The SAGE
Handbook of

Social Cognition

Edited by
Susan T. Fiske and
C. Neil Macrae

4

Consciousness, Metacognition, and the Unconscious

Piotr Winkielman & Jonathan W. Schooler

What are we aware of? What do we know about our own thoughts and feelings? These questions about consciousness and metacognition lie at the heart of social cognition – a field that relies on concepts and methods of cognitive psychology, cognitive science, and neuroscience to understand how people think about others and themselves. Importantly, these questions need answers not only because they are scientifically fascinating but also because it is practically important to examine how in daily life people manage to (sometimes) answer questions like: "What are you thinking about right now?", "How do you feel right now?", "How much do you want to drink?", "Do you want to smoke?", "Do you dislike this social group?", "Do you find this person attractive?", "Did you enjoy this piece of music?", "Did you notice this object?", "Do you understand this passage of text?", and "Were you influenced by this message?" These standard questions require the ability of the person to access, evaluate, and express the cognitive and affective contents of her own mind. And though these answers may often come to our mind effortlessly, giving the illusion of "self-transparency," the process of reaching them is actually tricky and may be fraught with error. After all, not all mental states are conscious. Furthermore, those states that are conscious can be so in various forms and to different degrees. Finally, translating our consciousness into a report may introduce a variety of distortions.

Our chapter reviews theorizing and empirical research on consciousness and metacognition, using the following organizational structure. We start with a few historical remarks, highlighting

the growing interest in psychology in questions of consciousness and metacognition. We then review selected findings on consciousness and metacognition, using an important distinction between mental states that are simply conscious and mental states that are accompanied by rich metacognitive representations of those states. We then move to a discussion of cognition and emotion in the absence of consciousness. Throughout, we address neural correlates of consciousness and metacognition and touch on questions concerning the relationship between social cognition and self-cognition.

HISTORICAL PERSPECTIVE

Just like the actual phenomena, the history of thinking about consciousness and metacognition undergoes a pattern of waxing and waning. Descartes famously argued that consciousness is the only undeniable reality. Everything else could be created by a malicious demon, but our very own thoughts and feelings are necessarily true. For Descartes, consciousness also comes with metacognitive transparency, making our own mental states self-apparent. It is worth noting this because the Cartesian view still underlies much of everyday intuitions about the mind and, as we describe later, is being challenged by modern social-cognitive research that argues that even one's own thoughts may be subject to misrepresentation, in the form of temporal and translational dissociations, or may even be completely

obscured, as in the case of unconscious goals and unconscious affect.

Following some 19th-century philosophers (e.g., Husserl, Brentano), early psychologists had focused on the nature of intentionality and the structure of experience (Boring, 1953). Their method – introspection – assumed that the critical constituents of mental life (thoughts, feelings, volitions) are in principle consciously accessible. They also argued psychology should be fundamentally interested in content-bearing intentional states (thoughts, intentions, goals), but not so much in non-conscious events (e.g., associative chains, reflexes, physiology), which are best left to biologists. As we discuss in detail later, these foundational notions are being challenged by modern researchers in social cognition who argue that states, such as goals, can operate unconsciously.

From the perspective of modern social cognition, another interesting historical aspect was the early interest in social foundations of self-awareness. This was particularly pronounced in the writing of symbolic interactionists (e.g., Mead, 1934) as well as some developmental psychologists (e.g., Vygotsky, 1962). For them, self-consciousness is partly created by the necessity to function in a social community. After all, social coordination requires the individual to be able to adjust to others, see things from their perspective, anticipate their reactions, self-regulate, and, critically, learn to adapt towards oneself the stance of a "generalized other." As we discuss later, these historical themes are reflected in contemporary social-cognitive research on self-monitoring, self-regulation, and self-awareness.

Though consciousness was of central importance for the early psychologists, still they admitted some role for unconscious processes. For example, Helmholtz proposed that vision involves unconscious inferences, whereas James wrote about habits and the subconscious (Kihlstrom, 2007). Behaviorism, and the ambition to make psychology "objective" and equal to other natural sciences, brought disfavor to the methods of introspection and to mentalistic concepts like consciousness. Along with this came the belief that behavior is ultimately under the control of the environment, rather than mysterious "internal forces." This behaviorist credo still guides some modern researchers in social cognition who tend to highlight how much of social behavior is under "environmental control," at least in the sense of "power of the situation" (e.g., Bargh, 2007). Behaviorists also assumed that providing mechanistic explanations of behavior will make concepts like "consciousness" and "volition" disappear from the psychological vocabulary, the way concepts like "phlogiston' or "life force" disappeared from modern scientific physical and

biological vocabulary. Elements of this view are also visible in modern social cognition when researchers argue that as psychological science identifies more and more simple, low-level, and thus unconscious mechanisms it will no longer need complex, high-level concepts, like "consciousness" (e.g., Bargh, 2007). This view is related to a position in philosophy called "eliminative materialism," which proposes replacing high-level mental concepts with references to low-level biological substrates (Churchland, 1981). For example, instead of talking about "seeing," science should talk about specific visual computations in the occipital lobe. However, note that this reductionist argument is logically problematic. For one, it confuses the identification of low-level mechanisms with a satisfying explanation of high-level concepts (Fodor, 1968). But, low-level explanations often offer an uninteresting or even obscuring reduction. For example, providing a complete low-level explanation of physical materials that money is made out of (paper, metal) is not interesting, and does not in any way make the explanations of money in terms of its functions superfluous. In fact, it is counter-productive as it occludes relevant properties (e.g., that a $1 coin functions exactly the same as $1 paper bill). Furthermore, the argument that consciousness can be reduced to smaller, mechanistic, unconscious parts is also problematic because consciousness is an emergent phenomenon that requires an interaction of many (possibly dumb) parts, and thus by definition not reducible to any single one of them. As an analogy, it is impossible to understand the concept of "driving," or how a car works by considering its parts separately.

In the general experimental psychology, the anti-consciousness behaviorist stance started to relax in the mid-1970s when researchers began to tackle issues like controlled and automatic processing (Shiffrin & Schneider, 1977), attentional selection (Kahneman, 1973), and unconscious perception (Marcel, 1983), leading to recognition that consciousness is "respectable, useful, and probably necessary" (Mandler, 1975). Now, the legitimacy of the topic is fully reestablished and "everyone who is conscious, is studying consciousness" (Churchland, 2005).

Social cognition researchers have always been in the game of understanding "consciousness," though, often with the goal of demonstrating its limits. Thus, a pioneering study showed that a stimulus' value can be enhanced via its unconscious mere exposure (Kunst-Wilson & Zajonc, 1980). Another pioneering study showed that people's conscious beliefs about the causes of their own behavior can be at odds with actual causes (Nisbett & Wilson, 1977). These early studies opened the way to a wealth of demonstrations of

unconscious influences on social perception, affect, reasoning, decision, and behavior (e.g., Bargh, 1989; Wegner, 2002; Wilson, 2002). Some researchers became so excited about these findings that they began to argue that "most of a person's everyday life is determined not by their conscious intentions and deliberate choices but by mental processes that are put into motion by features of the environment and that operate outside of conscious awareness and guidance" (Bargh & Chartrand, 1999, p. 462). Along with all the excitement about the unconscious in social cognition came fascination with all things "implicit" – perception, learning, attitudes, self-esteem, self-concepts, stereotypes, partisanship, goals, etc. (Greenwald et al., 2002). Perhaps capturing all this, it is telling that the latest *Handbook of Social Psychology* has a chapter on "Automaticity and the unconscious" but no chapter on consciousness (Fiske, Gilbert, & Lindzey, 2010).

However, there have also been attempts to highlight the limits of the unconscious mind. In fact, some cognitive psychologists seem to relish their role as an empirical and theoretical "police" on more spectacular abilities attributed to the unconscious in social cognition (Kihlstrom, 2008; Merikle & Reingold, 1998; Shanks, 2005, 2006). Some critiques of the scientific as well as public misconceptions about the unconscious also came from within social psychology (e.g., Greenwald, 1992). Interestingly, there has been fairly constant interest within social psychology in processes underlying conscious self-control (Metcalfe & Mischel, 1999) and in the questions of self-regulation and, more recently, "free will" (Baumeister, 2008; Schooler, 2010). Recently, social psychologists have also become more interested in the question of the social origins of consciousness – an interesting return to ideas of Mead and Vygotski (e.g., Baumeister & Masicampo, 2010).

A few historic remarks on metacognition

The history of ideas about metacognition is a bit less oscillatory. Some of this is because the basic concept of metacognition carries less philosophical baggage. After all, few doubt that people have some capacity to think about their own mental states, skills, and capacities. Consequently, there is little controversy in the science of metacognition, understood as an effort to uncover the relation between people's mental states and their beliefs about those mental states (Koriat, 2006).

Specifically, since the time of Piaget, psychologists have wanted to know how much children, and adults, know about their own cognitive processes and skills (comprehension, memory,

emotion, intelligence), and how they can improve them to make their cognitive functioning more efficient. Accordingly, research on the developmental and educational aspect of metacognition has a long history (Flavell, 1979). This history continues to have resonance today. For example, there has recently been some enthusiasm about a Vygotsky-inspired curriculum ("tools-of-the mind") which, among other things, teaches children self-knowledge and self-regulation skills in a social setting (Diamond et al., 2007).

Somewhat trickier problems arise when metacognition invokes concepts such as "a sense of uncertainty," "feeling-of-knowing," "tip-of-the-tongue," or "sense of agency" (Koriat, 2006; Nelson, 1996). These require stronger commitment to a view of metacognition as a particular kind of inward-directed, self-reflective, metaconscious capacity. This capacity is a necessarily complex skill, as it involves re-representing one's own current thoughts (i.e., in addition to thoughts, having thoughts about those thoughts). Also, this capacity is necessarily introspective, implying rich subjectivity, with the ability to discern subtle internal states (e.g., a feeling that one knows the answer, without being able to think of the answer right now). These mentalistic and introspective elements of metacognition make some researchers uncomfortable. But explanations of performance on tests of "metacognitive abilities" that do not postulate any kind of access to introspective "private" information have trouble explaining some basic data (e.g., why first-person observers are so much better in making metacognitive judgments than third-party observers; Jameson et al., 1993). Furthermore, as we will elaborate below, it is hard to non-introspectively explain the main metacognitive issues that we will discuss here, which have to do with a person waxing and waning out of "metacognitive" awareness as a function of her attention to the direction of her own thoughts (Schooler, 2002). As a result, the metacognitive perspective is increasingly gaining in popularity, including in social cognition. Since about the mid-1990s there have been many conferences, books, and papers dedicated to social metacognition and several comprehensive reviews are available (Jost, Kruglanski, & Nelson, 1998; Petty, Briñol, Tormala, & Wegener, 2007; Schwarz, 2004).

