Implicit Measures in Social and Personality Psychology

Bertram Gawronski

The University of Western Ontario, Canada

Jan De Houwer

Ghent University, Belgium

In press. Chapter to appear in: H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd edition). New York: Cambridge University Press.

Please address correspondence concerning this chapter to:

Bertram Gawronski, Ph.D.
Department of Psychology
The University of Western Ontario
Social Science Centre
London, Ontario N6A 5C2
Canada

Phone: +1 (519) 661-2111 Ext. 87672
Email: bgawrons@uwo.ca

Implicit Measures in Social and Personality Psychology

Self-report measures arguably represent one of the most important research tools in social and personality psychology. To measure people's attitudes, beliefs, and personality characteristics, it seems rather straightforward to simply ask them about their thoughts, feelings, and behaviors. Yet, researchers are well aware that people are sometimes unwilling or unable to provide accurate reports of their own psychological attributes. In socially sensitive domains, for example, responses on self-report measures are often distorted by social desirability and self-presentational concerns. Similarly, the value of self-report measures seems limited for psychological attributes that are introspectively inaccessible or outside of conscious awareness. To overcome these limitations, psychologists have developed alternative measurement instruments that reduce participants' ability to control their responses and do not require introspection for the assessment of psychological attributes. In social and personality psychology, such measurement instruments are commonly referred to as *implicit measures*, whereas traditional self-report measures are often described as *explicit measures*.

The main goal of the current chapter is to provide a general introduction to the use and meaning of implicit measures in social and personality psychology. Toward this end, we first explain what implicit measures are and in which sense they may be described as implicit. We then provide an overview of the currently available paradigms, including descriptions of their basic procedures and some recommendations on how to choose among the various measures. Expanding on this overview, we outline what kinds of insights can be gained from implicit measures for understanding the determinants of behavior, biases in information processing, and the formation and change of mental representations. In the final sections, we discuss some caveats regarding the interpretation of implicit measures and potential directions for future developments.

What Are Implicit Measures?

A central characteristic of implicit measures is that they aim to capture psychological attributes (e.g., attitudes, stereotypes, self-esteem) without requiring participants to report a subjective assessment of these attributes. However, there are a lot of such indirect measurement techniques and only few of them have been described as implicit. Thus, 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 procedures, whereas others use them to describe the nature of the psychological attributes 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 tap into conscious representations whereas implicit measures tap into unconscious representations (e.g., Greenwald & Banaji, 1995).

Although these conceptualizations are relatively common in the literature on implicit measures, we believe that it is conceptually more appropriate to classify different measures in terms of whether the to-be-measured psychological attribute 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 psychological attribute 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 psychological attribute on participants' responses is intentional, resource-dependent, conscious, or controllable (cf. Bargh, 1994; Moors & De Houwer, 2006). 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 a psychological attribute influences measurement outcomes rather than the measurement procedure itself (e.g., Petty et al., 2009) or the underlying psychological attribute (e.g., Greenwald & Banaji, 1995). Moreover, whereas the classification of measurement outcomes as implicit or explicit depends on the processes that underlie a given measurement procedure, measurement procedures may be classified as direct or indirect on the basis of their objective structural properties (De Houwer & Moors, 2010). Specifically, a measurement procedure can be described as direct when the measurement outcome is based on participants' self-assessment of the to-be-measured attribute (e.g., when participants' racial attitudes are inferred from their self-reported liking of Black people). Conversely, a measurement procedure can be described as indirect when the measurement outcome is not based on a 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 of attributes other than the to-be-measured attribute (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 procedures 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, descriptions of measures as *implicit* should be interpreted as tentative (for a review of relevant evidence, see De Houwer et al., 2009). We will return to this issue when we discuss caveats regarding the interpretation of implicit measures, in particular the joint contribution of automatic and controlled processes.

An Overview of Basic Paradigms

The use of implicit measures in social and personality psychology has its roots in the mid-1980s when researchers adopted sequential priming tasks from cognitive psychology to study the automatic activation of attitudes (Fazio, Sanbonmatsu, Powell, & Kardes, 1986) and stereotypes (Gaertner & McLaughlin, 1983). These studies provided the foundation for the development of Greenwald, McGhee, and Schwartz's (1998) implicit association test (IAT), which stimulated the current surge in the use of implicit measures. Over the past decade, the toolbox of available measurement instruments has grown substantially through the development of new paradigms and the refinement of existing tasks. In the following sections, we provide an overview of the currently available paradigms, including details on their task structure, reliability, and applicability.[1]

Implicit Association Test

One of the most frequently used paradigms is Greenwald et al.'s (1998) IAT. The IAT consists of two binary categorization tasks that are combined in a manner that is either compatible or incompatible with the to-be-measured psychological attributes. For example, in an IAT to assess preferences for White over Black people, participants are successively presented with positive and negative words and pictures of Black and White faces that have to be classified as positive and negative or as Black and White, respectively. In one of the two critical blocks, the

[1] For "cook-book" style instructions that include procedural information regarding the implementation of different paradigms (e.g., number of trials, presentation times, etc.), we recommend the chapters by Gawronski, Deutsch, and Banse (2011), Teige-Mocigemba, Klauer, and Sherman (2010), and Wentura and Degner (2010).

two categorization tasks are combined in such a way that participants have to respond to positive words and pictures of White faces with one key, and to negative words and pictures of Black faces with another key. In the other critical block, participants have to respond to positive words and pictures of Black faces with one key, and to negative words and pictures of White faces with another key. The basic idea underlying the IAT is that quick and accurate responses are facilitated when the key mapping in the task is compatible with a participant's preference (e.g., Black-negative; White-positive), but impaired when the key mapping is preference-incompatible (e.g., White-negative; Black-positive). Based on this consideration, the mean difference in participants' response latency (or error rates) in the two blocks is typically interpreted as an index of their preference for White over Black people or the other way round, depending on the calculation of the difference score (for details regarding the scoring of IAT data, see Greenwald, Nosek, & Banaji, 2003).

A typical IAT includes a total of five blocks. Two of the five blocks contribute the critical trials for the calculation of the so-called IAT score; the other three blocks include practice trials for the two critical blocks (see Table 1). For example, an IAT to measure preferences for White over Black people would begin with a first practice block in which participants are asked to categorize pictures of Black and White faces as fast and accurately as possible as Black versus White (*initial target-concept discrimination*). In a second practice block, participants are presented with positive and negative words which have to be categorized as pleasant versus unpleasant, again as quickly and accurately as possible (*initial attribute discrimination*). In the third block, the two categorization tasks are combined, such that participants are presented with words and pictures in alternating order, which have to be categorized according to the same key assignments as in the first two blocks (*initial combined* *task*). For example, participants may be asked to press a right-hand key every time they see a positive word or a picture of a White person and a left-hand key every time they see a negative word or a picture of a Black person. As with the first two blocks, participants are asked to respond as quickly and accurately as possible. The fourth block is almost equivalent to the first block, the only difference being that the key assignment for the two target categories is now reversed (*reversed target-concept discrimination*). Finally, the fifth block again combines the two categorization tasks, this time using the key assignments of the second and fourth blocks (*reversed combined task*). In the current example, this would imply that participants have to press a right-hand key every time they see a positive word or a picture of a Black person and a left-hand key every time they see a negative word or a picture of a White person.

