In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.). (2019). Measurement in social psychology (pp. 29-55). New York, NY: Taylor & Francis.

2

IMPLICIT MEASURES

Procedures, Use, and Interpretation

Bertram Gawronski and Adam Hahn

There is no doubt that self-report measures have provided invaluable insights for a wide range of psychological questions (see Jaccard and Blanton, Chapter 1, this volume). After all, a straightforward way to find out what is on a person's mind is to directly ask the person about his or her thoughts and feelings. Yet, self-report measures have been criticized for their inability to capture mental contents that people are either unwilling or unable to report. First, self-report measures are known to be susceptible to self-presentation and socially desirable responding (Crowne & Marlowe, 1960). Second, self-report measures are not well-suited to capture thoughts and feelings that are outside of conscious awareness (Greenwald & Banaji, 1995). To overcome these limitations, psychologists have developed performance-based instruments that (a) limit participants' ability to strategically control their responses, and (b) do not rely on introspection for the measurement of thoughts and feelings. Based on their indirect approach in the assessment of mental contents, these performance-based instruments are often referred to as *implicit measures*, whereas traditional self-report measures are described as *explicit measures*.

Despite the popularity of implicit measures as a tool to overcome the two well-known problems of explicit measures, an accumulating body of research suggests that the relation between implicit and explicit measures involves a much more complex set of factors that cannot be reduced to motivational distortions and lack of introspective access. In a nutshell, the available evidence indicates that (a) strategic control is just one among several factors that can lead to dissociations between implicit and explicit measures and (b) the thoughts and feelings captured by implicit measures are consciously accessible (see Gawronski, LeBel, & Peters, 2007). Together, these findings pose a challenge to the common practice of interpreting dissociations between implicit and explicit measures as indicators of socially desirable responding or lack of introspective awareness. Thus, to ensure accurate conclusions for theory development and real-world applications, it is imperative to use and interpret implicit measures in a manner that is consistent with the available evidence (for an overview, see Gawronski & Payne, 2010).

The current chapter provides an introduction to implicit measures that takes these issues into account. The overarching goal is to offer empirically based guidance for the appropriate use and interpretation of implicit measures. Toward this end, we first explain what it means for a measure to be implicit and then provide a brief overview of the most popular measurement instruments. Expanding on this overview, we discuss various factors that lead to converging versus diverging outcomes on implicit and explicit measures, and how implicit measures can complement explicit measures in individual difference and experimental designs. In the final section, we discuss some caveats against widespread, yet empirically unfounded, assumptions in research using implicit measures.[1]

What Is "Implicit" About Implicit Measures?

A frequent question in research using implicit measures concerns the meaning of the terms *implicit* and *explicit*. This issue is a common source of confusion, because some researchers use the terms to describe features of measurement instruments, whereas others use them to describe the psychological constructs assessed by particular measurement instruments. For example, it is sometimes argued that participants are aware of what is being assessed by an explicit measure but they are unaware of what is being assessed by an implicit measure (e.g., Petty, Fazio, & Briñol, 2009). Yet, other researchers assume that the two kinds of measures tap into distinct memory representations, such that explicit measures capture conscious representations whereas implicit measures capture unconscious representations (e.g., Greenwald & Banaji, 1995).

Although these conceptualizations are relatively common in the literature on implicit measures, we deem it more appropriate to classify different measures in terms of whether the to-be-measured mental content influences participants' responses on the task in an automatic fashion (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). Specifically, measurement outcomes may be described as *implicit* if the impact of the to-be-measured mental content on participants' responses is unintentional, resource-independent, unconscious, or uncontrollable. Conversely, measurement outcomes may be described as *explicit* if the impact of the to-be-measured mental content on participants' responses is intentional, resource-dependent, conscious, or controllable (cf. Bargh, 1994). For example, a measure of racial attitudes may be described as implicit if it reflects participants' racial attitudes even when they do not have the goal to express these attitudes (i.e., unintentional) or despite the goal to conceal these attitudes (i.e., uncontrollable).

An important aspect of this conceptualization is that the terms *implicit* and *explicit* describe the process by which mental contents influence measurement outcomes rather than the measurement instrument (cf. Petty et al., 2009) or the to-be-measured psychological construct (cf. Greenwald & Banaji, 1995). Moreover, whereas the classification of measurement outcomes as implicit or explicit depends on the processes that underlie a given measurement instrument, the instruments themselves may be classified as direct or indirect on the basis of their objective structural properties (De Houwer & Moors, 2010). Specifically, a measurement instrument can be described as direct when the measurement outcome is based on participants' self-assessment of the to-be-measured mental content (e.g., when participants' racial attitudes are inferred from their self-reported liking of Black people). Conversely, a measurement instrument can be described as indirect when the measurement outcome is not based on any self-assessment (e.g., when participants' racial attitudes are inferred from their reaction time performance in a speeded categorization task) or when it is based on a self-assessment that does not involve the to-be-measured mental content (e.g., when participants' racial attitudes are inferred from their self-reported liking of a neutral object that is quickly presented after a Black face). In line with this conceptualization, we use the terms *direct* and *indirect* to describe measurement instruments and the terms *explicit* and *implicit* to describe measurement outcomes. However, because claims about the automatic versus controlled nature of measurement outcomes have to be verified through empirical data, any descriptions of measures as *implicit* should be interpreted as tentative (for a review of relevant evidence, see De Houwer et al., 2009).

A popular way to conceptualize the mental contents captured by implicit measures refers to the idea of *mental association* (Greenwald et al., 2002). For example, the construct of *attitude* has been defined as a mental association between an object and its evaluation (Fazio, 2007). Expanding on this definition, *prejudice* can be defined as evaluative association involving a social group, and *self-esteem* as evaluative association involving the self. Similarly, *stereotypes* can be defined as semantic associations between a social group and stereotypical attributes, whereas the *self-concept* refers to semantic associations between the self and its attributes. In general, the concept of mental association is applicable to any kind of target objects (e.g., consumer products, political candidates) and their evaluative and semantic attributes. Implicit measures are based on the idea that activation of a mental concept can spread to other associated concepts in memory (Collins & Loftus, 1975). To the extent that the associative link between two concepts is sufficiently strong, spread of activation is assumed to occur automatically (i.e., unintentionally, unconsciously, efficiently, uncontrollably; see Bargh, 1994). Implicit measures make use of such automatic processes by assessing the effect of stimuli or stimulus features on participants' performance (e.g., response times, error rates) in responding to other stimuli or stimulus features. Although some theorists have proposed alternative frameworks that reject the notion of mental associations (e.g., De Houwer, 2014; Hughes, Barnes-Holmes, & De Houwer, 2011), associative theorizing has been a driving force in the development of implicit measures, and it still serves as a prominent framework for their application in basic and applied research.

Measurement Instruments

Although there are more than a dozen performance-based instruments whose measurement outcomes may be described as implicit, some of them tend to be more popular than others. In current section, we briefly explain the procedural details of the most frequently used instruments and provide a list of less frequently used instruments for the sake of comprehensiveness.

