What Do Implicit Measures Tell Us?

Scrutinizing the Validity of Three Common Assumptions

Bertram Gawronski, Etienne P. LeBel, and Kurt R. Peters

University of Western Ontario, London, Ontario, Canada

ABSTRACT—*Experimental paradigms designed to assess "implicit" representations are currently very popular in many areas of psychology. The present article addresses the validity of three widespread assumptions in research using these paradigms: that (a) implicit measures reflect unconscious or introspectively inaccessible representations; (b) the major difference between implicit measures and self-reports is that implicit measures are resistant or less susceptible to social desirability; and (c) implicit measures reflect highly stable, older representations that have their roots in long-term socialization experiences. Drawing on a review of the available evidence, we conclude that the validity of all three assumptions is equivocal and that theoretical interpretations should be adjusted accordingly. We discuss an alternative conceptualization that distinguishes between activation and validation processes.*

> 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*

Were the above quote phrased in third-person plural, one might be tempted to conclude that St. Augustine was a psychologist in the 21st century, complaining about the participants in his research. In fact, contemporary psychology is characterized by widespread skepticism against the use of self-report measures. One of the insights leading to this skepticism is that human beings have only limited introspective access to the processes that underlie their judgments and behavior (Nisbett & Wilson, 1977; Wilson & Dunn, 2004). Thus, when people are asked to report their mental states, they often rely on naive theories, which may or may not be accurate. Another concern regarding the usefulness of self-reports is that they are often biased by social desirability (Crowne & Marlowe, 1960; Paulhus, 1984). Thus, even if people have introspective access to their inner mental life, accurate assessment via self-report can be quite difficult in domains in which certain types of responses are socially undesirable.

Over the last decade, researchers have gained hope in overcoming these problems with the development of a new class of indirect measurement procedures. In contrast to standard self-report measures, the new indirect measures do not require introspective access to the mental representations they are designed to assess.[1] Instead, these measures rely on participants' performance in experimental paradigms, such as sequential priming (Neely, 1977) or response compatibility tasks (Kornblum, Hasbroucq, & Osman, 1990). The most prominent examples of the new indirect measures are probably Greenwald, McGhee, and Schwartz's (1998) Implicit Association Test (IAT) and the evaluative priming paradigm developed by Fazio and colleagues (Fazio, Jackson, Dunton, & Williams, 1995; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Other examples include De Houwer's (2003a) Extrinsic Affective Simon Task; Nosek and Banaji's (2001) Go/No-go Association Task; Payne, Cheng, Govorun, and Stewart's (2005) Affect Misattribution Procedure; and Wittenbrink, Judd, and Park's (1997) semantic priming paradigm (for reviews, see Petty, Fazio, & Briñol, in press; Wittenbrink & Schwarz,

Address correspondence to Bertram Gawronski, Department of Psychology, University of Western Ontario, Social Science Centre, London, Ontario, Canada N6A 5C2, e-mail: bgawrons@uwo.ca.

[1]The new indirect measures are often referred to as implicit measures. Following De Houwer (2006), we use the terms *direct* and *indirect* to describe features of measurement procedures and the terms *explicit* and *implicit* to describe features of the constructs assessed by a particular measurement procedure. Moreover, given the equivocal nature of the constructs assessed by indirect measures, we will use the terms *explicit* and *implicit* with quotation marks.

2007).[2] Even though most of these measures have their roots in social psychology, they have been adopted in almost all areas of psychology, such as personality psychology (e.g., Asendorpf, Banse, & Mücke, 2002), developmental psychology (e.g., Rutland, Cameron, Milne, & McGeorge, 2005), cross-cultural psychology (e.g., Kim, Sarason, & Sarason, 2006), health psychology (e.g., Sherman, Rose, Koch, Presson, & Chassin, 2003), clinical psychology (e.g., Teachman, Gregg, & Woody, 2001), consumer psychology (e.g., Maison, Greenwald, & Bruin, 2004), forensic psychology (e.g., Gray, McCulloch, Smith, Morris, & Snowden, 2003), and neuropsychology (e.g., Phelps et al., 2000).

Even though researchers in different areas employ the new indirect measures for very different epistemic interests, there are three relatively widespread assumptions regarding the functionality of these measures. First, it is often assumed that the new indirect measures provide access to unconscious, "implicit" mental representations that are not accessible to introspection or self-report (e.g., Asendorpf et al., 2002; Bacchus, Baldwin, & Packer, 2004; Bosson, Swann, & Pennebaker, 2000; Brunel, Tietje, & Greenwald, 2004; DeSteno, Dasgupta, Bartlett, & Cajdric, 2004; Devos & Banaji, 2005; Dijksterhuis, 2004; Ellwart, Rinck, & Becker, 2006; Jost, Pelham, & Carvallo, 2002; Teachman, Gapinski, Brownell, Rawlins, & Jeyaram, 2003). This assumption is based on Greenwald and Banaji's (1995) discussion of these measures as a window to "introspectively unidentified (or inaccurately identified) traces of past experience that mediate responses" (p. 5). Second, it is frequently assumed that the major difference between indirect measures and self-reports is that the latter are often biased by self-presentation or social desirability, whereas indirect measures are resistant—or at least less susceptible—to such motivational distortions (e.g., Dunham, Baron, & Banaji, in press; Fazio et al., 1995; Gray et al., 2003; Greenwald, Nosek, & Banaji, 2003; Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002; Jellison, McConnell, & Gabriel, 2004; Nier, 2005; Rutland et al., 2005; Wiers, Van Woerden, Smulders, & De Jong, 2002; Ziegert & Hanges, 2005). This interpretation is based on the observation that participants' attempts to fake responses seem to be less effective for the new indirect measures than they are for standard self-report measures (e.g., Banse, Seise, & Zerbes, 2001; Egloff & Schmukle, 2002; Kim, 2003; Steffens, 2004). Third, it is often assumed that the new indirect measures assess highly stable, old representations that have their roots in long-term socialization experiences (e.g., Conner & Feldman-Barrett, 2005; DeHart, Pelham, & Tennen, 2006; Dovidio, Kawakami, & Beach, 2001; Gregg, Seibt, & Banaji, 2006; Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003; Petty, Tormala, Briñol, & Jarvis, 2006; Rudman, 2004; Rydell & McConnell, 2006; Sinclair, Dunn, & Lowery, 2005; Wilson, Lindsey, & Schooler, 2000). This assumption is based on theories in attitude research, which state that recently acquired attitudes often do not overwrite old attitudes but instead coexist with older, ostensibly stable, implicit ones (e.g., Petty et al., 2006; Wilson et al., 2000).

Even though the status of these three assumptions is still controversial among scholars studying the mechanisms underlying indirect measurement procedures (for an overview, see Wittenbrink & Schwarz, 2007), their accuracy is often taken for granted in research applying these measures. The main goal of the present article is to scrutinize the validity of the three assumptions regarding the functionality of the new indirect measures. Drawing on a review of the available evidence, we argue that the validity of all three assumptions is equivocal and that theoretical interpretations should be adjusted accordingly. On the basis of this conclusion, we will propose an alternative conceptualization that does not incorporate the three characteristics commonly attributed to the new indirect measures but still highlights a functional difference between these measures and standard self-reports that we deem important for many questions addressed in psychological research.