The IAT is a very flexible task that can be used to assess almost any type of association between pairs of concepts. For example, by using evaluative attribute dimensions (e.g., pleasant vs. unpleasant) the IAT can be used to assess relative preferences between pairs of objects or categories. Alternatively, the evaluative attribute dimension may be replaced with a semantic dimension to assess semantic associations (e.g., stereotypical associations between Black and White people and the attributes of being athletic versus intelligent). The same flexibility applies to the use of target categories, which may include any pair of objects or categories that can be contrasted in a meaningful manner (e.g., male vs. female). Examples of previous applications include IATs designed to assess prejudice, stereotypes, attitudes toward consumer products, the self-concept, and self-esteem (for an overview, see Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). Another advantage of the IAT is that it typically shows reliability estimates that are comparable to the ones of traditional self-report measures (see Table 2).[2]

[2] Note that reliability estimates tend to be lower for second and subsequent IATs if more than one IAT is administered in the same session (Gawronski et al., 2011).

Nevertheless, the IAT has also been the target of methodological criticism (for a detailed discussion, see Teige-Mocigemba et al., 2010). A very common concern is that the task structure of the IAT is inherently comparative, which undermines its suitability for the assessment of associations to a single target concept or a single attribute. For example, the race IAT can be used to assess relative preferences for Whites over Blacks (or the other way round), but it is not possible to calculate separate indices for evaluations of Blacks and evaluations of Whites (Nosek, Greenwald, & Banaji, 2005). Another concern is that the presentation of compatible and incompatible trials in separate, consecutive blocks can distort measurement scores through various sources of systematic error variance (Teige-Mocigemba et al., 2010). To overcome these shortcomings, researchers have developed a number of procedural variants of the IAT. These variants include modifications that make the IAT amenable for assessing associations of a single target concept (Single Category IAT; Karpinski & Steinman, 2006) or a single attribute (Single Attribute IAT; Penke, Eichstaedt, & Asendorpf, 2006), variants that avoid blocked presentations of compatible and incompatible trials by combining them in a single block (Recoding Free IAT; Rothermund, Teige-Mocigemba, Gast, & Wentura, 2009; Single Block IAT; Teige-Mocigemba, Klauer, & Rothermund, 2008), and an abbreviated variant that is considerably shorter than the standard IAT (Brief IAT; Sriram & Greenwald, 2009). Although the suggested modifications seem quite promising, the currently available evidence is still too scarce to judge whether they retain the functional properties of the standard IAT. The only exception in this regard is the Single Category IAT (Karpinski & Steinman, 2006) which has demonstrated its usefulness in a considerable number of studies.

Evaluative Priming Task

The evaluative priming task employs the basic procedure of sequential priming to assess evaluative responses (Fazio et al., 1986). Toward this end, participants are briefly presented with a prime stimulus (e.g., a Black face) that is followed by a positive or negative target word. In the typical version of the task, participants are asked to quickly determine whether the target word is positive or negative by pressing one of two response keys (*evaluative decision task*). To the extent that the prime stimulus leads to faster responses to positive words (compared to a neutral baseline prime), the prime stimulus is assumed to be associated with positive valence. However, if the prime stimulus facilitates responses to negative words (compared to a neutral baseline prime), it is assumed to be associated with negative valence (for details regarding the calculation of priming scores, see Wittenbrink, 2007). The evaluative priming task can be used to assess evaluative responses to any type of object that can be presented as a prime stimulus in a sequential priming task, and it has been successfully used with prime presentations above the threshold of conscious awareness (i.e., supraliminal presentation) as well as extremely short prime presentations that remain below conscious awareness (i.e., subliminal presentation). Although the standard variant of the task employs evaluative decisions about positive and negative target words, procedural modifications that have been proposed include the pronunciation of positive and negative target words (Bargh, Chaiken, Raymond, & Hymes, 1996) and the naming of positive and negative pictures as target stimuli (Spruyt, Hermans, De Houwer, Vandekerckhove, & Eelen, 2007).

A major advantage of the evaluative priming task is that it allows researchers to calculate separate priming scores for different kinds of associations that are confounded in the IAT (Wittenbrink, 2007). For example, in an evaluative priming task using Black and White faces as primes and positive and negative words as targets, the inclusion of a neutral baseline prime (e.g., a grey square) makes it possible to separately measure (a) positive associations with White faces (defined as the difference in response latencies to positive words following White versus neutral primes), (b) positive associations with Black faces (defined as the difference in response latencies to positive words following Black versus neutral primes), (c) negative associations with White faces (defined as the difference in response latencies to negative words following White versus neutral primes), (d) negative associations with Black faces (defined as the difference in response latencies to negative words following Black versus neutral primes). Although research using the evaluative priming task has provided important insights into the mechanisms underlying attitude-behavior relations (for a review, see Fazio, 2007), a major problem is its low reliability, which rarely exceeds Cronbach's Alpha values of .50 (see Table 2).

Semantic Priming Tasks

A somewhat less common, though very similar paradigm, is Wittenbrink, Judd, and Park's (1997) semantic priming task. The basic procedure of this measure is analogous to Fazio et al.'s (1986) evaluative priming task, the only difference being that (a) participants are presented with meaningful words and meaningless letter strings as target stimuli and (b) participants' task is to determine as quickly as possible whether the letter string is a meaningful word or a meaningless non-word (*lexical decision task*). To the extent that the presentation of a given prime stimulus facilitates quick responses to a meaningful target word (compared to a baseline prime), the prime stimulus is assumed to be associated with the semantic meaning of the target word. For example, in an application of the task to racial stereotypes, Wittenbrink et al. (1997) found facilitated responses to trait words related to the stereotype of African Americans (e.g., athletic, hostile) when participants were primed with the word *Black* before the presentation of the target words. Different than Fazio et al.'s (1986) evaluative priming task, Wittenbrink et al.'s (1997) paradigm is primarily concerned with semantic associations between a target object and a semantic concept (e.g., associations between *self* and *extraverted*) rather than evaluative associations between a target object and its valence (e.g., associations between *self* and *positive*).

