

# Implicit Measures of Attitudes

Edited by  
BERND WITTENBRINK  
NORBERT SCHWARZ

2007



THE GUILFORD PRESS  
New York London

## 11 What Do We Know about Implicit Attitude Measures and What Do We Have to Learn?

Bertram Gawronski  
Galen V. Bodenhausen

I cannot totally grasp all that I am. . . . For that darkness is lamentable in which the possibilities in me are hidden from myself: so that my mind, questioning itself upon its own powers, feels that it cannot rightly trust its own report.

—ST. AUGUSTINE, *Confessions*

In his influential *Confessions*, St. Augustine lamented the evasiveness of full self-understanding, questioning the reliability of our intuitions about our own minds. Although introspection and self-report formed the foundation of the earliest approaches to studying psychology, the development of psychological research over the course of the 20th century led social psychologists steadily back to the view espoused by St. Augustine in the early 5th century. In particular, demonstrations of respondents' lack of introspective access to the causes of their own judgments and behavior (see, e.g., Nisbett & Wilson, 1977) forever shattered the illusion that self-reports would be a sufficient means for illuminating the workings of the mind. Alternatives to self-report began to be formulated, and at the dawn of the 21st century an explosion of research is exploring the usefulness of these new, implicit measures of mental contents and processes (see,

e.g., Lane, Banaji, Nosek, & Greenwald, Chapter 3, this volume; Wittenbrink, Chapter 2, this volume). While undeniably exciting, this plethora of new techniques for peering into our inner mental lives raises a number of important questions that will have to be thoroughly researched and satisfactorily answered in order for the promise of these new measures to be fully realized. What is the appropriate theoretical construct corresponding to each measure? Do these measures really provide access to unconscious mental processes? What do the different measures have in common, and what is unique to each particular measure? How are the different measures related to physiological correlates? It is toward these questions that we turn our attention.

### CONCEPTUAL CLARITY

The initial enthusiasm for implicit measures seemed to be accompanied by relatively loose and shifting conceptualizations of their meaning. Part of the confusion undoubtedly arises from more fundamental ambiguities and debates about the meaning of relevant theoretical constructs. For example, much of the work on implicit measures has focused on the assessment of attitudes. However, fundamental disagreements still exist concerning the appropriate conceptualization of "attitude." Some theorists view attitudes as enduring structures in long-term memory (e.g., Fazio, 1995), whereas others view them as momentarily constructed evaluations that integrate current contextual information with selective subsets of long-term memory (e.g., Schwarz & Bohner, 2001). Whereas the former approach conceives of attitudes as relatively static and defined by fixed structural properties, the latter emphasizes a dynamic process with minimal structural assumptions. Theorists from these different camps might very well make different assumptions about what an implicit measure of attitudes is capturing—the strength of a stable association in long-term memory on one hand, or the emergent net evaluative implications of contextually activated knowledge and situationally available input on the other. Thus, labeling something as an "attitude measure" does nothing to clarify the matter, because the term *attitude* means different things to different researchers. It is an open question whether implicit measures map well on to a particular conceptualization of "attitude," a question that must be addressed by systematic empirical investigation.

The importance of this issue becomes even more apparent in the context of early theorizing that implicit attitude measures provide direct access to stable evaluative representations that have their roots

in long-term socialization experiences (see, e.g., Dovidio, Kawakami, & Beach, 2001; Greenwald & Banaji, 1995; Rudman, 2004; Wilson, Lindsey, & Schooler, 2000). This assumption has been challenged over the last few years by accumulating evidence that implicit measures of attitudes are highly susceptible to contextual influences (for a review, see Blair, 2002). However, even though such findings may be interpreted as evidence for the attitudes-as-constructions account (Schwarz & Bohner, 2001), it is not entirely clear how implicit attitude measures are influenced by the context. Does the context lead to a shift in the measurement of an otherwise stable attitude, or does the context influence how the attitude itself is constructed on the spot? Again, the specific answer to these questions depends on the preferred conceptualization of attitudes as momentary constructions (Schwarz & Bohner, 2001) or stable evaluative representations (Fazio, 1995). Notwithstanding these ambiguities, the fact that implicit attitude measures show a strong susceptibility to contextual influences poses a serious challenge to the original expectation that these measures provide direct and unbiased access to stable evaluative representations in memory.

Another fundamental confusion, noted by several scholars (see especially De Houwer, 2006), concerns the meaning of the term *measure*, which can be used to refer to a measurement *procedure* or to the *outcome* of a measurement procedure. As De Houwer (2006) noted, a measurement procedure can be direct or indirect, whereas the term *implicit* is meaningful only with regard to the outcome of a measurement procedure. Conversely, it does not make sense to call the outcome of measurement procedure direct or indirect, just as it does not make sense to label a measurement procedure explicit or implicit. An important consideration is that whereas the direct versus indirect nature of a measurement procedure can be determined a priori by the objective properties of the task, the explicit versus implicit nature of the outcome of a measurement procedure needs to be established empirically.