UNCONSCIOUS REPRESENTATIONS

A major difference between the new indirect measures and traditional self-reports is that the new indirect measures do not require introspection for the assessment of mental representations. Thus, the new indirect measures certainly have the potential to tap unconscious representations that are inaccessible to introspection. Note, however, that this does not logically imply that indirect measures actually do assess unconscious representations. The latter assumption is an empirical claim that has to be scrutinized as such (De Houwer, 2006).[3]

One empirical finding that is often interpreted as supporting the unconscious nature of indirectly assessed "implicit" representations is that self-reports and indirectly assessed representations often show relatively low correlations (e.g., Banaji, Lemm, & Carpenter, 2001). Indeed, self-reports should be uncorrelated to the outcome of indirect measures when the latter reflect unconscious representations (unless there is reason to assume an incidental relation). However, low correlations between the two kinds of measures can be the result of many factors other than lack of introspective access (Gawronski & Bodenhausen, 2007; Gawronski, Hofmann, & Wilbur, 2006).

[2]The present article is primarily concerned with a particular class of measurement procedures that have been used as alternatives to questionnaire measures. Thus, the present article does not address other types of measures that have also been described as "implicit," such as measures of implicit memory or implicit learning.

[3]In this context, it is important to distinguish between participants' knowledge of what an indirect measure is assessing and participants' introspective access to the construct assessed by an indirect measure (Fazio & Olson, 2003). These two aspects can be independent, such that participants may know what a given measure is assessing, but they may have no introspective access to the construct assessed by this measure. Conversely, participants may have introspective access to the construct assessed by a given measure, but they may be unaware that this construct is assessed by a given measure. The present discussion is primarily concerned with the question of whether people have introspective access to the constructs assessed by indirect measures.

A trivial but nevertheless important reason for low correlations is measurement error. Psychometric factors are important because several studies have shown relatively low estimates of internal consistency for indirect measures (e.g., Banse, 1999, 2001; Bosson et al., 2000; Gawronski, 2002; Olson & Fazio, 2003; Teige, Schnabel, Banse, & Asendorpf, 2004). These findings suggest that correlations to self-report measures might often be attenuated by measurement error. In fact, when the impact of measurement error is controlled (for instance, with latent variable analyses), correlations between self-reports and indirectly assessed representations have been shown to be quite substantial (e.g., Cunningham, Preacher, & Banaji, 2001; Gawronski, 2002; see also Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005).

In addition, correlations between indirect measures and self-reports are often reduced by lack of conceptual correspondence. Addressing weak relations between attitudes and behavior, Ajzen and Fishbein (1977) noted that the actual impact of attitudes on behavior is often underestimated when the measures used do not correspond in terms of their specificity (e.g., attitudes toward the environment as a predictor of monetary donations to Greenpeace). The same is true for the relation between self-reports and indirectly assessed representations. For instance, the IAT (Greenwald et al., 1998) typically involves a comparison of two target objects, thus reflecting a measure of relative preference rather than absolute evaluations (for a single category variant of the IAT, see Karpinski & Steinman, 2006). Hence, it is not very surprising that correlations between the IAT and self-report measures are generally higher when the latter are relative rather than absolute (Hofmann, Gawronski, et al., 2005).

Even lower correlations may be expected when the two measures do not target the same object. For example, researchers examining racial prejudice often employ standardized self-report scales, such as the Modern Racism Scale (e.g., McConahay, 1986). Many of these scales assess participants' opinions regarding political issues such as affirmative action or the continued discrimination of ethnic minority members in modern society. In contrast, indirect measures of racial prejudice typically assess participants' responses to members of ethnic minority groups (e.g., faces of Black and White individuals). Thus, even though general evaluations of ethnic minority members may be systematically related to people's political opinions (Gawronski, Peters, Brochu, & Strack, 2006), correlations between measures targeting different attitude objects may show low correlations because of their conceptual distinctness, not because one is tapping unconscious representations. Consistent with these assumptions, a meta-analysis by Hofmann, Gawronski, et al. (2005) demonstrated that correlations between IAT measures (Greenwald et al., 1998) and standard self-report measures were generally higher when the two conceptually corresponded to each other than when they did not correspond conceptually.

Taken together, the available evidence suggests that correlations between self-reports and indirectly assessed representations are often attenuated by measurement error and lack of conceptual correspondence. Given these findings, it seems premature to interpret low correlations between the two kinds of measures as evidence for the unconscious or "implicit" nature of indirectly assessed representations. In fact, when the methodological distortions we mentioned are taken into account, the two types of measures typically show quite substantial correlations, which stands in contrast to the assumption that indirect measures assess unconscious or introspectively inaccessible representations (cf. Kihlstrom, 2004).

From a critical perspective, however, one could still argue that correlations between self-reports and indirectly assessed representations do not necessarily imply that the latter are introspectively accessible. For example, your personal reports of your best friends' attitudes toward country music may be highly correlated with their actual evaluations of country music. However, it would seem odd to interpret this correlation as evidence for introspective access to your friends' attitudes. Drawing on similar considerations, Nosek (2005) recently argued that high correlations between self-reports and indirectly assessed representations could be the result of self-perception processes (see Bem, 1967). According to this argument, indirectly assessed "implicit" representation may influence people's behavior, which may then serve as a basis for their self-reports. The result would be a high correlation between self-reports and indirectly assessed "implicit" representations even if the latter are unconscious or introspectively inaccessible.

Notwithstanding the logical plausibility of this argument, we consider a self-perception account as problematic for two reasons. First, in keeping with Occam's razor, it seems more parsimonious to propose a single process (i.e., introspection) rather than two processes (i.e., behavioral influence and self-perception of behavior) to account for correlations between self-reports and indirectly assessed representations (see Quine, 1963). Second, a self-perception account seems inconsistent with the available evidence. Olson and Fazio (2001), for example, used an evaluative conditioning procedure that repeatedly paired neutral conditioned stimuli (CS) with either positive or negative unconditioned stimuli (US). The evaluative conditioning manipulation showed corresponding effects on both indirectly assessed and self-reported evaluations, with the two measures being highly correlated. As participants had no opportunity to observe their behavior toward the CS before they reported their attitudes, self-perception cannot account for the obtained correlation between self-reported and indirectly assessed evaluations. Moreover, it is important to note that participants were not aware of the pairings of CS and US during the conditioning manipulation, which rules out demand effects as an alternative explanation.

