorporating also judgments derived from information about price, fuel consumption, and maintenance. Hence, the more an explicit representation includes additional propositions that are different from those implied by implicit representation, the more implicit and explicit representations will diverge. Consistent with the information integration hypothesis, implicit–explicit consistency has been found to be lower for those individuals who are high in need for cognition (Florack, Scarabis, & Bless, 2001) or high in working memory capacity (Hofmann, Gschwendner, Friese, Wiers, & Schmitt, 2008, Study 2). Conversely, it has been found that the correspondence between explicit and implicit dispositions increases when explicit self-reports are made spontaneously or under time pressure (Hofmann, Gawronski, et al., 2005; Koole, Dijksterhuis, & van Knippenberg, 2001; Ranganath, Smith, & Nosek, 2007). Taken together, factors that prevent additional information retrieval and integration may lead to self-inferences that are more in line with implicit representations.

Invalidation Processes

The process of information retrieval just described should yield gradual changes in implicit–explicit consistency as more and more information is integrated. In contrast to the dilution effect, certain factors may lead to a complete invalidation of potentially diagnostic cues. As a result of invalidation, the information contained in these cues is not utilized for self-inference at all. Specifically, information that would be potentially diagnostic of implicit representations may become tagged as false, inappropriate, or nondiagnostic and thus become rejected as a basis for an explicit judgment (e.g., Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). If an invalidation process is directed at mental or behavioral cues stemming from implicit representations, explicit and implicit representations are likely to diverge.

How are such invalidation processes brought about? Consistent (no pun intended) with theories of cognitive consistency (e.g., Festinger, 1957), invalidation may result if the propositional implication of an implicit representation (e.g., "I do not like obese people") is logically inconsistent with other currently represented propositions. Of particular interest in this regard are chronically available self-views (e.g., Gawronski & Bodenhausen, 2006; Gawronski, Peters, Brochu, & Strack, 2008). For instance, it has often been reported that people high in motivation to control prejudiced reactions (MCPR) show decreased implicit–explicit correspondence compared with

people low in MCPR (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Hofmann, Gschwendner, & Schmitt, 2005; Payne, Cheng, Govorun, & Stewart, 2005). Assuming that both groups do not differ on their level of implicit prejudice (Devine, 1989), this pattern of findings can be interpreted as the result of an invalidation process in individuals who view themselves as unprejudiced (e.g., "I am an egalitarian person"). Specifically, chronic self-views may lead people to invalidate diagnostic mental experiences (e.g., becoming aware of negative gut feelings toward the prejudiced group) or behavioral self-observations (e.g., becoming aware of the fact that one physically avoids members of the prejudiced group) that are inconsistent with these self-views.

A second source of inconsistency may be given by lay theories about the diagnostic value of certain mental or behavioral events (Nisbett & Wilson, 1977b). A number of largely untested research hypotheses can be derived from this idea. Regarding mental experiences, for instance, people may question whether intuitions and gut reactions can be trusted (Epstein, Pacini, Denes-Raj, & Heier, 1996). All else being equal, low trust in intuition should, therefore, lead to a rejection of mental experiences stemming from implicit dispositions. Accordingly, higher correspondence between implicit and explicit self-esteem has been obtained among people with high faith in their intuition (Jordan, Whitfield, & Zeigler-Hill, 2007). In the case of behavioral cues, people may need to be convinced that their nonverbal behavior may tell them something central about themselves and that the situation in which it is observed is valid enough to warrant a dispositional self-attribution (Hofmann et al., 2009).[7] Finally, some people may be more tolerable than others of inconsistencies between new potentially diagnostic information and existing top-down views about themselves (e.g., Cialdini, Trost, & Newsom, 1995; Kruglanski & Webster, 1996).