In this context, De Houwer (2006) pointed out another, related problem in the way the term *implicit* is used. In particular, he argued that researchers often fail to specify in what sense a particular measure should be regarded as implicit. As originally used in the memory literature on patients with amnesia (see, e.g., Warrington & Weiskrantz, 1968), "implicit" mental processes referred specifically to processes that operated in the absence of conscious awareness. Implicit memory thus was evident in task performance reflecting the residue of previous experience, in the absence of any explicit memory for that experience. In keeping with this precedent, it seems to be commonly the case in the social psychology literature that whenever something

is called “implicit,” it is assumed to be consciously inaccessible. But as De Houwer (2006) noted, this is an empirical assumption, and it is usually an untested one.

Besides the interpretation of the term *implicit* as “unconscious,” implicit attitude measures are sometimes assumed to reflect “automatic” attitudes. Bargh (1994) identified four separate senses in which information processing can be considered automatic, and lack of awareness is but one of these “four horsemen” of automaticity. Many so-called automatic processes happen with awareness (Moors & De Houwer, 2006). At the level of implicit attitude measures, it is thus important to specify which aspect(s) of automaticity one capitalizes on. It may be that awareness covaries with measures, such that for some measures, respondents are consciously aware of the assessed attitude. The same holds for other aspects of automaticity (i.e., controllability, spontaneity, and resource dependency). We argue that it is important to understand how these issues map onto the features of a particular task and whether they remain invariant across different content domains or interact with content. It may be, for example, that the controllability of responses varies across different tasks, across content domains, or as a function of motivation and practice at controlling particular kinds of responses. It is incumbent upon researchers to understand the scope of what they can claim about the implicitness of their measures and to tailor their theoretical conclusions accordingly.

Another major lesson for researchers that will lead to greater conceptual clarity is the loss of innocence regarding our aspiration to create “process-pure” measures. A process-pure measure would be one that cleanly and unambiguously indexes a single construct of interest. Many people using the Implicit Association Test (IAT), for example, functionally regard it as a process-pure measure of the strength of a mental association (or, more accurately, a pair of associations). However, performance on the IAT is clearly influenced by many other factors, some of which are theoretically uninteresting and potentially controllable (e.g., general perceptual-motor skills) and others of which are highly interesting in their own right. The application of process-dissociation techniques (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Payne, 2001) provides a way of systematically decomposing these theoretically relevant components of performance, and it represents one of the most promising and important directions for research employing implicit attitude measures. For example, Conrey and colleagues (2005), in their Quad-Model, have shown that performance on the IAT (and many other measures) reflects not only the automatic associative bias that

has been the focus of most IAT users, but also on several other processes, such as the discriminability of the stimulus, success at overcoming bias, and general guessing biases. Of the other processes in the Quad-Model, the theoretically most relevant is the process of overcoming bias. Multinomial modeling of IAT data has confirmed that the motivation/ability to overcome the automatic associative bias also influences IAT performance systematically. By applying a process-dissociation methodology to the IAT (among other measures), it becomes possible to obtain much more finely tuned estimates of the automatic associative bias by pulling out separate processes that influence overall performance. The important general lesson here is that no measure is process pure, and a major path toward clarity lies in the application of new techniques for decomposing task performance into more specific, conceptually meaningful subcomponents. By defining our terms and specifying our constructs as precisely as possible, rapid progress will be greatly facilitated.

## INTROSPECTION AND CONSCIOUSNESS

Does the outcome of an indirect measurement procedure necessarily reflect an implicit attitude in the sense that it is unavailable to introspection or self-report? Not necessarily. As noted by De Houwer (2006), whether or not the construct assessed by a given task is “implicit” (or unconscious) needs to be established empirically. Indeed, attitudes assessed with indirect measures are sometimes highly correlated with self-reported attitudes (e.g., Banse, Seise, & Zerbes, 2001; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Teachman, Gregg, & Woody, 2001). On one hand, these findings may indicate that the attitude reflected in the implicit measure is available to introspection. On the other hand, however, one could object that there might be a separate, implicit aspect to the attitude that just happens to coincide with the explicit aspect in the case of some attitudes. Notwithstanding these two possible interpretations, it seems more parsimonious to assume that the attitude in such cases is available to introspection; yet it may nevertheless influence some other automatic aspects of information processing that are picked up via implicit attitude measures. Someone who detests spiders may be well aware of that fact and may show a rapid, involuntary avoidance response to spiders on an indirect measure of attitudes. Should that response be labeled an “implicit” attitude that is unavailable to introspection, or is it best viewed as a spontaneous, unintentional consequence of a given attitude?

In a similar way, one could ask whether low correlations between self-reported and indirectly assessed attitudes indicate that the indirectly assessed attitude is “implicit” or unconscious. Again, the answer to this question is, not necessarily. Correlations between self-reported and indirectly assessed attitudes would naturally be expected to be low if the indirectly assessed attitude is unconscious (unless there is reason to assume a spurious relation). However, low correlations between the two kinds of measures can arise for a multitude of reasons other than a lack of introspective access (Gawronski, Hofmann, & Wilbur, 2006). In fact, there is now a large body of evidence showing that multiple factors determine whether “implicit” attitudes assessed with indirect measures are related to self-reported “explicit” attitudes. If these factors are controlled, explicit and implicit attitude measures typically show quite substantial correlations.