Sequential Priming Tasks

Historically, the first type of performance-based instruments that has been used to measure social-psychological constructs is based on the notion of sequential priming (for a review, see Wentura & Degner, 2010). In a typical sequential priming task, participants are briefly presented with a prime stimulus, which is followed by a target stimulus. Depending on the nature of the task, participants are asked to (a) classify the target as positive or negative (i.e., evaluative decision task; see Fazio, Jackson, Dunton, & Williams, 1995), (b) classify the target in terms of a semantic property (i.e., semantic decision task; see Banaji & Hardin, 1996), or (c) decide whether the target is a meaningful word or a meaningless letter string (i.e., lexical decision task; see Wittenbrink, Judd, & Park, 1997). The basic idea underlying sequential priming tasks is that quick and accurate responses to the target should be facilitated when the target is congruent with the mental contents that were activated by the prime stimulus. In contrast, quick and accurate responses to the target should be impaired when the target is incongruent with the mental contents that were activated by the prime stimulus.

For example, if a person has a positive attitude toward Donald Trump, this person should be faster and more accurate in identifying the valence of positive words when the person has been primed with an image of Donald Trump compared to priming trials with a neutral baseline stimulus (e.g., Lodge & Taber, 2005). Conversely, evaluative classifications of negative words should be slower and less accurate when the person has been primed with an image of Donald Trump compared to priming trials with a neutral baseline stimulus. Similarly, a person who holds strong gender stereotypes should show better performance in identifying the gender of female pronouns after being presented with stereotypically female professions (e.g., nurse) than stereotypically male professions (e.g., doctor), and vice versa (e.g., Banaji & Hardin, 1996). Finally, using a lexical decision task to assess racial stereotypes, a person may show facilitated classifications of target words related to positive and negative stereotypes of African Americans (e.g., athletic, criminal) after being primed with Black faces compared to priming trials with a neutral baseline stimulus (e.g., Wittenbrink et al., 1997). Although sequential priming tasks are among the most widely used instruments in research using implicit measures, they have been criticized for their low reliability, which rarely exceed Cronbach's Alpha values of .50 (Gawronski & De Houwer, 2014).

Implicit Association Test (and Variants)

The most prominent instrument in research using implicit measures is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). In the critical blocks of the IAT, participants are asked to complete two binary categorization tasks that are combined in a manner that is either congruent or incongruent with the to-be-measured mental content. For example, in the commonly used race IAT, participants may be asked to categorize pictures of Black and White faces in terms of their race and positive and negative words in terms of their valence. In one critical block of the task, participants are asked to press one response key for Black faces and negative words and another response key for White faces and positive words (i.e., prejudice-congruent block). In the other critical block, participants are asked to complete the same categorization tasks with a reversed key assignment for the faces, such that they have to press one response key for White faces and negative words and the other response key for Black faces and positive words (i.e., prejudice-incongruent block). The basic idea underlying the IAT is that responses in the task should be facilitated when two mentally associated concepts are mapped onto the same response key. For example, a person who has more favorable associations with Whites than Blacks should show faster and more accurate responses when White faces share the same response key with positive words and Black faces and share the same response key with negative words, compared with the reversed mapping.

IAT scores are inherently relative in the sense that they conflate four conceptually independent constructs. For example, in the race IAT, a participant's performance is jointly determined by the strength of White-positive, Black-positive, White-negative, and Black-negative associations (see Blanton, Jaccard, Gonzales, & Christie, 2006). This limitation makes the IAT inferior to sequential priming tasks, which permit the calculation of separate priming scores for each of the four associations if the tasks include appropriate baseline primes (see Wentura & Degner, 2010). Yet, the IAT is superior in terms of its internal consistency, which is typically in the range of .70 to .90 (Gawronski & De Houwer, 2014). At the same time, the IAT has been criticized for its blocked presentation of "congruent" and "incongruent" trials, which has been linked to several sources of systematic measurement error (see Teige-Mocigemba, Klauer, & Sherman, 2010). To address these and various other limitations, researchers have developed several variants of the standard IAT that avoid blocked presentations of congruent and incongruent trials, permit non-relative measurements for individual targets and attributes, and reduce the overall length of the task. These IAT variants include the Recoding-Free IAT (IAT-RF; Rothermund, Teige-Mocigemba, Gast, &

Wentura, 2009), the Single-Block IAT (SB-IAT; Teige-Mocigemba, Klauer, & Rothermund, 2008), the Single-Category IAT (SC-IAT; Karpinski & Steinman, 2006), the Single-Attribute IAT (SA-IAT; Penke, Eichstaedt, & Asendorpf, 2006), and the Brief IAT (BIAT; Sriram & Greenwald, 2009).

Go/No-Go Association Task

Another task that has been developed with the goal of overcoming the relative nature of measurement scores in the standard IAT is the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001). On the GNAT, participants are asked to press a button (*go*) in response to some stimuli, and to withhold a response (*no go*) to other stimuli. Different types of stimuli are then paired with the "go" response on different blocks of the task. For example, in one block of a GNAT to measure racial attitudes, participants may be asked to press the "go" button when they see a picture of a Black face or a positive word, and not respond to any other stimuli (which may include pictures of White faces, negative words, and distractor stimuli). In another block, participants may be asked to press the "go" button for pictures of Black faces and negative words, and not respond to any other stimuli. The same task may be repeated in two additional blocks for White instead of Black faces. Because GNAT scores are calculated on the basis of participants' error rates (rather than response times) using signal detection theory (Green & Swets, 1966), the GNAT typically includes a response deadline (e.g., 600 ms) to increase the number of systematic errors. The GNAT has shown lower reliability estimates compared with the standard IAT (Gawronski & De Houwer, 2014). Yet, a clear advantage is the possibility to calculate GNAT scores for individual target objects (e.g., attitudes toward Blacks) instead of relative scores involving two target objects (e.g., relative preference for Whites of Blacks).

Extrinsic Affective Priming Task

Another measure that has been designed to address structural limitations of the IAT is the Extrinsic Affective Simon Task (EAST; De Houwer, 2003). On the EAST, participants are presented with target words (e.g., Pepsi) that are shown in two different colors (e.g., yellow vs. blue) and positive and negative words in white color. Participants are asked to respond to the colored words in terms of their color and to the white words in terms of their valence. In the critical block of the task, participants are asked to respond to positive white words and words of one color (e.g., yellow) with the same key and to negative white words and words of the other color (e.g., blue) with another key (or vice versa). Because the target words are presented in different colors over the course of the task, each target is sometimes paired with the response key for positive words and sometimes with the response key for negative words. The critical question is whether participants respond faster and more accurately to a given target depending on whether its color requires a response with the "positive" or the "negative" key. A major advantage of the EAST is that it does not include blocked presentations of congruent and incongruent trials, which resolves the problems associated with the blocked structure of the IAT (see Teige-Mocigemba et al., 2010). Yet, the EAST has been shown to be inferior to the IAT in terms of its reliability and construct validity, which has been attributed to the feature that participants do not have to process the semantic meaning of the target words (De Houwer & De Bruycker, 2007a). To address this limitation, De Houwer and De Bruycker (2007b) have developed a modified variant of the EAST that ensures semantic processing of the target words, which they called the Identification-EAST (ID-EAST). Although the EAST has originally been designed to measure evaluative associations, some studies have demonstrated its applicability to the measurement of semantic associations (e.g., Teige, Schnabel, Banse, & Asendorpf, 2004).