Definition and distinctions

But what exactly is "çonsciousness" – with and without the "meta"? Some writers on consciousness devote pages to the elucidation of different meanings (e.g., Zeeman, 2002), but we will highlight only two different senses in which the term appears in psychological literature.

This will also allow us to briefly comment on some current debates in social cognition.

Conscious as "awake and mindful"
The word "conscious" can refer to a global state of an individual. One use of this word is similar to "awake" or "vigilant," as opposed to "asleep" or "comatose." The sleepy–vigilant dimension is typically investigated by neurologists, though some interesting social cognition studies have shown that anesthetized patients form implicit, but not explicit, memory for events during surgery (Kihlstrom et al., 1990). A more relevant meaning of "conscious" as a description of a global state refers to a "mindful," as opposed to a "robot-like," dimension. In that sense, being conscious is the ability to have subjective experiences, wishes, desires, and complex thoughts, and to perform flexible, self-initiated, purposeful behaviors. For example, patients in a persistent vegetative state (PVS) maintain regular sleep–wake cycles, respond to simple stimulation (e.g., withdraw their hand from sharp objects), yet are not considered conscious and possessive of "personhood" by medical experts. This is because PVS patients are unable to make choices, process complex information, show flexible behavior, and initiate purposive actions (Laureys et al., 2002). Interestingly, some recent work that investigated everyday criteria for attribution of a "mind" found that college students basically use two dimensions in their decisions – capacity for agency and capacity for experience – and that these dimension predict how much the "organism" is valued (Gray, Gray, & Wegner, 2007).

It is interesting to contrast the above medical, legal, and everyday view that conscious experience and purposeful action is essential for "personhood" with the dominant view in social cognition that minimizes the role of consciousness in complex thought, choice, and purposive behavior. We actually doubt that, despite some radical "anti-consciousness" declarations in articles and chapters, modern social cognition researchers privately believe there is little distinction between people and complex robots. We also noticed that in informal discussion those researchers readily concede that only *some* thoughts, attitudes, goals or decisions are unconscious, and that only *sometimes* a sense of voluntary control is illusory. We will return to this issue later.

Conscious as "subjectively experienced" and "available for report and intentional use"
A major cluster of meanings for "conscious" centers on the subjective status of a particular mental content (perception, thought, or feeling). First, being conscious of *x* means having *x* represented

in subjective experience – it "feels like something" to be in a conscious state of, say, seeing red (as opposed to just unconsciously reacting to red). Second, being conscious of *x* means having *x* potentially available to report and to use in intentional control of behavior. It is in that second sense that psychologists are interested in whether there are unconscious perceptions, memories, goals, attitudes, or emotions (Bargh, 1989; Greenwald, 1992; Winkielman & Berridge, 2004).

Metaconsciousness
The interest in what makes certain mental content "conscious," and what makes it available for report and control, binds together the research on consciousness with research on metacognition. As mentioned earlier, much of the work under the heading of metacognition simply asks about a person's beliefs about her own comprehension, memory, intelligence, etc. However, other researchers, including one of us in particular (J.W.S.), are interested in metacognition as it reveals the ability (and failures) of human thought to represent itself. Specifically, about a decade ago, Schooler (2001, 2002, Schooler & Schreiber, 2004) proposed a distinction between conscious thoughts that occur without the additional element of explicit self-reflection, and thoughts that are accompanied by the explicit representation of having a thought. That is, mental content could be "experientially conscious," a constituent of ongoing experience, without being explicitly reflected upon. One example of this is the experience of mind-wandering while reading where people can temporarily fail to notice that their eyes are moving across the page but their mind is completely elsewhere. But, mental content can also be "metaconscious" (or "meta-aware") and explicitly represented as a content of one's own consciousness. As for example, when one suddenly realizes that one has been mind-wandering instead of paying attention to what was being read. It is this type of consciousness that is typically assessed when an experimenter asks participants questions like, "What are you thinking about now?", "What goal are you currently pursuing?" or "How happy do you feel right now?" An example of metaconscious affect would be feeling happy and at the same time having an articulated thought: "I am happy now" (Schooler, Ariely, & Loewenstein, 2003; Schooler & Mauss, 2010).

The above distinction between consciousness and metaconsciousness is allied with some terms that have been introduced over the years. For example, it roughly corresponds to the distinction between "first-order" and "second-order" consciousness. For example, Lambie and Marcel (2002) argued that individuals with alexithymia have a first-order experience of emotions but lack

a second-order awareness of the fact that they are experiencing the emotions. However, it is also worth noting also that others use the term "second-order consciousness" in a manner that does not directly map on to the notion of metaconsciousness as used by Schooler and colleagues. For example, philosopher Ned Block (1995) introduced a distinction between phenomenal consciousness (first order) and access consciousness (second order), which resembles a distinction between "perceptual vs conceptual" content. This diverges from the present view, in which both perceptual and conceptual content can become metaconscious. Similarly, a popular distinction in social cognition concerns "construal levels" (see Chapter 12), but again we propose that information from any construal level can become "metaconscious." Finally, Rosenthal (1986) has emphasized a philosophical distinction between first-order and higher-order mental states. In his view, consciousness (of any kind) only occurs when a mental representation is accompanied by a higher-order state explicitly articulating the content of a first-order thought. This is different from our view that first-order mental states can be conscious, even when they are not accompanied by explicit knowledge of their occurrence. For further discussion of these issues, see Winkielman and Schooler (2011).

Functions of consciousness and metaconsciousness

A central assumption in social cognition is that mental information is represented on several levels. Accordingly, much research attention focuses on understanding how these different levels relate to each other (Smith & DeCoster, 2000; Strack & Deutsch, 2004). Here, we ask what distinguishes unconscious, conscious, and metaconscious representations. This question touches on a more general problem of the purpose of consciousness – a problem that has received a variety of functional and mechanistic answers in the psychological literature. In general, researchers have emphasized the idea that consciousness is associated with (i) special access to mental content, and (ii) special functions that can be performed on this content.

Conscious access
Several theories posit that consciousness is a representational system characterized by special access to mental content. One useful framework is the Global Workspace Theory, which proposes

that consciousness functions to allow communication and coordination between the many isolated, parallel sub-processors in the human mind (Baars, 1988). Consciousness constitutes a "global workspace" where various local processors can "broadcast" their outputs and talk to each other in a common internal code. As a result, the previously independent and isolated local processes can coordinate, sequence, and structure their actions, thus helping the organism achieving its goals. For example, by representing tactile, visual, and auditory processes in a common matrix, the putative global workspace allows for the identification of novel cross-modal and cross-temporal connections (e.g., "The sequence of musical notes I just heard has the same pattern as the sequence of colored lights I saw before."). Of course, like almost anything else, cross-modal integration can be automatized into unconscious, suggesting that access in the global workspace might be only needed to initially connect novel sensations and responses. More importantly, the global accessibility of conscious representations makes them available for verbal report and for high-level processes such as conscious judgment, reasoning, and the planning and guiding of action.

But what gives representations conscious or "global" access? Cognitive researchers often emphasize the role of representation "strength" (Cleermans, 2005). The notion of "strength" captures the idea that representations require a certain stability and quality before they can enter working memory, where they can be actively maintained, and become accessible for potential report. One determinant of "strength" is activation, which in turn is determined by many factors, such as stimulus energy (longer presented items are more likely to become conscious than briefly presented items) and recency (more recent items are more likely to become conscious than older items), and so forth. Representational strength is also influenced by focused attention – a perceptual amplifier and selector of events (conscious and non-conscious) that fall within its scope. Thus, an objectively very weak stimulus can reach consciousness, if it receives attentional processing and there is little perceptual competition (Breitmeyer & Ogmen, 2006). Interestingly, recent research shows, somewhat paradoxically, that focused conscious attention may be necessary for some unconscious processes (Koch & Tsuchiya, 2007). For example, subliminal priming is enhanced by attentional cuing of location (Sumner et al., 2006) and limbic responses are stronger if brief affective stimuli fall in the scope of focused attention (Pessoa et al., 2002). These observations may explain why so many successful subliminal priming paradigms in social cognition require

that the subject is paying attention to a specific area on the screen (even if the prime remains invisible). It may also explain why many social-cognitive studies on unconscious processes use "unobtrusive" rather than subliminal priming. In those studies, participants are exposed to stimuli in a definitely conscious, attended, and prolonged fashion (e.g., as a part of a sentence-unscrambling task or a crossword puzzle), with the "unconscious" element being the relevance of the task to subsequent judgment or the importance of a particular stimulus dimension. In short, focused attention might be a precondition for many unconscious effects.

Another factor that modulates whether or not mental content is conscious has to do with anatomical and functional disconnection. Thus, a visual representation in blindsight patients can be strong (e.g., it can drive pointing behavior), but remains unconscious because it is restricted to lower visual pathways (Weiskrantz, 1986). Similarly, habits (e.g., biking) may involve representations that are very robust, but unconscious, because they are only instantiated in the motor system (Cleermans, 2005). A functional disconnection may occur when input fails to cohere with currently processed information. For example, a distinct, prolonged, unusual, and dynamic event (e.g., a gorilla slowly walking through a room of people passing balls to each other) can remain unconscious, when participants "look" at the scene and closely attend to its more familiar features (Simons & Chabris, 1999). One explanation of this "blindness" is that the event is incompatible with the current mental model (i.e., generalized schema) of the situation or with participants' current perceptual goals.

Finally, there is some exciting neuroscientific research on the mechanisms of conscious access. Some evidence suggests that consciousness represents a form of multiregional activation, which is perhaps integrated by oscillatory activity (Tononi, 2004). For example, conscious perception of a stimulus is associated with synchronous activation of higher associative cortices, particularly parietal, prefrontal, and anterior cingulate areas, whereas unconscious perception is associated only with a local activation (Dehaene et al., 2006). Synchronized oscillations may serve as a mechanism for binding information in the "global workspace" discussed earlier. Consistent with these ideas, clinical work has shown that the previously mentioned patients in a PVS (awake but unconscious) show only localized, modality-specific responses to stimuli, whereas patients in a minimally conscious state show coherent responses across multiple sensory and associative systems (Laureys et al., 2002).

Conscious thinking
Some argue that consciousness enables higher-order, meaning-based, logically well-structured processing of information (Block, 1995; Searle, 1997). In contrast, the unconscious is restricted to simpler, associative forms of processing. This distinction resembles, but does not completely overlap with, "dual process" theories in social cognition. For example, Strack and Deutsch (2004) suggest that social cognition is carried out by two systems: a reflective system that relies on knowledge about facts and values, and an impulsive system based on associative links and motivational orientations. The differential information base upon which the two systems rely determines the types of responses they engender. The reflective system, drawing on propositions about the world, leads to responses based on rational considerations. In contrast, the impulsive system, drawing on associations and impulses, leads to non-reasoned actions.