First, it is often assumed that indirect attitude measures are less affected by individuals’ deliberate attempts to control their responses than self-report measures. Such motivationally driven influences are particularly pronounced in socially sensitive domains where social desirability may affect self-reported, but not indirectly assessed attitudes. Consistent with this assumption, several studies have demonstrated that self-reported and indirectly assessed attitudes toward racial minority groups are highly correlated when individual differences in the motivation to control prejudiced reactions are controlled (e.g., Akrami & Ekehammer, 2005; Banse & Gawronski, 2003; Dunton & Fazio, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Gawronski, Gschke, & Banse, 2003; Hofmann, Gschwendner, & Schmitt, 2005). In a similar vein, Nier (2005) demonstrated that correlations between self-reported and indirectly assessed attitudes toward African Americans were significantly higher when participants believed that inaccurate self-reports could be detected by means of a lie detector.

Second, correlations between self-reported and indirectly assessed attitudes have been shown to depend on the degree of cognitive deliberation. Consistent with this assumption, Florack, Scarabis, and Bless (2001) demonstrated that individuals with a strong dispositional tendency to engage in cognitive deliberation (i.e., high need for cognition; see Cacioppo, Petty, Feinstein, & Jarvis, 1996) showed lower correlations between self-reported and indirectly assessed attitudes than individuals with a low tendency to engage in deliberation. In a similar vein, a meta-analysis by Hofmann, Gawronski, Gschwendner, Le, and Schmitt (2005) found that correlations between self-reported and indirectly assessed attitudes generally increase as a function of the spontaneity of self-reports.

Third, self-reports may or may not correspond to indirect measures with regard to the specific aspect of the attitude that is being assessed. As such, correlations between the two are sometimes reduced simply because of such underlying conceptual differences. Banse and colleagues (2001), for example, demonstrated that indirectly assessed attitudes toward homosexuals show higher correlations with self-reported attitudes when the latter involve self-reports on affective responses (e.g., “I feel uncomfortable nearby two men kissing each other”) than when they involve self-reports on normative beliefs (e.g., “Gay men should not work with children or adolescents”). These results were corroborated in Hofmann, Gawronski, and colleagues’ (2005) meta-analysis, showing that attitudes assessed with the IAT show higher correlations with affective as compared to cognitive self-report measures. In addition, Hofmann et al. found that low correlations can also be due to mismatches in dimensionality (see also Nosek, 2005). The IAT, for example, generally involves a comparison between two attitude objects, thus representing relative rather than absolute evaluations. Thus, it is not very surprising that correlations between the IAT and explicit self-reports are generally higher when the latter involve the same relative rather than absolute evaluations.

Finally, implicit attitude measures often exhibit low internal consistencies (e.g., Banse, 1999; Bosson, Swann, & Pennebaker, 2000; Gawronski, 2002; Olson & Fazio, 2003). Thus, their correlations to self-reported attitudes are often reduced by measurement error. Consistent with this assumption, Cunningham, Preacher, and Banaji (2001) found substantial correlations between self-reported and indirectly assessed attitudes when the impact of measurement error was controlled with latent variable analyses (see also Gawronski, 2002; Hofmann, Gawronski, et al., 2005).

Taken together, these results suggest that indirectly assessed and self-reported attitudes show substantial correlations under certain conditions. If the relevant factors are controlled, the two kinds of attitude measures typically show quite substantial correlations. From this perspective, it seems unwarranted to claim that “implicit” attitudes assessed with indirect measures reflect unconscious attitudes that are unavailable to introspection. To be sure, indirect attitude measures usually do not require introspection for the assessment of an attitude. However, that does not imply that the assessed attitude is unavailable to introspection.

In addition to the assumption that indirectly assessed attitudes themselves are unconscious, there is another interpretation of the term *unconscious* that has been proposed in the context of implicit

attitude measures (Greenwald & Banaji, 1995). Rather than referring to the attitude itself, the term *unconscious* could also refer to the influence an attitude has on other mental processes. In other words, indirectly assessed attitudes may be consciously accessible, but they may influence other psychological processes outside of conscious awareness (see Nisbett & Wilson, 1977). Even though research on this question is rather limited, there seems to be at least some evidence for this assumption. For example, Gawronski and colleagues (2003) demonstrated that people are sometimes unaware of how indirectly assessed attitudes influence their interpretation of ambiguous information. In this study, 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., Darley & Gross, 1983; Duncan, 1976; Dunning & Sherman, 1997; Kunda & Sherman-Williams, 1993; Sagar & Schofield, 1980), German participants evaluated the behavior more negatively when the target was Turkish than when he was German. However, this effect was moderated by indirectly assessed attitudes, such that the target's category membership influenced the interpretation of ambiguous behavior only for participants with negative attitudes toward Turkish people and not for those with neutral attitudes (see also Hugenberg & Bodenhausen, 2003). More important, the influence of indirectly assessed attitudes on the interpretation of ambiguous behavior was *not* moderated by participants' motivation to control prejudiced reactions. Instead, motivation to control prejudice affected only the relation between self-reported and indirectly assessed attitudes toward Turkish people, such that self-reported and indirectly assessed attitudes were highly correlated for participants low in motivation, but not for those high in motivation to control prejudice (see also Akrami & Ekehhammer, 2005; Banse & Gawronski, 2003; Dunton & Fazio, 1997; Fazio et al., 1995; Hofmann, Gschwendner, & Schmitt, 2005). Self-reported attitudes had no impact on the interpretation of ambiguous behavior. Thus, given that participants were generally able to control the influence of indirectly assessed attitudes on their interpretation of ambiguous behavior (i.e., participants were not under time pressure or otherwise cognitively depleted), these results are consistent with the assumption that people are sometimes unaware of the impact of "implicit" attitudes on the interpretation of ambiguous behavior. In other words, whereas indirectly assessed attitudes seem to be conscious in the sense that they are introspectively accessible to self-report, they still seem to involve an unconscious component, such that they can influence other men-