Affect Misattribution Procedure

The Affect Misattribution Procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005) was designed to combine the structural advantages of sequential priming tasks with the superior psychometric properties of the IAT (for a review, see Payne & Lundberg, 2014). Two central differences to traditional priming tasks are that (a) the target stimuli in the AMP are ambiguous and (b) participants are asked to report their subjective evaluations of the targets. The basic idea is that participants may misattribute the affective feelings elicited by primes to the neutral targets, and therefore judge the targets more favorably when they were primed with a positive stimulus than when they were primed with a negative stimulus. For example, in an AMP to measure racial attitudes, participants may be asked to indicate whether they find Chinese ideographs visually more pleasant or visually less pleasant than average after being primed with pictures of Black versus White faces. A preference for Whites over Blacks would be indicated by a tendency to evaluate the Chinese ideographs more favorably when the ideographs followed the presentation of a White face than when they followed the presentation of a Black face. Interestingly, priming effects in the AMP emerge even when participants are explicitly informed about the nature of the task and instructed not to let the prime stimuli influence their evaluations of the targets (Payne et al., 2005).

Although the AMP has shown satisfactory reliability estimates that are comparable to those of the IAT (Gawronski & De Houwer, 2014; Payne & Lundberg, 2014), the task has been criticized for being susceptible to intentional use of the primes in evaluations of the targets (Bar-Anan & Nosek, 2012). However, this criticism has been refuted by research showing that correlations between AMP effects and self-reported intentional use of the primes reflect retrospective confabulations of intentionality (i.e., participants infer that they must have had such an intention when asked afterwards) rather than actual effects of intentional processes (e.g., Gawronski & Ye, 2015; Payne et al., 2013). Although the AMP was originally designed to measure evaluative associations, modified procedures have been used to measure semantic associations (e.g., Krieglmeyer & Sherman, 2012; Sava et al., 2012).

Approach-Avoidance Tasks

Approach-avoidance tasks are based on the idea that positive stimuli should elicit spontaneous approach reactions, whereas negative stimuli should elicit spontaneous avoidance reactions. In line with this idea, Solarz (1960) found that participants were faster at pushing a lever towards them (approach) in response to positive as opposed to negative stimuli, and pushing it away from them (avoidance) for negative as opposed to positive stimuli. Chen and Bargh (1999) expanded on this finding by instructing participants to make either an approach or an avoidance movement as soon as a stimulus appeared on screen. They then calculated participants' response time to a given stimulus depending on whether they had to show an approach or an avoidance movement in response to that stimulus. Their results showed that participants were faster in making an approach movement in response to positive compared to negative stimuli. Conversely, participants were faster in making an avoidance movement in response to negative compared to positive stimuli.

Initial accounts of approach-avoidance tasks interpreted the obtained response patterns as reflecting direct links between particular motor actions and motivational orientations (e.g., contraction of arm extensor = avoidance; contraction of arm flexor muscle = approach). However, in contrast to these accounts, more recent findings suggest that congruency effects in approach-avoidance tasks depend on the evaluative meaning that is ascribed to a particular motor action in the task. For example, Eder and Rothermund (2008) found that participants were faster in moving a lever backward in response to positive words than negative words when this movement was described as "pull" (positive) and the opposite movement as "push" (negative). In contrast, participants were faster in moving a lever backward in response to negative words than positive words when this movement was described as "downward" (negative) and the opposite movement as "upward" (positive). Corresponding patterns emerged for forward movements. These results suggest that the labels used to describe particular motor actions in approach-avoidance tasks are essential for accurate interpretations of their measurement outcomes. Although some versions of approach-avoidance tasks have shown satisfactory estimates of internal consistency, their reliability varies considerably depending on the variant that is used (Krieglmeyer & Deutsch, 2010).

Other Instruments

Although the reviewed instruments are the most popular examples in the current list of available measures, there are several other instruments with unique features that make them better suited for particular research questions. Although we do not have the space to explain the procedural details of these instruments here, we briefly list them for the sake of comprehensiveness. For example, the Action Interference Paradigm (AIP; Banse, Gawronski, Rebetez, Gutt, & Morton, 2010) has been developed for research with very young children who may not be able to follow the complex instructions of other tasks. The Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010) and the Relational Responding Task (RRP; De Houwer, Heider, Spruyt, Roets, & Hughes, 2015) have been designed to measure automatically activated propositions (rather than automatically activated associations). Other instruments have targeted various methodological limitations of existing measures (e.g., blocked structure, relative measurement, low reliability), including the Evaluative Movement Assessment (EMA; Brendl, Markman, & Messner, 2005), the Implicit Association Procedure (IAP; Schnabel, Banse, & Asendorpf, 2006), and the Sorting Paired Features Task (SPFT; Bar-Anan, Nosek, & Vianello, 2009).

Convergence vs. Divergence Between Implicit and Explicit Measures

The broader idea underlying the use of implicit measures is that they provide information that cannot be gained from explicit measures. This idea is prominently reflected in (a) research on the relation between implicit and explicit measures, (b) research using implicit and explicit measures to predict behavior, and (c) experimental research using implicit and explicit measures as dependent variables.

Relations Between Implicit and Explicit Measures

Correlations between implicit and explicit measures tend to be relatively low overall. Several meta-analyses have found average correlations in the range of .20 to .25 (e.g., Cameron, Brown-Iannuzzi, & Payne, 2012; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). These correlations have been interpreted as evidence that implicit and explicit measures capture related, yet conceptually distinct, constructs (e.g., Nosek & Smyth, 2007). However, such interpretations provide little insight into what these constructs are and why they are weakly related. More seriously, there is evidence that the average correlations obtained in meta-analyses underestimate their actual relation, in that the average correlations are suppressed by various methodological factors. One of the most essential factors in this regard is the low internal consistency of many implicit measures (see Gawronski & De Houwer, 2014). To the extent that the internal consistency of an implicit measure is relatively low, its correlation with explicit measures will be attenuated by measurement error (e.g., Cunningham, Preacher, & Banaji, 2001). Yet, such attenuated correlations may not necessarily reflect distinct psychological constructs.