Does processing of meaning require consciousness? This question is a subject of long debate, which touches on tricky issues of the relation between semantic cognition and associationism (McClelland & Rogers, 2003). It is now widely accepted that subliminally presented pictures and words can activate related semantic and affective categories (Greenwald, Draine, & Abrams, 1996; Marcel, 1983). Even subliminally presented single digits can activate magnitude information (Dehaene et al., 2006). Thus, there is no doubt that complex content can be unconsciously activated across meaning dimension. However, the evidence for unconscious semantic processing, rather than automatic activation, is sparse. For example, unconscious priming responds to partial- rather than whole-word information (e.g., fragments of affectively negative words "smut" and "hile" prime the affectively positive word "smile"), is not sensitive to basic operations like negations ("not," "un-," or "dis-"), and cannot process two-digit numbers (Abrahms & Greenwald, 2000). One may wonder whether these limitations arise because subliminal presentations afford very weak stimulus input. However, similar results hold when the input is conscious and only conscious processing capacity is reduced. Thus, processing relational information such as negation ("no disease") or causality ("smoke causes fire") requires conscious capacity, whereas processing information about association does not (Deutsch, Gawronski, & Strack, 2006; Hummel & Holyoak, 2003). In a straightforward but telling demonstration of this point, DeWall et al. (2008) presented participants with a standard set of graduate record examination (GRE) analytical problems, and asked them to solve these problems under typical

conditions or under cognitive load. Not surprisingly, loaded participants did much worse.

Note that even if some complex cognitive skills can ultimately be automatized, it does not follow that the initial acquisition and mastery of these skills is possible without consciousness. For example, most adults can do basic multiplication table automatically, via associative recall (2 × 2 is 4). However, no one believes that the unconscious actually does multiplication. It is generally thought that highly trained operations become automatic over time and can eventually be performed by "dumb" associative retrieval (Logan, 1998; Rickard, 2005; Smith & DeCoster, 2000). As an example, solving a novel mathematical problem, like 78 × 56, is not helped by "thinking about it unconsciously," under cognitive load, or by "sleeping on it."

All this seems to suggest that the unconscious is rather "dumb" (Loftus & Klinger, 1992). Yet, this image has recently been challenged by claims that complex decision are often better made by "unconscious thought" (Dijksterhuis et al., 2006). For example, in one study, participants were presented with 12 attributes for each of four different cars (e.g., "car A has a cup holder," "car B is safe"). That is, participants learned about 48 attributes total. One car had 75% positive attributes, two had 50% positive attributes, and one had 25% positive attributes. Note that the presentation of car–attribute pairs was all mixed, which makes it rather hard to keep track of all the information. Participants were then asked to choose the best car. One group of participants (termed "conscious thinkers") made their decision after 4 minutes of deliberation, and another group (termed "unconscious thinkers") after 4 minutes of engaging in a distracting anagram-solving task. Interestingly, the results showed that 60% of "unconscious thinkers" chose the ostensibly "best" car (i.e., the one with the greatest number of positive attributes), while only 20% of the "conscious thinkers" did so. For the authors, these results show that the unconscious thinking not only facilitates decisions but also is actually better than conscious thinking. To quote: "It should benefit the individual to think consciously about simple matters and to delegate thinking about more complex matters to the unconscious" (Dijksterhuis et al., 2006, p. 1007).

However, other interpretations are possible. First, it is not clear why distraction by anagrams eliminates conscious thought, rather than reducing its amount. If so, perhaps distraction is advantageous because it helps to prevent overthinking (cf. Rey, Goldstein, & Perruchet, 2009). This is not unlike other "verbal overshadowing" effects, to which we return later (Wilson & Schooler, 1991). Also, reduction in capacity may encourage a reliance on simple heuristics that are more

effective in that particular case. In fact, when the "best" solution to a problem is to simply count the number of positive (+) vs negative (–) attributes, engaging in deeper processing that focuses on the attribute meaning (cup holder vs safety) might lead to a suboptimal decision making (Gigerenzer & Goldstein, 1996). Accordingly, the benefits of "unconscious thought" may only apply to so-called "linear integration problems," where the attribute content (cup holder vs safety) does not matter, or can be consciously translated into attribute weights before unconscious "thought." Also, note that it is strange that giving people 4 minutes to think consciously about a simple choice produces such a low response – only 20% correct in the conscious thinking group! After all, it is not that complicated to figure out that a car with 12 (75%) positive attributes is better than a car with 4 (25%) positive attributes. This suggests that the problem encountered by "conscious thinkers" may simply lie in greater confusion about the original attributes and their assignments to cars – perhaps because recall is susceptible to primacy or recency effects and interference by the intermediate task (Shanks, 2006). It may also suggest that conscious thinkers have no problem at all. After all, there are some normative issues with this task, as it is not clear whether a response that simply maximizes + vs – is optimal for most subjects in these types of task (e.g., Why is it best to treat cup holders and safety as equivalent attributes?). In short, while recent evidence does suggest some limitations to extensive deliberation, the degree to which this research implicates truly intelligent unconscious processing remains to be determined. Some doubts about the power of "unconscious thought" are also raised by the fact that the finding of superiority of the decisions in the distraction conditions, over control conditions, is empirically fragile (Acker, 2008). Perhaps reflecting these critiques, more recent writings tone down the radical earlier claims about the general superiority of unconscious processes for decision making (Bongers & Dijksterhuis, 2009).

Conscious control

Consciousness is associated not only with special access to mental content but also with special operations that can be performed on this content. Several of these operations fall under the umbrella name "control," thus linking consciousness to what cognitive scientists call "executive functions" (Norman & Shallice, 1986). One aspect of control is selection. Thus, conscious content can be preferentially attended to and maintained in working memory or discarded if not needed. Another aspect of control is intentionality. Action can be deliberately started and stopped, or can be

delayed until appropriate conditions materialize. Scheduling conflicts can be resolved, and new hierarchies can be established. Finally, with control comes flexibility. Thus, extant mental content can be redeployed in adaptive, non-routine ways, and old response chains can be broken up and rearranged. This simple point was recently elegantly demonstrated in a study where participants had to come up with novel titles, musical improvisations, or interesting drawings. Not surprisingly, participants under cognitive load produced repetitive, inflexible, and uninspiring works (Baumeister, Schmeichel, DeWall, & Vohs, 2007).

One interesting aspect of conscious control is its restricted capacity. Thus, only few elements can be consciously manipulated at any one time, operations must be performed in a serial rather than parallel fashion, and processing is subject to severe bottlenecks (Pashler, 1998). In fact, it is hard to be overwhelmed by the power of the unconscious given how many accidents are caused by ill-advised attempts to multitask (e.g., talking on the cell phone while driving; Levy, Pashler, & Boer, 2006).

Of course, not all forms of control are conscious. The world is filled with very complex mechanical devices that automatically monitor the conditions of subordinate processes and adjust their operation (Shinskey, 1979). The human body has many systems of complex control loops (e.g., homeostatic temperature and blood sugar mechanisms). Furthermore, several "mental" processes automatically adjust their operation based on contextual conditions (Carver & Scheier, 1990). Thus, people unconsciously regulate eye movements to facilitate text processing (Reichle, Pollatsek, Fisher, & Rayner, 1998) and unconsciously adjust hand movements to capture the desired object (Triesch et al., 2003). Finally, people are typically unaware of several layers of control required for coherent speaking and writing.

Reflecting the preceding, and other considerations, Bargh (1989) pointed out that it should be possible to dissociate consciousness and control in social cognition. Subsequently, various forms of automatic control have been proposed (Fitzsimons & Bargh, 2004). One case is the pursuit of "unconscious goals" (see also Chapter 5). Evidence comes from studies where individuals who are primed subliminally or unobtrusively with goal-related words (e.g., "cooperate," "achieve," "memorize") show corresponding adjustment in their behavior (e.g., show more helpful behavior, solve more problems, or remember more details). Furthermore, these adjustments appear to be sensitive to conditions under which the goal is appropriate and to track success at goal pursuit. These findings are interesting, but note that the type of "control" explored in these studies is rather

different than in research on executive functions. First, the "unconscious goal" paradigms rely on an unobtrusive activation of preformulated, standard goals, rather than on the formulation of novel goals. Second, those goals do not require participants to overcome stronger alternative behavior (e.g., go against a prepotent tendency), but operate in situations where behavioral choices are already predetermined (participants can either cooperate or compete, with the likelihood of either action relatively equal). Third, the outcomes are fairly unimportant and do not require participants to reflect on the meaning or consequences of their actions. Accordingly, we suspect that many effects attributed to unconscious goals simply reflect the influence of primes on the interpretation of a vague experimental situation, including giving participants an idea of what and how much they are supposed to do (see Förster, Liebermann, & Friedman, 2007 for discussion). Furthermore, in some "unconscious goal" paradigms participants could also be actually conscious of the goal, but just confused about its source. As we discuss later, there are also multiple ways in which goals could be conscious, but not verbally reported because of temporal and translation dissociations.

Finally, while unconscious goal activation clearly operates under some conditions, unconscious goals have yet to be shown to possess anything approaching the potency or flexibility of conscious goals. To illustrate this, imagine a following experiment. Participants are in a room with both food and drink. Participants are first told once, consciously, that their goal is to eat. Next, participants are given an unconscious priming procedure with multiple words related to the goal of drinking. We predict that very few, if any participants, would behave in accordance with the more recent, but unconscious "goal" to drink.

Indeed, in addition to offering a skewed perspective on the role of consciousness in mediating behaviors, the present trend towards attributing the bulk of human action to unconscious mechanisms may have undesirable effects on people's self-regulatory ability. Evidence supporting this concern comes from Vohs and Schooler (2008), who asked some participants to read an excerpt from Francis Crick's *The Astonishing Hypothesis* that articulates the view that conscious control is an epiphenomenon; i.e., that people lack any meaningful sort of free will. Compared to controls, participants exposed to the message that conscious control is illusory behaved more immorally on a passive cheating task. Moreover, their increased cheating was mediated by decreased belief in free will. In a second experiment, exposure to deterministic statements led participants to overpay themselves on a cognitive test, relative to participants who were exposed to

statements endorsing free will. Of course, such findings do not speak to the actual efficacy of conscious control. Nevertheless, they do raise concerns about the impact that a scientific dismissal of conscious control might have on the population at large, and thus further highlight the importance of not overstating the degree to which science has shown consciousness to be impotent (Schooler, 2010; Shariff, Vohs, & Schooler, 2008).