It is possible that validation processes that involve the same constellations over and over again may themselves become so highly automatized that invalidation occurs outside of conscious awareness (i.e., before people become consciously aware of discordant implicit processing outputs) (e.g., Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). This suggests the fascinating possibility of unconscious motivational barriers to introspection (Wilson & Dunn, 2004), similar to the Freudian idea of repression (Erdelyi, 1993). To our knowledge, these ideas are a still scientifically largely unchartered territory (see Northoff, Bermpohl, Schoeneich, & Boeker, 2007, for a recently proposed neuroscientific approach).

A Final Note on Measurement

When making inferences about the interplay of implicit and explicit representations, it is important to keep in mind that the measures used will only be proxies—never pure reflections—of the hypothetical constructs they are supposed to assess. Rather, both implicit and explicit measures inevitably suffer from method-specific variance and reliability constraints that impinge on the quality of measurement (paths g and h in the model). Explicit self-report measures are known to be sensitive to the way questions are posed (Schwarz, 1999). Also, self-reports are generally more susceptible to faking than implicit measures (e.g., Asendorpf et al., 2002). Implicit measures likewise suffer from method-specific sources of contamination (e.g., Mierke & Klauer, 2003). In addition, some implicit measures such as affective or sequential priming exhibit unsatisfactory reliabilities (Cunningham, Preacher, & Banaji, 2001; Olson & Fazio, 2003). Furthermore, implicit–explicit correlations may be attenuated by a lack of structural or conceptual correspondence between measures (Gschwendner, Hofmann, & Schmitt, 2008; Hofmann, Gschwendner, Nosek, et al., 2005; Payne, Burkley, & Stokes, 2008) such as when using black and white faces in an Implicit Association Test and questions addressing perceived discrimination in the Modern Racism Scale (Brigham, 1993). Taken together, implicit–explicit correlations can be attenuated because of method-specific variance, measurement error, or lack of correspondence. Interpreting the size of implicit–explicit correlations in an absolute manner may, therefore, lead to false conclusions about the "true" relationship at the level of representations. For this reason, approaches contrasting implicit–explicit consistency as a function of specific experimental manipulations or personality dispositions in a relative manner are to be preferred. Nevertheless, researchers should use corresponding measures and make sure that the situational or dispositional moderators of interest are not confounded with sources of method-specific variance, changes in the reliability of measurement, or both.

SUMMARY AND CONCLUSIONS

Recent social cognition research has provided extensive support for a fundamental insight of

early pioneers of psychology: that a vast amount of everyday human information processing occurs unconsciously. The amount of evidence attesting to the workings of the adaptive unconscious is so massive that it has become a challenging task to take up the cudgels for consciousness as a mental faculty that complements nonconscious processing in important ways. Drawing on global workspace theory (e.g., Baars, 1997; Dehaene & Naccache, 2001), we have argued that consciousness implies the global availability of information to a large range of participating modular systems that otherwise would not gain access to each other's content. Information that has entered the global workspace enjoys the advantage of being actively represented for a longer duration so that it can be subject to all kinds of transformations and serve as the basis for the generation and execution of new, flexibly adjusted action plans. From this perspective, consciousness appears to be most needed when the organism encounters novel situations or problems that demand a careful analysis. Often, however, conscious processing may be superfluous and be effectively bypassed by powerful (but relatively inflexible) automatic processing, resulting in smooth, efficient, but potentially biased behavior.

We have also argued that, without consciousness, there would be no sense of self, at least in the human sense with which we are all familiar. Only conscious processing appears to enable the long-term temporal integration of experience into a coherent narrative about oneself. Because this constructive process of self-inference draws on a limited informational basis, explicit self-views may often be dissociated from the contents of the adaptive unconscious, rendering people strangers to themselves (Wilson, 2002). These dissociations become particularly striking when people are explicitly instructed to provide explanations for their often unconsciously driven feelings, judgments, and behaviors.