tal processes outside of conscious awareness. Notwithstanding this finding, much more research is needed to clarify which particular processes are influenced by indirectly assessed attitudes, and which of these influences do or do not occur outside people's conscious awareness.

### COMMONALITIES AND DIFFERENCES BETWEEN IMPLICIT MEASURES

Originally it was a matter of some vexation when two different indirect measures of a particular attitude (e.g., IAT and sequential priming measures of racial attitudes) were found not to correlate particularly strongly (or even not at all). After all, it is not uncommon to find that a Likert scale and a semantic differential scale measuring attitudes toward the same attitude object correlate extremely strongly. If different self-report measures of attitudes correlate in this manner, why should indirect measures not likewise correspond? If the starting assumption is that there is some fixed entity called an "implicit attitude" that can be measured in a variety of different ways, then it would indeed seem disheartening if different measures of this same entity did not correlate with one another. However, as we have argued, if the emphasis is shifted to the idea that it is the *outcome* of a particular measurement procedure that is implicit, and that the implicitness or automaticity of this outcome may depend on specific features of the measurement procedure, then the important question becomes, which automatic aspect(s) of attitude activation can be captured by a particular task? Is it spontaneous approach/avoidance tendencies or the activation of evaluative associations in memory (Neumann, Hülsebeck, & Seibt, 2004)? Is it the activation of conceptual or evaluative knowledge in memory (Wittenbrink, Judd, & Park, 2001)? Or is it something else? Measures should be expected to correlate strongly only to the extent that they each tap similar information-processing consequences of attitude activation. Different self-report measures typically do assess the very same aspect of processing (i.e., deliberative evaluation of the attitude object), while different indirect measures are likely to assess a broader range of disparate processes. Indeed, recent research has begun to examine these differences, leading toward a typology of indirect attitude measures that will provide researchers with a rich arsenal to draw from in attacking the mysteries of the black box. Ultimately, the target and range appropriate for each of these research tools need to be specified.

Several proposals have been offered to describe and account for the different emphases of the different indirect attitude measures. Olson and Fazio (2003), for example, proposed that available measures differ in the extent to which they tap general categorical responses versus exemplar-based responses. The attitude version of the IAT, for example, requires respondents to categorize a set of exemplars (e.g., names or faces) in terms of their membership in the relevant attitudinal category (e.g., a racial group). As such, the cognitive focus of the task is on the category. This situation can be contrasted with a common version of the affective priming task, which examines the effects of individual exemplars (e.g., faces) on the processing of subsequently encountered words. While the category to which the exemplar belongs is expected to influence this processing, the cognitive focus of the task is not on the category per se. Thus, differences in the way exemplars are evaluated, versus evaluation of the category as a whole, underlie the low correspondence between affective priming measures and the IAT, according to Olson and Fazio. They found that by modifying the affective priming task in a way that encouraged categorization of the exemplars, correlations with the IAT were greatly enhanced. Mitchell, Nosek, and Banaji (2003) provided further evidence that the IAT does indeed operate at the level of categorical evaluations. In their studies, the same set of exemplars (e.g., African American athletes) produced evidence of both negative and positive associations, depending on whether they were categorized by race or by occupation. These studies make it clear that a focus on the category per se, rather than on individual exemplars of the category, can produce marked shifts in implicit attitude measures.

De Houwer (2003) has proposed another useful distinction between various indirect measures. A basic structural property of many indirect measurement procedures is that they involve examining whether information processing is facilitated (speeded) or impaired (slowed) by the presentation of an attitude object (see Lane et al., Chapter 3, this volume; Wittenbrink, Chapter 2, this volume). The crucial question revolves around whether or not the processes triggered by the activation of the attitude are compatible or incompatible with other processing requirements. De Houwer (2003) argued that there are at least two general kinds of compatibility (see also Kornblum, Hasbroucq, & Osman, 1990) and that different tasks vary in the extent to which they draw upon one or the other of these two phenomena. One of these, *response compatibility*, has to do with how the (natural or induced) response tendencies associated with the attitude object map onto the response requirements of the task (see

also Stroop, 1935). This issue is important in the IAT, for example (see Lane et al., Chapter 3, this volume). In a race IAT, African American exemplars are supposed to be categorized as "Black." Thus, when an African American face (or name) appears on the screen, there will be a (task-appropriate) tendency to respond by pressing the "Black" button. However, for a respondent who harbors negative associations toward African Americans, there may also be a spontaneous response tendency toward the "Unpleasant" button. For the so-called compatible trial block, the "Unpleasant" button and the "Black" button are one and the same, so the two response tendencies are compatible and performance should be facilitated by the fact of their coinciding. For the incompatible trial block, however, the "Unpleasant" response is *not* the same as the "Black" response, so these two response tendencies will be in conflict. As such, performance should suffer (i.e., slower reactions, more errors).