META-CONSCIOUS MONITORING

As we have discussed, some forms of control might be automatized and unconscious, but others clearly involve consciousness. In fact, one form of control may require explicitly articulating the content of the conscious state to bring it into metaconsciousness. Thus, periodically, the mind encounters situations which require more resource-dependent conscious monitoring processes. In effect, this occurs anytime one explicitly attempts to answer the question, "What am I thinking or feeling?" Given that the answer requires an explicit representation of one's state, over and above the state itself, it offers individuals the opportunity to step out of the situation, which may be critical for many of the innovative behaviors of which individuals are capable. However, it also raises the possibility that the re-description process individuals might get it wrong.

There are two kinds of dissociations between levels of mental representation that follow from the claim that metaconsciousness involves the intermittent re-representation of the contents of consciousness (Schooler, 2002). *Temporal dissociations* occur when metaconsciousness temporarily fails to take stock of the current contents of thought (e.g., failing to notice that one is mind-wandering during reading). *Translation dissociations* occur if the meta-representation process misrepresents the original experience. Such dissociations are particularly likely when one verbally reflects on non-verbal experiences or attempts to takes stock of ambiguous experiences. Several interesting social-cognitive phenomena illustrate these metacognitive dissociations.

Temporal dissociations

Mind-wandering
We suspect that all readers have had the experience of suddenly realizing that, despite their best intentions and the fact that their eyes have continued to move across the page, they have no idea what they have been reading. Such examples

suggests that tacit monitoring systems failed to catch the mind's drifting, and that it takes a higher-level explicit monitoring process to take stock of the specific contents of thought and alert one to the fact that they have wandered off task. Over the last several years, Schooler and colleagues have used the mind-wandering phenomenon to examine the function of meta-awareness in a domain where mind-wandering is antithetical to success (see Smallwood & Schooler, 2006 for a review). Specifically, Schooler, Reichle, and Halpern (2005) developed a paradigm to identify temporal lapses of meta-awareness during the attentionally demanding task of reading. In this research, participants read passages of text and indicated every time they caught their minds zoning out. They were then asked whether they had been aware that they had been zoning out prior to reporting it. In a second condition, participants were additionally probed intermittently and asked to indicate whether they had been zoning out at that moment. The results revealed that participants: (1) frequently caught themselves zoning out during reading; (2) were still often caught zoning out by the probes; and (3) frequently reported that they had been unaware that they had been zoning out, particularly when they were caught by the probes. These findings demonstrate that individuals frequently lack meta-awareness of drifting off, even when they are specifically instructed to be vigilant for such lapses.

Additional studies have elucidated the distinct processes associated with the occurrence of mind-wandering (as revealed by probe-caught episodes) and meta-awareness of mind-wandering (as revealed by self-caught episodes). These studies also examined the role that self-awareness may play in the transition from consciousness to metaconsciousness. For example, Sayette, Reichle and Schooler (2009) examined the impact of alcohol on participants' self-caught and probe-caught mind-wandering during reading. Alcohol was an excellent candidate as a variable likely to impact meta-awareness, given its well-established capacity to reduce people's more general self-awareness (e.g. Hull, 1981). In this study, half of the participants received a real alcoholic beverage, whereas the other half of participants received a placebo that they believed contained alcohol. Participants then engaged in a reading task, in which they read *War and Peace* while their mind-wandering was assessed using both the probe-caught and self-caught procedures. Results showed that alcohol had different effects on self-caught vs probe-caught mind-wandering. Despite mind-wandering more than twice as often as participants in the placebo group (as revealed by the probe measure), participants in the alcohol group were no more likely (and, indeed, were slightly less likely) to

catch themselves in the act. One can argue that participants in the alcohol group should have had many more opportunities to catch themselves, but they did not catch themselves more often than the sober participants. Apparently, they were impaired in their ability to notice mind-wandering episodes, whereas sober participants were more capable of detecting mind-wandering when it occurred.

A related study by Sayette, Schooler, & Reichle (2010) further illustrates the value of the self-caught/probe-caught methodology for assessing the conditions that impact on meta-awareness. This study also explored the role of motivation in meta-awareness. Specifically, this study focused on the effect of cigarette craving on people's mind-wandering and their meta-awareness thereof. Like alcohol consumption, craving is a factor that might reasonably be expected to impact on meta-awareness. This is because craving may engage working memory and self-regulation resources that otherwise might be directed toward noticing that one has become distracted, thereby simultaneously increasing the occurrence of one's mind-wandering and decreasing the probability of catching it. To explore this issue, Sayette et al. employed the reading task described above with smokers, but in this case half of the participants were induced into a craving state by refraining from smoking for at least 6 hours prior to the experiment. All participants were then given the reading task with both self-caught and probe-caught assessment of mind-wandering. The results revealed that craving significantly impacted on participants' meta-awareness of their mind-wandering. Despite mind-wandering more than three times as often as participants in the no-craving condition, participants in the craving condition were not more likely to spontaneously catch themselves mind-wandering.

These findings suggest that craving, like alcohol, disrupts individuals' meta-awareness of the current contents of thought. In so doing, they also highlight possible reasons why it is so difficult to self-regulate during craving states. While conventional wisdom holds that individuals are fully aware of their cravings, some researchers suggest that cravings can occur unconsciously (Robinson & Berridge, 1993). This debate assumes that cravings must be either conscious or unconscious. The alternative framework suggested here, in which consciousness is divided into experiential consciousness (contents of experience) and metaconsciousness (explicit awareness of the contents of consciousness) suggests that individuals can be conscious of craving but lack metaconsciousness of the fact that they are craving. Such a mental state, in turn, might contribute to relapses. Considerable research suggests that relapse behaviors may occur absent-mindedly, with individuals

failing to explicitly notice that they are relapsing (Tiffany, 1990). Traditionally, the occurrence of absent-minded relapses was taken to suggest that cravings need not trigger relapse. However, the view suggested by the present perspective is that absent-minded relapses may be associated with craving states in which one lacks metaconsciousness of the cravings (see Cheyene, Carriere, & Smilek, 2006). By simultaneously promoting absent-mindedness and decreasing meta-awareness, the unnoticed craving state may induce a unique condition in which individuals are maximally likely to engage in a relapse behavior and minimally likely to notice themselves doing so.

Another approach for investigating the role of meta-awareness in mind-wandering is simply asking participants after probe-caught mind-wandering episodes whether or not they had been previously aware that their minds had drifted. Although it is not self-evident that participants would necessarily be able to accurately reflect on the prior meta-awareness of their mental states, a variety of studies suggest that they can accurately discern whether or not they were meta-aware of their mind-wandering, as evidenced by the consistent differences between mind-wandering episodes characterized as occurring with vs without meta-awareness. For example, in one study participants read text one word at a time and were warned that it would periodically stop making sense – become gibberish (Schooler, McSpadden, Reichle, & Smallwood, 2010). We found that although participants typically noticed the occurrence of gibberish as soon as it began, if they failed to notice it right away, they often continued for several sentences before they did. Moreover, if participants were probed following extended durations of missing gibberish, they were markedly more likely to report mind-wandering without meta-awareness than if they were probed at random intervals. Similarly, Smallwood et al. (2008) found that if individuals reported mind-wandering without meta-awareness prior to encountering clues in a Sherlock Holmes story, they were markedly less likely to solve the who-dunnit. Mind-wandering with meta-awareness proved far less disruptive to the construction of a mental model.

When mind-wandering episodes are characterized as having proceeded without meta-awareness, they are not only more disruptive to task performance but also more neurocognitively distinct from on-task performance. In a recent study (Christoff et al., 2009), we scanned participants during a non-demanding vigilance task in which they had to respond to every digit that appeared on the screen with the exception of a specific rarely presented target item for which they had to withhold a response. Because the targets are relatively rare,

it is common to mind-wander during this task and then accidently respond when a target is presented. In this study, participants were periodically probed and asked whether they had been mind-wandering, and if so whether or not they had been aware of that fact. Two sets of regions were more active prior to periods in which individuals reported having been mind-wandering relative to having been on-task. One was the default network (including the medial prefrontal cortex (PFC), posterior cingulate cortex/precuneus region and the temporoparietal junction) which is known to become increasingly recruited with decreasing external task demands. This result is interesting, as the default network is often implicated in stimulus-independent thought (Mason et al., 2007). The second set of regions was the executive network, including the dorsal anterior cingulate cortex (ACC) and the dorsolateral PFC (DLPFC), which becomes activated when individuals engage in demanding mental activity. Of particular interest to the present discussion was the further finding that the discrepancy between areas of neural activity associated with on-task vs off-task thought was markedly more pronounced when individuals lacked meta-awareness of the fact that they were mind-wandering.

The fact that participants' awareness of their mind-wandering impacted on the magnitude of discrepancy between on-task vs off-task thinking suggests potentially important implications for conceptualizing metaconsciousness. First, it reveals the robustness of the distinction between those mental states that are accompanied by meta-awareness and those that are not. One might easily have conjectured that participants would have a hard time reliably characterizing their mind-wandering as having been associated with meta-awareness. However, these findings suggest that this distinction was easy for participants to make. After all, mind-wandering episodes accompanied by meta-awareness (vs no meta-awareness) were associated with different patterns of brain activation. Second, one might have thought that when participants were mind-wandering without realizing it, that their brains would show relatively little activity compared to being on-task. To the contrary, it was during periods of mind-wandering without meta-awareness that individuals showed the maximum distinctive regions of brain activation. Apparently, while mind-wandering without meta-awareness may be especially counterproductive to ongoing task performance, the mind is far from being empty and rather is engaging in vigorous brain activity.

Well-being appraisals
We often fail to explicitly notice our own emotional states (e.g., sullenness, cheerfulness) until

someone points them out to us. If we commonly lack metaconsciousness of affective states, the induction of continuous metaconsciousness may alter the quality of affective experience. Schooler et al. (2003) explored this issue by asking participants to report on-line happiness while listening to hedonically ambiguous music (Stravinsky's *Rite of Spring*). The results showed that continuous hedonic monitoring reduced individuals' post-music ratings of happiness, relative to a condition in which participants listened to music without monitoring. The fact that hedonic monitoring altered participants' experience suggests that (by default) individuals are, at most, only intermittently metaconscious of their affective state.

Automaticity
Automatic behaviors are often assumed to be unconscious (Bargh, 1997; Wood et al., 2002). However, there is a peculiarity to this designation. Consider a person driving automatically while engaging in some secondary task (e.g., talking). Although such driving is compromised, one still experiences the road at some level. Thus, a more appropriate characterization of the consciousness of automatic behaviors may be that they are experienced but lack metaconsciousness, the latter only taking hold when individuals run into difficulty.