Another variant of semantic priming that is procedurally closer to Fazio et al.'s (1986) evaluative priming task includes only meaningful words as target stimuli, with participants being asked to categorize the target words in terms of their semantic rather than evaluative meaning (*semantic decision task*). For example, Banaji and Hardin (1996) presented participants with prime words referring to stereotypically male or stereotypically female occupations (e.g., nurse, doctor), which were followed by male or female pronouns (e.g., he, she). Participants' task was to classify the pronouns as male or female as quickly as possible. Results showed that participants were faster in responding to the male and female pronouns on stereotype-compatible trials (e.g., nurse-she, doctor-he) than stereotype-incompatible trials (e.g., nurse-he, doctor-she). An important difference between the two kinds of priming tasks is that lexical classifications (i.e., word vs. non-word) tend to be substantially faster than evaluative or semantic classifications, which leads to smaller effect sizes in priming tasks using lexical classifications. Because priming effects on lexical classifications are often in the range of only a few milliseconds, they are particularly prone to measurement error (e.g., due to distraction), which poses a challenge to the reliability of semantic priming paradigms using lexical decision tasks.

Affect Misattribution Procedure

A relatively recent, but already very popular measure, is Payne, Cheng, Govorun, and Stewart's (2005) affect misattribution procedure (AMP). In this task, participants are briefly presented with a prime stimulus, which is followed by a brief presentation of a neutral Chinese ideograph. The Chinese ideograph is then replaced by a black-and-white pattern mask, and participants' task is to indicate whether they consider the Chinese ideograph as visually more pleasant or visually less pleasant than the average Chinese ideograph. The typical finding is that the neutral Chinese ideographs tend to be evaluated more favorably when participants have been primed with a positive stimulus than when they have been primed with a negative stimulus. Although responses in the AMP may seem rather easy to control, priming effects in the AMP have been shown to emerge even when participants are instructed not to let the prime stimuli influence the evaluation of the ideographs and even when they were given detailed information about how the prime stimuli may influence their responses on the task (Payne et al., 2005).

As with Fazio et al.'s (1986) evaluative priming task, the AMP can be used to assess evaluative responses toward any kind of stimuli that can be used as primes in the task. Yet, a major advantage of the AMP is that it shows higher effect sizes and reliability estimates that are comparable to the ones of traditional self-report measures (see Table 2). Combined with the procedural advantages of sequential priming (e.g., compatible and incompatible trials being intermixed rather than blocked), these features make the AMP one of the most promising alternatives to the IAT to date. Recently, researchers have also started to investigate the usefulness of the AMP to measure semantic associations, which broadens its potential applicability. For example, using a modified version of the AMP, Gawronski and Ye (2011) asked participants to guess whether the Chinese ideographs referred to a male or a female name. As primes they used words referring to stereotypically male occupations (e.g., doctor) or stereotypically female occupations (e.g., nurse). Results showed that participants were more likely to guess "male" than "female" when they were primed with stereotypically male occupation than when they were primed with a stereotypically female occupation. Moreover, priming scores were systematically related to self-report measures of hostile and benevolent sexism (Glick & Fiske, 1996), but not perceptions of gender discrimination, suggesting that the priming effects resulting from gender-stereotypical occupations are genuinely related to the endorsement of sexist attitudes instead of reflecting mere knowledge of unequal gender distributions in these occupations. An important caveat is that participants may sometimes base their responses on intentional evaluations of the prime stimuli instead of the neutral Chinese ideographs, which can undermine the implicit nature of the task (Bar-Anan & Nosek, in press).

Go/No-go Association Task

Nosek and Banaji's (2001) go/no-go association task (GNAT) was inspired by the basic structure of the IAT with an attempt to make the task amenable for the assessment of associations involving a single target concept (e.g., evaluations of Black people) rather than two target concepts (e.g., relative preferences for White over Black people). Toward this end, participants are asked to show a *go* response to different kinds of target stimuli (e.g., by pressing the space bar) and a *no-go* response to distracter stimuli (i.e., no button press). In one block of the task, the targets include stimuli related to the target concept of interest (e.g., Black faces) and stimuli related to one pole of a given attribute dimension (e.g., positive words); the distracters typically include stimuli related to the other pole of the attribute dimension (e.g., negative words). In a second block, the classification of the particular attribute poles as targets and distracters is reversed (e.g., *go* for Black faces and negative words, and *no-go* for positive words). GNAT trials typically include a response deadline, such that participants are asked to show a *go* response to the targets before the expiration of that deadline (e.g., 600 msec). Error rates are analyzed by means of signal detection theory (Green & Swets, 1966), such that differences in sensitivity scores (*d'*) between the two pairings of *go* trials (e.g., Black-positive vs. Black-negative) are interpreted as an index of associations between the target concept of interest and the respective attributes. Like the IAT, the GNAT is quite flexible in its application, in that targets and distracters may include a variety of concepts and attributes, including evaluative and semantic attributes associated with individuals, groups, and non-social objects (e.g., partner evaluations, self-concept, racial prejudice, consumer preferences). The average reliability of the GNAT is lower compared to the Single Category IAT and the AMP, but still higher compared to the evaluative priming task (see Table 2). A potential problem of the GNAT is that it retains the original block-structure of the IAT, which has been linked to various sources of systematic measurement error (Teige-Mocigemba et al., 2010).

Extrinsic Affective Simon Task

Another procedure that has been designed to resolve procedural limitations of the IAT is the extrinsic affective Simon task (EAST; De Houwer, 2003). In the critical block of the task, participants are presented with target words (e.g., *beer*) that are shown in two different colors (e.g., yellow vs. blue) and with positive and negative words that are shown in white ink color. Participants are instructed to categorize the presented words in terms of their valence when they are shown in white ink color, and to categorize them in terms of their ink color when they are colored. For example, in an EAST designed to measure evaluations of alcoholic beverages, participants may be presented with positive and negative words in white ink (e.g., spider, sunrise) and with names of alcoholic and non-alcoholic beverages (e.g., beer, soda) that are presented in yellow ink on some trials and in blue ink on others. Participants' task is to press a left-hand key when they see a white word of negative valence or a word printed in blue ink and to press a right-hand key when they see a white word of positive valence or a word printed in yellow ink. To the extent that participants show faster (or more accurate) responses to a colored word (e.g., *beer*) when the required response to this word is combined with a positive as compared to a negative response, it is inferred that participants showed a positive response to the object depicted by the colored word. Although the EAST was originally designed as a measure of evaluative responses, a number of studies have demonstrated its applicability to other domains, such as the assessment of self-related associations (e.g., Teige, Schnabel, Banse, & Asendorpf, 2004).

A typical EAST includes a total of three blocks, two practice blocks and one critical block. In the first block, participants are presented with the colored target words, which have to be categorized in terms of their ink color. In the second block, participants are presented with positive and negative words in white ink which have to be categorized in terms of their valence. In the critical third block, the two categorization tasks are combined, such that participants are presented with white and colored words in alternating order. Participants' task is to categorize the words in terms of their valence when they are presented in white ink and to categorize the words in terms of their ink color if they are colored.

