In S. Vazire, & T. D. Wilson (Eds.). (2012). Handbook of self-knowledge. New York: Guilford Press.

CHAPTER 3

Self-Insight from a Dual-Process Perspective

BERTRAM GAWRONSKI
GALEN V. BODENHAUSEN

Many cultures consider self-insight an important virtue for a fulfilled and authentic life (Wilson & Dunn, 2004). Challenging the feasibility of this enterprise, however, research in psychology has uncovered a plethora of obstacles that can undermine accurate self-knowledge, many of which are reviewed in this handbook. This chapter provides a conceptual analysis of these obstacles from a dual-process perspective. Over the past decades, dual-process approaches have provided theoretical guidance for virtually all areas of psychology, offering conceptual integrations of existing evidence and novel predictions of previously undetected phenomena (for a review, see Gawronski & Creighton, in press). Yet, despite their popularity, there have been few attempts to analyze the mental underpinnings of self-knowledge from a dual-process perspective (for a notable exception, see Epstein, 1994). The main goal of this chapter is to fill this gap.

Toward this end, we refrain from providing an exhaustive review of specific dual-process theories. Instead, we use the common foundation of dual-process theories—the distinction between automatic and controlled processes—to illustrate the range and the limits of people's insights into the causes, the contents, and the effects of their mental associations. Our focus on mental associations is based on recent definitions of major social-psychological constructs as associations between two concepts in memory (e.g., Greenwald et al., 2002). For example, the construct of *attitude* has been defined as the mental association between an object and its evaluation (Fazio, 1995). Correspondingly, *self-esteem* can be defined as the association between the self and its evaluation, just as *prejudice* can be defined as the association between a social group and its evaluation (Greenwald et al., 2002). With regard to nonevaluative constructs, *self-concept* can be defined as the association between the self and its attributes (e.g., self–extraverted), just as *stereotypes* can be conceptualized

as associations between a social category and stereotypical attributes (e.g., women–warm). At a more complex level, goals can be conceptualized as a combination of two associations, namely, an association between means and an end-state (Kruglanski et al., 2002), and an additional association between the end-state and its evaluation (Custers & Aarts, 2005). In the following sections, we first discuss the notion of automaticity and control as the common foundation of dual-process theories. Expanding on this discussion, we then outline various implications of this distinction for people's insights into the causes, the contents, and the effects of their mental associations, including attitudes, prejudices, stereotypes, self-esteem, self-concepts, and goals.[1]

Automaticity and Control

Dual-process theories have their roots in the assumption that mental processes can be characterized on the basis of whether they operate in an automatic or controlled fashion. The defining features of automatic processes are that (1) they do not involve conscious awareness; (2) they do not require a person's intention to be started; (3) they operate even under limited cognitive resources; and (4) they cannot be stopped or altered voluntarily. Conversely, controlled processes (1) operate under conscious awareness; (2) require a person's intention to be started; (3) fail to operate when cognitive resources are limited; and (4) can be stopped or altered voluntarily (Bargh, 1994; Moors & De Houwer, 2006). As we outline below, all of these features play a significant role for specific aspects of self-insight. Yet the most important characteristic for the current analysis is the first one: conscious awareness. With regard to people's insight into their mental associations, the object of awareness can be further divided into three different components: (1) awareness of the causes of one's mental associations, (2) awareness of the contents of one's mental associations, and (3) awareness of the effects of one's mental associations (Gawronski, Hofmann, & Wilbur, 2006). Specifically, individuals may or may not know why they have certain mental associations; they may or may not know that they have certain kinds of mental associations; and they may or may not know how their mental associations influence their behavior (see Figure 3.1).

Insight into the Causes of One's Mental Associations

Inaccurate beliefs about the causes of one's mental associations can have significant consequences if judgments and decisions are based on these beliefs. For example, Wilson argued that people are often unaware of the true causes of their preferences (e.g., Wilson, Dunn, Kraft, & Lisle, 1989). Thus, when analyzing reasons for their preferences, people tend to rely on reasons that are accessible and easy to communicate. Yet these reasons do not always reflect the true causes of their preferences, leading them to shift their preferences toward those that are in line with the generated reasons. To the extent that these momentarily constructed preferences are taken as a basis for choices and decisions, the quality of these decisions can be suboptimal in terms of subjective (e.g., Wilson et al., 1993) and objective (e.g., Wilson & Schooler, 1991) standards. In a study by Wilson and colleagues (1993), for example, participants were