In summary, we argue that low correlations between self-reports and indirectly assessed representations are inconclusive

in regard to the potential unconsciousness of indirectly assessed representations. In fact, the available evidence suggests that correlations between self-reports and indirectly assessed representations are quite substantial when methodological factors (e.g., measurement error, lack of conceptual correspondence) are taken into account. These results stand in contrast to the assumption that participants have no introspective access to the representations assessed by indirect measures. To be sure, the new indirect measures do not presuppose introspective access for the assessment of mental representations. However, this does not logically imply that the mental representations assessed with indirect measures are indeed unconscious.

SOCIAL DESIRABILITY

Some researchers appreciate the conscious nature of indirectly assessed representations and argue that the primary difference between self-reports and indirect measures resides in their differing susceptibility to social desirability. Specifically, it is assumed that self-reports are often biased by self-presentation or social desirability, whereas indirect measures are resistant—or at least less susceptible—to such motivational distortions. This assumption is most prominently reflected in the idea that indirect measures may provide a "bona fide pipeline" (Fazio et al., 1995) to people's true representations (see also Asendorpf et al., 2002; Brunel et al., 2004; Dunham et al., in press; Ellwart et al., 2006; Gray et al., 2003; Greenwald et al., 2003; Hermans et al., 2002; Jellison et al., 2004; Nier, 2005; Rutland et al., 2005; Teachman et al., 2003; Wiers et al., 2002; Ziegert & Hanges, 2005).

From an empirical perspective, the social-desirability argument implies two things. First, it implies that correlations between self-reports and indirectly assessed representations should be moderated by social desirability, such that correlations between the two measures decrease as a function of increasing social desirability. Second, it implies that deliberate attempts to influence responses on indirect measures are ineffective (or at least less effective than attempts to influence responses on self-report measures). We will discuss both assumptions in turn.

Moderator Effects on Correlations

Even though it is often assumed that social desirability moderates the relation between self-reports and indirectly assessed representations, the available evidence is much less straightforward than one may expect. Egloff and Schmukle (2003) tested the impact of social desirability on the relation between indirect and self-report measures of anxiety but failed to find any moderator effect for a standard social desirability measure adapted from Crowne and Marlowe (1960) or for a more differentiated social desirability measure distinguishing between self-deception and impression management (Paulhus, 1984). Similar null effects are reported by Hofmann, Gschwendner, and Schmitt (2005) for ethnic prejudice and a social desirability measure adapted from Crowne and Marlowe (1960). Riketta (2006) tested the effects of self-deception and impression management (Paulhus, 1984) for two different indirect measures of self-esteem: a variant of the IAT (Greenwald & Farnham, 2000) and the name letter effect (Koole, Dijksterhuis, & Van Knippenberg, 2001). In contrast to Egloff and Schmukle (2002), Riketta did find a significant moderator effect of self-deception. However, the particular direction of this effect was opposite for the two indirect measures such that higher levels of self-deception reduced correlations for the IAT but increased correlations for the name letter effect.

Nosek (2005) used multilevel modeling analyses to investigate the impact of object-related self-presentation concerns on the relation between directly and indirectly assessed evaluations of 57 different attitude objects. Supporting the assumption that self-presentation moderates the relation between the two kinds of measures, Nosek found that correlations were higher when self-presentation concerns were low than when they were high. However, these findings stand in contrast to meta-analytic results by Hofmann, Gawronski, et al. (2005). These researchers asked independent coders to rate the level of social desirability associated with each of the 53 study topics included in their meta-analysis and then used these ratings to predict the obtained correlations. For a sample of 151 study correlations, the relation between social desirability and study correlations failed to reach statistical significance ($\beta = -.02$). The only study characteristic that reliably predicted correlations was spontaneity; correlations between self-reports and indirectly assessed representations significantly increased as a function of increasing spontaneity during self-report. In fact, when they controlled for spontaneity in the course of making a judgment, Hofmann, Gawronski, et al. (2005) found a positive rather than negative relation between social desirability and study correlations. That is, correlations between the IAT and self-report measures increased rather than decreased as a function of increasing social desirability.

Given the ambiguous evidence for social-desirability effects, one could argue that social desirability may be too general to capture motivational distortions in self-reports and that specific motivations may be more effective in accounting for discrepancies between self-reports and indirect measures. Consistent with this assumption, several studies have found that individual differences in the motivation to control prejudiced reactions moderate the relation between self-reported and indirectly assessed evaluations of ethnic minority groups (e.g., Akrami & Ekehammar, 2005; Dunton & Fazio, 1997; Fazio et al., 1995; Gawronski, Geschke, & Banse, 2003; Hofmann, Gschwendner, & Schmitt, 2005; Payne, 2001; Payne et al., 2005). More precisely, self-reported and indirectly assessed evaluations of ethnic minority groups were highly correlated for participants with a low motivation to control prejudiced reactions. However,

correlations were typically close to zero for participants with a high motivation to control prejudiced reactions. Banse and Gawronski (2003) compared this more specific motivation with social desirability and argued that social desirability may often be too vague to allow precise predictions about the direction of motivational distortions in self-report measures. For instance, in a social context in which racial prejudice is seen as undesirable, higher levels of social desirability should reduce self-reported negativity toward ethnic minority groups. However, in a social context in which racial prejudice is the norm, social desirability should lead to more self-reported negativity against ethnic minority groups. In other words, whereas the particular direction of motivational influences seems to be context independent for relatively specific motivations (e.g., motivation to control prejudiced reactions), the direction of social desirability effects may be context dependent in the sense that they depend on momentarily salient social norms (see also Plant & Devine, 1998).

Despite the available evidence for more specific motivational influences, it is important to note that motivational factors are neither necessary nor sufficient to moderate the relation between self-reports and indirectly assessed representations. On the one hand, motivational factors are not necessary, as correlations between self-reports and indirectly assessed representations can also be influenced by cognitive factors. In a study by Gawronski and LeBel (2007), for example, participants completed both a self-report measure and an indirect measure of evaluations of Coke and Pepsi. Half of the participants were asked to focus on their feelings toward Coke and Pepsi before they completed the self-report measure. The remaining half were asked to think about why they preferred one soft drink over the other (see Millar & Tesser, 1986; Wilson & Dunn, 1986). Results indicate that correlations between self-reported and indirectly assessed evaluations were significantly higher when participants focused on their feelings ($r = .51$) than when they were asked to think about reasons for their preference ($r = .19$). Similar effects were obtained for self-reported and indirectly assessed preferences for Europe versus Asia as well as self-reported and indirectly assessed preferences for owned versus nonowned objects. Given that the employed introspection manipulation is independent of motivational factors, these results suggest that nonmotivational, cognitive factors (i.e., focus of introspection) can be sufficient to moderate the relation between self-reports and indirectly assessed representations. In other words, motivational factors are not necessary to influence correlations between the two kinds of measures.