A very similar logic is at work in the affective priming task proposed by Fazio and colleagues (Fazio et al., 1995; see Wittenbrink, Chapter 2, this volume). In this task, the exemplar prime (e.g., a Black face) may elicit a spontaneous evaluative response tendency. Irrespective of this tendency, the task requirement is to categorize the target words presented after the primes in terms of their valence. For trials in which the spontaneous evaluative response to the prime coincides with the valence of the target word, performance should be facilitated because both the prime and the target potentiate the same response. For trials in which the spontaneous evaluative response to the prime is different from the valence of the target word, performance should be inhibited because of the incompatibility of the response tendencies elicited by prime and target. From the standpoint of this structural characteristic, the IAT and the affective priming task are quite similar.

The other form of compatibility that De Houwer (2003) described is *stimulus compatibility*. The fundamental issue here concerns how semantically similar two sets of stimuli are. The lexical decision task used by Wittenbrink, Judd, and Park (1997) is a good example of an indirect measure that relies on stimulus compatibility (see Wittenbrink, Chapter 2, this volume). In this task, a prime stimulus (e.g., an African American face) precedes a standard lexical decision task, in which respondents must simply categorize a letter string as a word or nonword. The words used as target stimuli in the lexical decision task vary in their stereotypical relation to the prime stimulus (e.g., *hostile* vs. *friendly* as target words that are consistent vs. inconsistent with the negative stereotype of African Americans). For in-

stance, if an African American face activates stereotype-consistent concepts in memory, this reaction will either coincide with or contradict the semantic notion of the target word. When the semantic connotations of the prime and target word coincide, responses should be faster, but they should be slower when they are contradictory. It is important to note that as meaningful target words require responses with the same key irrespective of their semantic connotation, (in)compatibility in the lexical decision task is defined on the level of stimulus features rather than on the level of response tendencies; any response tendency that is created by the prime will be irrelevant to the task-required response, which is “word” irrespective of whether the target word is stereotype consistent or stereotype inconsistent.

De Houwer’s (2003) structural analysis leads to the interesting prediction that performance on the IAT and the affective priming task may look more similar to one another than performance on affective versus semantic priming tasks. Indeed, this prediction was borne out in a series of studies by Gawronski and Bodenhausen (2005). These studies also documented another way in which indirect measures differ from one another—in terms of their sensitivity to metacognitive inferences. In research on self-reported attitudes, there is extensive evidence that individuals use metacognitive information to inform their judgments. In this research, two types of influences on judgment are contrasted: the content of activated knowledge versus the experienced ease with which this knowledge was retrieved (for a review, see Schwarz, Bless, Wänke, & Winkielman, 2003). In a typical study, participants are asked to generate a certain number of instances fitting a particular category (e.g., likeable Canadians). For instance, the more likeable Canadians one calls to mind, the more one’s activated knowledge is consistent with a positive evaluation of Canadians. However, because it is typically harder to think of more instances than to think of fewer, experiencing the difficulty of trying to think of many likeable Canadians may imply that there are not very many of them, consistent with a more negative evaluation of Canadians. In the realm of self-reported judgments, experienced ease is commonly observed to determine the tone of overall judgments, unless its diagnosticity is explicitly discredited.

Gawronski and Bodenhausen (2005) extended this research to the realm of indirect attitude measures. A common assumption about indirect measures is that they simply tap the activation level of associations in memory. From this standpoint, the more exemplars of a certain type one calls to mind, the more one’s implicit evaluation will move in the direction of those exemplars, based on the activated

knowledge. Whether or not it was easy or difficult to generate the exemplars should be irrelevant. Thus, performance on a race IAT should reflect more negative evaluations after participants have generated many dislikeable African Americans, as compared to generating just a few dislikeable African Americans. In actuality, however, results were just the opposite. IAT scores reflected the pattern that would be expected if retrieval experiences were guiding responses. Subsequent studies replicated this pattern and showed that it also generalized to an affective priming task as well. Considered from the standpoint of De Houwer’s (2003) distinction between response compatibility and stimulus compatibility, we reasoned that a direct reflection of activated knowledge would be most likely in tasks that involve stimulus compatibility. Again, this prediction was borne out. In a direct comparison of two sequential priming tasks assessing implicit prejudice against African Americans, implicit prejudice depended on the experienced ease of retrieving dislikeable African Americans when the task included an evaluative decision task (i.e., response compatibility). In contrast, implicit prejudice was influenced by the overall amount of activated exemplars when the task included a lexical decision task (i.e., stimulus compatibility).