Unwanted thoughts
Wegner (1994) suggested that individuals possess an implicit monitoring system that tracks unwanted thoughts (e.g., of a white bear) in order to veer away from them. But what exactly is this system monitoring? Wegner suggests that it is monitoring the contents of preconsciousness (i.e., thoughts that are near, but below, the threshold of consciousness). However, another, and perhaps more intuitive, possibility is that these systems actually monitor the contents of consciousness itself: that is, perhaps individuals can consciously think about a white bear, without explicitly realizing that they are doing so. In this case, the monitoring system can catch the unwanted thought and raise it to the level of meta-awareness, in effect saying: "There you go again, thinking about that unwanted thought." Recent evidence for this account comes from a study in which participants were asked to try not to think about a previous romantic relationship while reading or while simply sitting quietly (Fishman, Smallwood, & Schooler, 2006). As in standard unwanted-thought paradigms, participants were asked to self-report every time they noticed an unwanted thought coming to mind. In addition, however, they were periodically randomly asked whether at that particular moment they were having the unwanted thought.

The results revealed that participants frequently experienced "unnoticed unwanted thoughts" about their previous relationship, which they experienced but failed to notice until they were probed. Furthermore, these unnoticed unwanted thoughts were detrimental to participants' performance on a test of the reading material, suggesting again that they were conscious. Intriguingly, participants for whom the unwanted thoughts carried emotional weight (i.e., they still wished they were in the relationship) were less likely than participants who no longer wanted to be in the relationship to notice the thoughts themselves, and more likely to be caught having the thought. This suggests that cognitive defenses do not banish disturbing thoughts to the unconscious, but rather prevent us from reflecting on them (Schooler, 2001).

Translation dissociations

The idea that metaconsciousness involves a re-representation of the contents of consciousness suggests that, as with any recoding process, some information may get lost or distorted in translation. The likelihood of noise perturbing the translation process may be particularly great when individuals: (1) verbally reflect on inherently non-verbal experiences; (2) are motivated to misrepresent their experience; or (3) possess a lay theory that is inconsistent with their actual experience.

Verbal reflection
Some experiences are inherently difficult to put into words: the structure of a face, the taste of a wine, the complex tonalities of Stravinsky, and the intuitions leading to insights. If individuals attempt to verbalize these inherently non-verbal and holistic experiences, the resulting re-representations may fail to do justice to the original experience. Schooler and Engstler-Schooler (1990) examined the effects of describing faces, which, because of their holistic nature, are notoriously difficult to capture in words. Participants viewed a face and subsequently either described it in detail or engaged in an unrelated verbal activity. When given a recognition test that included a different photograph of the target face, along with similar distractors, verbalization participants performed substantially worse than controls. This effect of verbalization, termed "verbal overshadowing," has been found in variety of other domains of visual memory (Schooler, Fiore, & Brandimonte, 1997), including colors (Schooler & Engstler-Schooler, 1990), shapes (Brandimonte, Schooler, & Gabbino, 1997), as well as other modalities such as audition (Schooler et al., 1997) and taste (Melcher & Schooler, 1996). Similar disruptions

resulting from verbal reflection have also been observed in various other domains hypothesized to rely on non-verbal cognition. Thinking aloud during problem solving can disrupt the intuitive processes associated with insight problem solving while having no effect on the logical processes associated with analytical problem solving (Schooler, Ohlsson, & Brooks, 1993). Verbally reflecting on the basis of affective judgments can interfere with the quality of affective decision making, as assessed both by the opinions of experts (Wilson & Schooler, 1991) and by post-choice satisfaction (Wilson, Lisle, Schooler, Hodges, Klaaren, & LaFleur, 1993). Verbally articulating the basis of the match between analogical stories can reduce people's sensitivity to meaningful deep-structure relationships, while increasing their emphasis on superficial surface–structure relationships (Sieck, Quinn, & Schooler, 1999). Of course, in many cases verbal analysis can be helpful. This occurs when experiences are readily translated into words, due either to the nature of the task (e.g., logical problem solving, Schooler et al., 1993) or to individuals' unique verbal expertise (e.g., wine experts, Melcher & Schooler, 1996). However, the process of articulating experiences can sometimes result in translation dissociations, where meta-awareness misrepresents conscious content.

Motivation
In some situations, people may be motivated to misrepresent their experiences to themselves. For example, homophobic individuals may not want to recognize when they are aroused by depictions of homosexual acts (Adams, Wright, & Lohr, 1996): that is, individuals may consciously experience the arousal but, because of their motivation, fail to become meta-aware of it (see also Lambie & Marcel, 2002). Our perspective also suggests a different view of repression. Freud argued that repression prevented unwanted feelings from coming to consciousness, but we would say that it primarily prevents such feelings from reaching meta-awareness (Schooler, 2001; Schooler & Schreiber, 2004).

Stereotyping
The distinction between conscious and metaconscious states also provides a way of potentially reconceptualizing existing findings in the domain of stereotyping. For example, several researchers work with the notion of "aversive racists," defined as individuals who reveal evidence of implicit racism but are not conscious of their racist tendencies (e.g., Gaertner & Dovidio, 1986; Son Hing et al., 2008). This idea speaks directly to the disparities that can emerge when discrepant

motivations exist at different levels of consciousness. Aversive racists are identified empirically as being those individuals who score high on racism when gauged with implicit measures – i.e., the Implicit Association Test (IAT) – but low when gauged with explicit measures. Evidence for the importance of this distinction comes from the examination of aversive racists' evaluations of stories depicting other-race target individuals, who vary with respect to the degree to which low liking ratings can be attributed to something else besides race. When aversive racists have no excuse for holding negative attitudes towards other-race individuals (e.g., when the target person is characterized as acting politely), then they behave very much like individuals with no racist tendencies. However, when there is an opportunity to justify their discriminatory behavior in a manner that does not necessarily invoke the label of "racist" (e.g., when the target individual behaves in a slightly unfriendly manner), these individuals do act like racists. Son Hing et al. (2008) suggest that aversive racists behave in this fashion because they hold non-conscious racist views that are inconsistent with their conscious views and can only rely on their racist tendencies when they can avoid construing them as such. However, the distinction between consciousness and metaconsciousness raises another possibility: namely, that when individuals experience racist tendencies, they simply do not recognize this experience due to motivation not to take stock of racist reactions. Accordingly, when confronted with the behaviors of an individual towards whom they have racist attitudes, aversive racists experience negative affect. If a justification for this affect exists that is consistent with their views of themselves (i.e., that the individual behaved somewhat rudely), then they embrace this affect. However, when no such outlet is available, they ignore it. Critical to this account, however, is the notion that aversive racists are actually experiencing the affect; it is simply a matter of whether or not they are prepared to allow themselves to take stock of it. Thus, a reasonable alternative way to characterize aversive racists is to suggest that they experience racism but lack explicit awareness of this experience – or, in the terms of metaconsciousness theory, that they exhibit translation dissociations due to a motivation not to acknowledge their racist tendencies.

Implicit attitudes

The distinction between consciousness and metaconsciousness may also have important implications for the study of implicit attitudes more generally (Greenwald et al., 2002). According

to the standard view, attitudes are either consciously held (as revealed by standard attitudinal measures) or else inaccessible to consciousness (as revealed by implicit measures such as the IAT). However, the distinction between consciousness and metaconsciousness raises the possibility that some measures that have been characterized as implicit might instead be conscious but lacking in metaconsciousness. According to this view, there may be some situations in which individuals experience negative attitudes but do not acknowledge them as such, as was suggested as an interpretation of the earlier described findings by Song Hing and colleagues. Some implicit measures may in fact tap such experienced but non-reported attitudes. In contrast, other implicit measures may tap attitudes that are truly unconscious and never experienced. This account may explain why various tests of implicit attitudes such as the IAT versus implicit priming can often be uncorrelated (Gawronski & Payne, 2010). Some tests may be measuring truly unconscious attitudes whereas others may be measuring attitudes that are experienced but occur without meta-awareness.

Faulty theories

Finally, translation dissociations can occur if individuals have a faulty theory about what they should be feeling in a particular situation, which then colors their appraisal of their actual experience. Individuals' theories of how they think they should feel may also color their retrospective appraisal of prior experiences.

We have explored this possibility in the fascinating but controversial domain of childhood memory and childhood happiness. It is likely that at least some aspects of one's answer to the question "How happy was your childhood?" are constructed using one's naïve theories of how one "must have felt" (Skurnik, Schwarz, & Winkielman, 2000). One cue to this assessment comes from cognitive outcomes, such as one's memory for the period, with the bridge linking memory and happiness provided by naïve theories. To investigate this process, Winkielman and Schwarz (2001) first asked participants to recall few (easy task) or many (difficult task) examples from their childhood. Participants were then asked to evaluate the quality of their childhood. After completion of the recall task, but prior to rating their childhood happiness, participants were provided with two different theories. Some participants were told that psychologists have found that a poor childhood memory indicates an unhappy childhood, because many unpleasant experiences are

purged from memory. Others were told that psychologists have found that a poor childhood memory indicates a happy childhood, because many unpleasant experiences are ruminated upon. For both groups, it was emphasized that these are poorly supported hypotheses and that the relevant evidence is limited to small and unusual clinical samples, making it worthwhile to test these hypotheses with a general college population. As predicted, participants' ratings of their childhood depended on the naïve theory offered to them and the subjective difficulty of retrieving memories. When participants found it difficult to retrieve childhood events, participants who were told that happy events fade from memory evaluated their childhood as happier than did participants who were told that had events fade from memory. The opposite happened when the retrieval was easy.

Distortions in retrospective meta-awareness may also be one reason why individuals come to characterize childhood abuse experiences as having been previously forgotten (Schooler, 2001). Recent evidence suggests that characterizations of abuse experiences as having been long forgotten and suddenly remembered may at least sometimes be the result of metacognitive failures in which individuals overestimate the degree to which the memory was previously forgotten (Geraerts et al., 2008, Schooler, 2001). Factors that may contribute to such distortions are faulty theories in which individuals assume that they would surely recall having thought about past trauma, unless it was actively repressed. Evidence in partial support of this view comes from a survey study (Joslyn & Schooler, 2006) in which participants' recollections of prior inappropriate sexual experiences were influenced by the manner in which questions about the experience were asked. Participants were asked to define what experiences constitute sexual abuse either before or after they were asked if those experiences had ever happened to them. Individuals who defined sexual abuse before considering whether they had been abused themselves were more likely to identify themselves as having been abused, and reported being more upset at the time that the event occurred. Apparently, thinking about the experience in the context of sexual abuse invoked theories about how upsetting the experience must have been, which in turn colored the way they recalled feeling at the time. If, as adults, individuals believe they were more traumatized by abuse than they actually were as children, then the experience may not have been as significant at the time, and thus no special forgetting mechanisms may have been required for the experience to have been forgotten.