On the other hand, motivational factors do not seem to be to sufficient to influence the relation between self-reports and indirectly assessed representations. Regarding the aforementioned influence of motivation to control prejudiced reactions, Gawronski, Peters, et al. (2006) argued that the egalitarian goals underlying such motivations influence self-reported evaluations of a given ethnic minority group only when that group is considered to be a target of discrimination (see also Franco & Maass, 1999). Consistent with this assumption, Gawronski, Peters, et al. found that high levels of the motivation to control prejudiced reactions (Dunton & Fazio, 1997) reduced the relation between self-reported and indirectly assessed evaluations of Black people only when participants considered Black people to be a target of racial discrimination (see McConahay, 1986). However, for participants who considered Black people to have gained equal status with Whites, correlations between self-reported and indirectly assessed evaluations of Black people were generally high irrespective of motivation to control prejudiced reactions. In other words, motivational factors were not sufficient to influence correlations between the two kinds of measures.

As a whole, the available evidence suggests that motivational factors are neither necessary nor sufficient to influence correlations between self-reports and indirectly assessed representations. Instead, correlations between the two kinds of measures seem to depend on a complex interplay of cognitive and motivational factors that goes far beyond self-presentation and social desirability (for a comprehensive review, see Hofmann, Gschwendner, Nosek, & Schmitt, 2005). Thus, the common assumption that correlations between self-reports and indirect measures primarily depend on self-presentation, social desirability, or other types of motivational distortions is an unwarranted oversimplification.

Controlled Influences on Indirect Measures

A second implication of the social desirability argument is that indirect measures are resistant to deliberate attempts to control responses (e.g., Asendorpf et al., 2002; Brunel et al., 2004; Dunham et al., in press; Ellwart et al., 2006; Gray et al., 2003; Greenwald et al., 2003; Hermans et al., 2002; Jellison et al., 2004; Nier, 2005; Rutland et al., 2005; Teachman et al., 2003; Wiers et al., 2002; Ziegert & Hanges, 2005). This interpretation is based on the observation that attempts to fake responses seem to be less effective for the new indirect measures than for standard self-report measures (e.g., Banse et al., 2001; Egloff & Schmukle, 2002; Kim, 2003; Steffens, 2004). When distinguishing between the activation and the suppression of an unwanted response tendency, one can again interpret the claim that responses on indirect measures are resistant to control in two different ways. First, one could argue that the activation of unwanted response tendencies in indirect measures is uncontrollable (lack of proactive control). Second, one could argue that, once activated, unwanted response tendencies cannot be suppressed (lack of retroactive control). The available evidence for both assumptions is equivocal.

The assumption that performance on indirect measures is immune to proactive control has been challenged by research showing that mental imagery (Blair, Ma, & Lenton, 2001) or deliberate retrieval of information from memory (Gawronski & Bodenhausen, 2005) can influence the scores revealed by these measures. In the study by Blair et al., for example, participants

were asked to think of either a stereotypical or a counterstereotypical woman. Participants in a control condition were asked to think of a gender-neutral topic (e.g., vacation). After the imagination task, all participants completed an indirect measure of gender stereotyping. Results showed that indirectly assessed gender stereotyping was significantly reduced for participants who imagined a counterstereotypical woman. These results suggest that the activation of the construct-relevant response tendencies employed in indirect measures can be controlled proactively by deliberately activating specific representations in memory.

The assumption that indirect measures are immune to retroactive control has been challenged by recent attempts to disentangle the contribution of multiple processes in indirect measures (e.g., Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Payne, 2001). In their quad model, Conrey et al. proposed a total of four different factors that all contribute to performance on indirect measures: the likelihood that automatic bias is activated by a stimulus, the likelihood that the correct response can be determined, the likelihood that automatic bias is overcome in favor of the correct response, and the likelihood that a general guessing bias drives the response. Through the use of multinomial modeling techniques (for a review, see Batchelder & Riefer, 1999), the quad model allows one to quantify the relative contribution of each of the four factors on task performance on the indirect measure. Applied to the present question, the most important factor is overcoming bias, which directly represents the notion of retroactive control of an already-activated response tendency. Conrey et al. conducted a series of studies using the quad model, which all showed that overcoming bias has a significant impact on participants' performance on indirect measures unless their cognitive capacity is diminished (see also Klauer & Teige-Mocigemba, in press; Lowery, Hardin, & Sinclair, 2001; Payne, 2001).[4]

Taken together, the available evidence indicates that indirect measures are not immune to deliberate attempts to control responses. On the one hand, the activation of unwanted response tendencies can be controlled proactively by deliberately retrieving information from memory (e.g., Blair et al., 2001; Gawronski & Bodenhausen, 2005). On the other hand, even when unwanted responses tendencies are activated, retroactive attempts to control for the impact of these tendencies on task performance have been shown to be effective under conditions of sufficient cognitive capacity (e.g., Conrey et al., 2005; Klauer & Teige-Mocigemba, in press; Lowery et al., 2001; Payne, 2001).

[4]Note that Conrey et al.'s (2005) quad model is designed for a particular type of task, namely measures that are based on the interference of two independent response tendencies (see De Houwer, 2003b; Gawronski & Bodenhausen, 2005). Thus, even though the present arguments apply to all kinds of measures that include a response interference component (e.g., De Houwer, 2003a; Fazio et al., 1995; Greenwald et al., 1998, Nosek & Banaji, 2001), they may not be applicable to tasks that are based on different mechanisms (e.g., Payne et al., 2005; Wittenbrink et al., 1997).

OLD REPRESENTATIONS

Another common assumption in research using the new indirect measures is that these measures assess highly stable, old representations that have their roots in long-term socialization experiences (e.g., Conner & Feldman-Barrett, 2005; DeHart et al., 2006; Dovidio et al., 2001; Gregg et al., 2006; Jordan et al., 2003; Petty et al., 2006; Rudman, 2004; Rydell & McConnell, 2006; Sinclair et al., 2005; Wilson et al., 2000). Researchers have based this assumption on theories in attitude research stating that recently acquired attitudes often do not overwrite old attitudes but instead coexist with old, ostensibly stable, implicit ones (e.g., Petty et al., 2006; Wilson et al., 2000). The common conclusion derived from these theories is that indirect measures tap highly stable, old attitudes, whereas standard self-report measures tap newly acquired attitudes, at least as long as people are motivated and able to retrieve their new attitudes from memory (Wilson et al., 2000).