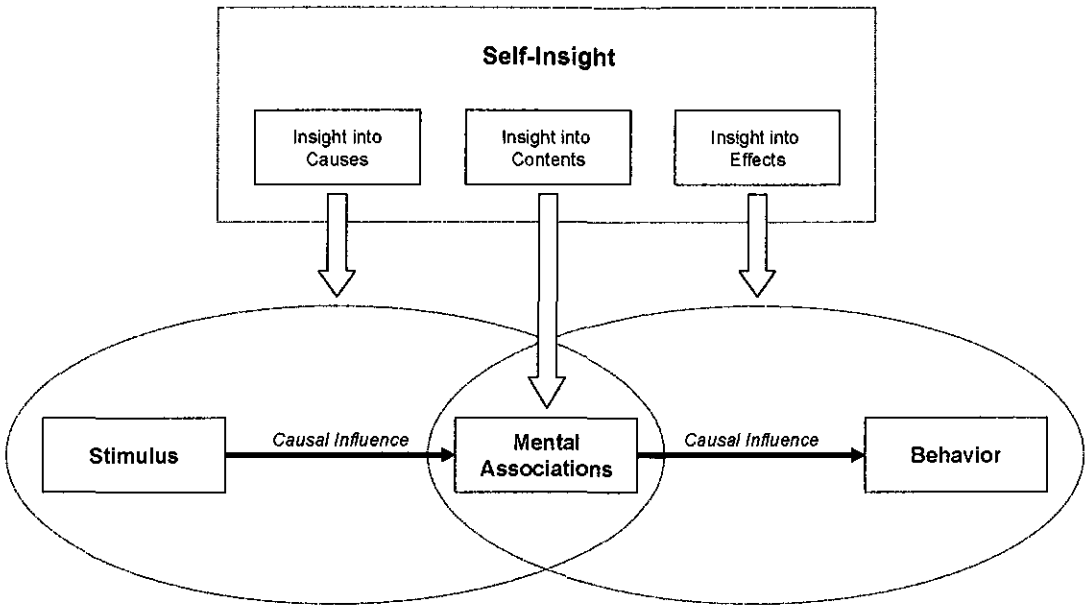

FIGURE 3.1. Different components of self-insight pertaining to the causes, contents, and effects of one's mental associations.

asked to choose between two kinds of posters. Half of the participants were additionally asked to think about reasons for their preference; participants in a control group were not asked to think about reasons. Those who were asked to think about reasons not only showed different preferences compared with those who were not asked to think about reasons but they were also less satisfied with their choice when they were contacted 3 weeks after the study.

From a theoretical perspective, accurate knowledge of the causes of one's mental associations requires awareness of three distinct components: (1) the causally relevant stimuli; (2) the mental associations themselves; and (3) the causal relation between the two (see Figure 3.1). Thus, insight into the causes of one's mental associations will be limited if people lack awareness of any one of these components.

Unawareness of Causally Influential Stimuli

In many cases, our mental associations are the product of conscious learning processes, for example, when we read a newspaper article about unhealthy ingredients of certain food products (*propositional learning*). Yet, in other cases, new associations are formed unintentionally outside of conscious awareness (*associative learning*). For example, research on evaluative conditioning (EC) has shown that repeated pairings of a formerly neutral conditioned stimulus (CS) with a positive or negative unconditioned stimulus (US) lead to changes in the evaluation of the CS in line with the valence of the US (for a meta-analysis, see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). A common interpretation of these effects is that the CS–US pairings create a mental association between the CS and the US in memory (e.g., Gawronski & Bodenhausen, 2006, 2011). As a result, encountering the CS at future occasions activates the representation of the US, thereby producing an evaluative response

to the CS that matches the one to the US (e.g., Baeyens, Eelen, Van den Bergh, & Crombez, 1992; Walther, Gawronski, Blank, & Langer, 2009). Importantly, some studies have shown EC effects even when the CS–US pairings involved subliminal presentations of the CS (e.g., Dijksterhuis, 2004; Gawronski & LeBel, 2008; Knight, Nguyen, & Bandettini, 2003) or the US (e.g., De Houwer, Hendrickx, & Baeyens, 1997; Krosnick, Betz, Jussim, & Lynn, 1992; Rydell, McConnell, Mackie, & Strain, 2006). These results suggest that mental associations can be formed without awareness of the relevant stimuli, implying that people can have mental associations without knowing their causal origin.

Whereas research on EC is concerned with the *formation* of mental associations, priming effects refer to the *activation* of existing associations. Similar to EC research using subliminal presentations of the CS or the US, a large body of research shows that subliminal presentations of stimuli can activate mental associations in memory (e.g., Wittenbrink, Judd, & Park, 1997). One of the most prominent examples in this regard is Devine's (1989) research on automatic and controlled processes in prejudice and stereotyping. In her study, participants were subliminally primed either with evaluatively neutral words or with words related to the stereotype of African Americans, and then read a brief story about a target who behaved ambiguously (cf. Bargh & Pietromonaco, 1982). Results showed that the behavior of the target was interpreted more negatively when participants were primed with the stereotype of African Americans than when they were primed with neutral words. Applied to the present question, these results suggest that participants had particular thoughts while reading the description of the behavior without being aware of the stimuli that were responsible for these thoughts. As a result, they misattributed their thoughts to the behavior of the target, thereby leading to more negative interpretations of the ambiguous behavior when they were primed with the stereotype of African Americans.