COGNITION AND AFFECT WITHOUT CONSCIOUSNESS

So far we have emphasized that many mental states are actually conscious, though lacking metacognitive representations. However, there are clearly cases when a mental state has a demonstrable influence on behavior, but cannot be directly accessed by consciousness. It is now almost universally accepted that many perceptual and cognitive operations can occur without individuals being conscious (Kihlstrom, 2007). One classic example comes from research on so-called blindsight patients with damaged primary visual cortex (area V1 of striate cortex) but intact subcortical visual pathways. These patients can discriminate simple visual features (e.g., location, or shape), as revealed in pointing and guessing behavior, while denying any awareness of the discriminated features (Weiskranz, 1986). In that case, the mental representation (e.g., "x is a square") is genuinely unconscious – the patients truly do not know that they "know" what shape was presented. Another classic example is patients with visual agnosia who cannot consciously recognize visual features, but have largely intact sensorimotor abilities, and can pursue actions in response to these "unconscious" features (Milner & Goodale, 1995). As discussed earlier, the social cognition literature is rich with demonstrations of unconscious processing in behavioral experiments (Bargh, 1997).

But what about affect? There is certainly research suggesting that briefly presented affective stimuli can work as unconscious triggers of conscious affective states (Kihlstrom, 2007; Öhman, Flykt, & Lundqvist, 2000; Zajonc, 1994). But can affect itself be consciously inaccessible? In recent years, one strand of our work has focused on exploring a dissociation which occurs when a person is in a demonstrable affective state (as evidenced by its impact on behavior, physiology, and cognition) without having conscious access to that state (Winkielman & Berridge, 2004). The idea of "unconscious affect" may seem initially strange – after all, how can there be feelings that are not felt? Note, though, that evolutionarily speaking, conscious representation of affect in the form of a "feeling" is a late achievement compared with the ability to respond affectively to relevant stimuli, which is present in animals that extend deep into our evolutionary ancestry, such as fish and reptiles. Accordingly, the basic affective neurocircuitry is contained in the subcortical brain, and can operate even in the absence of cortex (Berridge, 2003). A reader interested in a recent review of neuroscientific evidence for "unconscious emotion" may want to

consult Winkielman, Berridge, and Sher (2011). However, evolutionary and neuroscientific considerations can only be suggestive of unconscious affect in typical humans. We thus embarked on a program of psychological research to test this possibility using standard experimental paradigms with normal college participants.

Impact of subliminal affective stimuli on behavior, but not on subjective experience

One way of testing unconscious emotion involves separating the impact of affective stimuli on behavior from their impact on conscious feelings. This was done in a series of studies by Winkielman, Berridge, and Wilbarger (2005). In Study 1 participants were first presented with a series of subliminal emotional facial expressions – happy, neutral, or angry. Immediately after this affect induction, participants were given two counterbalanced tasks. One task required participants to self-report on conscious feelings of valence and arousal – a measure of introspective access to the current affective state. The other task was a measure of behavioral impact of the current affective state and asked participants to take a pitcher of lemonade-like beverage to pour into their cup as much as they wanted and to drink as much as they wanted.

The results of this study illustrate that subliminal emotional expressions can influence people's actual consumption behavior. Subliminal happy facial expressions caused participants to pour more into their cup, and to drink more than angry facial expressions. Importantly, participants reported no conscious awareness of any intervening change in their subjective state, as measured by their reports of valence and arousal: that is, they did not report feeling more pleasant (or aroused) after happy facial expressions than angry expressions.

This study suggests that consciously inaccessible affective states can drive behavior. However, how does unconscious affect accomplish this result? After all, many steps of the consumption behavior are consciously mediated, in the sense that they require the ability to understand verbal instructions, form an intention, and execute complex movements. To understand this process, we tested the idea that unconscious affect directly modifies the perceived value of presented options, without changing conscious feelings. In Study 2, we presented people with the same series of subliminal happy or angry faces. Then some participants were given just a single sip of the fruit beverage, and were asked to rate its perceived

value. Other participants rated various shades of their current feelings on a 20-item scale. The results showed that the subliminal expressions influenced the perceived value of the drink, with happy faces leading to higher ratings of willingness to pay and the desire to drink. Again, no changes in feelings were found. In sum, this study supported the idea that unconscious affect works via change in the perception of the desirability and value of presented options, without manifesting itself as a change in subjective experience.

It is also worth highlighting that in both studies described above, the effect of prime was amplified by thirst (Winkielman et al., 2005). This is consistent with other work from social psychology suggesting that unconscious cues interact with affective and motivational states in determining goal-oriented behavior (see Chapter 5; see also Custers & Aarts, 2010; Ferguson, 2007). However, our interpretation is that thirst does not represent an "unconscious goal," but rather a low-level motivational amplifier of incoming affective cues (Winkielman et al., 2011).

True dissociation or lack of meta-awareness?

One can ask whether participants in the above studies had no experience of their affective reaction (true unconscious affect) or whether they simply lacked meta-awareness of conscious affective states (experienced but unrealized affect). After all, it is possible that participants were not attending on-line to their feelings, or did not consider their subliminally biased feelings as a potential impairment to their judgments and thus ignored them. Other studies from our lab have addressed this possibility.

In one series of studies, participants were subliminally flashed facial expressions of happiness and anger that were masked by to-be-rated Chinese ideographs (Winkielman, Zajonc, & Schwarz, 1997). In addition, the studies employed various attributional manipulations in which some participants were informed about the possibility of change in their affective experience and offered possible causes of such change (irrelevant "other" pictures, irrelevant background music). If participants' feelings are indeed consciously accessible and form the basis of their judgments, such attributional manipulations should trigger corrective processes, such as discounting and augmenting (Schwarz & Clore, 1983). However, the results of these studies showed no evidence of any discounting or augmenting effects, as predicted by the attributional account (Winkielman et al., 1997). Furthermore, there was also no evidence

for feelings in participants' self-reports of experience, again consistent with the idea that the facially triggered affect was unconscious.

Results of a recent study reaffirmed this conclusion with another paradigm (Bornemann, Winkielman, & van der Meer, 2011). The study investigated whether people can somehow "feel" their reactions to briefly presented emotional stimuli by deliberately focusing on their internal subjective state. Specifically, participants were briefly flashed happy, neutral or angry faces and were asked to identify their valence. One group of participants was instructed to do this task while focusing on their feelings. One control group was instructed to use a visual focus strategy, and another group received no strategy instructions. The results showed no beneficial effect of feeling-focus instruction on detection rates, suggesting that the affective responses to faces were unconsciously unavailable, despite participants trying to use them.

Physiological consequences: Unfelt but genuine affect

One issue often raised about these findings is the nature of the unconscious affective states. Are the unconscious states elicited by subtle and brief stimuli (like faces) simply "evaluative," in the sense of changes in activation of value-related, but cold concepts like "goodness" or "badness"? Or are they genuinely "hot" – in the sense of being represented across multiple physiological and psychological systems? This is a difficult question, especially since unconscious affective states are likely to be weaker and less differentiated (Clore, 1994). Still, some recent evidence from our lab suggests that unconscious affect involves genuine physiological changes and is distinguishable from pure evaluative states.

In the just-described study by Bornemann et al. (2011), in addition to behavioral responses, we monitored participants' physiological activity using facial electromyography (EMG). The results revealed distinct physiological responses for different stimulus valences. Angry faces produced the strongest reactions on the frown-generating corrugator supercilii, and happy faces produced the lowest reactions. This suggests that briefly presented and unfelt faces generate at least some muscular reactions.

However, one could argue that the facial EMG responses to faces could simply represent motor mimicry. Thus, in other studies we assessed the reactions to unconsciously presented facial pictures using physiological measures that serve as an index of activation of a low-level positive

affective system, such as post-auricular startle reflex (Starr, Lin, & Winkielman, 2007). Participants showed more post-auricular startle reflex to unconscious happy rather than angry faces, suggesting genuine, albeit weak, activation of the low-level affect system. Recently, we followed up on this finding and presented participants with affective pictures (faces or pictures) designed to induce an affective state, and evaluative words (adjectives and nouns) selected to activate relevant semantic dimensions. As expected, faces and pictures elicited stronger physiological responses than valence-matched words. Critically, changes in behavior were driven more strongly by pictures and faces, consistent with the notion that the behavioral changes reflect genuine "hot" affect, rather than cold evaluation (Starr, Winkielman, Golgolushko, 2008).

In sum, a range of findings from behavioral and physiological experiments suggest that one can obtain genuine dissociation between an underlying affective process and its conscious, experiential awareness. As such, they give credence to the notion of "unconscious affect." Importantly, though, the idea of "unconscious emotion" does not imply that conscious feelings are an unnecessary "icing on the emotional cake" (LeDoux, 1996). Conscious happiness, anxiety, anger, guilt, and sadness are critical in people's life. They may as well be what makes life worth living. As an example, most people probably would not spend money on substances that make them only "unconsciously happy," but result in "happy" behavior. In contrast, they are clearly willing to spend on substances, like alcohol or drugs, which influence conscious states, without doing much good to behavior. Besides recreational reasons, conscious emotions are actually useful in judgments and decisions, giving the decision makers valuable feedback that they might, but are not forced to, explicitly consider in making choices (Winkielman, Knutson, Paulus, & Trujillo, 2007).

Unconscious or not metaconscious?

The preceding discussion argued for some cases where mental states, like affect, can remain genuinely unconscious. However, this discussion also illustrates how hard it is to distinguish empirically between processes that are genuinely unconscious or conscious but not meta-aware. This is tricky, as a failure of verbal report can result from either an absence of experience or an absence of meta-awareness. However, future studies may help to adjudicate between them. For example, if unreported states are indeed represented in consciousness, then in principle they should be influenced

by manipulations targeting consciousness, such as cognitive load or explicit monitoring. Experiences in the absence of meta-awareness can also be revealed retrospectively. For example, it is possible to catch conscious, but not meta-aware, states with the external probe procedure, which, as described earlier, was successfully employed in research on zoning out and unnoticed, unwanted thoughts. In principle, similar strategies could be used in other paradigms. For example, perhaps individuals who fail to spontaneously report a goal (e.g., competition) could be caught consciously experiencing such goal states, if probed at the right time. It may also be possible to refine individuals' ability to carefully scrutinize their prior state. For example, if individuals are experiencing something without concurrent meta-awareness, then in principle it may be possible to have them later recall and metacognitively represent their prior state, when some additional source of self-insight (e.g., mindfulness training) is provided or biases due to motivation are removed. Thus, individuals going through the break-up of a romantic relationship may retrospectively recognize past experiences of jealousy or anger that had previously escaped meta-awareness. Of course, retrospective analyses have their own pitfalls, as it is possible to infer states that may not have actually been experienced at the time (Joslyn & Schooler, 2006). However, if individuals are capable of retrospectively reporting states for which they lack a basis for inference (e.g., determining whether they were subject to subliminally presented mood manipulations), then the conclusion that the state was experienced seems reasonable. Ultimately, determination of whether or not unreported states are genuinely unconscious, or experienced but not meta-aware, will come down to an assessment of the preponderance of evidence in each case.