Are people always ignorant about their adaptive unconscious? We believe that self-insight into implicit dispositions may be best viewed as a gradual thing. Using the relationship between implicit and explicit measures as a criterion, we proposed a model according to which self-insight into implicit dispositions may be the result of a self-inference process. The accuracy of this process will be highest if valid cues for implicit dispositions exist (cue validity), are detected (cue detection), and are used (cue utilization) as a basis for an explicit inference about one's attitude or self. Although there is evidence that valid cues do exist in the form of internal phenomenal experiences such as gut feelings or self-observed nonverbal behaviors, cue detection and cue utilization may be hampered by a host of variables related to attentional focus, information integration, and invalidation processes. Perhaps most centrally, chronic self-views and false lay theories may reduce the accuracy of self-inferences.

Because introspection can go awry so easily, limited but at the same time variable self-insight into the adaptive unconscious can be expected. As the interplay between implicit and explicit cognition becomes increasingly well understood, so will be the conditions under which people can increase their self-knowledge about the inner workings of their minds. Whether such self-insight is always beneficial to the individual is a different story about which it may be worthwhile to introspect a little more.

ACKNOWLEDGMENTS

We thank Axel Cleeremans, Malte Friese, Nico Frijda, Regina Krieglmeyer, Robert Kordts, Katie Lancaster, Justin Storbeck, Sascha Topolinski, and Reinout Wiers for valuable comments.

NOTES

1. For a more fine-grained list, see, for instance, Jack and Shallice (2001). Also, for reasons of space, we do not delve into the complex relationship between consciousness and learning (e.g., Cleeremans & Jiménez, 2002).
2. The correct answer to the anagram can be found in the title of this chapter.
3. As a reviewer pointed out, the theater metaphor seems more open to Dennett's (1991) criticism of the "Cartesian theater" in which he asks "But who is the audience?" The newscast metaphor seems somewhat less open to this kind of criticism because it suggests some modules broadcasting information that other modules might access, so there is no homuncular audience but only modules as both generators and consumers of information.
4. Whether a given process receives top-down amplification may depend on a complex interplay of the current state of affairs. Rather than invoking some kind of homunculus or Cartesian master who decides what is and what is not attended to at will, Dennett (2001) and others warn against taking the term *top-down* too literally. Instead, attentional amplification should be seen as a competitive process that is heavily constrained by the activation of currently active processors representing current states, processing goals, rewards, and needs of the organism. Their combined influ-

ence on the emerging net pattern of activation may be lumped together as a top-down influence for convenience, but such a view is surely quite different from that of an internal observer sitting "on top" of everything (Dennett, 2001).

5. The present framework shares many commonalities with the associative–propositional evaluation (APE) model proposed by Gawronski and Bodenhausen (2006). Because of its focus on introspective access, the present framework is only concerned with the bottom-up link from implicit to explicit representations and does not incorporate possible reverse, top-down influences from explicit to implicit cognition (Gawronski & Bodenhausen, 2006). Moreover, whereas the APE model is exclusively concerned with evaluations, the present framework provides a background for other types of implicit representations (e.g., implicit self-concept) and other types of processes (e.g., behavioral causation and behavioral observation) as well. Finally, whereas the APE model assumes no introspective limits to associative evaluations, the present framework explicitly takes into account access-related factors.
6. The assumption here is that implicit associations have already been formed by previous encounters with the attitude object. However, even in the case of novel objects, gut feelings may accumulate more quickly than explicit evaluations and, if detected, be used for conscious processing (see Wagar & Dixon, 2006).
7. In fact, the inclusion of a state-inference measure in the work on self-perception discussed previously (Hofmann et al., 2009, Study 3) revealed that self-perceivers made state inferences in accordance with their nonverbal behavior but were hesitant to use these state inferences as a basis for a more general trait inference about themselves. Neutral observers, in contrast, readily turned their state inferences into corresponding trait inferences about the target persons. Thus, a change in perspective (i.e., viewing oneself from the outside) was not sufficient here to overcome the strong tendency of actors to interpret their own behavior in situational terms (Jones & Nisbett, 1972), even though doing so could have resulted in novel insights about themselves.

REFERENCES

Acker, F. (2008). New findings on unconscious versus conscious thought in decision making: Additional empirical data and meta-analysis. *Judgment and Decision Making, 3,* 292–303.