Similar effects have been found for the unconscious activation of goals. For example, Ferguson (2008) subliminally primed participants with words related to the goal of being thin or neutral control words that were unrelated to the goal of being thin. Afterwards, all participants completed a measure of automatic evaluations of diet-related words. Participants in the control condition showed relatively neutral responses to the diet-related words regardless of their dieting skills (which were assessed prior to the study). In contrast, participants in the goal-priming condition showed positive responses to the diet-related words when their dieting skills were high, but negative responses when their dieting skills were low. As with Devine's (1989) research, these results suggest that goals can be activated by goal-related stimuli without conscious awareness of these stimuli.

Unawareness of Causal Link

In many situations, people are consciously aware of the momentarily present stimuli, but they may not be aware of the causal impact of these stimuli on their mental associations. For example, research on EC effects often uses procedures with supraliminal presentations, in which participants are consciously aware of both the CS and the US. Yet when subsequently asked to report which CS was paired with which US, participants are sometimes unable to identify correctly the specific CS–US contingencies.

Even though EC effects tend to be larger when participants have conscious knowledge of the relevant CS–US pairings (for a meta-analysis, see Hofmann et al., 2010), a considerable body of research has found significant EC effects in the absence of contingency awareness (e.g., Baeyens, Eelen, & Van den Bergh, 1990; Olson & Fazio, 2001; Sweldens, Van Osselaer, & Janiszewski, 2010; Walther & Nagengast, 2006). Given that CS–US contingencies represent an important component in the causal link that is responsible for the newly formed associations, these results suggests that people can have mental associations without being aware of the causal link between their mental associations and consciously encoded stimuli.

Similar considerations apply to the activation of existing associations in priming effects. Instead of presenting the causally effective prime stimuli subliminally, many studies use procedures in which participants are consciously aware of the relevant primes. Yet they often remain unaware of the causal impact of the prime stimuli on their thoughts and behavior. For example, to activate the stereotype of older adults, Bargh, Chen, and Burrows (1996) used a scrambled-sentence task that was described as a test of language proficiency. For half of the participants, the task included words related to the older adults stereotype (e.g., *retired*). For the remaining half, the task included neutral words unrelated to the older adults stereotype (e.g., *thirsty*). The well-known finding is that participants walked slower down the hall at the end of the study when they were primed with stereotype-related words than when they were primed with stereotype-unrelated words. Note that in this task participants were fully aware of the words that were supposed to activate the stereotype of older adults. However, when participants were asked whether they thought that the words in the scrambled-sentence task might have affected them in any way, none of them believed that the words had any impact on their thoughts and behavior.

Similar procedures have been used in research on goal pursuit. For example, in Bargh, Gollwitzer, Lee-Chai, Barndollar, and Trötschel's (2001) seminal demonstration of unconscious goal priming, participants completed a word-search puzzle that included either words related to a high-performance goal (e.g., *succeed*) or neutral words that were unrelated to performance (e.g., *carpet*). Results showed that participants primed with performance-related words showed enhanced performance on a subsequent task compared with participants primed with neutral words. As with Bargh and colleagues' (1996) scrambled-sentence task, participants in this study were consciously aware of the words that were presented in the word-search puzzle. However, none of them thought that the words influenced their motivation or performance in the subsequent task. In other words, participants were consciously aware of the stimuli that influenced their momentary goals, but they were unaware of the causal impact these words had on their goals.

Unawareness of Mental Associations

A common interpretation in research on priming effects is that participants are not aware of the primed associations when these associations have been activated outside of conscious awareness. For example, in the literature on unconscious goal priming, it is often assumed that participants are not aware of their momentary goals if they (1) are not aware of the stimuli that activated these goals (e.g., Ferguson, 2008) or (2) are not aware of the causal impact of consciously perceived stimuli on their

momentary goals (e.g., Bargh et al., 2001). This may well be true, but it is important to note that awareness of one's goals per se is conceptually distinct from awareness of the stimuli that influence one's goals, or awareness of the impact of these stimuli on one's goals. To establish empirically the unconsciousness of a goal, one would have to ask participants about their momentary goals rather than about the stimuli used to activate the goal or the perceived causal impact of these stimuli. Unfortunately, such measures are rarely included in research on unconscious goal pursuit, which makes inferences about the unconsciousness of the relevant goals premature (for a notable exception, see Bar-Anan, Wilson, & Hassin, 2010). This problem also applies to studies in which participants are asked about their goal-related behavior. For example, in a study by Bargh and colleagues (2001), participants were primed with cooperation-related words or neutral control words. Results showed that participants in the goal-priming condition showed more cooperation in a subsequent resource dilemma task compared with participants in the control condition. Yet participants' self-reports on how much they cooperated during the dilemma task were unrelated to their actual cooperation. In a strict sense, these findings show that participants' self-perceptions of their own behavior often deviate from their actual behavior. However, they remain silent about whether participants were aware or unaware of the primed goal. Again, to establish the unconsciousness of the goal per se, one would have to ask participants about the goal itself (e.g., "How important is it for you to cooperate during the task?"), not about their retrospective self-perception of their behavior. After all, there is an important difference between not knowing that one has a particular goal and having inaccurate perceptions of one's behavior.