Although the EAST resolves many of the procedural limitations of the IAT, its average reliability is less than satisfying (see Table 2). De Houwer and De Bruycker (2007) speculated that the low reliability of the EAST is due to the fact that participants do not have to process the meaning of the colored target stimuli for the color-based responses in the task. To overcome this limitation, they developed a modified version of the EAST in which participants are forced to process the meaning of the target stimuli. The identification EAST (ID-EAST) includes presentations of target and attribute words in upper and lower cases. Positive and negative attribute words have to be categorized in terms of their valence irrespective of whether they are displayed in upper or lower cases; the target words have to be categorized depending on whether they are presented in upper or lower cases. For example, in an ID-EAST designed to measure evaluative responses to beer, participants may be presented with positive and negative words and the word *beer* in either upper or lower cases. Participants' task would be to categorize the attribute words in terms of their valence by pressing one of two response keys. However, for the word *beer*, participants would be instructed to press one response key when it is presented in upper cases and the opposite key when it is presented in lower cases. Because the attribute words are also presented in upper and lower cases, participants are therefore required to process the semantic meaning of the word *beer* before they are able to identify the correct response key. This procedural modification increased the reliability of the EAST, although it is still somewhat lower than the average reliabilities of the IAT and the AMP (see Table 2).

Approach-Avoidance Tasks

Another set of paradigms can be subsumed under the general label *approach-avoidance tasks*. The general assumption underlying these tasks is that positive stimuli facilitate approach reactions and inhibit avoidance reactions, whereas negative stimuli facilitate avoidance reactions and inhibit approach reactions. In the first empirical demonstration of such effects, Solarz (1960) found that participants were faster pulling a lever toward them (approach) in response to positive compared to negative words. Conversely, participants were faster pushing a lever away from them (avoidance) in response to negative compared to positive words. Expanding on these findings, Chen and Bargh (1999) showed that these effects emerge even if the required response is unrelated to the valence of the stimuli (e.g., approach as soon as a word appears on the screen vs. avoid as soon as a word appears on the screen). However, in contrast to earlier interpretations of these effects as being due to direct, inflexible links between motivational orientations and particular motor actions (contraction of flexor muscle = approach; contraction of extensor muscle = avoidance), accumulating evidence suggests that congruency effects in approach-avoidance tasks depend on the evaluative meaning that is assigned to a particular motor action in the task. For example, Eder and Rothermund (2008) found that participants are faster pulling a lever (flexor contraction) in response to positive words and faster pushing a lever (extensor contraction) in response to negative words when the required motor responses were described as pull (i.e., positive meaning attributed to flexor contraction) and push (i.e., negative meaning attributed to extensor contraction). However, these effects were reversed when the same motor responses were described as upward (i.e., positive meaning attributed to extensor contraction) and downward (i.e., negative meaning attributed to flexor contraction). These results indicate that the particular descriptions of the required motor actions can influence the direction of congruency effects in approach-avoidance tasks. Hence, carefully designed instructions with unambiguous response labels are important to avoid misinterpretations of the resulting scores.

Although most studies have used variations of the abovementioned standard paradigm, noteworthy modifications include the Evaluative Movement Assessment (EMA), which includes left-right responses and visual depictions of their respective meanings (Brendl, Markman, & Messner, 2005), and the Implicit Association Procedure (IAP), in which motor movements are used to assess self-related associations (Schnabel, Banse, & Asendorpf, 2006). An important caveat regarding the use of approach-avoidance tasks is that their reliabilities vary substantially as a function of specific task characteristics (see Table 2). For example, reliability estimates are lower for tasks in which stimulus valence is response-irrelevant compared with tasks in which stimulus valence is response-relevant (e.g., Field, Caren, Fernie, & De Houwer, in press; Krieglmeyer & Deutsch, 2010). Moreover, reliability estimates for the EMA tend to be lower for between-participant comparisons of evaluations of the same object compared to within-participant comparisons of preferences for different objects (see Table 2).

Sorting Paired Features Task

A relatively novel procedure is the sorting paired features (SPF) task, which measures four separate associations in a single response block (Bar-Anan, Nosek, & Vianello, 2009). By using combinations of two simultaneously presented stimuli and four (instead of two) response options, the SPF task breaks the four associations that are confounded in the standard IAT into separate indices. For example, in an application of the SPF task to measure racial prejudice, participants may be presented with pairs of faces and words that involve (a) a White face and a positive word, (b) a Black face and a positive word, (c) a White face and a negative word, and (d) a Black face and a negative word. Participants' task is to press one of four response keys depending on the particular stimulus combination. Across four blocks of the task, the response key assignment is set up in a manner such that one stimulus dimension is mapped along a vertical response dimension (e.g., positive-right, negative-left), whereas the other stimulus dimension is mapped onto a horizontal response dimension (e.g., white-up, black-down). These mappings are counterbalanced across the four blocks, such that each pair of categories is mapped once with each of the four response keys over the course the task.

For example, in a first block of the race SPF task, combinations of White faces and positive words may require a response with the upper right key (e.g., O); combinations of White faces and negative words may require a response with the upper left key (e.g., W); combinations Black faces and positive words may require a response with the lower right key (e.g., C); and combinations Black faces and negative words may require a response with the lower left key (e.g., M). The key assignment for one stimulus dimension may then be switched in the second block, such that stimulus combinations with positive words go to the left and stimulus combinations with negative words got to the right, while keeping the response dimension for the target category constant (i.e., White-up, Black-down). The third and fourth block would then use the two valence mappings with the opposite mapping for the target category (i.e., White-down, Black-up). An index of the association between two concepts is calculated by subtracting a participant's mean response latency on all trials with the relevant stimulus combination (e.g., White-positive) from this participant's mean latency on all types of trials (e.g., White-positive; White-negative; Black-positive; Black-negative), divided by the standard deviation of the participant's response latencies on all trials. In their original presentation of the SPF task, Bar-Anan et al. (2009) report internal consistencies (Spearman-Brown) of the four individual scores ranging between .39 and .71, and test-retest reliabilities between .51 and .60. So far, the SPF has been successfully applied to assess race-related associations and associations related to political attitudes (e.g., Democrats vs. Republicans), although additional research seems desirable to corroborate the validity of the task.

Implicit Relational Assessment Procedure

The implicit relational assessment procedure (IRAP) was developed by Barnes-Holmes and colleagues based on their behavior-analytic theory of human language and thinking (for a review, see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010). On each trial of an IRAP, participants are presented with two stimuli on the screen (e.g., a picture of an overweight person and a positive word) and participants are trained to identify as quickly as possible which of two keys they are required to press in response to a particular stimulus combination. The two response options are labeled to refer to different ways in which the two stimuli might be related (e.g., similar vs. opposite). Typically, participants are faster when the correct response is in line with their beliefs about how the two stimuli are related than when the correct response contradicts their beliefs about the relation between the two stimuli (for details regarding the scoring of IRAP data, see Barnes-Holmes et al., 2010).