The proposed matching of measurement procedures with "old" versus "new" attitudes implies that indirectly assessed evaluations should exhibit a higher level of robustness against experimental attempts to change attitudes than do self-reported evaluations. In contrast to this assumption, however, a closer look at the available evidence reveals a rather mixed picture (for a review, see Gawronski & Bodenhausen, 2006a). To our knowledge, there are only two published studies that have indeed shown the predicted asymmetry, such that experimental attempts to change attitudes influenced self-reported but not indirectly assessed evaluations (e.g., Gawronski & Strack, 2004; Gregg et al., 2006). Several other studies found corresponding changes in self-reported and indirectly assessed evaluations (e.g., Gawronski, Walther, & Blank, 2005; Olson & Fazio, 2001; Richeson & Nussbaum, 2004), which is also consistent with the proposed mapping in terms of old versus new attitudes. If a given experimental procedure is indeed able to overwrite the old evaluative representation of an attitude object, this change should be reflected in both self-reports and indirect measures. However, a common finding that is quite difficult to explain with the proposed mapping is a change in indirectly assessed but not self-reported evaluations (e.g., Dasgupta & Greenwald, 2001; Gawronski & LeBel, 2007; Karpinski & Hilton, 2001; Olson & Fazio, 2006). Olson and Fazio (2006), for example, tested the effectiveness of evaluative conditioning in changing racial prejudice against African Americans. In contrast to the assumption that indirect measures assess old attitudes, evaluative conditioning showed a significant effect on indirectly assessed but not self-reported evaluations of African Americans. A similar finding was obtained by Karpinski and Hilton (2001). In their study, participants were presented with repeated pairings of the words *youth* and *elderly* with either positive or negative words and then asked to complete an indirect and a self-report measure of ageism. As with Olson and Fazio's (2006) evaluative conditioning manipulation, the word pairings influenced only

indirectly assessed but not self-reported evaluations of young and old people.

Additional evidence for changes in indirectly assessed but not self-reported evaluations comes from a study by Dasgupta and Greenwald (2001). In their study, participants were presented with pictures of either admired Black and disliked White individuals (e.g., Michael Jordan, Charles Manson) or admired White and disliked Black people (e.g., Tom Hanks, O.J. Simpson). Immediately afterwards, all participants completed an indirect measure and several self-report measures of prejudice against African Americans. Results showed a significant effect of exemplar valence on indirectly assessed but not self-reported evaluations. More precisely, indirectly assessed evaluations showed a lower preference for Whites over Blacks when participants were previously exposed to admired Black and disliked White individuals than when they were exposed to admired White and disliked Black individuals. However, self-reported evaluations of Blacks versus Whites were generally unaffected by the exemplar-exposure manipulation.

Additional counterevidence against the assumption that indirect measures assess highly stable representations comes from research demonstrating a high level of context sensitivity for these measures (for reviews, see Blair, 2002; Gawronski & Bodenhausen, 2006a). Even though most of these studies did not include a corresponding self-report measure, they consistently indicate that indirect measures are highly sensitive to differences in the momentary context. Wittenbrink, Judd, and Park (2001), for example, showed that indirectly assessed evaluations of African American individuals differ as a function of the background context in which these individuals are presented (e.g., at a barbeque, against a graffiti wall). Expanding on these findings, Barden, Maddux, Petty, and Brewer (2004) have shown that these changes do not simply depend on the valence of the context but on the social role of a given individual. For instance, a Black individual elicited highly negative responses on an indirect measure when seen in a prison context in dress suggesting the role of a prisoner, but the same Black person elicited highly positive responses when seen in a prison context in dress suggesting the role of a lawyer. Other contextual factors that have been shown to influence indirect measures include the relative salience of social categories (e.g., Kühnen et al., 2001; Mitchell, Nosek, & Banaji, 2003; Pratto & Shih, 2000; Steele & Ambady, 2006), anticipated social roles in an upcoming interaction (e.g., Richeson & Ambady, 2001, 2003), food and nicotine deprivation (e.g., Ferguson & Bargh, 2004; Seibt, Häfner, & Deutsch, 2007; Sherman et al., 2003), and mood states (e.g., DeSteno et al., 2004; Gemar, Segal, Sagrati, & Kennedy, 2001).

Together, these results indicate that the proposed equation of indirect measures versus self-report measures with old versus new representations (respectively) is empirically unfounded. First, indirect measures have shown a relatively high level of context sensitivity, which stands in contrast to the assumption that indirect measures reflect highly stable representations. Second, several studies that compared experimental effects on self-reports and indirectly assessed representations obtained a lower (rather than higher) level of robustness for indirectly assessed representations.

WHAT CAN INDIRECT MEASURES TELL US?

If indirect measures do not assess unconscious representations, if susceptibility to social desirability and other motivational distortions is not the primary difference between self-report and indirect measures, and if indirect measures do not assess highly stable, old representations, then what exactly do indirect measures tell us? Gawronski and Bodenhausen (2006a, 2006b, in press) recently presented a new conceptualization that does not rely on the three assumptions about indirect measures but still attributes a significant difference to indirect and self-report measures. The central assumption in this model is that indirect measures provide a proxy for the *activation* of associations in memory, whereas self-report measures reflect the outcome of *validation* processes (see Strack & Deutsch, 2004). The crucial difference between activation and validation processes is that the activation of associations can occur independently of whether a person considers these associations as accurate or inaccurate. Validation processes, in contrast, are generally concerned with assessing the (subjective) truth or falsity of activated information. The truth or falsity of a given proposition, in turn, is assumed to be assessed by the consistency of this proposition with all other information that is momentarily considered for a particular judgment (Gawronski, Strack, & Bodenhausen, in press; see also Festinger, 1957). If the proposition is consistent with the other momentarily considered information, it may be used for a corresponding judgment (e.g., in a self-report measure). However, if it is inconsistent with other information, the resulting inconsistency may lead to a rejection of this proposition as a valid basis for a corresponding judgment. It is important to note that the mere rejection of a given proposition as false (i.e., the ascription of a negative truth value) does not necessarily lead to a deactivation of the associations the proposition is based on (Deutsch, Gawronski, & Strack, 2006). Thus, inconsistency-related rejections of propositions typically affect only judgments assessed with self-report measures but not the activation of associations assessed with indirect measures (e.g., Gawronski, Peters, et al., 2006; Gawronski & Strack, 2004; see also Forehand & Perkins, 2005; Gregg et al., 2006).

Such dissociations between the outcomes of activation and validation processes can be particularly important when it comes to the prediction of behavior. Strack and Deutsch (2004) provided an extensive discussion of how activated associations can influence behavior irrespective of whether these associations are considered accurate or inaccurate. Consistent with this assumption, previous research has shown that indirect measures predict various types of spontaneous behaviors that cannot be predicted by corresponding self-report measures (e.g., As-

endorpf et al., 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio et al., 1995; Hofmann, Rauch, & Gawronski, 2007; Neumann, Hülsenbeck, & Seibt, 2004). In a similar vein, several studies have shown that indirect measures predict biases in information processing that cannot be predicted by self-report measures (e.g., Gawronski et al., 2003; Hugenberg & Bodenhausen, 2003, 2004). Aside from these unique influences, activated associations and propositional conclusions can also have interactive effects when they diverge. For instance, Briñol, Petty, and Wheeler (2006) demonstrated that people with large discrepancies between "implicit" (associative) and "explicit" (propositional) self-concepts tend to elaborate self-related information more extensively than do people with small discrepancies (see also Petty et al., 2006). In a similar vein, Jordan and colleagues have shown that combinations of high "explicit" and low "implicit" self-esteem promote narcissistic and defensive behaviors (Jordan, Spencer, & Zanna, 2005; Jordan et al., 2003). Thus, even though indirect measures may not be suitable to capture unconscious representations, to eliminate motivational distortions, or to assess highly stable, old representations, they can be very informative when it comes to understanding the processes that underlie human behavior.