To be fair, it is important to note that these issues seem less controversial in research on unconscious aspects of associative learning, in which the (un)consciousness of the resulting association is rarely conflated with (un)conscious aspects of the learning process that is responsible for this association. For example, most studies on EC effects assess evaluative representations by means of self-report measures that simply ask participants how much they like or dislike the CS (cf. Hofmann et al., 2010). In these studies, the relevant evaluative association is assumed to be consciously accessible even when certain aspects of the learning process that led to this association remain outside of conscious awareness (e.g., subliminal presentation of the stimuli, lack of contingency awareness). The important message is that unawareness of the relevant stimuli or the causal effect of these stimuli does not say anything about people's awareness of the relevant mental associations. The latter question is discussed in more detail in the following section.

Insight into the Contents of One's Mental Associations

One of the most central questions in the context of self-insight is whether people can have certain kinds of mental associations without being aware that they have these associations. In other words, is it possible that people have unconscious attitudes, unconscious prejudices, unconscious self-esteem, unconscious stereotypes, unconscious self-concepts, or unconscious goals (cf. Blanton & Jaccard, 2008; Buhrmester, Blanton, & Swann, 2011)? As illustrated in Figure 3.1, insight into the contents of one's mental associations—including the relevant concepts and the associative links

between them—not only constitutes a key component in its own right, but also represents a central part in people's insight into the causes and the effects of their mental associations. If one is unaware of the existence of a particular mental association, it is logically impossible to know where this association is coming from and what effects it has on one's behavior.

In the dual-process literature, there are two classes of theories that make different assumptions about conscious and unconscious mental contents. Whereas some theories assume that conscious and unconscious contents are based on distinct memory structures that operate independently (e.g., Banaji, 2001; Rydell & McConnell, 2006), other theories assume that conscious and unconscious contents are based on the same memory structures, with unconscious contents being characterized by activation levels that do not pass the threshold of conscious awareness (e.g., Gawronski & Bodenhausen, 2006; Strack & Deutsch, 2004). Even though the debate between dual-representation and single-representation theories seems difficult to resolve on empirical grounds (Greenwald & Nosek, 2009), it is often assumed that unconscious associations can be captured with implicit measurement procedures, such as sequential priming tasks (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Payne, Cheng, Govorun, & Stewart, 2005; Wittenbrink et al., 1997) and the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). A common assumption in this research is that explicit self-report measures tap mental contents that are consciously accessible, whereas implicit measures provide access to associations that are introspectively inaccessible.

Of course, implicit measures do not presuppose introspective access for the assessment of mental contents. However, whether the mental contents captured by implicit measures are indeed unconscious is an empirical question that has to be tested as such (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). A frequently cited finding in support of the unconsciousness claim is that implicit and explicit measures often show rather low correspondence (Banaji, 2001). Yet further scrutiny of the available evidence suggests that the mental associations assessed by implicit measures are indeed consciously accessible, and that various other factors account for the frequently obtained dissociations between implicit and explicit measures (Fazio, 2007; Gawronski et al., 2006). In the domain of attitudes, for example, several studies have shown that implicit and explicit measures show rather high correspondence if participants focus on their gut feelings when reporting an evaluation (e.g., Banse, Seise, & Zerbes, 2001; Gawronski & LeBel, 2008; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Ranganath, Smith, & Nosek, 2008; Scarabis, Florack, & Gosejohann, 2006; Smith & Nosek, 2011). These results are difficult to reconcile with the claim that implicit measures tap mental associations that are generally inaccessible to introspection. Yet they are in line with dual-process theories that assume implicit measures tap spontaneous gut responses resulting from mental associations that are activated unintentionally upon the encounter of an attitude object (for a review, see Hofmann, Gschwendner, Nosek, & Schmitt, 2005). These gut responses, in turn, may serve as a basis for explicit judgments, unless the individual is motivated and able to deliberate on individual attributes of the object (Fazio, 2007) or the gut response is inconsistent with other momentarily considered information (Gawronski & Bodenhausen, 2006, 2011).

Even though the available evidence is consistent with dual-process accounts that

emphasize other features of automaticity rather than the notion of awareness (e.g., intentionality, controllability; for a discussion, see Payne & Gawronski, 2010), it is worth noting that correspondence between implicit and explicit measures may be the result of at least three processes that have different implications for people's insight into the contents of their mental associations (see Hofmann & Wilson, 2010). First, it is possible that people have direct introspective access to their mental associations. In this case, there would be no a priori limits to people's ability to know their mental associations, and any lack of knowledge may be regarded as the product of insufficient motivation to introspect on one's associations. Second, there may be cases in which people have no direct access to their mental associations, but they may have indirect access through the subjective experiences that result from these associations. In such cases, people would have indirect access to contents that elicit subjective experiences (e.g., evaluative associations that elicit affective gut feelings), but there could be limits to the ability to know contents that do not elicit subjective experiences (e.g., purely semantic associations that do not elicit any feelings). Third, there may be cases in which people have no direct access to their associations, but they may have indirect access through self-perceptions of the behaviors that result from these associations (Bem, 1972). As with the first case, there would be no a priori limits to people's ability to know their mental associations, although knowledge of mental contents may be limited either when people fail to pay attention to their behavior (Carver & Scheier, 1981; Wicklund, 1975) or when their subjective interpretation of their behavior is biased (Jones & Nisbett, 1972). Yet what is essentially required is that people have accurate theories about what kinds of behaviors are caused by what kinds of associations. We return to the question of naive theories in the section on insights into the effects of mental associations.