Further evidence for the differential role of stimulus compatibility and response compatibility in indirect attitude measures is implied by research that investigated the impact of multiple primes on sequential priming effects. Balota and Paul (1996), for example, found that two sequentially presented prime stimuli resulted in additive effects in a typical semantic priming paradigm using a lexical decision task. This result is consistent with an interpretation of sequential priming effects in terms of spreading activation (see Collins & Loftus, 1975), implying that increasing stimulation should increase the activation level of corresponding associations in memory. Interestingly, such additive effects turn into contrast effects when the stimulus compatibility structure implied in the lexical decision task is changed into a response compatibility structure. Gawronski, Deutsch, and Seidel (2005), for example, found that affective priming effects in Fazio and colleagues’ (1995) paradigm were *more* pronounced when the evaluative prime stimulus was preceded by a context prime of the *opposite* valence. However, affective priming effects were *less* pronounced when the prime stimulus was preceded by a context prime of the *same* valence. This finding stands in contrast to the assumption that affective priming effects in evaluative decision tasks have their roots in the same spreading activation mechanisms that are responsible for semantic priming effects in lexical decision tasks (see, e.g., Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Hermans,

De Houwer, & Eelen, 1994). However, it is consistent with claims that affective priming effects are driven by a response compatibility mechanism (see, e.g., De Houwer, Hermans, Rothermund, & Wentura, 2002; Klauer, Roßnagel, & Musch, 1997; Klinger, Burton, & Pitts, 2000; Wentura, 1999), implying that affective priming effects should be influenced by participants' ability to ignore task-irrelevant features, in this case the valence of the prime (see Besner & Stolz, 1999; Besner, Stolz, & Boutilier, 1997). Applied to the present question, the valence of the second prime may be more salient, and thus more difficult to ignore, when it is evaluatively inconsistent with the valence of the first prime. However, the valence of the second prime may be less salient, and thus easier to ignore, when it is evaluatively consistent with the valence of the first prime (Gawronski, Deutsch, & Strack, 2005).

In summary, the available evidence indicates that the widespread equation of the outcome of different measurement procedures with a single "implicit attitude" construct is quite problematic. Different measures are characterized by very different task structures (e.g., exemplar- vs. category-related responses; stimulus vs. response compatibility). As such, measures that may appear similar on the surface can produce very different results, depending on their underlying task structure. Thus, future research employing indirect attitude measures are well advised to take structural differences between tasks into account in order to avoid theoretical misinterpretations of the obtained results.

### PHYSIOLOGICAL CORRELATES

So far, our discussion has primarily addressed measures that are based on response latencies such as the IAT (see Lane et al., Chapter 3, this volume) or sequential priming tasks (see Wittenbrink, Chapter 2, this volume). Even though the vast number of studies using these measures already provide important insights into their underlying mechanisms, and thus into the nature of the constructs assessed with these tasks, many issues are still unresolved, some of which were discussed in the preceding sections of this chapter. A fruitful complement in this endeavor could be the use of physiological measures (see Ito & Cacioppo, Chapter 5, this volume). For instance, by searching for physiological correlates of the performance in indirect attitude measures, physiological measures could provide further insights into both the commonalities and differences between different measures or task properties. In fact, several studies have already investigated

the relation between implicit attitude measures and physiological responses. One of the first studies in this area, for example, found that performance on the race IAT was significantly related to amygdala activation in response to Black faces (Phelps et al., 2000). In a similar vein, Chee, Sriram, Soon, and Lee (2000) demonstrated that performance on a flower–insects IAT was significantly related to activation in the left dorsolateral prefrontal cortex. Milne and Grafman (2001) extended these findings by showing that lesions in the ventromedial prefrontal cortex eliminate implicit gender stereotyping in the IAT. A more complex pattern in the relation between implicit attitude measures and physiological responses was demonstrated by Richeson and colleagues (2003). In their study, race bias in the IAT predicted activation in the right dorsolateral prefrontal cortex in response to Black faces, and this activation mediated the relation between race bias in the IAT and performance impairments in the Stroop task resulting from interactions with a Black confederate. Finally, Wheeler and Fiske (2005) have shown that both automatic stereotype activation and amygdala activation in response to Black faces was significantly reduced when participants were asked to focus on a nonracial category (see also Mitchell et al., 2003).

Even though these findings provide first insights into how implicit attitude measures are related to activity in different areas of the brain, the origin and the nature of the obtained relations remain obscure as long as they cannot be matched to the specific psychological mechanisms underlying indirect attitude measures. For example, the available data on the IAT raise the question of why IAT performance is related to both amygdala activation (Phelps et al., 2000) and activation in the dorsolateral prefrontal cortex (Chee et al., 2000; Richeson et al., 2003). Given that amygdala activation reflects automatically activated negativity, whereas activation in the dorsolateral prefrontal cortex reflects deliberate inhibition, these findings suggest that IAT performance is influenced by both automatic and controlled processes. This assumption is consistent with Conrey and colleagues' (2005) claim that performance on indirect attitude measures is not process pure, but influenced by multiple processes. In terms of Conrey and colleagues' Quad-Model, amygdala activation may be related to the associative bias component, whereas activation in the dorsolateral prefrontal cortex may be related to the component reflecting success at overcoming associative bias. However, the mapping of the processes proposed by Conrey et al. and specific physiological correlates is still hypothetical at this stage. Thus, future research on the relation between indirect attitude measures and physiological correlates could provide an even stronger contribution by



considering the specific mechanisms underlying different kinds of measures.