Anderson, N. H. (Ed.). (1981). *Foundations of information integration theory.* New York: Academic Press.

Armstrong, D. M. (1980). What is consciousness? In D. M. Armstrong (Ed.), *The nature of mind and other essays* (pp. 55–67). Ithaca, NY: Cornell University Press.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83,* 380–393.

Atkinson, A., Thomas, M., & Cleeremans, A. (2000). Consciousness: Mapping the theoretical landscape. *Trends in Cognitive Sciences, 4,* 372–382.

Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind.* New York: Oxford University Press.

Baddeley, A. D. (1986). *Working memory.* Oxford, UK: Clarendon Press.

Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford, UK: Oxford University Press.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In J. R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 2, pp. 1–40). Hillsdale, NJ: Erlbaum.

Bargh, J. A. (2005). Bypassing the will: Toward demystifying the nonconscious control of social behavior. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 37–58). Oxford, UK: Oxford University Press.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71,* 230–244.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Troetschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology, 81,* 1014–1027.

Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology, 46,* 610–620.

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

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). New York: Academic Press.

Berridge, K. C. (2003). Pleasures of the brain. *Brain and Cognition, 52,* 106–128.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18,* 227–287.

Bongers, K. C. A., Dijksterhuis, A., & Spears, R. (2008). *On the role of consciousness in goal-pursuit.* Unpublished manuscript.

Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology, 23,* 1933–1967.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley: University of California Press.

Carpenter, W. B. (1874). *Principles of mental physiology, with their applications to the training and discipline of the mind, and the study of its morbid conditions.* New York: Appleton.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76,* 893–910.

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995).

Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology, 69,* 318–328.

Cleeremans, A., & Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R. M. French & A. Cleeremans (Eds.), *Implicit learning and consciousness* (pp. 1–40). Hove, UK: Psychology Press.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measurement: Consistency, stability, and convergent validity. *Psychological Science, 12,* 163–170.

Damasio, A. R. (1999). *The feeling of what happens. Body and emotion in the making of consciousness.* San Diego, CA: Harcourt.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition, 79,* 1–37.

Dennett, D. C. (1991). *Consciousness explained.* Boston: Little, Brown.

Dennett, D. C. (2001). Are we explaining consciousness yet? *Cognition, 79,* 221–237.

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology, 91,* 385–405.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5–18.

Dietrich, A. (2007). *Introduction to consciousness.* Houndmills, UK: Palgrave.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311,* 1005–1007.

Dijksterhuis, A., & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science, 1,* 95–109.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82,* 62–68.

Eastwick, P. W., & Finkel, E. J. (2008). Sex differences in mate preferences revisited: Do people know what they initially desire in a romantic partner? *Journal of Personality and Social Psychology, 94,* 245–264.

Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an Implicit Association Test for assessing anxiety. *Journal of Personality and Social Psychology, 83,* 1441–1455.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology, 71,* 390–405.

Erdelyi, M. H. (1993). Repression: The mechanism and the defense. In D. M. Wegner & J. W. Pennebaker (Eds.), *Handbook of mental control* (pp. 126–148). Englewood-Cliffs, NJ: Prentice Hall.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255–278.

Farthing, G. W. (1992). *The psychology of consciousness.* Englewood Cliffs, NJ: Prentice Hall.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013–1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F., R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50,* 229–238.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 42,* 522–527.

Festinger, L. (Ed.). (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Florack, A., Scarabis, M., & Bless, H. (2001). When do associations matter?: The use of automatic associations towards ethnic groups in person judgments. *Journal of Experimental Social Psychology, 37,* 518–524.

Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology.* Cambridge, MA: MIT Press.

Freud, S. (1960). Das Unbewusste [The unconscious]. In S. Freud (Ed.), *Das Ich und das Es und andere metapsychologische Schriften* (pp. 73–104). Frankfurt: Fischer. (Original work published 1915)

Freud, S. (1968). *A general introduction to psychoanalysis* (J. Riviere, Trans.). New York: Washington Square Press. (Original work published 1924)

Frijda, N. H. (2005). Emotion experience. *Cognition and Emotion, 19,* 473–497.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception.* San Diego, CA: Academic Press.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692–731.