As already mentioned above, the proposed conceptualization of indirect and self-report measures in terms of activation and validation takes a different perspective on the three assumptions discussed in the present article. First, the proposed conceptualization does not assume that activated associations assessed with indirect measures are generally unconscious or introspectively inaccessible (Gawronski, Hofmann, & Wilbur, 2006). In fact, we argue that most of these associations are consciously accessible and that they are used as a basis for explicit judgments unless their propositional implication is inconsistent with other momentarily considered information.

Second, the proposed conceptualization considers cognitive processes rather than motivational processes to be the primary determinant of correlations between self-reports and indirectly assessed associations. Specifically, we argue that the likelihood of propositional inconsistency typically increases as a function of the amount of information that is considered for a particular judgment. Given that the amount of considered information typically increases as a function of cognitive elaboration, correlations between the two kinds of measures should decrease as a function of cognitive elaboration (e.g., Florack, Scarabis, & Bless, 2001; Hofmann, Gawronski, et al., 2005; Koole et al., 2001). Note, however, that the primary determinant of reliance on activated associations is not cognitive elaboration per se but rather the consistency or inconsistency of the propositional information implied by these associations with all other propositional information. Thus, if enhanced cognitive elaboration can resolve the inconsistency between a given proposition and other momentarily considered information, cognitive elaboration should actually increase rather than decrease correlations between the two kinds of measures. Moreover, motivational factors may influence the relation between the two kinds of measures via propositionally represented goals and processes of motivated reasoning. Such motivational influences, however, are indirect rather than direct, in that they are mediated by the validation of propositions (Gawronski & Bodenhausen, in press).

Finally, the proposed model does not imply that either indirect measures or self-report measures should reveal a higher level of stability. In fact, either type of measure may show a higher or lower level of stability depending on (a) the particular nature of the present influence and (b) whether direct changes in one type of process lead to indirect changes in the other (Gawronski & Bodenhausen, 2006a, in press). For example, evaluative conditioning can be assumed to influence the structure of associations in memory, thereby influencing the activation of associations in response to a given stimulus (for a review, see De Houwer, Thomas, & Baeyens, 2001). If these newly formed associations are consistent with other momentarily considered information, they should be used as a basis for a corresponding judgment. The result is a change in self-reported and indirectly assessed evaluations, with changes in self-reported evaluations being mediated by changes in indirectly assessed evaluations (e.g., Olson & Fazio, 2001; see Gawronski & Bodenhausen, 2006a). If, however, newly formed associations are rejected as a basis for a corresponding judgment, evaluative conditioning may influence only indirectly assessed but not self-reported evaluations (e.g., Gawronski & LeBel, 2007; see also Karpinski & Hilton, 2001; Olson & Fazio, 2006). In addition, it is important to note that the activation of associations in memory generally depends on both chronic and situational factors (Bargh, Bond, Lombardi, & Tota, 1986). Thus, the high sensitivity of indirect measures to contextual factors is not very surprising if such measures are considered as a proxy of the activation level of associations in memory.

CAVEAT

Notwithstanding the consistency of our conceptualization with the available evidence, it is important to note that indirect measures only provide a proxy of association activation and are not a direct reflection of activation. As Conrey et al. (2005) have shown in their research using the quad model, indirect measures are not process-pure reflections of activated associations. Instead, performance on these measures is influenced by several different, simultaneously operating processes. Thus, research using indirect measures should take the lack of process purity into account by testing the role of different processes for a particular outcome (e.g., Conrey et al., 2005). If such tests are not feasible, then alternative interpretations should at least be discussed.

Another important issue concerns potential differences between indirect measures. One reviewer pointed out that different kinds of indirect measures often show rather low correlations with each other (e.g., Olson & Fazio, 2003; Sherman et al.,

2003), supposedly challenging the assumption that indirect measures of the same construct represent a coherent category. We generally agree with the contention that indirect measures do not represent a coherent category (for a review, see Gawronski & Bodenhausen, 2007). However, we disagree with the claim that different types of indirect measures typically assess the same construct. Specifically, we argue that indirect measures do not form a coherent category because many of them do *not* assess the same construct. In line with this claim, Olson and Fazio (2003) argued that indirect measures differ in the extent to which they tap category-related versus exemplar-related associations. In the attitude version of the IAT (Greenwald et al., 1998), for example, participants are required to categorize a set of exemplars (e.g., names or faces) in terms of their membership in a relevant attitudinal category (e.g., a racial group). This situation differs from the one typically involved in Fazio et al.'s (1995) affective priming task, in which participants are not explicitly required to process the category membership of the presented stimuli. Even though low correlations between indirect measures can also be due to low measurement reliability (Cunningham et al., 2001), the difference between category-related versus exemplar-related responses is important when it comes to understanding low correlations between indirect measures. Olson and Fazio (2003), for example, found that correlations between the IAT and the affective priming task were significantly higher when participants were required to process the category membership of the presented exemplars in the affective priming task. In a similar vein, individual features of the presented exemplars have been shown to influence responses in the affective priming task (e.g., Livingston & Brewer, 2002), whereas the influence of such features is often overridden by the category applied to the presented exemplars in the IAT (e.g., De Houwer, 2001; Mitchell et al., 2003). Similar to our conclusion regarding correlations to self-report measures in the context of unconscious representations (see above), these results suggest that different kinds of indirect measures may differ with regard to the constructs they assess, such that some assess associations related to individual features of exemplars whereas others assess associations related to the categories that are applied to the presented exemplars (Gawronski & Bodenhausen, 2006b). Thus, when interpreting data obtained with indirect measures, it seems important to consider which type of associations (e.g., category related vs. exemplar related) are assessed by the employed measure.

SUMMARY

The main goal of the present article was to scrutinize the validity of three widespread assumptions in research adopting the new indirect measures: (a) Indirect measures provide access to unconscious, "implicit" mental representations that are not accessible to introspection or self-report; (b) the major difference between indirect measures and self-reports is that the latter are often biased by social desirability, whereas indirect measures are resistant—or at least less susceptible—to such motivational distortions; and (c) indirect measures assess highly stable, old representations that have their roots in long-term socialization experiences. Based on a review of the available evidence, we argued that the validity of all three assumptions is equivocal.

First, there is no empirical evidence that mental representations assessed with indirect measures are unconscious. In fact, indirect measures tend to show quite substantial correlations with self-report measures when methodological factors (e.g., measurement error, conceptual correspondence) are taken into account. Second, the available evidence indicates social desirability and other motivational factors are neither necessary nor sufficient to influence the relation between indirectly assessed and self-reported representations. In addition, performance on indirect measures has been shown to be susceptible to proactive control by means of deliberate activation of information in memory as well as retroactive control via suppression of unintended response tendencies. Both findings indicate that the difference between self-report and indirect measures is far more complex than just a matter of self-presentation, social desirability, or other motivational distortions. Third, indirect measures have shown an unexpectedly high level of context sensitivity and sometimes have even exhibited a higher sensitivity to experimental manipulations than have self-report measures. These findings stand in contrast to the assumption that indirect measures assess highly stable, old representations.