Other factors that contribute to low correlations can be broadly interpreted in terms of the correspondence principle in research on attitude-behavior relations (Ajzen & Fishbein, 1977). In general, correlations between implicit and explicit measures tend to be higher if the two kinds of measures correspond in terms of their dimensionality and content. For example, Hofmann et al. (2005) found that implicit measures reflecting relative preferences for one group over another tend to show higher correlations to explicit measures of the same relative preference compared to explicit measures of absolute evaluations. Similarly, implicit measures of race bias using Black and White faces as stimuli tend to show higher correlations to evaluative ratings of the same faces compared to evaluative ratings of anti-discrimination policies and perceptions of racial discrimination (e.g., Payne, Burkley, & Stokes, 2008).[2] Thus, without correspondence at the measurement level, it seems premature to interpret low correlations as evidence for distinct constructs at the conceptual level.

In addition to these methodological factors, there are a number of psychological factors that influence correlations between implicit and explicit measures. Overall, correlations tend to be larger for self-reported feelings, affective reactions, and "gut" responses compared to judgments that are more cognitive in nature (e.g., Gawronski & LeBel, 2008; Smith & Nosek, 2011). For example, in a study by Banse, Seise, and Zerbes (2001), scores of a gay-straight IAT showed higher correlations to self-reported affective reactions towards gay people (e.g., self-reported affect when seeing two men kissing each other) compared to self-reported cognitive reactions (e.g., agreement with the statement that gay men should not be allowed to work with children). Implicit and explicit measures also show higher correlations when participants are given less time to think about their judgments than when they are encouraged to deliberate about their response (e.g., Ranganath, Smith, & Nosek, 2008).

Theoretically, varying relations between implicit and explicit measures have been explained in terms of the *activation* of mental contents versus the *application* of activated contents for overt judgments (for a review, see Hofmann, Gschwendner, Nosek, & Schmitt, 2005). For example, the MODE model (Motivation and Opportunity as DEterminants) assumes that implicit measures capture the activation of automatic associations in response to an object (Fazio, 2007). Depending on a person's motivation and opportunity, the person may engage in deliberate processing to scrutinize specific attributes of the object. In this case, people are assumed to base their judgments on the nature of relevant attributes instead of automatically activated associations. Hence, to the extent that both the motivation and the opportunity for deliberate processing are high, correlations between implicit and explicit measures should be low. Yet, when either the motivation or the opportunity for deliberate processing are low, people are assumed to rely on their automatic reactions, leading to higher correlations between implicit and explicit measures.

A similar explanation is offered by the associative-propositional evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011). According to the APE model, implicit measures reflect the activation of mental associations on the basis of feature similarity and spatiotemporal contiguity. In contrast, explicit measures are assumed to reflect the outcome of propositional processes that assess the validity of activated mental contents for overt judgments. A central assumption of the APE model is that the propositional validation of activated mental contents involves an assessment of consistency, in that inconsistency requires a reassessment and potential revision of one's beliefs. Thus, correspondence between implicit and explicit measures is assumed to depend on whether the association captured by an implicit measure is consistent with other information that is considered for a self-reported judgment. To the extent that it is consistent with other salient information, it is usually regarded as valid and therefore used as a basis for self-reported judgments. However, if it is inconsistent with other salient information, people may reject this association in order to restore cognitive consistency (e.g., Gawronski, Peters, Brochu, & Strack, 2008; Gawronski & Strack, 2004). Thus, a central difference to the MODE model is that deliberate processing may not necessarily reduce the relation between implicit and explicit measures. Instead, the APE model predicts that such reductions should occur only when the additionally considered information is inconsistent with the association captured by the implicit measure. To the extent that deliberate processing involves a selective search for information that supports the validity of this association, deliberate processing may in fact increase rather than decrease the relation between implicit and explicit measures (e.g., Galdi, Gawronski, Arcuri, & Friese, 2012; Peters & Gawronski, 2011).

Prediction of Behavior With Implicit and Explicit Measures

A common question about implicit measures is whether they predict behavior, with several independent meta-analyses suggesting different conclusions (e.g., Cameron et al., 2012; Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). Although this question is perfectly justified, it does not reflect the more nuanced theoretical views that have guided research on the prediction of behavior with implicit and explicit measures (for reviews, see Friese, Hofmann, & Schmitt, 2008; Perugini, Richetin, & Zogmaister, 2010). Instead of testing zero-order relations between implicit measures and behavioral criteria, a substantial body of research aimed at gaining a deeper understanding of predictive relations by focusing on the following three questions: (a) What kinds of behaviors do implicit and explicit measures predict? (b) Under which conditions do implicit and explicit measures predict behavior? (c) For whom do implicit and explicit measures predict behavior?

Inspired by the assumptions of dual-process theories (e.g., Fazio, 1990), one of the earliest findings was that implicit measures tend to outperform explicit measures in the prediction of spontaneous behavior (e.g., eye gaze in interracial interactions predicted by implicit measures of racial prejudice), whereas explicit measures tend to outperform implicit measures in the prediction of deliberate behavior (e.g., content of verbal responses in interracial interactions predicted by explicit measures of racial prejudice). This double dissociation has been replicated in a variety of domains with several different measures (e.g., Asendorpf, Banse, & Mücke, 2002; Dovidio, Kawakami, & Gaertner, 2002; Fazio et al., 1995).

Expanding on the idea that the predictive validity of implicit and explicit measures is determined by automatic versus controlled features of the to-be-predicted behavior (Fazio, 2007; Strack & Deutsch, 2004), several studies have investigated contextual conditions under which implicit versus explicit measures are superior in predicting a given behavior. The main finding of this research is that explicit measures outperform implicit measures in the prediction of a given behavior under conditions of unconstrained processing resources, whereas implicit measures outperform explicit measures under conditions of constrained processing resources. For example, Hofmann, Rauch, and Gawronski (2007) found that candy consumption under conditions of cognitive depletion showed a stronger relation to an implicit measure of candy attitudes, whereas candy consumption under control conditions showed stronger relations to an explicit measure (see also Friese, Hofmann, & Wänke, 2008). Similar findings have been obtained for the prediction of interpersonal behavior in interracial interactions (Hofmann, Gschwendner, Castelli, & Schmitt, 2008).

Adopting an individual difference approach, Hofmann, Gschwendner, Friese, Wiers, and Schmitt (2008) found a similar moderation pattern for individual differences in working memory capacity (WMC). In a series of studies, Hofmann and colleagues found that implicit measures outperform explicit measures in the prediction of a given behavior for people with low WMC, whereas explicit measures outperform implicit measures for people with high WMC. The broader idea underlying this research is that individual differences in WMC and situationally available resources are functionally equivalent, such that the implementation of behavioral decisions via reflective processes requires cognitive resources (Strack & Deutsch, 2004). To the extent that cognitive resources are scarce, behavior will be determined by impulsive tendencies that result from automatically activated associations. This idea also resonates with another individual difference factor that has been found to moderate the prediction of behavior: a person's preferred thinking style. Several studies have shown that explicit measures are better predictors of behavior for people with a preference for a deliberative thinking style, whereas implicit measures are better predictors of behavior for people with a preference for an intuitive thinking style (e.g., Richetin, Perugini, Adjali, & Hurling, 2007).