SUMMARY AND IMPLICATIONS

This chapter discussed the conscious, metaconscious, and the unconscious, focusing on the domain of social cognition. Our goal here was to offer not only a review of theories and findings from our and related labs but also to problematize some common assumptions and provide a fresh perspective on classic issues. We began by emphasizing the resurgence of consciousness in contemporary science, and contrasted this view with eliminativism of mainstream social cognition. We then discussed what makes mental events conscious, and highlighted the role of consciousness in complex thought and action. Following this, we distinguished between conscious and metaconscious states. We considered various disconnections

between conscious and metaconscious representations, highlighting both temporal and translation dissociations. We also highlighted the possibility that some seemingly unconscious states may in fact be conscious states lacking in meta-awareness, and showed how this possibility affects the interpretation of some major social-cognitive phenomena. Finally, we considered some genuinely unconscious mental events, including the curious phenomenon of "unconscious affect."

What are some implications of all this for social cognition? Let's start with some concrete recommendations. First, as we have discussed throughout, researchers need to test more thoroughly if a mental content – a thought, a goal, or a feeling – is possibly experienced but not metacognitively available before declaring it unconscious. Besides psychological tools, there are now plenty of physiological tools available that can facilitate such determination. Second, given that metacognition necessarily involves a process of translation, often into a verbal form, of participants' thoughts and feelings, researchers should use methods which capture the "first-order" nature of these states . This may again involve behavioral techniques as well as psychophysiology. More theoretically, as we have emphasized throughout, many of today's popular frameworks depend (both in the sense of conceptual foundations as well as popularity) on the assumption that the process is unconscious (e.g., unconscious goals, unconscious thought, unconscious emotion, etc.). Challenging this assumption may require major reformulation of these theories.

CONCLUSION

Over the last several years there has been remarkable progress in the psychology and neuroscience of consciousness and metacognition. Clearly, some believe that certain features of consciousness, like subjectivity, qualia, and intentionality, will never submit to scientific scrutiny (Searle, 1997). Others believe that as research progresses, "consciousness" will disappear from the scientific vocabulary (Rey, 1983). In contrast, we believe that future scientific advances will only increase our appreciation of the marvels of consciousness and metacognition. We hope that social cognitive researchers will continue to contribute vigorously to this progress.

ACKNOWLEDGMENT

REFERENCES

Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science, 11*, 118–124.

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.

Adams, H. E., Wright, L. W., & Lohr, B. A. (1996). Is homophobia associated with homosexual arousal? *Journal of Abnormal Psychology, 105*, 440–445.

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

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford Press.

Bargh, J. A. (1997). Advances in social cognition. In R. S. Wyer, Jr. (Ed.), *The automaticity of everyday life* (pp. 1–61). Mahwah, NJ: Erlbaum.

Bargh, J. A. (2007). Social psychological approaches to consciousness. In P. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness.* New York: Cambridge University Press.

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

Baumeister, R. F. (2008). Free will in scientific psychology. *Perspectives on Psychological Science, 3*, 14–19.

Baumeister, R. F., & Masicampo, E. J. (2010). Conscious thought is for facilitating social and cultural interactions: How mental simulations serve the animal–culture interface. *Psychological Review, 117*, 945–971.

Baumeister, R. F., Schmeichel, B. J., DeWall, C. N., & Vohs, K. D. (2007). Is the conscious self a help, a hindrance, or an irrelevance to the creativity process? In A. M. Colombus (Ed.), *Advances in psychology research* (Vol. 53, pp. 137–152). New York: Nova.

Berridge, K. C. (2003). Comparing the emotional brain of humans and other animals. In: R. J. Davidson, H. H. Goldsmith, & K. Scherer (Eds.), *Handbook of affective sciences* (pp. 25–51). New York: Oxford University Press.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences 18*, 27–287.

Bongers, K. C. A., & Dijksterhuis, A. (2009). Consciousness as a trouble-shooting device? The role of consciousness in goal-pursuit. In E. Morsella, J. A. Bargh, & P. Gollwitzer (Eds.), *The Oxford handbook of human action* (pp. 589–604). New York: Oxford University Press.

Boring, E. G. (1953). A history of introspection. *Psychological Bulletin, 50*, 169–189.

Bornemann, B., Winkielman, P., & van der Meer (2011). Can you feel what you don't see? Using bodily feedback to detect briefly presented emotional stimuli. *International Journal of Psychophysiology.* doi:10.1016/j.ijpsycho.2011.04.007

Brandimonte, M. A., Schooler, J. W., & Gabbino, P. (1997). Attenuating verbal overshadowing through visual

retrieval cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 915–931.

Breitmeyer, B. G., & Ogmen, H. (2006). *Visual masking: Time slices through conscious and unconscious Vision* (2nd ed.). New York: Oxford University Press.

Carver, C. S., & Scheier, M. S. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review, 197*, 19–35.

Cheyene J. A., Carriere J. S. A., & Smilek D. (2006). Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness & Cognition, 3*, 578–592.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences, 106*, 8719–8724.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*(2), 67–90.

Churchland, P. S. (2005). Brain wide shut. *New Scientist*, 30 April, pp. 46–49.

Cleeremans, A. (2005). Computational correlates of consciousness. *Progress in Brain Research, 150*, 81–98.

Clore, G. L. (1994). Why emotions are never unconscious. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 285–290). New York: Oxford University Press.

Crick, F. (1994). *The astonishing hypothesis.* New York: Scribner's.

Custers, R., & Aarts, H. (2010). The unconscious will: How the pursuit of goals operates outside of conscious awareness. *Science, 329*, 47–50.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences, 10*, 204–211.

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.

DeWall, C. N., Baumeister, R. F., & Masicampo, E. J. (2008). Evidence that logical reasoning depends on conscious processing, *Consciousness and Cognition, 17*, 628–645.

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387–1388.

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.

Ferguson, M. J. (2007). On the automatic evaluation of end-states. *Journal of Personality and Social Psychology, 92*, 596–611.

Fishman, D., Smallwood, J., & Schooler, J. W. (2006). Unwanted and meta-unknown. Unpublished manuscript.

Fiske, S. T., Gilbert, D. T., & Lindzey, G. (Eds.) (2010). *Handbook of social psychology* (5th ed.). New York: Wiley.

Fitzsimons, G. M., & Bargh, J. A (2004). Automatic self-regulation. In Baumeister, R. F. & Vohs, K. D. (Eds.),

Handbook of self-regulation: Research, theory and applications (pp. 151–170). New York: Guilford Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34,* 906–911.

Fodor, J. (1968). *Psychological explanation.* New York: Random House.

Förster, J., Liberman, N., & Friedman, R. (2007). Seven principles of automatic goal pursuit: A systematic approach to distinguishing goal priming from priming of non-goal constructs. *Personality and Social Psychology Review, 11,* 211–233.

Gaertner, S. L., & J. F. Dovidio. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination and racism: Theory and research* (pp. 61–89). Orlando, FL: Academic Press,

Gawronski, B., & Payne, B. K. (Eds.) (2010). *Handbook of implicit social cognition: Measurement, theory, and applications.* New York: Guilford Press.

Geraerts, E., Lindsay, D. S., Merckelbach, H., Jelicic, M., Raymaekers, L., Arnold, M. M., et al. (2008). Cognitive mechanisms underlying recovered memory experiences of childhood sexual abuse. *Psychological Science, 20,* 1.

Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650–669.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315,* 619.

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

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3–25.

Greenwald, A. G., Draine, S. C., & Abrams, R. L. (1996). Three cognitive markers of unconscious semantic activation. *Science, 273,* 1699–1702.

Hull, J. G. (1981). A self-awareness model of the causes and effects of alcohol consumption. *Journal of Abnormal Psychology, 90,* 586–600.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110,* 220–264.

Jameson, A., Nelson, T. O., Leonesio, R. J., & Narens, L. (1993). The feeling of another person's knowing. *Journal of Memory and Language, 32,* 320–335.

Joslyn, S., & Schooler, J.W. (2006). Influences of the present on the past: The impact of interpretation on memory for abuse. In L.G. Nilsson & N. Ohta (Eds.), *Memory and society: Psychological perspectives.* New York: Psychology Press.

Jost, J. T., Kruglanski, A. W., & Nelson, T. O. (1998). Social metacognition: An expansionist review. *Personality and Social Psychology Review, 2,* 137–154.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Kihlstrom, J. F. (2007). The psychological unconscious. O. John, R. Robins, & L. Pervin (Eds.), *Handbook of*

Personality: Theory and Research (3rd ed.). New York: Guilford Press.

Kihlstrom, J. F. (2008). The automaticity juggernaut. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are we free? Psychology and free will* (pp. 155–180). New York: Oxford University Press.

Kihlstrom, J. F., Schacter, D. L., Cork, R. L., Hurt, C. A., & Behr, S. E. (1990). Implicit and explicit memory following surgical anesthesia. *Psychological Science, 1,* 303–306.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Science. 11,* 16–22.

Koriat, A. (2006). Metacognition and consciousness. In: *Cambridge handbook of consciousness.* New York: Cambridge University Press.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science, 207,* 557–558.

Lambie, J. A., & Marcel, A. J. (2002). Consciousness and the varieties of emotion experience: A theoretical framework. *Psychological Review, 109,* 219–259.

Laureys, S., Antoine, S., Boly, M., Elincx, S., Faymonville, M. E., & J. Berre, J., et al. (2002). Brain function in the vegetative state. *Acta Neurol Belg, 102,* 177–185.

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

Levy, J., Pashler, H., & Boer, E. (2006). Central interference in driving: Is there any stopping the psychological refractory period? *Psychological Science, 17,* 228–235.

Loftus, E. F., & Klinger, M. R. (1992). Is the unconscious smart or dumb? *American Psychologist, 47,* 761–765.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Mandler, G. (1975). Consciousness: Respectable, useful, and probably necessary. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 229–254). Hillsdale, NJ: Lawrence Erlbaum Associates.

Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15,* 197–237.

Mason, M. F., Norton, M. I., Van, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: The default network and stimulus independent thought. *Science, 315,* 393–395.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience,*

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

Melcher, J., & Schooler, J. W. (1996). The misremembrance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste. *The Journal of Memory and Language, 35,* 231–245.