For example, participants might be presented with a picture of a slim person and the word *good*, a picture of a slim person and the word *bad*, a picture of an overweight person and the word *good*, or a picture of an overweight person and the word *bad*. Depending on the specific picture-word combination, participants are trained to press either a key that indicates that the picture and the word are similar or a key that indicates that the picture and the word are opposite. Specifically, participants may have to press the *similar* key for slim-good and overweight-bad combinations and the *opposite* key for slim-bad and overweight-good combinations in some blocks of the task. In other blocks, participants may have to press the *similar* key for slim-bad and overweight-good combinations and the *opposite* key for slim-good and overweight-bad combinations. Whereas in the first type of blocks, the relational meaning of the required key responses is compatible with the attitudinal beliefs of those participants who like slim people or dislike overweight people, the relational meaning in the second type of blocks is compatible with the attitudinal beliefs of participants who like overweight people or dislike slim people. Although the task structure of the IRAP has some resemblance to the IAT, in that it combines associations between two target objects and two attributes, the IRAP has been shown to be amenable to the measurement of attitudes toward individual objects in a non-relative manner (e.g., Roddy, Stewart, & Barnes-Holmes, 2011).

A unique characteristic of the IRAP is that it is designed to capture propositional beliefs rather than mere associations. Whereas associations link two concepts without specifying the particular way in which these concepts are related, propositional beliefs do specify the way in which concepts are related (Hughes, Barnes-Holmes, & De Houwer, 2011). For example, a person might simultaneously believe that he *is* bad and that he *wants to be* good. An implicit measure that captures mere associations would not be able to differentiate between these two beliefs. Instead, it would show evidence for associative links between *self* and *bad* and, at the same time, between *self* and *good*. In the IRAP, these beliefs can be differentiated by using different types of stimulus combinations (e.g., the expressions *I am* and *I am not* versus the expressions *I want to be* and *I do not want to be* presented in combination with the words *good* and *bad*; Nicholson & Barnes-Holmes, in press). Although the IRAP has been primarily used to measure evaluative beliefs (e.g., *being slim is good*), it is also amenable to the assessment of semantic beliefs (e.g., *I am able to approach spiders*; Nicholson & Barnes-Holmes, in press). Reliability estimates, however, differ substantially between studies, ranging from values as low as .23 to values as high as .81. Although little is known about procedural variables that moderate the reliability of IRAP effects, some studies suggest that the reliability of the IRAP increases with decreases in the response deadline (Barnes-Holmes et al., 2010).

Action Interference Paradigm

The action interference paradigm (AIP) has been developed for research involving very young children, who might get overwhelmed by the complex task requirements of other paradigms. For example, in one application to study the development of gender stereotypes, Banse, Gawronski, Rebetez, Gutt, and Morton (2010) told young children that Santa Claus needs their help in delivering Christmas presents to other children. In a first block of the task, the children were told that the first family had a boy and a girl and that the boy would like to get trucks and the girl would like to get dolls. The children were then shown pictures of trucks and dolls on the screen, and they were asked to give the presents to the kids as quickly as possible by pressing the buttons of a response box that were marked with pictures of the boy and the girl. In a second block, the children were told that they are now at the house of another family, which also had a boy and a girl. However, this boy would like to get dolls and the girl would like to get trucks. The children were then shown the same pictures of trucks and dolls, and they were asked to press the response buttons that were marked with the pictures of another boy and girl. Controlling for various procedural features, Banse et al. (2010) found that children were faster in making stereotype-compatible assignments (i.e., boy-truck, girl-doll) compared to stereotype-incompatible assignments (i.e., boy-doll, girl-truck), which was interpreted as evidence for spontaneous gender stereotyping in children.

Among the paradigms reviewed in the current chapter, the AIP is the most content-specific measure, in that the original variant is particularly designed for the assessment of gender-stereotypes. Nevertheless, it seems possible to modify the AIP for the assessment of other constructs. For example, to assess evaluative responses in the domain of racial prejudice, the gender categories could be replaced by racial categories and the assignment task may involve the distribution of desirable and undesirable objects to Black and White children. However, it is important to point out that applications of the AIP to other domains require a different framing of the task in the instructions. In addition, it is worth noting that the internal consistency of the AIP is relatively low with Cronbach’s Alpha values in the range of .30 and .50 (Gawronski et al., 2011).

How to Choose a Measurement Procedure

Given the large number of available paradigms, a common question by novices is which of them they should choose for their own research. In making this choice, we believe that it is important to consider that measurement procedures are tools and different types of research questions require different kinds of tools. Thus, instead of recommending a particular paradigm as the “best” one, we try to provide some heuristics that might be useful in identifying the most suitable paradigm for a particular research question.

A first issue is that the reviewed paradigms differ considerably with regard to their flexibility. Whereas some tasks have been developed to assess either semantic or evaluative representations, others are more specific in the type of questions for which they can be used (see Table 2). Thus, a first constraint on the choice of a particular measure is whether one’s research question involves semantic or evaluative representations. Similarly, whereas some measures are suitable to measure representations involving individual targets and individual attributes, other paradigms involve comparisons between pairs of targets and pairs of attributes (see Table 2). Thus, to maximize the conceptual overlap between research design and implicit measurement scores, it is important to consider whether one’s research question involves a comparison between pairs of targets and pairs of attributes. For example, the comparative structure of the IAT seems less problematic if one is interested in how gender-stereotypical associations influence impressions of men versus women who engage in stereotype-congruent versus stereotype-incongruent behaviors (e.g., Gawronski, Ehrenberg, Banse, Zukova, & Klauer, 2003). However, the IAT seems less suitable if one is interested in evaluative responses toward a particular target person, which are easier to capture with sequential priming tasks (e.g., Rydell & Gawronski, 2009).

Another important consideration is the wide range of reliability estimates that have been reported for different implicit measures (see Table 2). Whereas some paradigms have consistently shown satisfying reliability estimates across different applications, others suffer from large variations or clearly unsatisfactory psychometric properties. Although concerns about low reliability tend to be more common in personality psychology than in social psychology, low internal consistency can be a problem in both individual difference and experimental designs. On the one hand, low internal consistency can distort the rank order of participants with regard to 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).[3]

Finally, it is important to point out that none of the reviewed measures is perfect, and that any choice between these tasks involves a trade-off between desirable and undesirable features. In addition to structural aspects and reliability estimates, examples of other relevant features include the overall length of the task and its suitability for populations that may be less experienced with computer-based tasks than undergraduate students (e.g., children, older adults). Of course, the relative importance of these features depends one's research question, which makes it difficult to make strong recommendations on a priori grounds. Nevertheless, we hope that our review and the above heuristics are helpful in making informed decisions about which measure might be most useful for a given research question.