Given the available evidence, it would seem adequate to adjust theoretical interpretations of data obtained with indirect measures. As an alternative framework, we proposed a conceptualization in terms of *activation* and *validation* processes, which does not incorporate any of the three claims about indirect measures. Nevertheless, the proposed conceptualization attributes an important difference to the two kinds of measures when it comes to understanding the interplay of activation and validation processes in influencing judgments and behavior. Even though this model has already been applied to study processes of attitude formation and change (for a review, see Gawronski & Bodenhausen, 2006a), applications in other areas are needed to test its range and limits. In addition, more research is needed to clarify the mechanisms that underlie performance on indirect measures, particularly the interplay of multiple, confounded processes (Conrey et al., 2005). Only empirical data can determine the accuracy of theoretical interpretations, and the present analysis is a good example that such interpretations sometimes turn out to be inaccurate.

Acknowledgments—This article 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. We are grateful to Rainer Banse and Wilhelm Hofmann for helpful comments on an earlier draft of this article.

REFERENCES

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888–918.

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.

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

Bacchus, J.R., Baldwin, M.W., & Packer, D.J. (2004). Increasing implicit self-esteem through classical conditioning. *Psychological Science*, *15*, 498–502.

Banaji, M.R., Lemm, K.M., & Carpenter, S.J. (2001). The social unconscious. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 134–158). Oxford, United Kingdom: Blackwell.

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. (2001). Affective priming with liked and disliked persons: Prime visibility determines congruency and incongruency effects. *Cognition & Emotion*, *15*, 501–520.

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.

Barden, J., Maddux, W.W., Petty, R.E., & Brewer, M.B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, *87*, 5–22.

Bargh, J.A., Bond, R.N., Lombardi, W.J., & Tota, M.E. (1986). The additive nature of chronic and temporary sources of construct accessibility. *Journal of Personality and Social Psychology*, *50*, 869–878.

Batchelder, W.H., & Riefer, D.M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*, 57–86.

Bem, D.J. (1967). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). New York: Academic Press.

Blair, I.V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*, 242–261.

Blair, I.V., Ma, J., & Lenton, A. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*, 828–841.

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.

Briñol, P., Petty, R.E., & Wheeler, S.C. (2006). Discrepancies between explicit and implicit self-concepts: Consequences for information processing. *Journal of Personality and Social Psychology*, *91*, 154–170.

Brunel, F.F., Tietje, B.C., & Greenwald, A.G. (2004). Is the Implicit Association Test a valid and valuable measure of implicit consumer social cognition? *Journal of Consumer Psychology*, *14*, 385–404.

Conner, T., & Feldman-Barrett, L. (2005). Implicit self-attitudes predict spontaneous affect in everyday life. *Emotion*, *5*, 476–488.

Conrey, 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.

Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349–354.

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

Dasgupta, N., & Greenwald, A.G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.

DeHart, T., Pelham, B.W., & Tennen, H. (2006). What lies beneath: Parenting style and implicit self-esteem. *Journal of Experimental Social Psychology*, *42*, 1–17.

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, *37*, 443–451.

De Houwer, J. (2003a). The extrinsic affective Simon task. *Experimental Psychology*, *50*, 77–85.

De Houwer, J. (2003b). 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., Thomas, S., & Baeyens, F. (2001) Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, *127*, 853–869.

DeSteno, D.A., Dasgupta, N., Bartlett, M.Y., & Cajdric, A. (2004). Prejudice from thin air: The effect of emotion on automatic intergroup attitudes. *Psychological Science*, *15*, 319–324.

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

Devos, T., & Banaji, M.R. (2005). American = White? *Journal of Personality and Social Psychology*, *88*, 447–466.

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., 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.

Dunham, Y., Baron, A.S., & Banaji M.R. (in press). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development*.

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.

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

Egloff, B., & Schmukle, S.C. (2003). Does social desirability moderate the relationship between implicit and explicit anxiety measures? *Personality and Individual Differences*, *35*, 1697–1706.

Ellwart, T., Rinck, M., & Becker, E.S. (2006). From fear to love: Individual differences in implicit spider associations. *Emotion*, *6*, 18–27.

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., & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297–327.

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.

Ferguson, M.J., & Bargh, J.A. (2004). Liking is for doing: The effects of goal pursuit on automatic evaluation. *Journal of Personality and Social Psychology*, *87*, 557–572.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row Peterson.

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

Forehand, M.R., & Perkins, A. (2005). Implicit assimilation and explicit contrast: A set/reset model of response to celebrity voice-overs. *Journal of Consumer Research*, *32*, 435–441.

Franco, F.M., & Maass, A. (1999). Intentional control over prejudice: When the choice of the measure matters. *European Journal of Social Psychology*, *29*, 469–477.

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., & Bodenhausen, 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., & Bodenhausen, G.V. (2006a). 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. (2006b). Associative and propositional processes in evaluation: Conceptual, empirical, and meta-theoretical issues. Reply to Albarracín, Hart, and McCulloch (2006), Kruglanski and Dechesne (2006), and Petty and Briñol (2006). *Psychological Bulletin*, *132*, 745–750.

Gawronski, B., & Bodenhausen, G.V. (2007). What do we know about implicit attitude measures and what do we have to learn? In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 265–286). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G.V. (in press). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE Model. *Social Cognition*.

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. (2007). *Dual processes in attitude change: Understanding the relative robustness of "implicit" and "explicit" evaluations*. Manuscript submitted for publication.

Gawronski, B., Peters, K.R., Brochu, P., & Strack, F. (2006). *Understanding the relations between different forms of racial prejudice: The interplay of associative and propositional processes*. Manuscript submitted for publication.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology*, *40*, 535–542.

Gawronski, B., Strack, F., & Bodenhausen, G.V. (in press). Attitudes and cognitive consistency: The role of associative and propositional processes. In R.E. Petty, R.H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new wave of implicit measures*. Mahwah, NJ: Erlbaum

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology*, *41*, 618–626.

Gemar, M.C., Segal, Z.V., Sagrati, S., & Kennedy, S.J. (2001). Mood-induced changes on the Implicit Association Test in recovered depressed patients. *Journal of Abnormal Psychology*, *110*, 282–289.

Gray, N.S., McCulloch, M.J., Smith, J., Morris, M., & Snowden, R.J. (2003). Violence viewed by psychopathic murderers: Adapting a revealing test may expose those psychopaths who are most likely to kill. *Nature*, *423*, 497–498.

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

Greenwald, A.G., & Farnham, S.D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*, 1022–1038.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.K.L. (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., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.

Gregg, A.P., Seibt, B., & Banaji, M.R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*, 1–20.

Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy*, *40*, 217–234.

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 measure. *Personality and Social Psychology Bulletin*, *31*, 1369–1385.