Gawronski, B., Hofmann, W., & Wilbur, C. (2006). Are implicit attitudes unconscious? *Consciousness and Cognition, 15,* 485–499.

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology, 44,* 1355–1361.

Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin, 34,* 648–665.

Gilbert, D. T., & Wilson, T. D. (2007). Prospecting: Experiencing the future. *Science, 317,* 1351–1354.

Greenwald, A. G. (1992). New Look 3: Unconscious cognition reclaimed. *American Psychologist, 47,* 766–779.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464–1480.

Gschwendner, T., Hofmann, W., & Schmitt, M. (2008). Convergent and predictive validity of implicit and explicit anxiety measures as a function of specificity

similarity and content similarity. *European Journal of Psychological Assessment, 24,* 254–262.
Hassin, R. R., Aarts, H., Eitam, B., Custers, R., & Kleiman, T. (2009). Non-conscious goal pursuit and the effortful control of behavior. In J. A. Bargh & E. Morsella (Eds.), *The psychology of action* (Vol. 2, pp. 549–568). New York: Oxford University Press.
Hassin, R. R., Uleman, J. S., & Bargh, J. A. (2005). *The new unconscious.* New York: Oxford University Press.
Hixon, J. G., & Swann, W. B. (1993). When does introspection bear fruit?: Self-reflection, self-insight, and interpersonal choices. *Journal of Personality and Social Psychology, 64,* 35–43.
Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369–1385.
Hofmann, W., Gschwendner, T., Friese, M., Wiers, R. W., & Schmitt, M. (2008). Working memory capacity and self-regulation: Toward an individual differences perspective on behavior determination by automatic versus controlled processes. *Journal of Personality and Social Psychology, 95,* 962–977.
Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology, 16,* 335–390.
Hofmann, W., Gschwendner, T., & Schmitt, M. (2005). On implicit-explicit consistency: The moderating role of individual differences in awareness and adjustment. *European Journal of Personality, 19,* 25–49.
Hofmann, W., Gschwendner, T., & Schmitt, M. (2009). The road to the implicit self not taken: Discrepancies between self- and observer-inferences about implicit dispositions from nonverbal behavioral cues. *European Journal of Personality, 23,* 343–366.
Jack, A. I., & Shallice, T. (2001). Introspective physicalism as an approach to the science of consciousness. *Cognition, 79,* 161–196.
James, W. (1890). *The principles of psychology.* New York: Holt.
Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perception of the causes of behavior. In E. E. Jones, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.
Jordan, C. H., Whitfield, M., & Zeigler-Hill, V. (2007). Intuition and the correspondence between implicit and explicit self-esteem. *Journal of Personality and Social Psychology, 93,* 1067–1079.
Joynt, R. J. (1981). Are two heads better than one? *Behavioral and Brain Sciences, 4,* 108–109.
Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183.
Karpinski, A., Steinman, R. B., & Hilton, J. L. (2005). Attitude importance as a moderator of the relationship between implicit and explicit attitude measures. *Personality and Social Psychology Bulletin, 31,* 949–962.
Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology, 80,* 669–685.
Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: "Seizing" and "freezing." *Psychological Review, 103,* 263–283.
Lambie, J. A., & Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: A theoretical framework. *Psychological Review, 109,* 219–259.
Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7,* 12–18.
LeDoux, J. (1996). *The emotional brain: The mysterious underpinnings of emotional life.* New York: Simon & Schuster.
Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies, 6,* 47–55.
Libet, B., Gleason, C. A., Wright, E. W., & Perl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain, 106,* 623–642.
Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15,* 197–237.
Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35,* 63–78.
Merikle, P. M., Joordens, S., & Stolz, J. A. (1995). Measuring the relative magnitude of unconscious influences. *Consciousness and Cognition, 4,* 422–439.
Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85,* 1180–1192.
Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132,* 297–326.
Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77,* 167–184.
Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology, 64,* 723–739.
Nisbett, R. E., & Wilson, T. D. (1977a). The halo effect: Evidence for unconscious alternation of judgments. *Journal of Personality and Social Psychology, 35,* 250–256.
Nisbett, R. E., & Wilson, T. D. (1977b). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.
Northoff, G., Bermpohl, F., Schoeneich, F., & Boeker, H. (2007). How does our brain constitute defense mechanisms?: First-person neuroscience and psychoanalysis. *Psychotherapy and Psychosomatics, 76,* 141–153.
Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134,* 565–584.
Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science, 14,* 636–639.