[3] There is still no consensus about how estimates of internal consistency should be calculated for implicit measures (cf. Williams & Kaufmann, in press). We recommend to split all critical trials of the task into two test-halves (e.g., first versus second half of all trials of an evaluative priming task) and to calculate two separate measurement scores on the basis of the two test-blocks (e.g., one priming score on the basis of the first half and another one on the basis of the second half). The two scores can then be used to calculate a split-half coefficient or a Cronbach's Alpha value. Note that it is not appropriate to calculate reliability estimates on the basis of the raw data from different types of trials (e.g., mean responses latencies on different kinds of prime-target combinations). Such estimates would reflect the internal consistency of responses on different types of trials (e.g., internal consistency of responses latencies for positive and negative words), not the internal consistency of the implicit measurement score (e.g., internal consistency of evaluative priming effect).

What Can We Learn from Implicit Measures?

The number of studies using implicit measures has grown exponentially over the past decade and their findings have influenced virtually every area of psychology (for an overview, see Gawronski & Payne, 2010). A popular theme in these studies concerns dissociations between explicit and implicit measures. Such dissociations are often interpreted with reference to dual-process theories, in that the different measures are assumed to reflect the operation of distinct mental processes (e.g., Fazio, 2007; Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006; Strack & Deutsch, 2004). In the following sections, we provide a brief overview of the insights that can be gained from these dissociations with regard to the prediction of behavior, the prediction of biases in information processing, and the formation and change of mental representations.

Implicit Measures as a Tool for Predicting Behavior

Two of the first questions that have been asked about implicit measures were: (1) Do implicit measures predict behavior? (2) Do implicit measures add anything to the prediction of behavior over and above explicit measures? Both questions were soon answered positively and research quickly moved beyond zero-order and additive relations to investigate the conditions under which explicit and implicit measures predict behavior (for reviews, see Friese, Hofmann, & Schmitt, 2008; Perugini, Richetin, & Zogmaister, 2010). Inspired by theorizing on attitude-behavior relations, 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; Fazio, Jackson, Dunton & Williams, 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, several recent studies found that a given behavior showed stronger relations to explicit measures compared with implicit measures under conditions of unconstrained processing resources. Yet, the same behavior showed stronger relations to implicit measures than explicit measures when processing resources were depleted. For example, candy consumption under cognitive depletion has been shown to be related to an implicit measure of candy attitudes, but to an explicit measure of candy attitudes under control conditions (e.g., Hofmann, Rauch, & Gawronski, 2007). Similar findings have been obtained for the motivation to engage in elaborate cognitive processing (e.g., Scarabis, Florack, & Gosejohann, 2006). Adopting an individual difference approach, a number of studies have demonstrated that explicit measures are better predictors of behavior for people with a preference for rational thinking styles, whereas implicit measures are better predictors of behavior for people with a preference for intuitive thinking styles (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 defensive behaviors, such as favoring one's ingroup over outgroups and dissonance-related attitude change (e.g., Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003).

Perugini et al. (2010) have provided a conceptual summary of different patterns in the prediction of behavior by implicit measures (see Figure 1). These patterns include: (1) single association patterns in which implicit measures, but not explicit measures, predict the relevant behavior; (2) additive patterns in which implicit and explicit measures jointly predict the relevant behavior; (3) double dissociation patterns in which implicit and explicit measures uniquely predict different kinds of behavior; (4) moderation patterns in which implicit and explicit measures predict the relevant behavior under different conditions; (5) multiplicative patterns in which implicit and explicit measures interactively predict the relevant behavior. All of these patterns have been demonstrated in the literature and they are generally consistent with current dual-process theorizing (e.g., Fazio, 2007; Strack & Deutsch, 2004). However, their boundary conditions are still not well understood, which makes it difficult to predict particular outcomes in an a priori manner. Thus, an important task for future research is to identify the particular conditions under which each of these patterns occurs.

Implicit Measures as a Tool for Predicting Biases in Information Processing

Although double dissociation patterns in the prediction of spontaneous and deliberate behavior are well established in the literature, there are several studies in which implicit measures outperformed explicit measures in the prediction of deliberate judgments, even when there is evidence for the construct validity of the explicit measure. These findings suggest that the representations captured by implicit measures may bias the processing of available information, which can influence deliberate judgments that are based on this information. One example in this regard is the interpretation of ambiguous information. Previous research has shown that contextual cues can distort the interpretation of ambiguous information in a manner that is consistent with the subjective meaning of the contextual cues. For example, in the domain of racial prejudice, the same ambiguous behavior is often interpreted in a positive manner when the actor is White, but negatively when the actor is Black (e.g., Sagar & Schofield, 1980). Although self-reported interpretations of ambiguous behavior may be regarded as an example of deliberate behavior, interpretational biases have been found to reveal stronger relations to implicit measures compared with explicit measures (e.g., Gawronski, Geschke, & Banse, 2003; Hugenberg & Bodenhausen, 2003). This asymmetry has been interpreted as evidence that biases in the interpretation of ambiguous information are driven by the associations that are automatically activated by contextual cues rather than by perceivers' explicitly held beliefs.

Another example of bias in information processing is selective information search. A common finding in the literature on cognitive dissonance is that people selectively expose themselves to information that is consistent with their self-reported attitudes (for a meta-analysis, see Hart et al., 2009). Although this bias has been shown to be reduced for attitudes that are not held with conviction, research using implicit measures has found that even undecided individuals have a tendency to selectively expose themselves to particular information (Galdi, Gawronski, Arcuri, & Friese, in press). Whereas selective exposure in decided participants showed stronger relations to explicit compared with implicit measures, selective exposure in undecided individuals showed stronger relations to implicit compared with explicit measures. Such biases in information processing explain why implicit measures are capable of predicting future choices and decisions that seem highly deliberate, such as voting behavior and other political decisions (e.g., Galdi, Arcuri, & Gawronski, 2008; Payne, Krosnick, Pasek, Lelkes, Akhtar, & Tompson, 2010). For example, undecided voters may selectively expose themselves to information that is consistent with their implicit preference, and this biased set of information may ultimately provide the basis for their deliberate decision to vote for a particular candidate. Thus, to the extent that deliberate choices are based on the information that is available to an individual and the representations captured by implicit measures predict processing biases in the acquisition of this information (e.g., biased interpretation, selective exposure), implicit measures can be expected to make a unique contribution to the prediction of future decisions even when these decisions are highly deliberate.