### WHAT DO WE HAVE TO LEARN?

Given the explosion of research on *implicit attitudes* and *automatic evaluations* (see Wittenbrink, Chapter 2, this volume, Figure 2.1; see also Musch & Klauer, 2003), it is probably not an overstatement to claim that the development of indirect attitude measures brought about, if not a scientific revolution, then at least a substantial reorientation of priorities and perspectives. Certainly, this change has left us not only with important new insights into the psychology of evaluation, but also with some unresolved controversies. These controversies raise a lot of challenging questions that call for answers. Some of the questions we consider to be particularly important are discussed in this chapter. These questions concern (1) the general nature of attitudes as constructions versus stable representations, (2) the "implicitness" of the constructs assessed by indirect attitude measures, (3) the particular processes that are influenced by indirectly assessed attitudes and whether these influences do or do not occur outside of conscious awareness, (4) commonalities and differences between measures as a function of objective task characteristics, and (5) physiological correlates of the processes underlying indirect attitude measures.

Notwithstanding the importance of these issues, the question that we regard as the most important one has not yet been addressed. This question is related to the lack of connection between large-scale theories on the determinants of judgments and behavior, and small-scale theories about what exactly indirect attitude measures assess, how they work, and how they are related to each other. The problem addressed by this question is not so much that we would lack either large-scale theories on human behavior or small-scale theories on indirect attitude measures. Rather, the problem is that the concepts proposed by these two classes of theories do not map onto each other, such that the conceptual terminology of one class of theories could be unambiguously translated into the terminology of the other. For instance, when talking about indirect attitude measures, large-scale theories typically employ abstract theoretical concepts, such as "implicit attitudes" (e.g., Wilson et al., 2000) or "evaluative associations" (e.g., Strack & Deutsch, 2004). Small-scale theories, in contrast, usually refer to concrete operational concepts, such as "response compatibility" (De Houwer, 2003) or "salience asymmetries"

(Rothermund & Wentura, 2004). The challenge that researchers face right now is to find ways to combine the two approaches, such that the relevant concepts can be mapped onto each other and a coherent nomological network can be constructed. A first step in this direction may be Conrey and colleagues' (2005) Quad-Model. This model not only covers some of the theoretical constructs typically proposed by large scale-theories (e.g., automatic attitude activation, cognitive control), but also addresses the mechanisms underlying a specific type of indirect attitude measures (i.e., measures based on response compatibility).

A second question that is directly related to this first one is how physiological processes map onto the constructs proposed by large- and small-scale theories. As outlined above, there is an accumulating body of research that investigated the relation between implicit attitude measures and physiological correlates (e.g., Chee et al., 2000; Milne & Grafman, 2001; Phelps et al., 2000; Richeson et al., 2003; Vanman, Saltz, Nathan, & Warren, 2004; Wheeler & Fiske, 2005). However, compared with the vast number of studies on the mechanisms underlying indirect attitude measures, the available evidence regarding physiological correlates is still very limited. We believe that the proposed mapping of large-scale and small-scale theories could be substantially enriched if this endeavor is associated with a mapping to physiological correlates. Such a multifocus approach would thus include (1) large-scale theories of affect, cognition, and behavior, (2) small-scale theories concerning the mechanisms underlying indirect attitude measures, and (3) theories of physiological functioning. The investigation and integration of this full range of conceptual issues constitutes an ambitious agenda for the next generation of research on "implicit attitudes."

### ACKNOWLEDGMENTS

Preparation of this chapter was supported by grants from the Canada Research Chairs Program, the Social Sciences and Humanities Research Council of Canada, and the Academic Development Fund of the University of Western Ontario to Bertram Gawronski.

### REFERENCES

- Akrami, N., & Ekehammar, B. (2005). The association between implicit and explicit prejudice: The moderating role of motivation to control prejudiced reactions. *Scandinavian Journal of Psychology*, *46*, 361-366.

- Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 827–845.
- Banse, R. (1999). Automatic evaluation of self and significant others: Affective priming in close relationships. *Journal of Social and Personal Relationships*, 16, 803–821.
- Banse, R., & Gawronski, B. (2003). Die Skala Motivation zu vorurteilsfreiem Verhalten: Skaleneigenschaften und Validierung [The scale motivation to act without prejudice: Psychometric properties and validity]. *Diagnostica*, 49, 4–13.
- 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.
- 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.
- Besner, D., & Stolz, J. A. (1999). Unconsciously controlled processing: The Stroop effect reconsidered. *Psychonomic Bulletin and Review*, 6, 449–455.
- Besner, D., Stolz, J. A., & Boutillier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin and Review*, 4, 221–225.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6, 242–261.
- Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, 79, 631–643.
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, B. W. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in the need for cognition. *Psychological Bulletin*, 119, 197–253.
- Chee, M. L., Sriram, N., Soon, C. S., & Lee, K. M. (2000). Dorsolateral prefrontal cortex and the implicit association of concepts and attributes. *Neuroreport for Rapid Communication of Neuroscience Research*, 11, 135–140.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428.
- Conroy, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The Quad-Model of implicit task performance. *Journal of Personality and Social Psychology*, 89, 469–487.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measurement: Consistency, stability, and convergent validity. *Psychological Science*, 12, 163–170.
- Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44, 20–33.
- De Houwer, J. (2003). A structural analysis of indirect measures of attitudes. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219–244). Mahwah, NJ: Erlbaum.
- De Houwer, J. (2006). What are implicit measures and why are we using them? In