Deviating from approaches in which implicit and explicit measures are seen as competitors in the prediction of behavior, several studies have investigated interactive relations between the two kinds of measures. The general assumption underlying these studies is that discrepancies between implicit and explicit measures are indicative of an unpleasant psychological state that people aim to reduce

(Rydell, McConnell, & Mackie, 2008). For example, people showing large discrepancies on implicit and explicit measures of a particular psychological attribute (e.g., attitude, self-concept) have been shown to elaborate attribute-related information more extensively than people with small discrepancies (e.g., Briñol, Petty, & Wheeler, 2006). In a similar vein, combinations of high self-esteem on explicit measures and low self-esteem on implicit measures have been shown to predict narcissistic and defensive behaviors (e.g., Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003).

Implicit and Explicit Measures as Dependent Variables in Experimental Designs

The available evidence for dissociations in the prediction of behavior raised the question of what determines the outcomes on implicit and explicit measures. This question has been particularly dominant in research on attitude formation and change, which has shown various dissociations in the antecedents of attitudes captured by implicit and explicit measures (for a review, see Gawronski & Bodenhausen, 2006). Whereas some studies found effects on explicit, but not implicit, measures (e.g., Gawronski & Strack, 2004; Gregg, Seibt, & Banaji, 2006), others showed effects on implicit, but not explicit, measures (e.g., Gibson, 2008; Olson & Fazio, 2006). Yet, other studies found corresponding effects on explicit and implicit measures (e.g., Olson & Fazio, 2001; Whitfield & Jordan, 2009). These inconsistent patterns posed a challenge to traditional theories of attitude formation and change, which inspired the development of new theories that have been designed to explain potential dissociations between implicit and explicit measures of attitudes (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Petty, Briñol, & DeMarree, 2007).

One example is the associative-propositional evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011), which distinguishes between the activation of associations in memory (*associative process*) and the validation of momentarily activated information (*propositional process*). According to the APE model, processes of association activation are driven by principles of similarity and contiguity; processes of propositional validation are assumed to be guided by principles of cognitive consistency. The distinction between associative and propositional processes is further linked to implicit and explicit measures, such that implicit measures are assumed to reflect the outcomes of associative processes, whereas explicit measures are assumed to reflect the outcomes of propositional processes. Drawing on several assumptions about mutual interactions between associative and propositional processes, the APE model implies precise predictions regarding the conditions under which a given factor should lead to (a) changes on explicit but not implicit measures; (b) changes on implicit but not explicit measures; (c) corresponding changes on explicit and implicit measures, with changes on implicit measures being mediated by changes on explicit measures; and (d) corresponding changes on explicit and implicit measures, with changes on explicit measures being mediated by changes on implicit measures.

For example, consistent with the predictions of the APE model, cognitive dissonance has been shown to change explicit, but not implicit, evaluations (e.g., Gawronski & Strack, 2004). Conversely, repeated pairings of a neutral conditioned stimulus (CS) with a valenced unconditioned stimulus (US) have been shown to change implicit evaluations of the CS. Yet, explicit evaluations were influenced only when participants were instructed to focus on their feelings toward the CS, which presumably led to a validation of the affective reaction resulting from the newly formed associations (e.g., Gawronski & LeBel, 2008). The central implication of this research is that implicit measures can be more or less resistant to external influences than explicit measures, with their relative resistance depending on whether (a) a given factor targets the content of mental associations (leading to changes on implicit measures) or the perceived validity of activated contents (leading to changes on explicit measures), and (b) proximal changes in one of the two processes lead distal changes in the other process (i.e., when a newly formed association is perceived as valid or when propositional inferences influence the structure of mental associations).

Some Caveats

In the final section of this chapter, we discuss some caveats against widespread assumptions in research using implicit measures. Although the accuracy of these assumptions is often taken for granted, they are either conceptually problematic or inconsistent with the available evidence. Thus, it seems prudent to take these issues into account to ensure appropriate interpretations of the data obtained with implicit measures.

The Metric of Implicit Measures Is Arbitrary

Many of the scoring procedures for implicit measures involve the calculation of difference scores, in which latencies or error rates on "compatible" trials are compared with the latencies or error rates on "incompatible" trials (or neutral baseline trials). The resulting numerical values are often used to infer a psychological attribute on one side of a bipolar continuum if the resulting score is higher than zero (e.g., a preference for Whites over Blacks) and a psychological attribute on the other side of the continuum if the score is lower than zero (e.g., a preference for Blacks over Whites), with a value of zero being interpreted as a neutral reference point. Although such metric interpretations are very common (for a discussion, see Blanton & Jaccard, 2006), they are conceptually problematic because incidental features of the stimulus materials have been shown to influence both the size and the direction of implicit measurement scores (e.g., Bluemke & Friese, 2006; Bluemke & Friese, 2009; Scherer & Lambert, 2009; Steffens & Plewe, 2001).

Because it is virtually impossible to quantify the contribution of such material effects, absolute interpretations of implicit measurement scores are therefore not feasible regardless of whether they involve characteristics of individual participants (e.g., participant X shows a preference for Whites over Blacks) or samples (e.g., 80% of the sample showed a preference for Whites over Blacks).

Yet, it is important to note that most research questions in social and personality psychology do not require absolute interpretations, but instead are based on relative interpretations of measurement scores. The latter applies to experimental designs in which measurement scores are compared across different groups (e.g., participants in the experimental group show higher scores compared to participants in the control group) as well as individual difference designs in which measurement scores are compared across different participants (e.g., participant A has a higher score compared to participant B). Hence, the abovementioned problems do not necessarily undermine the usefulness of implicit measures in social and personality psychology, although they do prohibit the widespread practice of absolute interpretations of measurement scores of individual participants or samples.

Implicit Measures Do Not Provide a Window to the Unconscious

A common assumption in research using implicit measures is that they provide a window to the unconscious, including unconscious attitudes, unconscious prejudice, unconscious stereotypes, unconscious self-esteem, etc. (e.g., Bosson, Swann, & Pennebaker, 2000; Cunningham, Nezlek, & Banaji, 2004; Rudman, Greenwald, Mellott, & Schwartz, 1999). Such claims are based on the notion that implicit measures rely on performance-related indicators, and therefore do not require introspective access for the assessment of mental contents. However, this methodological fact does not permit the reverse inference that the mental contents captured by implicit measures are introspectively inaccessible (cf. Gawronski & Bodenhausen, 2015). Any such claims are empirical hypotheses that require supportive evidence. Importantly, the available evidence clearly contradicts the assumption that the mental contents captured by implicit measures are unconscious (for a review, see Gawronski, Hofmann, & Wilbur, 2006). Using multiple IATs capturing attitudes toward different social groups, Hahn, Judd, Hirsh, and Blair (2014) found that participants were quite accurate in predicting the patterns of their IAT scores. The median within-subjects correlation between predicted and actual scores across four studies (total $N = 430$) was $r = .68$ (average $r = .54$). Interestingly, the same analysis applied to the relation between explicit measures and IAT scores showed much lower correlations (average $r = .20$), similar to the ones typically observed in this area (see Hofmann et al., 2005). These findings pose a challenge to the claim that implicit measures provide a window to the unconscious. Yet, they are consistent with theories that explain dissociations between implicit and explicit measures in terms of other processes that involve a deliberate rejection of consciously accessible associations (e.g., Fazio, 2007; Gawronski & Bodenhausen, 2006).