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge?: The role of structural fit. *Journal of Personality and Social Psychology, 94,* 16–31.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89,* 277–293.

Payne, J. W., Samper, A., Bettman, J. R., & Luce, M. F. (2008). Boundary conditions on unconscious thought in complex decision making. *Psychological Science, 19,* 1118–1123.

Petty, R. E., Briñol, P., & Tormala, Z. L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. *Journal of Personality and Social Psychology, 82,* 722–741.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., & Gore, J. C. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12,* 729–738.

Posner, M. I. (1994). Attention: The mechanisms of consciousness. *Proceedings of the National Academy of Sciences USA, 91,* 7398–7403.

Ranganath, K., Smith, C., & Nosek, B. (2007). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement. *Journal of Experimental Social Psychology, 44,* 386–396.

Reber, R., & Schwarz, N. (2001). The hot fringes of consciousness: Perceptual fluency and affect. *Consciousness and Emotion, 2,* 223–231.

Rosenthal, D. (1993). Thinking that one thinks. In M. Davis & G. Humphreys (Eds.), *Consciousness* (pp. 15–36). Oxford, UK: Blackwell.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110,* 145–172.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 501–518.

Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology, 85,* 33–46.

Schooler, J. W. (2002). Re-representing consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences, 6,* 339–344.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54,* 93–105.

Schwarz, N., & Clore, G. L. (1996). Feelings and phenomenal experiences. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 433–465). New York: Guilford Press.

Shallice, T. (1988). *From neuropsychology to mental structure.* Cambridge, MA: Cambridge University Press.

Smith, C. T., & Nosek, B. A. (2007). *Affective focus increases the concordance between implicit and explicit attitudes.* Unpublished manuscript.

Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs, 74,* 1–29.

Storbeck, J., & Clore, G. L. (2008). Affective arousal as information: How affective arousal influences judgments, learning, and memory. *Social and Personality Psychology Compass, 2,* 1824–1843.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220–247.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science, 282,* 1846–1851.

Topolinski, S., & Strack, F. (2009). The architecture of intuition: Fluency and affect determine intuitive judgments of semantic and visual coherence and judgments of grammaticality in artificial grammar learning. *Journal of Experimental Psychology: General, 138,* 39–63.

Vogeley, K., Kurthen, M., Falkai, P., & Maier, W. (1999). Essential functions of the human self model are implemented in the prefrontal cortex. *Consciousness and Cognition, 8,* 343–363.

Wagar, B. M., & Dixon, M. (2006). Affective guidance in the Iowa Gambling Task. *Cognitive, Affective, and Behavioural Neuroscience, 6,* 277–290.

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

Weisskrantz, L. (1997). *Consciousness lost and found: A neurophysiological explanation.* New York: Oxford University Press.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInterney, 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.

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

Wilson, T. D., & Dunn, E. (2004). Self-knowledge: Its limits, value, and potential for improvement. *Annual Review of Psychology, 55,* 493–518.

Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 22, pp. 287–343). New York: Academic Press.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107,* 101–126.

Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin, 31,* 121–135.

Zemack-Rugar, Y., Bettman, J. R., & Fitzsimons, G. J. (2007). The effects of nonconsciously priming emotion concepts on behavior. *Journal of Personality and Social Psychology, 93,* 927–939.