Another important question in the context of attitudes concerns the correspondence between people's beliefs about what they like or dislike and their actual evaluative responses. Drawing on the distinction between implicit and explicit measures, one could argue that implicit measures capture people's actual responses to an attitude object, whereas explicit measures tap people's beliefs about what they like or dislike. From this perspective, correspondence between the two would indicate that people's beliefs about their attitudes are accurate, whereas dissociations between the two kinds of measures would indicate inaccurate self-beliefs. Even though this conceptualization resonates with interpretations of implicit measures as a window to people's "true self," we propose that self-beliefs and actual evaluative responses are rooted in two distinct types of mental associations, both of which could be assessed with implicit measures. Whereas self-beliefs could be considered as a particular aspect of one's self-concept, involving mental links between the self and the attribute of liking or disliking a particular object (e.g., an association between the concepts *self* and *liking baseball*), actual evaluative responses are presumably rooted in mental links between the object and positive or negative characteristics of that object (e.g., an association between the concepts *baseball* and *fun*). From this perspective, implicit measures may be designed to tap either one or the other type of association, and either type of association may influence self-reported liking under particular conditions.[2] Yet the two kinds of associations may have distinct antecedents, in that object-related associations stem from direct experiences with or communicated information about the object, whereas self-related associations are the product of self-

perceptions of one's evaluative responses (see Gawronski et al., 2008). Even though the two kinds of associations may often show a high level of correspondence, there may be cases in which they dissociate, for example, when evaluative responses rooted in object-related associations are attributed to situational factors instead of internal attributes (e.g., Hofmann, Gschwendner, & Schmitt, 2009). In such cases, people's self-referential beliefs about whether they like or dislike a given attitude object may deviate from their actual evaluative response to that object. An illustrative example is the notion of aversive racism, in that aversive racists are assumed to experience negative feelings in response to racial outgroups, while being convinced that they have positive attitudes toward these groups (Dovidio & Gaertner, 2004).

Insight into the Effects of One's Mental Associations

A third question in the context of self-insight concerns the extent to which people are aware of the behavioral effects of their mental associations. In dual-process frameworks, awareness of the effects of mental associations is typically studied by means of behavioral control. A common assumption in dual-process theories is that people control for biasing influences on their behavior when three conditions are met. First, they have to be *motivated* to control their behavior for biasing influences. Second, they have to be *able* to engage in behavioral control. Third, they have to be *aware* of the biasing influence (e.g., Hall & Payne, 2010; Strack & Hannover, 1996; Wegener & Petty, 1997; Wilson & Brekke, 1994). To the extent that both the first and the second conditions are met, people are assumed to be unaware of the biasing influence if they fail to control their behavior despite their motivation and ability to do so. In such cases, lack of awareness may again involve one of three distinct components: (1) the mental associations themselves; (2) the relevant behavior; and (3) the causal relation between the two (see Figure 3.1). Thus, insight into the effects of mental associations can be limited if people are unaware of any of these components.

Unawareness of Mental Associations

To the extent that people can be unaware of momentarily activated associations and these associations nevertheless influence behavior, people may misattribute their behavior to other factors, which could lead to inaccurate self-beliefs. In line with this contention, Bar-Anan and colleagues (2010) showed that participants who were primed with a particular goal (e.g., affiliation) did not differ from control participants on a self-report measure of goal strength. Yet participants in the goal-priming condition were more likely to choose activities that were conducive to that goal. Importantly, when asked to explain their choices, participants misattributed their preferences to plausible reasons that were accessible and easy to communicate (e.g., stable dispositions) rather than their momentary goals, and the inaccurate self-beliefs resulting from these misattributions influenced subsequent choices in a manner that was consistent with their newly formed self-beliefs. These results are consistent with the proposition that awareness of the content of one's mental associations is an essential precondition for understanding their effects on one's behavior.

Unawareness of Behavior

A common implication of many dual-process theories is that implicit measures should be better predictors of spontaneous behavior, whereas explicit measures should show superior performance in the prediction of deliberate behavior (e.g., Fazio, 2007; Strack & Deutsch, 2004; Wilson, Lindsey, & Schooler, 2000). This prediction has been confirmed in a large number of studies that classified different kinds of behaviors on theoretical grounds as either spontaneous or deliberate (for reviews, see Friese, Hofmann, & Schmitt, 2008; Perugini, Richetin, & Zogmaister, 2010). The typical interpretation of these findings is that the spontaneous behaviors in these studies are difficult to control and therefore influenced by the automatically activated associations assessed by implicit measures. In other words, participants are assumed to be motivated to control their behavior, but they are unable to do so. Yet an alternative interpretation is that people are able to control at least some of the behaviors classified as spontaneous, but that they are unaware of how their mental associations affect these behaviors (see Gawronski et al., 2006). For example, speaking time (McConnell & Leibold, 2001) or spatial distance (Fazio et al., 1995) in social interactions with a black person seem relatively easy to control. However, people may be unaware of how these behaviors are affected by their evaluative associations regarding black people. As a result, they may not attempt to control these behaviors even if they have the motivation and the ability to do so (Strack & Hannover, 1996).