Dissociations Do Not Necessarily Reflect Motivated Distortions

Another common assumption in research using implicit measures is that they resolve the well-known problems of social desirability. This assumption is based on the notion that it is much more difficult to strategically influence one's scores on an implicit measure compared to one's scores on an explicit measure. Although it is correct that motivated distortions on explicit measures can lead to dissociations between implicit and explicit measures, the validity of this proposition does not permit the reverse conclusion that any dissociation reflects motivated distortions on explicit measures (cf. Gawronski & Bodenhausen, 2015). After all, dissociations can also result from cognitive processes, such as the deliberate analysis of specific attributes (see Fazio, 2007) or the consideration of additional information that is inconsistent with automatically activated associations (see Gawronski & Bodenhausen, 2006). Although either of these processes may elicit motivational concerns, they can lead to dissociations between implicit and explicit measures for purely cognitive reasons (e.g., Gawronski & LeBel, 2008).

Implicit Measures Do Not Provide Superior Access to Old Representations

Some theories suggest that implicit measures reflect highly stable, old representations whereas explicit measures reflect recently acquired, new representations (e.g., Petty, Tormala, Briñol, & Jarvis, 2006; Rudman, 2004; Wilson, Lindsey, & Schooler, 2000). The central idea underlying these theories is that previously formed representations may not be erased from memory when people acquire new information that is inconsistent with these representations. To the extent that earlier acquired knowledge is often highly overlearned, older representations are assumed to be activated automatically upon encounter of a relevant stimulus. In contrast, more recently acquired knowledge is usually less well learned, which implies that the retrieval of newer representations requires controlled processing. Based on these assumptions, implicit measures have been claimed to reflect highly stable, old representations whereas explicit measures reflect more recently acquired, new representations.

Conceptually, these assumptions imply two related, yet empirically distinct, predictions: (a) implicit measures are more resistant to change than explicit measures; (b) implicit measures are more stable over time than explicit measures. Both predictions are at odds with the available evidence. The first prediction stands in contrast to the large body of studies showing experimentally induced changes on implicit, but not explicit, measures (e.g., Gawronski & LeBel, 2008; Gibson, 2008; Olson & Fazio, 2006). The second prediction stands in contrast to the finding that implicit measures tend to show lower stability over time than explicit measures (Gawronski, Morrison, Phills, & Galdi, 2017). Together, these findings pose a challenge to the widespread assumption that implicit measures reflect highly stable, old representations.

Implicit Measures Are Not Immune to Context Effects

Another common assumption about implicit measures is that they can help researchers to resolve the problem of context effects on self-reports. Research on response processes in self-report measures has identified a wide range of contextual factors that can undermine accurate assessments (for a review, see Schwarz, 1999). With the development of performance-based instruments that do not rely on self-assessments, many researchers expected to gain direct access to people's "true" characteristics without contamination by contextual factors. However, the available evidence suggests that implicit measures are at least as susceptible to contextual influences as explicit measures (for reviews, see Blair, 2002; Gawronski & Sritharan, 2010). Theoretically, most of these context effects can be explained with the distinction between activation and application discussed earlier in this chapter. The basic idea is that contextual factors may influence either (a) the activation of mental contents, which should lead to context effects on implicit measures or (b) the application of activated contents for overt judgments, which should lead to context effects on explicit measures. The bottom-line is that neither implicit nor explicit measures are immune to context effects, which poses a challenge to the idea that implicit measures provide context-independent reflections of people's "true" characteristics.

Implicit Measures Do Not Speak to the Automaticity of an Experimental Effect

A defining characteristic of implicit measures is that the to-be-measured mental content influences measurement outcomes in an automatic fashion (see De Houwer et al., 2009). Based on this assumption, implicit measures are sometimes included as dependent measures in experimental studies to test whether the employed manipulation influences the observed outcomes in an automatic fashion. However, such interpretations conflate the impact of mental contents on measurement outcomes with the impact of experimental manipulations on mental contents (Gawronski & De Houwer, 2014). To illustrate this difference, consider a study by Peters and Gawronski (2011) in which participants were asked to recall past behaviors reflecting either extraversion or introversion, and then to complete an IAT designed to measure associations between the self and extraversion (versus introversion). Results showed that IAT scores of self-extraversion associations were higher when participants were asked to recall extraverted behaviors than when they were asked to recall introverted behaviors. Based on the (flawed) assumption that implicit measures can be used to identify automatic effects of experimental manipulations, one might be tempted to conclude that recalling past behaviors influenced self-associations in an automatic fashion. However, the task of recalling past behaviors was fully conscious, intentional, and controllable. Thus, a more appropriate conclusion is that (a) the experimental manipulation influenced the activation of self-associations in a non-automatic fashion, and (b) the activated self-associations influenced participants' responses on the IAT in an automatic fashion. Whereas the former refers to the effect of the experimental manipulation on mental contents, the latter refers to the effect of mental contents on measurement outcomes. The distinction between implicit and explicit measures speaks only to the latter effect, but it does not provide any insight about the nature of the former effect.

Implicit Measures Are Not Process-Pure

As we noted earlier in this chapter, implicit measures are often assumed to provide direct proxies for mental associations. However, in a strict sense, implicit measures reflect behavioral responses, and these responses should not be equated with their presumed underlying mental constructs (De Houwer, Gawronski, & Barnes-Holmes, 2013). Although the impact of mental associations on implicit measures is rarely disputed (for a notable exception, see De Houwer, 2014), a considerable body of research suggests that implicit measures do not provide process-pure reflections of mental associations (Sherman, Klauer, & Allen, 2010). To disentangle the contributions of multiple qualitatively distinct processes to implicit measures, theorists have developed formal models that provide quantitative estimates of these processes, including applications of process dissociation (Payne & Bishara, 2009), multinomial modeling (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Meissner & Rothermund, 2013; Stahl & Degner, 2007), and diffusion modeling (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007).