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

Hofmann, W., Gschwendner, T., & Schmitt, M. (2005). On implicit-explicit consistency: The moderating role of individual differences in awareness and adjustment. *European Journal of Personality*, *19*, 25–49.

Hofmann, W., Rauch, W., & Gawronski, B. (2007). And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. *Journal of Experimental Social Psychology*, *43*, 497–504.

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

Hugenberg, K., & Bodenhausen, G.V. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in racial categorization. *Psychological Science*, *15*, 342–345.

Jellison, W.A., McConnell, A.R., & Gabriel, S. (2004). Implicit and explicit measures of sexual orientation attitudes: Ingroup preference and related behaviors and beliefs among gay and straight men. *Personality and Social Psychology Bulletin*, *30*, 629–642.

Jordan, C.H., Spencer, S.J., & Zanna, M. (2005). Types of high self-esteem and prejudice: How implicit self-esteem relates to ethnic discrimination among high explicit self-esteem individuals. *Personality and Social Psychology Bulletin*, *31*, 693–702.

Jordan, C.H., Spencer, S.J., Zanna, M.P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology*, *85*, 969–978.

Jost, J.T., Pelham, B.W., & Carvallo, M.R. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology*, *38*, 586–602.

Karpinski, A., & Hilton, J.L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 774–788.

Karpinski, A., & Steinman, R.B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*, 16–32.

Kihlstrom, J.F. (2004). Implicit methods in social psychology. In C. Sansone, C. Morf, & A. Panter (Eds.), *The Sage handbook of methods in social psychology* (pp. 195–212). Thousand Oaks, CA: Sage.

Kim, D. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, *66*, 83–96.

Kim, D., Sarason, B.R., & Sarason, I.G. (2006). Implicit social cognition and culture: Explicit and implicit psychological acculturation, and distress of Korean-American young adults. *Journal of Social and Clinical Psychology*, *25*, 1–32.

Klauer, K.C., & Teige-Mocigemba, S. (in press). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology*.

Koole, S.L., Dijksterhuis, A., & Van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, *80*, 669–685.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility—a model and taxonomy. *Psychological Review*, *97*, 253–270.

Kühnen, U., Schießl, M., Bauer, N., Paulig, N., Pöhlmann, C., & Schmidthals, K. (2001). How robust is the IAT? Measuring and manipulating implicit attitudes of East- and West-Germans. *Zeitschrift für Experimentelle Psychologie*, *48*, 135–144.

Livingston, R.W., & Brewer, M.B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology*, *82*, 5–18.

Lowery, B.S., Hardin, C.D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*, 842–855.

Maison, D., Greenwald, A.G., & Bruin, R. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology*, *14*, 405–415.

McConahay, J.B. (1986). Modern racism, ambivalence, and the modern racism scale. In J.D. Dovidio & S.L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). Orlando, FL: Academic Press.

Mitchell, J.P., Nosek, B.A., & Banaji, M.R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*, 455–469.

Millar, M.G., & Tesser, A. (1986). Effects of affective versus cognitive focus on the attitude-behavior relationship. *Journal of Personality and Social Psychology*, *51*, 270–276.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.

Neumann, R., Hülsenbeck, 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 evaluation. *Journal of Experimental Psychology: General*, *134*, 565–584.

Nosek, B.A., & Banaji, M.R. (2001). The go/no-go association task. *Social Cognition*, *19*, 625–666.

Olson, M.A., & Fazio, R.H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*, 413–417.

Olson, M.A., & Fazio, R.H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, *14*, 636–639.

Olson, M.A., & Fazio, R.H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*, 421–433.

Paulhus, D.L., (1984). Two component models of social desirable responding. *Journal of Personality and Social Psychology*, *46*, 598–609.

Payne, 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.

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.

Petty, R.E., Fazio, R.H., & Briñol, P. (Eds.). (in press). *Attitudes: Insights from the new implicit measures*. Mahwah, NJ: Erlbaum.

Petty, R.E., Tormala, Z.L., Briñol, P., & Jarvis, W.B.G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST Model. *Journal of Personality and Social Psychology*, *90*, 21–41.

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

Plant, E.A., & Devine, P.G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811–832.

Pratto, F., & Shih, M. (2000). Social dominance orientation and group context in implicit group prejudice. *Psychological Science*, *11*, 515–518.

Quine, W.V.O. (1963). On simple theories of a complex world. *Synthese*, *15*, 103–106.

Richeson, J.A., & Ambady, N. (2001). Who's in charge? Effects of situational roles on automatic gender bias. *Sex Roles*, *44*, 493–512.

Richeson, J.A., & Ambady, N. (2003). Effects of situational power on automatic racial prejudice. *Journal of Experimental Social Psychology*, *39*, 177–183.

Richeson, J.A, & Nussbaum, R.J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, *40*, 417–423.

Riketta, M. (2006). Gender and socially desirable responding as moderators of the correlation between implicit and explicit self-esteem. *Current Research in Social Psychology*, *11*, 14–28.

Rudman, L.A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science*, *13*, 79–82.

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development*, *76*, 451–466.

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.

Seibt, B., Häfner, M., & Deutsch, R. (2007). Prepared to eat: How immediate affective and motivational responses to food cues are influenced by food deprivation. *European Journal of Social Psychology*, *37*, 359–379.

Sherman, S.J., Rose, J.S., Koch, K., Presson, C.C., & Chassin, L., (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology*, *22*, 13–39.

Sinclair, S., Dunn, E., & Lowery, B.S. (2005). The relationship between parental racial attitudes and children's implicit attitudes. *Journal of Experimental Social Psychology*, *41*, 283–289.

Steele, J.R., & Ambady, N. (2006). "Math is hard!" The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology*, *42*, 428–436.

Steffens, M.C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology*, *51*, 165–179.

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

Teachman, B.A., Gapinski, K.D., Brownell, K.D., Rawlins, M., & Jeyaram, S. (2003). Demonstrations of implicit anti-fat bias: The impact of providing causal information and evoking empathy. *Health Psychology*, *22*, 68–78.

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.

Teige, S., Schnabel, K., Banse, R., & Asendorpf, J.B. (2004). Assessment of multiple implicit self-concept dimensions using the Extrinsic Affective Simon Task. *European Journal of Personality*, *18*, 495–520.

Wiers, R.W., Van Woerden, N., Smulders, F.T.Y., & De Jong, P.J. (2002). Implicit and explicit alcohol related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology*, *111*, 648–658.

Wilson, T.D., & Dunn, D.S. (1986). Effects of introspection on attitude-behavior consistency: Analyzing reasons versus focusing on feelings. *Journal of Personality and Social Psychology*, *22*, 249–263.

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., 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). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*, 815–827.

Wittenbrink, B., & Schwarz, N. (Eds.). (2007). *Implicit measures of attitudes*. New York: Guilford Press.

Ziegert, J.C., & Hanges, P.J. (2005). Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology*, *90*, 553–562.