Unawareness of Causal Link

Even if people are consciously aware of their mental associations and the behaviors resulting from these associations, they may sometimes be unaware of the causal link between the two. One example in this regard is the impact of mental associations on the interpretation of ambiguous behavior. In a study by Gawronski, Geschke, and Banse (2003), for example, German participants were asked to form an impression of either a German or a Turkish individual on the basis of evaluatively ambiguous behavior. Consistent with previous research (e.g., Duncan, 1976; Sagar & Schofield, 1980), participants evaluated the behavior more negatively when the target was Turkish than when the target was German. However, this effect was moderated by participants' evaluative associations regarding Turks and Germans, such that the target's category membership influenced the interpretation of ambiguous behavior only for participants with a strong associative preference for Germans over Turks (see also Hugenberg & Bodenhausen, 2003). Importantly, the influence of evaluative associations was unaffected by participants' motivation to control prejudiced reactions. Instead, motivation to control prejudice moderated only the impact of evaluative associations on self-reported evaluations of Turkish people in general; that is, evaluative associations and self-reported evaluations showed a positive correlation only for participants low, but not for those high, in motivation to control prejudice (see Dunton & Fazio, 1997). Self-reported evaluations had no impact on the interpretation of ambiguous behavior. Thus, given that participants were generally able to control the influence of their evaluative associations on the interpretation of ambiguous behavior (i.e., participants were not under time pressure or otherwise cognitively depleted), these results suggest that participants were unaware of the impact of their evaluative

associations on the interpretation of ambiguous behavior. In other words, evaluative associations influenced behavioral interpretations irrespective of participants' motivation and their ability to control for this influence (Strack & Hannover, 1996).

Another important factor in this context is the accuracy of people's naive theories about the causal impact of their mental associations on behavior (Strack, 1992; Wegener & Petty, 1997; Wilson & Brekke, 1994). Even if people are consciously aware of their thoughts and their behavior, causal relations between the two are not directly observable but have to be inferred from observed covariations. According to Wegner and Wheatley (1999), such inferences about mental causation are guided by three general principles. First, the thought should precede the behavior with a sufficiently short interval (*priority*). Second, the content of the thought should be compatible with the behavior (*consistency*). Third, the thought should be the only apparent cause of the behavior in that situation (*exclusivity*). Even though the presence of these conditions seems rather easy to establish in many situations, there can be conditions under which their assessment is hindered, thereby leading to inaccurate inferences of mental causation. Such distortions can go either way, in that they may lead to overestimations (e.g., Wegner, Sparrow, & Winerman, 2004) or underestimations (e.g., Wegner, Fuller, & Sparrow, 2003) of mental causation. For example, in a study by Wegner and colleagues (2003) participants were asked to give freely chosen, random answers to a set of yes–no questions. Even though participants were convinced that their responses were entirely random, they answered more questions correctly than would have been expected by chance, and this effect was more pronounced for easy compared with difficult questions. Interestingly, these effects generalized to situations when participants were asked to answer yes–no questions by sensing the inclinations of a confederate who did not even know the questions. Thus, one could argue that participants were aware of their thoughts of the correct answers, as well as their behavioral response to the question. Yet it seems that they were unaware of the causal link between the two.

To the extent that people draw inaccurate inferences about the causal links between their thoughts and their behaviors, the naive theories based on these inferences have the potential to further undermine people's attempts to control for biasing influences of their mental associations. Thus, extending the list of prerequisites for effective and contextually appropriate behavioral control, dual-process theorists argued that people need to have not only the necessary motivation, ability, and awareness but also an accurate theory of *how* their behavior is biased (Strack, 1992; Wegener & Petty, 1997; Wilson & Brekke, 1994). If their naive theories of mental causation are inaccurate, people may adjust their behavior in line with the implications of these theories even when their behavior is unbiased. More seriously, people may sometimes adjust their behavior in the wrong direction, thereby promoting rather than reducing bias (e.g., Petty & Wegener, 1993).

Relations between Different Components of Self-Insight

A final important question concerns the relation between the proposed components of self-insight. This issue has been a common source of confusion, in that the different components are often conflated in theoretical interpretations of empirical findings.

For example, in research on goal pursuit, effects of unconscious goal priming are often interpreted as evidence that people can pursue goals of which they are unaware (e.g., Bargh et al., 2001; Ferguson, 2008). However, as we have argued in this chapter, the fact that people can be unaware of the cause of a momentary goal does not mean that they are unaware of the goal itself. Similar confusion has been caused by different interpretations of Greenwald and Banaji's (1995) definition of *implicit attitudes* as "introspectively unidentified (or inaccurately unidentified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects" (p. 8). Whereas some researchers have interpreted this definition as implying unawareness of the causes of an attitude, others have referred to this definition in claiming unawareness of the attitude itself. Again, as we have outlined in this chapter, the two aspects of self-insight are conceptually distinct, and the former type of unawareness does not necessarily imply the latter.