An illustrative example is Conrey et al.'s (2005) quad-model, which distinguishes between four qualitatively distinct processes underlying responses on implicit measures: (a) activation of an association (AC); (b) detection of the correct response required by the task (D); (c) success at overcoming associative bias (OB); and (d) guessing (G). Research using the quad-model has provided more fine-grained insights into the mechanisms underlying previous findings obtained with implicit measures. Whereas some effects have been shown to be genuinely related to underlying associations (e.g., changes on implicit measures of racial bias that result from extended training to associate racial groups with positive or negative attributes; see Calanchini, Gonsalkorale, Sherman, & Klauer, 2013), others stem from non-associative processes, such as successful versus unsuccessful inhibition of activated associations (e.g., increases in implicit measures of racial bias after alcohol consumption; see Sherman et al., 2008).

The Reliability of Implicit Measures Varies Widely Across Instruments

A final issue concerns the reliability of implicit measures. Unfortunately, measurement error is an issue of concern for several of the reviewed measures, showing estimates of internal consistency that seem unsatisfactory from a psychometric point of view (for a summary, see Gawronski & De Houwer, 2014). The only two measures that have consistently shown acceptable estimates of internal consistency (e.g., Cronbach's Alpha values in the range of .70 to .90) are the IAT (Greenwald et al., 1998) and the AMP (Payne et al., 2005). Most other measures (e.g., GNAT, EAST) have shown estimates of internal consistency that are slightly lower than what might be deemed acceptable (e.g., Cronbach's Alpha values in the range of .50 to .70). The lowest estimates of internal consistency have been observed for sequential priming tasks (e.g., Cronbach's Alpha values below .50). Although concerns about reliability tend to be more common in personality psychology than in social psychology, low internal consistency can be a problem in both correlational and experimental designs. On the one hand, low internal consistency can distort the rank order of participants in terms of a particular construct, which reduces correlations to other measures (e.g., in studies on the prediction of behavior). On the other hand, low internal consistency can reduce the probability of identifying effects of experimental manipulations (e.g., in studies on attitude change), which includes both initial demonstrations of an experimental effect and replications of previously obtained effects (LeBel & Paunonen, 2011). Thus, regardless of whether implicit measures are used in correlational or experimental designs, it seems prudent to take their varying levels of internal consistency into account.

Conclusions

Historically, the use of implicit measures is rooted in the idea that they overcome the well-known limitations of explicit measures in capturing thoughts and feelings that people are either unwilling or unable to report. As should be clear from the evidence reviewed in this chapter, the relation between implicit and explicit measures is much more complex, in that it cannot be reduced to social desirability or lack of awareness. Moreover, many widespread assumptions about implicit measures are either conceptually problematic or inconsistent with the available evidence. However, these conclusions do not imply that implicit measures are useless. If implicit measures are used and interpreted in a manner that is consistent with the available evidence, they can provide valuable insights into the processes underlying social judgment. In addition, they can serve as a useful complement in the prediction of behavior and in research on the formation and change of mental representations. Conceptually, dissociations between implicit and explicit measures in any of these applications can be interpreted as reflecting differences between (a) the activation of mental contents and (b) the application of activated contents for overt judgments. Given that the distinction between activation and application is relevant for almost any question regarding the mental processes underlying judgments and behavior, implicit measures still represent one of the most significant additions to the tool-box of psychological instruments.

Notes

1 Because of its shared concern with implicit measures, their use, and their conceptual meaning, the current chapter has overlap with previous publications by the authors addressing the same issues (e.g., Gawronski, 2009; Gawronski & De Houwer, 2014; Gawronski, Deutsch, & Banse, 2011; Gawronski et al., 2007; Hahn & Gawronski, 2015, 2018).
2 Judgments of anti-discrimination policies and perceptions of racial discrimination are central themes in the Modern Racism Scale (McConahay, 1986), which is often used as an explicit measure in research using implicit measures of racial bias.

References

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

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

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*, 136–141.

Banse, R., Gawronski, B., Rebetez, C., Gutt, H., & Bruce Morton, J. (2010). The development of spontaneous gender stereotyping in childhood: Relations to stereotype knowledge and stereotype flexibility. *Developmental Science*, *13*, 298–306.

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

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the Affective Misattribution Procedure. *Personality and Social Psychology Bulletin*, *38*, 1194–1208.

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task: A measure of association strengths. *Experimental Psychology*, *56*, 329–343.

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.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record*, *60*, 527–542.

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

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27–41.

Blanton, H., Jaccard, J., Gonzales, P. M., & Christie, C. (2006). Decoding the Implicit Association Test: Implications for criterion prediction. *Journal of Experimental Social Psychology*, *42*, 192–212.

Bluemke, M., & Fiedler, K. (2009). Base rate effects on the IAT. *Consciousness and Cognition*, *18*, 1029–1038.

Bluemke, M., & Friese, M. (2006). Do irrelevant features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, *42*, 163–176.

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.

Brendl, C. M., Markman, A. B., & Messner, C. (2005). Indirectly measuring evaluations of several attitude objects in relation to a neutral reference point. *Journal of Experimental Social Psychology*, *41*, 346–368.

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.

Calanchini, J., Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2013). Counter-prejudicial training reduces activation of biased associations and enhances response monitoring. *European Journal of Social Psychology*, *43*, 321–325.

Cameron, C. D., Brown-Iannuzzi, J. L., & Payne, B. K. (2012). Sequential priming measures of implicit social cognition: A meta-analysis of associations with behavior and explicit attitudes. *Personality and Social Psychology Review*, *16*, 330–350.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, *25*, 215–224.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

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., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality and Social Psychology Bulletin*, *30*, 1332–1346.

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

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

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*, 342–353.

De Houwer, J., & De Bruycker, E. (2007a). The implicit association test outperforms the extrinsic affective simon task as an implicit measure of interindividual differences in attitudes. *British Journal of Social Psychology*, *46*, 401–421.

De Houwer, J., & De Bruycker, E. (2007b). The identification-EAST as a valid measure of implicit attitudes toward alcohol-related stimuli. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*, 133–143.

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*, 252–287.

De Houwer, J., Heider, N., Spruyt, A., Roets, A., & Hughes, S. (2015). The relational responding task: Toward a new implicit measure of beliefs. *Frontiers in Psychology*, *6*, 319.

De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and differences. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.

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

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

Eder, A. B., & Rothermund, K. (2008). When do motor behaviors (mis)match affective stimuli? An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology: General*, *137*, 262.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, *23*, 75–109.

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

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

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

Friese, M., Hofmann, W., & Wänke, M. (2008). When impulses take over: Moderated predictive validity of explicit and implicit attitude measures in predicting food choice and consumption behavior. *British Journal of Social Psychology*, *47*, 397–419.

Galdi, S., Gawronski, B., Arcuri, L., & Friese, M. (2012). Selective exposure in decided and undecided individuals: Differential relations to automatic associations and conscious beliefs. *Personality and Social Psychology Bulletin*, *38*, 559–569.

Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology*, *50*, 141–150.

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

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

Gawronski, B., & Bodenhausen, G. V. (2015). Theory evaluation. In B. Gawronski & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 3–23). New York, NY: Guilford Press.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 283–310). New York, NY: Cambridge University Press.