- R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11–28). Thousand Oaks, CA: Sage.
- De Houwer, J., Hermans, D., Rothermund, K., & Wentura, D. (2002). Affective priming of semantic categorisation responses. *Cognition and Emotion*, 16, 643–666.
- Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 175–197). Malden, MA: Blackwell.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, 33, 510–540.
- 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.
- Dunning, D., & Sherman, D. A. (1997). Stereotypes and tacit inference. *Journal of Personality and Social Psychology*, 73, 459–471.
- 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.
- 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., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013–1027.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229–238.
- Florack, A., Scarabis, M., & Bless, H. (2001). When do associations matter? The use of automatic associations towards ethnic groups in person judgments. *Journal of Experimental Social Psychology*, 37, 518–524.
- Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology*, 49, 171–180.
- Gawronski, B., & Bollenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation versus retrieval experiences. *Journal of Personality and Social Psychology*, 89, 672–685.
- Gawronski, B., Deutsch, R., & Seidel, O. (2005). Contextual influences on implicit evaluation: Additive versus contrastive effects of evaluative context stimuli in affective priming. *Personality and Social Psychology Bulletin*, 31, 1226–1236.
- Gawronski, B., Deutsch, R., & Strack, F. (2005). Approach/avoidance-related motor actions and the processing of affective stimuli: Incongruity effects in automatic attention allocation. *Social Cognition*, 23, 182–203.
- Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression for-

- mation: 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.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27.
- Hermans, D., De Houwer, J., Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition and Emotion*, 8, 515–533.
- Hofmann, W., Gawronski, B., Gschwendner, T., Lee, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measure. *Personality and Social Psychology Bulletin*, 31, 1369–1385.
- Hofmann, W., Gschwendner, T., & Schmitt, M. (2005). On implicit–explicit consistency: The moderating role of individual differences in awareness and adjustment. *European Journal of Personality*, 19, 25–49.
- Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, 14, 640–643.
- Klauer, K. C., Rofsnagel, C., & Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 246–255.
- Klinger, M. R., Burton, P. C., & Pitts, G. S. (2000). Mechanisms of unconscious priming: I. Response competition, not spreading activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 441–455.
- Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus–response compatibility—a model and taxonomy. *Psychological Review*, 97, 253–270.
- Kunda, Z., & Sherman-Williams, B. (1993). Stereotypes and the construal of individuating information. *Personality and Social Psychology Bulletin*, 19, 90–99.
- Milne, E., & Grafman, J. (2001). Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *Journal of Neuroscience*, 21, 1–6.
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, 132, 455–469.
- Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, 132, 297–326.
- Musch, J., & Klauer, K. C. (Eds.). (2003). *Psychology of evaluation: Affective processes in cognition and emotion*. Mahwah, NJ: Erlbaum.
- Neumann, R., Hülkenbeck, K., & Seibt, B. (2004). Attitudes toward people with AIDS and avoidance behavior: Automatic and reflective bases of behavior. *Journal of Experimental Social Psychology*, 40, 543–550.
- Nier, J. A. (2005). How dissociated are implicit and explicit racial attitudes? A bogus pipeline approach. *Group Processes and Intergroup Relations*, 8, 39–52.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.

- Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit attitudes. *Journal of Experimental Psychology: General*, 134, 565–584.
- Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, 14, 636–639.
- Fayne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181–192.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, 12, 729–738.
- Richeson, J. A., Baird, A. A., Gordon, H. L., Heatherton, T. F., Wyland, C. L., Trawalter, S., et al. (2003). An fMRI investigation of the impact of interracial contact on executive function. *Nature Neuroscience*, 6, 1323–1328.
- Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology*, 133, 139–165.
- Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science*, 13, 79–82.
- 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.
- Schwarz, N., Bless, H., Wänke, M., & Winkielman, P. (2003). Accessibility revisited. In G. V. Bodenhausen & A. J. Lambert (Eds.), *Foundations of social cognition* (pp. 51–77). Mahwah, NJ: Erlbaum.
- Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 436–457). Malden, MA: Blackwell.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220–247.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 28, 643–662.
- Teachman, B. A., Gregg, A. P., & Woody, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology*, 110, 226–235.
- Vanman, E. J., Saltz, J. L., Nathan, L. R., & Warren, J. A. (2004). Racial discrimination by low-prejudiced Whites: Facial movements as implicit measure of attitudes related to behavior. *Psychological Science*, 15, 711–714.
- Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature*, 217, 972–974.
- Wentura, D. (1999). Activation and inhibition of affective information: Evidence for negative priming in the evaluation task. *Cognition and Emotion*, 13, 65–91.
- Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice: Social-cognitive goals affect amygdala and stereotype activation. *Psychological Science*, 16, 56–63.

- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101-126.
- 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.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Evaluative versus conceptual judgments in automatic stereotyping and prejudice. *Journal of Experimental Social Psychology*, *37*, 244-252.