In the preceding sections, we have argued that awareness of the contents of one's mental associations represents an important precondition for people's insight into both the causes and the effects of their mental associations (see Figure 3.1). As for the causes of one's mental associations, accurate knowledge presupposes (1) awareness of the relevant stimuli, (2) awareness of the mental associations themselves, and (3) awareness of the causal link between the two. Correspondingly, accurate knowledge about the effects of one's mental associations presupposes (1) awareness of the mental associations themselves, (2) awareness of the relevant behavior, and (3) awareness of the causal link between the two. Thus, accurate knowledge of the contents of one's mental associations represents a *necessary* precondition for insight into the causes and the effects of one's mental associations. Yet knowing the contents of one's mental associations is *insufficient* for accurate knowledge about the causes and the effects of these associations. For example, people may well be aware of their personal preferences, but they may be unaware of where these preferences come from or how these preferences influence their behavior (Nisbett & Wilson, 1997). From this perspective, comprehensive insight into one's mind includes all three components discussed in this chapter: the causes, the contents, and the effects of one's mental associations.

Conclusion

Our main goal in this chapter was to provide a conceptual analysis of self-insight from the perspective of dual-process theories. Even though dual-process theories differ in many regards (Gawronski & Creighton, in press), their shared distinction between automatic and controlled processes offers a valuable framework for understanding the mental underpinnings of self-insight. Toward this end, we have distinguished between insight into the causes, the contents, and the effects of one's mental associations, all of which are required for a comprehensive understanding of the working of one's mind. The implications of our analysis are applicable to any kind of mental association, including attitudes (i.e., object-evaluation associations), prejudice (i.e., group-evaluation associations), self-esteem (i.e., self-evaluation association), stereotypes (i.e., group-attribute association), self-concepts (i.e., self-attribute associations), and goals (i.e., means–ends-evaluation associations). Thus, we hope that our analysis will be useful for all researchers interested in self-knowledge, irrespective of their content area.

ACKNOWLEDGMENTS

Preparation of this chapter has been supported by the Canada Research Chairs Program (Grant No. 202555) and the Social Sciences and Humanities Research Council of Canada (Grant No. 410-2008-2247).

NOTES

1. Note that our analysis focuses on associations between mentally represented concepts, and therefore does not include self-knowledge of emotions or mood states. It also does not capture the role of cognitive feelings, such as processing fluency. Self-knowledge of emotions and mood states is discussed in more detail by Clore and Robinson (Chapter 12, this volume); cognitive feelings are discussed by Hofree and Winkielman (Chapter 13, this volume).

2. The distinction between actual evaluative responses and the self-concept of one's attitude may also explain differences between the standard IAT (Greenwald et al., 1998) and the personalized IAT (Olson & Fazio, 2004), in that the standard IAT assesses object-evaluation associations, whereas the personalized IAT assesses associations between the self and the attribute of liking or disliking a given object (for a discussion, see Gawronski, Peters, & LeBel, 2008).

REFERENCES

Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning. *Cognition and Emotion, 4*, 3–18.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1992). The content of learning in human evaluative conditioning: Acquired valence is sensitive to US revaluation. *Learning and Motivation, 23*, 200–224.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in remembering Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48*, 145–160.

Bar-Anan, Y., Wilson, T. D., & Hassin, R. R. (2010). Inaccurate self-knowledge formation as a result of automatic behavior. *Journal of Experimental Social Psychology, 46*, 884–894.

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

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

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

Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology, 43*, 437–449.

Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology, 6*, 1–62.

Blanton, H., & Jaccard, J. (2008). Unconscious racism: A concept in pursuit of a measure. *Annual Review of Sociology, 34*, 277–297.

Buhrmester, M. D., Blanton, H., & Swann, W. B. (2011). Implicit self-esteem: Nature, measurement, and a new way forward. *Journal of Personality and Social Psychology, 100*, 365–385.

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

Custers, R., & Aarts, H. (2005). Positive affect as implicit motivator: On the nonconscious operation of behavioral goals. *Journal of Personality and Social Psychology, 89*, 129–142.

De Houwer, J., Hendrickx, H., & Baeyens, F. (1997). Evaluative learning with subliminally presented stimuli. *Consciousness and Cognition, 6*, 87–107.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*, 347–368.

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

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86*, 345–355.

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. *Advances in Experimental Social Psychology, 36*, 1–52.

Duncan, B. L. (1976). Differential perception and attribution of intergroup violence: Testing the lower limits of stereotyping of Blacks. *Journal of Personality and Social Psychology, 34*, 590–598.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 23*, 316–326.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709–724.

Fazio, R. H. (1995). Attitudes as object–evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength* (pp. 247–282). Mahwah, NJ: Erlbaum.