Merikle, P. M., & Reingold, E. M. (1998). On demonstrating unconscious perception. *Journal of Experimental Psychology: General, 127,* 304–310.

Metcalfe, J., & Mischel, W. (1999). A hot/cool system analysis of delay of gratification: Dynamics of willpower. *Psychological Review, 106,* 3–19.

Milner, D., & Goodale, M. (1995). *The visual brain in action.* Oxford: Oxford University Press.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*(2), 102–116.

Nisbett, R. E., & Wilson, D. S. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–253.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory.* New York: Plenum Press.

Öhman, A., Flykt, A., & Lundqvist, D. (2000). Unconscious emotion: Evolutionary perspectives, psychophysiological data and neuropsychological mechanisms. In R. D. Lane, L. Nadel & G. Ahern (Eds.), *Cognitive neuroscience of emotion* (pp. 296–327). New York: Oxford University Press.

Pashler, H. E. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Petty, R. E., Briñol, P., Tormala, Z. L., & Wegener, D. T. (2007). The role of meta-cognition in social judgment. In E. T. Higgins & A. W. Kruglanski, (Eds.), *Social psychology: A handbook of basic principles* (2nd ed., pp. 254–284). New York: Guilford Press.

Pessoa, L., McKenna M., Gutierrez E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proc Natl Acad Sci USA, 99,* 11458–11463.

Reichle, E., Pollatsek, A., Fisher, D.L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125–157.

Rey, A., Goldstein, R. M., & Perruchet, P. (2009). Does unconscious thought improve complex decision making? *Psychological Research, 73,* 372–379. doi:10.1007/s00426-008-0156-4

Rey, G. (1983). A reason for doubting the existence of consciousness. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. 3, pp. 1–39). New York: Plenum Press.

Rickard, T. C. (2005). A revised identical elements model of arithmetic fact representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 250–257.

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Research Reviews, 18,* 247–291.

Rosenthal, D. (1986). Two concepts of consciousness. *Philosophical Studies, 49,* 329–359.

Sayette, M. A., Reichle, E. D., & Schooler, J. W. (2009). Lost in the sauce: The effects of alcohol on mind-wandering. *Psychological Science, 20,* 747–752.

Sayette, M. A., Schooler, J. W., & Reichle, E. D. (2010). Out for a smoke: The impact of cigarette craving on zoning-out during reading. *Psychological Science, 21,* 26–30.

Schooler, J. W. (2001). Discovering memories in the light of meta-awareness. *The Journal of Aggression, Maltreatment and Trauma, 4,* 105–136.

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

Schooler, J. W. (2010). What science tells us about free will. In R. F. Baumeister, A. R. Mele, & K. D. Vohs (Eds.), *Free will and consciousness: How might they work?* (pp. 191–218). Oxford: Oxford University Press.

Schooler, J. W., Ariely, D., & Loewenstein, G. (2003). The pursuit and assessment of happiness can be self-defeating. In I. Brocas & J. Carrillo (Eds.), *The psychology of economic decisions.* Oxford: Oxford University Press.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 17,* 36–71.

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At a loss from words: Verbal overshadowing of perceptual memories. In D. L. Medin (Ed.), *The psychology of learning and motivation* (pp. 293–334). San Diego, CA: Academic Press.Schooler, J. W. & Mauss, I. B. (2010). To be happy and to know it: The experience and meta-awareness of pleasure. In K. Berridge & M. Kringlebach (Eds.), *Pleasures of the brain.* Oxford: Oxford University Press.

Schooler J. W., McSpadden M., Reichle E. D., Smallwood J. (2010). Unnoticed nonsense: Mind-wandering can prevent people from realizing that they are reading gibberish. Manuscript submitted for publication.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122,* 166–183.

Schooler, J. W., Reichle, E. D., & Halpern, D. V. (2005). Zoning-out during reading: Evidence for dissociations between experience and meta-consciousness. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 204–226). Cambridge, MA: MIT Press.

Schooler, J., & Schreiber, C.A. (2004). Experience, meta-consciousness, and the paradox of introspection. *Journal of Consciousness Studies, 11*(7–8), 17–39.

Schwarz, N. (2004). Meta-cognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14,* 332–348.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45,* 513–523.

Searle, J. (1997). *The mystery of consciousness.* New York: New York Review Press.

Shanks, D. R. (2005). Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202–220). London: Sage.

Shanks, D. R. (2006). Are complex choices better made unconsciously? *Science, 313,* 716.

Shariff, A., Schooler, J. W., & Vohs, K. (2008). The hazards of claiming to have solved the hard problem of free will. In J. Baer and R. Baumeister (Eds.), *Psychology and free will* (pp. 181–204). Oxford: Oxford University Press.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190.

Shinskey, F. G. (1979). *Process control systems* (2nd ed.). New York: McGraw-Hill.

Skurnik, I., Schwarz, N., & Winkielman, P. (2000). Drawing inferences from feelings: The role of naive beliefs. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior* (pp. 162–175). Philadelphia: Psychology Press.

Sieck, W. R., Quinn, C. N., & Schooler, J. W. (1999). Justification effects on the judgment of analogy. *Memory and Cognition, 27,* 844–855.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28,* 1059–1074.

Smallwood, J., McSpadden, M. C., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition, 36,* 1144–1150.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132,* 946–958.

Smith, E. R., & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4,* 108–131.

Son Hing, L. S., Chung-Yan, G. A., Hamilton, L. K., & Zanna, M. P. (2008). A two-dimensional model that employs explicit and implicit attitudes to characterize prejudice. *Journal of Personality and Social Psychology, 94,* 971–987.

Starr, M. J., Lin, J., & Winkielman, P. (2007). The impact of unconscious facial expressions on consumption behavior involves changes in positive affect: Evidence from EMG and appetitive reflex-modulation. Poster presented at 47th Annual Meeting of Society for Psychophysiological Research, Savannah, GA.

Starr, M. J., Winkielman, P., & Gogolushko, K. (2008). Influence of affective pictures and words on consumption behavior and facial expressions. Poster presented at Society for Psychophysiological Research, Austin, TX.

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

Sumner, P., Tsai, P-C., Yu, K., & Nachev, P. (2006). Attentional modulation of sensorimotor processes in the absence of perceptual awareness. *PNAS, 103,* 10520–10525

Tiffany, S. T. (1990). A cognitive model of drug urges and drug-use behavior: Role of automatic and nonautomatic processes. *Psychological Review, 97,* 147–168.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience, 5,* 42.

Triesch J, Ballard D. H., Hayhoe, M. M., Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision, 3,* 86–94.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science, 19,* 49–54.

Vygotsky, L. S. (1962). *Thought and language.* Cambridge, MA: MIT Press.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101,* 34–52.

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

Weiskrantz, L. (1986). *Blindsight: A case study and its implications.* Oxford: Oxford University Press.

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

Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin, 19,* 331–339.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60,* 181–192.

Winkielman, P. & Berridge, K. C. (2004). Unconscious emotion. *Current Directions in Psychological Science, 13,* 120–123.

Winkielman, P., Berridge, K., & Sher, S. (2011). Emotion, consciousness, and social behavior. In J. Decety & J. T. Cacioppo (Eds.), *Handbook of social neuroscience* (pp 195–211). New York: Oxford University Press.

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, 1,* 121–135.

Winkielman, P., Knutson, B., Paulus, M. P., & Trujillo, J. T. (2007). Affective influence on decisions: Moving towards the core mechanisms. *Review of General Psychology, 11,* 179–192.

Winkielman, P. & Schooler, J. (2011). Splitting consciousness: Unconscious, conscious, and metaconscious processes in social cognition. *European Review of Social Psychology, 22,* 1–35.

Winkielman, P., & Schwarz, N. (2001). How pleasant was your childhood? Beliefs about memory shape inferences from experienced difficulty of recall. *Psychological Science, 12,* 176–179.

Winkielman, P., Zajonc, R. B., & Schwarz, N. (1997). Subliminal affective priming resists attributional interventions. *Cognition and Emotion, 11,* 433–465.

Wood, W., Quinn, J., & Kashy, D. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology, 83,* 1281–1297.

Zajonc, R. B. (1994). Evidence for nonconscious emotions. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 293–297). New York: Oxford University Press.

Zeeman, A. (2002). *Consciousness: A user's guide.* New Haven, CT: Yale University Press.

5

Goals, Motivated Social Cognition, and Behavior

Henk Aarts

Human behavior is sensitive to learning, is influenced by past experiences, and tends to be organized and structured in the service of future action. Research in the tradition of behaviorism has shown that behavior follows from rigid responses to stimuli that are reinforced by rewards. According to this work, the environment organizes and determines human behavior. However, acting on fixed stimulus–response rules, such as saying "yeah" when someone is knocking on the door, is not the whole story. Our behavior is more flexible to deal with the varying circumstances we encounter in daily social life. Such flexibility relies on our capacity to mentally represent what we want and do, and to control behavior in line with the representations. Accordingly, a substantial component of human behavior is directed at goals that motivate and control the behavioral system in a dynamic environment.

This chapter provides an analysis of the role of goals in social behavior. Fortunately, there are a few recent excellent volumes on goals (Aarts & Elliot, 2012; Moskowitz & Grant, 2009; Shah & Gardner, 2008), so there is no need to review the literature on this topic in all its details here again. Instead, this chapter aims to offer an examination of the general principles that govern goal pursuit. Goal-directed behavior has been mainly conceptualized and studied as the product of our conscious mind. That is, goal setting and control is believed to rely on consciousness, because people are often consciously aware of the goals they pursue. However, the discovery that decisions start in the unconscious (Libet et al., 1983; Soon et al., 2008), and the importance of unconscious processes in social cognition and behavior (Bargh, 2007) has questioned the causal status of consciousness in goal-directed behavior. Here we examine the origin and control of conscious goal-directed behavior and the possibility that goals operate outside awareness. However, we first briefly discuss a few research programs suggesting that the role of goals in social cognition is not taken for granted, and that we should be careful in considering the goal concept as an explanatory tool for social behavior.

THE DEBATE OF GOALS IN SOCIAL COGNITION

The idea that our behavior is goal-directed appears to be well-accepted by most contemporary researchers, but this was (and, in some instances, still is) not always the case in the study of social cognition. This dispute about the role of goals is rooted primarily in the cognitive revolution, in which there was no room for motivation, and cognition was seen as the more parsimonious account for behavior.

In the study on reasoning, the notion that goals affect attitudes (Festinger, 1957), attributions (Heider, 1958), and beliefs (Kruglanski, 1996) has been put forth by some psychologists and challenged by others. For instance, goals have been posited to lead people to make self-serving attributions for success and failure, even though such attributions do not reflect the actual cause