Implicit Measures as a Tool for Understanding the Formation and Change of Mental Representations

Given the available evidence for dissociations in studies using implicit and explicit measures as predictor variables, an interesting question concerns potential dissociations when implicit and explicit measures are used as dependent variables. 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 measures but not on implicit measures (e.g., Gregg, Seibt, & Banaji, 2006), others showed effects on implicit measures but not explicit measures (e.g., Olson & Fazio, 2006). Yet, other studies found corresponding effects on both explicit and implicit measures (e.g., Whitfield & Jordan, 2009). These inconsistent patterns posed a challenge to traditional theories of attitude formation and change, which inspired the development of novel 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 Gawronski and Bodenhausen's (2006, 2011) associative-propositional evaluation (APE) model, 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 logical consistency. This distinction between associative and propositional processes is further linked to implicit and explicit measures, such that implicit measures are assumed to reflect the behavioral outcome of associative processes, whereas explicit measures are assumed to reflect the behavioral outcome of propositional processes. Drawing on several assumptions about mutual interactions between associative and propositional processes, the APE model has generated a number of novel 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, whereas explicit evaluations were affected only when participants were instructed to introspect on their gut feelings toward the CS (e.g., Gawronski & LeBel, 2008). Although the APE model is just one among several theories that aim to account for dissociations in the antecedents of implicit and explicit measures (e.g., Rydell & McConnell, 2006; Petty et al., 2007), research including both kinds of measures as dependent variables can help provide deeper insights into the formation and change of mental representations.

Some Caveats Regarding the Interpretation of Implicit Measures

As we outlined in the preceding section, implicit measures have provided important insights into the determinants of behavior, biases in information processing, and the formation and change of mental representations. At the same time, there are a number of misconceptions about the type of information implicit measures can provide (Gawronski, 2009). In the current section, we discuss several assumptions that are quite common in the interpretation of implicit measures, yet questionable on the basis of the available evidence.

Conscious vs. Unconscious Representations

A very common assumption is that indirect measurement procedures provide a window into unconscious representations, whereas direct self-report measures reflect conscious representations (e.g., Greenwald & Banaji, 1995). The central idea underlying this assumption is that self-report measures require introspective access to the to-be-measured memory contents, which undermines their suitability for the measurement of memory contents that are unconscious. In contrast, indirect measures do not presuppose introspective access for the measurement of memory contents, which makes them amenable for the assessment of unconscious memory contents. It is important to note that any such claims represent empirical hypotheses that have to be tested as such. To be sure, it is true that indirect measures do not require introspective access for the assessment of memory contents. However, this does not imply that the memory contents that are assessed by these measures are indeed unconscious.

A common argument in support of the unconsciousness claim is that the two types of measures often show rather low correlations. To be sure, if the memory contents captured by an indirect measure are unconscious, their correspondence to self-report measures may be low. However, dissociations between different measures can be due to multiple other factors that do not imply lack of introspective access (for a review, see Hofmann, Gschwendner, Nosek, & Schmitt, 2005). For example, research on prejudice has shown that correlations between self-report and evaluative priming measures are higher when participants' motivation to control prejudiced reactions is low than when it is high (e.g., Dunton & Fazio, 1997), and the same effects have been shown for the IAT (e.g., Gawronski et al., 2003). Moreover, several studies in the domain of attitudes have shown that correlations between the two kinds of measures are higher when participants focus on their gut feelings toward the attitude object (e.g., Gawronski & LeBel, 2008). Taken together, these results suggest that low correspondence between direct measures and indirect measures may not be due to a lack of introspective access to the memory contents captured by the latter type of measure. Instead, their correspondence may be determined by a variety of other factors, such as motivational influences and introspective mindsets during judgment. Thus, interpretations of the two kinds of measures as providing access to conscious versus unconscious representations are difficult to reconcile with the available evidence (for a review, see Gawronski, Hofmann, & Wilbur, 2006).

Old vs. New Representations

Another common assumption is that implicit measures reflect highly stable, old representations that have not been replaced by more recently acquired, new representations. The central idea underlying this assumption is that previously formed representations are not 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 newer representations require controlled processes to be retrieved from memory. With regard to attitudes, for example, it is often assumed that people can have two distinct attitudes toward the same object, an earlier acquired "implicit" attitude that is activated automatically upon encounter of a relevant stimulus, and a more recently acquired "explicit" attitude that requires conscious effort to be retrieved from memory (e.g., Wilson, Lindsey, & Schooler, 2000). This distinction between (old) implicit and (new) explicit representations is often mapped onto particular kinds of measures, such that indirect measures are assumed to tap into earlier, acquired implicit representations, whereas direct self-report measures are claimed to capture more recently acquired, explicit representations (e.g., Rudman, 2004).

As with interpretations in terms of conscious versus unconscious representations, the claim that different kinds of measurement procedures are differentially sensitive to old versus newly formed representations is an empirical hypothesis that needs to be verified with relevant data. Consistent with this claim, there is some evidence showing an impact of recent experiences on explicit, but not implicit, measures (e.g., Gawronski & Strack, 2004; Gregg et al., 2006). However, there is also a large body of research showing the opposite pattern (e.g., Gawronski & LeBel, 2008; Olson & Fazio, 2006). The latter findings are difficult to reconcile with claims that implicit measures tap into highly overlearned, old representations, and that explicit measures reflect recently acquired, new representations.

Dissociations between Explicit and Implicit Measures

Implicit measures become particularly interesting when they show dissociations with explicit measures. However, when interpreting such dissociations it is important to consider a number of potential confounds that may hamper straightforward interpretations of the obtained results. One of the most common confounds is a mismatch in the relevant target object. For example, researchers interested in racial prejudice often use the race IAT as a measure of implicit prejudice and the Modern Racism Scale (McConahay, 1986) as a measure of explicit prejudice. Yet, dissociations between the two measures may not necessarily reflect two discrepant racial attitudes, given that the two measures assess evaluative responses to different kinds of objects. Whereas the race IAT captures evaluative responses to Black and White faces, the Modern Racism Scale measures perceptions of racial discrimination and evaluative responses to antidiscrimination policies. This concern echoes Ajzen and Fishbein’s (1977) correspondence principle in attitude-behavior relations, according to which measures of attitudes and behavior should match with regard to the relevant attitude object. In fact, correlations between implicit and explicit measures are considerably higher when their respective contents match than when their contents mismatch (Hofmann, Gawronski, et al., 2005).