Fazio, R. H. (2007). Attitudes as object–evaluation associations of varying strength. *Social Cognition, 25*, 603–637.

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

Ferguson, M. J. (2008). On becoming ready to pursue a goal you don't know you have: Effects of nonconscious goals on evaluative readiness. *Journal of Personality and Social Psychology, 95*, 1268–1294.

Friese, M., Hofmann, W., & Schmitt, M. (2008). When and why do implicit measures predict behaviour?: Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology, 19*, 285–338.

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

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44*, 59–127.

Gawronski, B., & Creighton, L. A. (in press). Dual-process theories. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition.* New York: Oxford University Press.

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology, 33,* 573–589.

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

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

Gawronski, B., Peters, K. R., & LeBel, E. P. (2008). What makes mental associations personal or extra-personal?: Conceptual issues in the methodological debate about implicit attitude measures. *Social and Personality Psychology Compass, 2,* 1002–1023.

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

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

Greenwald, A. G., & Nosek, B. A. (2009). Attitudinal dissociation: What does it mean? In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 65–82). Hillsdale, NJ: Erlbaum.

Hall, D. L., & Payne, B. K. (2010). Unconscious influences of attitudes and challenges to self-control. In R. R. Hassin, K. N. Ochsner, & Y. Trope (Eds.), *Self-control in society, mind, and brain* (pp. 221–242). New York: Oxford University Press.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136,* 390–421.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369–1385.

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit–explicit consistency? *European Review of Social Psychology, 16,* 335–390.

Hofmann, W., Gschwendner, T., & Schmitt, M. (2009). The road to the unconscious self not taken: Discrepancies between self- and observer-inferences about implicit dispositions from nonverbal behavioral cues. *European Journal of Personality, 23,* 343–366.

Hofmann, W., & Wilson, T. D. (2010). Consciousness, introspection, and the adaptive unconscious. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 197–215). New York: Guilford Press.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14,* 640–643.

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.

Knight, D. C., Nguyen, H. T., & Bandettini, P. A. (2003). Expression of conditional fear with and without awareness. *Proceedings of the National Academy of Sciences USA, 100,* 15280–15283.

Krosnick, J. A., Betz, A. L., Jussim, L. J., & Lynn, A. R. (1992). Subliminal conditioning of attitudes. *Personality and Social Psychology Bulletin, 18,* 152–162.

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W. Y., & Sleeth-Keppler, D. (2002). A theory of goal-systems. *Advances in Experimental Social Psychology, 34*, 331–378.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*, 435–442.

Moors, A., & De Houwer, J. (2006). Automaticity: A conceptual and theoretical analysis. *Psychological Bulletin, 132*, 297–326.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413–147.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653–667.

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

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1–15). New York: Guilford Press.

Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 255–277). New York: Guilford Press.

Petty, R. E., & Wegener, D. T. (1993). Flexible correction processes in social judgment: Correcting for context-induced contrast. *Journal of Experimental Social Psychology, 29*, 137–165.

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology, 44*, 386–396.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91*, 995–1008.

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science, 17*, 954–958.

Sagar, H. A., & Schofield, J. W. (1980). Racial and behavioral cues in black and white children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology, 39*, 590–598.

Scarabis, M., Florack, A., & Gosejohann, S. (2006). When consumers follow their feelings: The impact of affective or cognitive focus on the basis of consumer choice. *Psychology and Marketing, 23*, 1015–1034.

Smith, C. T., & Nosek, B. A. (2011). Affective focus increases the concordance between implicit and explicit attitudes. *Social Psychology, 42*, 300–313.

Strack, F. (1992). The different routes to social judgments: Experiential versus informational strategies. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 249–275). Hillsdale, NJ: Erlbaum.

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

Strack, F., & Hannover, B. (1996). Awareness of influence as a precondition for implementing

correctional goals. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 579–596). New York: Guilford Press.

Sweldens, S., Van Osselaer, S., & Janiszewski, C. (2010). Evaluative conditioning procedures and the resilience of conditioned brand attitudes. *Journal of Consumer Research, 37*, 473–489.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the backdoor: The US-revaluation effect. *Cognition and Emotion, 23*, 889–917.

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four picture recognition test. *Journal of Experimental Psychology: Animal Behavior Processes, 32*, 454–459.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. *Advances in Experimental Social Psychology, 29*, 141–208.

Wegner, D. M., Fuller, V. A., & Sparrow, B. (2003). Clever hands: Uncontrolled intelligence in facilitated communication. *Journal of Personality and Social Psychology, 85*, 5–19.

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology, 86*, 838–848.

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist, 54*, 480–492.

Wicklund, R. A. (1975). Objective self-awareness. *Advances in Experimental Social Psychology, 8*, 233–275.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116*, 117–142.

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

Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. *Advances in Experimental Social Psychology, 22*, 287–343.

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

Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspection can reduce post-choice satisfaction. *Personality and Social Psychology, 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.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationships with questionnaire measures. *Journal of Personality and Social Psychology, 72*, 262–274.