Gawronski, B., Deutsch, R., & Banse, R. (2011). Response interference tasks as indirect measures of automatic associations. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive methods in social psychology* (pp. 78–123). New York, NY: Guilford Press.

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

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

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, *2*, 181–193.

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, *43*, 300–312.

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

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

Gawronski, B., & Sritharan, R. (2010). Formation, change, and contextualization of mental associations: Determinants and principles of variations in implicit measures. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 216–240). New York, NY: Guilford Press.

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., & Ye, Y. (2015). Prevention of intention invention in the affect misattribution procedure. *Social Psychological and Personality Science*, *6*, 101–108.

Gibson, B. (2008). Can evaluative conditioning change attitudes toward mature brands? New evidence from the Implicit Association Test. *Journal of Consumer Research*, *35*, 178–188.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley-Blackwell.

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

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

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

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.

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.

Hahn, A., & Gawronski, B. (2015). Implicit social cognition. In J. D. Wright (Ed.), *The international encyclopedia of the social and behavioral sciences* (2nd ed., pp. 714–720). Oxford: Elsevier.

Hahn, A., & Gawronski, B. (2018). Implicit social cognition. In J. T. Wixted (Ed.), *The Stevens' handbook of experimental psychology and cognitive neuroscience* (4th ed., pp. 395–427). Malden, MA: Wiley-Blackwell.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, *143*, 1369.

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

Hofmann, W., Gschwendner, T., Castelli, L., & Schmitt, M. (2008). Implicit and explicit attitudes and interracial interaction: The moderating role of situationally available control resources. *Group Processes and Intergroup Relations*, *11*, 69–87.

Hofmann, W., Gschwendner, T., Friese, M., Wiers, R., & Schmitt, M. (2008). Working memory capacity and self-regulatory behavior: Towards and individual differences perspective on behavior determination by automatic versus controlled processes. *Journal of Personality and Social Psychology*, *95*, 962–977.

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

Hofmann, W., 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.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record*, *61*, 465–496.

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.

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.

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process-components of the Implicit Association Test: A diffusion model analysis. *Journal of Personality and Social Psychology*, *93*, 353–368.

Krieglmeyer, R., & Deutsch, R. (2010). Comparing measures of approach-avoidance behavior: The manikin task vs. two versions of the joystick task. *Cognition and Emotion*, *24*, 810–828.

Krieglmeyer, R., & Sherman, J. W. (2012). Disentangling stereotype activation and stereotype application in the stereotype misperception task. *Journal of Personality and Social Psychology*, *103*, 205–224.

LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, *37*, 570–583.

Lodge, M., & Taber, C. S. (2005). Automaticity of affect for political leaders, groups, and issues: An experimental test of the hot cognition hypothesis. *Political Psychology*, *26*, 455–482.

McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–126). New York, NY: Academic Press.

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology*, *104*, 45–69.

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

Nosek, B. A., & Smyth, F. L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, *54*, 14–29.

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

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

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*, 171–192.

Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, *20*, 272–314.

Payne, B. K., Brown-Iannuzzi, J., Burkley, M., Arbuckle, N. L., Cooley, E., Cameron, C. D., & Lundberg, K. B. (2013). Intention invention and the affect misattribution procedure: Reply to Bar-Anan and Nosek (2012). *Personality and Social Psychology Bulletin*, *39*, 375–386.

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

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

Payne, B. K., & Lundberg, K. B. (2014). The affect misattribution procedure: Ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass*, *8*, 672–686.

Penke, L., Eichstaedt, J., & Asendorpf, J. B. (2006). Single Attribute Implicit Association Tests (SA-IAT) for the assessment of unipolar constructs: The case of sociosexuality. *Experimental Psychology*, *53*, 283–291.

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

Peters, K. R., & Gawronski, B. (2011). Mutual influences between the implicit and explicit self-concepts: The role of memory activation and motivated reasoning. *Journal of Experimental Social Psychology*, *47*, 436–442.

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The Meta-Cognitive Model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition*, *25*, 657–686.

Petty, R. E., Fazio, R. H., & Briñol, P. (2009). The new implicit measures: An overview. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 3–18). New York, NY: Psychology Press.

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.

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

Richetin, J., Perugini, M., Adjali, I., & Hurling, R. (2007). The moderator role of intuitive versus deliberative decision making for the predictive validity of implicit and explicit measures. *European Journal of Personality*, *21*, 529–546.

Rothermund, K., Teige-Mocigemba, S., Gast, A., & Wentura, D. (2009). Minimizing the influence of recoding in the Implicit Association Test: The Recoding-Free Implicit Association Test (IAT-RF). *The Quarterly Journal of Experimental Psychology*, *62*, 84–98.

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

Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition*, *17*, 437–465.

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

Rydell, R. J., McConnell, A. R., & Mackie, D. M. (2008). Consequences of discrepant explicit and implicit attitudes: Cognitive dissonance and increased information processing. *Journal of Experimental Social Psychology*, *44*, 1526–1532.

Sava, F. A., Maricutoiu, L. P., Rusu, S., Macsinga, I., Virga, D., Cheng, C. M., & Payne, B. K. (2012). An inkblot for the implicit assessment of personality: The semantic misattribution procedure. *European Journal of Personality*, *26*, 613–628.

Scherer, L. D., & Lambert, A. J. (2009). Contrast effects in priming paradigms: Implications for theory and research on implicit attitudes. *Journal of Personality and Social Psychology*, *97*, 383–403.

Schnabel, K., Banse, R., & Asendorpf, J. (2006). Employing automatic approach and avoidance tendencies for the assessment of implicit personality self-concept: The Implicit Association Procedure (IAP). *Experimental Psychology*, *53*, 69–76.

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

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. A., & Groom, C. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*, 314–335.

Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Mathematical modeling of implicit social cognition: The machine in the ghost. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 156–175). New York, NY: Guilford Press.

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

Solarz, A. K. (1960). Latency of instrumental responses as a function of compatibility with the meaning of eliciting verbal signs. *Journal of Experimental Psychology*, *59*, 239–245.

Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology*, *56*, 283–294.

Stahl, C., & Degner, J. (2007). Assessing automatic activation of valence: A multinomial model of EAST performance. *Experimental Psychology*, *54*, 99–112.

Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie*, *48*, 123–134.

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

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.

Teige-Mocigemba, S., Klauer, K. C., & Rothermund, K. (2008). Minimizing method-specific variance in the IAT: A single block IAT. *European Journal of Psychological Assessment*, *24*, 237–245.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). A practical guide to Implicit Association Tests and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook* *of implicit social cognition: Measurement, theory, and applications* (pp. 117–139). New York, NY: Guilford Press.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). New York, NY: Guilford Press.

Whitfield, M., & Jordan, C. H. (2009). Mutual influences of explicit and implicit attitudes. *Journal of Experimental Social Psychology*, *45*, 748–759.

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.