In addition to content-related confounds, dissociations between implicit and explicit measures may also be due to structural task differences (Payne, Burkely, & Stokes, 2008). For example, whereas explicit measures are typically based on participants’ responses on rating scales, most implicit measures are based on response latencies or error rates. Hence, even if the two kinds of measures match with regard to their content (e.g., responses to Black and White faces), dissociations could also be due to differences in the particular aspects of participants’ responses that are used to derive the relevant measurement scores (e.g., ratings vs. latencies). To overcome this limitation, Payne et al. (2008) presented an extended variant of the AMP that increases the structural fit between implicit and explicit measures of the same construct. The basic structure of the task is similar to Payne et al.'s (2005) original AMP. Yet, the measure is administered in two different ways: an indirect variant for the assessment of implicit measurement outcomes and a direct variant for the assessment of explicit measurement outcomes. Whereas in the indirect variant participants are asked to evaluate the neutral Chinese ideographs and to ignore the prime stimuli, the direct variant asks participants to evaluate the prime stimuli and to ignore the Chinese ideographs. Thus, the two tasks provide measurement outcomes that are comparable not only with regard to the relevant target object (e.g., Black and White faces), but also with regard to basic structural features, such as the presentation format and the nature of the relevant responses. Although the two AMP variants showed meaningful differences that are compatible with current theorizing about implicit measures (e.g., the relation between explicit and implicit prejudice scores being moderated by motivation to control prejudiced reactions), their zero-order correlation was substantially higher compared to the low correlation that is typically found when there is a structural misfit between measures.

Reliability also has to be considered when interpreting dissociations between implicit and explicit measures. Whereas some implicit measures consistently show reliability estimates that are comparable to the ones revealed by explicit measures, others suffer from relatively low reliabilities (see Table 2). Thus, dissociations between implicit and explicit measures may sometimes be due to large proportions of measurement error in the implicit measure. Consistent with this concern, Cunningham, Preacher, and Banaji (2001) showed that correlations between implicit and explicit measures are considerably higher when measurement error is taken into account. Because low reliability can also reduce the probability of identifying effects of experimental manipulations (LeBel & Paunonen, 2011), the same concerns apply to studies that compare the relative impact of a given factor on implicit and explicit measures.

Social Desirability, Faking, and Lie Detection

A common assumption in research using implicit measures is that they resolve the well-known problems of social desirability. This assumption is based on the premise that responses on indirect measurement procedures are more difficult to control than responses on direct measurement procedures. However, several issues have to be considered in this context.

First, it is certainly possible to use implicit measures to rule out social desirability as an alternative explanation for effects obtained with explicit measures. To the extent that both measures show the same effects, it seems rather unlikely that the pattern revealed by the explicit measure is driven by social desirability. However, it is important to note that dissociations between implicit and explicit measures do not necessarily reflect an influence of social desirability on the explicit measure. As we argued earlier in this chapter, dissociations between the two kinds of measures can be due to multiple factors over and above social desirability (for a review, see Hofmann, Gschwendner, et al., 2005).

Second, it is important to note that responses on indirect measurement procedures are not entirely immune to faking. Although intentional distortions tend to be more difficult on indirect measures compared with direct measures, there is evidence that responses on indirect measures are susceptible to strategic influences to a certain extent (e.g., Klauer & Teige-Mocigemba, 2007; Steffens, 2004).

Third, even if responses on indirect measurement procedures were entirely immune to faking, this does not mean that their measurement outcomes could be used as a lie detector (e.g., Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008). To illustrate, consider the use of implicit measures of child-sex associations to identify convicted child molesters (e.g., Gray, Brown, MacCulloch, Smith, & Snowden, 2005). Several studies found that implicit measures are indeed successful in discriminating between pedophiles and non-pedophiles. However, child-sex associations may have their roots in a number of factors other than pedophilia, for example when a person has been the target of sexual abuse as a child. Because implicit measures are typically unable to distinguish between different sources of mental representations, claims that implicit measures could be used as a lie detector should be treated with caution.

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 assessment of psychological attributes (for a review, see Schwarz, 1999; Krosnick, Visser, & Lavrakis, this volume). With the development of indirect measurement procedures that do not rely on self-assessments, many researchers expected to gain direct access to people’s “true” personal 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 a review, see Gawronski & Sritharan, 2010). For example, several studies using implicit measures have shown that responses to the same person (e.g., racial minority member) can vary as a function of the context (e.g., family barbeque vs. graffiti wall) in which this person is presented (e.g., Wittenbrink, Judd, & Park, 2001).

Some researchers interpreted these findings as evidence that responses on any type of psychological measure, be it direct or indirect, do not reflect stable trait-like characteristics, but instead are constructed on the spot on the basis of momentarily accessible information (e.g., Schwarz, 2007). Other researchers have argued that contextual influences do not reflect a change in the response to a given object, but a change of the target object itself (e.g., Fazio, 2007). For example, evaluative responses to Michael Jordan may differ depending on whether he is categorized as an athlete or African American, and momentarily available context cues (e.g., basketball court vs. graffiti wall) may influence how he is categorized in the first place (e.g., Mitchell, Nosek, & Banaji, 2003). According to this view, the relevant category representations may be highly stable although contextual factors may influence which category representation becomes relevant in a given context. A third class of models takes a position in-between the two opposing camps, arguing that the same object may activate different patterns of stored associations in memory depending on the context in which the object is encountered (e.g., Gawronski & Bodenhausen, 2006). Drawing on the concept of pattern matching in memory retrieval, which associations are activated in a given situation is assumed to depend on the match between momentary input stimuli and the existing structure of associations in memory. Although it is rather difficult to distinguish among the three accounts on the basis of the currently available evidence, the bottom-line is that implicit measures are highly sensitive to contextual influences, which challenges the idea that implicit measures provide context-independent assessments of people's "true" representations.

"Automatic" Effects of Experimental Manipulations

A common assumption underlying the use of implicit measures is that the to-be-measured psychological attribute influences measurement outcomes automatically (cf. 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 a particular psychological attribute in an automatic fashion. However, such interpretations conflate the impact of the psychological attribute on measurement outcomes with the impact of the experimental manipulation on the psychological attribute (see Figure 2). Although such conflations are relatively common in the literature, they are not justified. After all, the implicitness of a given measure speaks only to the automaticity of the impact of the to-be-measured psychological attribute on the measurement outcome (Path B in Figure 2); it does not speak to the effect of an experimental manipulation on the psychological attribute (Path A in Figure 2).

To illustrate this issue, 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/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. At first glance, one might be tempted to conclude that recalling past behaviors influenced self-representations in an automatic fashion. However, the task of recalling past behaviors was fully conscious, intentional, and controllable, which implies that the experimental manipulation influenced self-representations in a non-automatic fashion. Of course, it is certainly possible that other experimental manipulations may influence self-representations unconsciously, unintentionally, and uncontrollably. This possibility, however, does not allow one to draw the reverse conclusion that implicit measures can be used to demonstrate the automatic nature of an experimental effect. For example, increased levels of self-esteem on the IAT as a result of personal threat do not necessarily indicate that threat defense mechanisms operate automatically (e.g., Rudman, Dohn, & Fairchild, 2007). Such inferences require additional manipulations, for example the use of a cognitive load task to investigate the resource (in)dependency of threat defense.

Absolute vs. Relative Interpretations

Another important issue concerns metric interpretations of implicit measurement scores. 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 continuum if the resulting score is higher than zero (e.g., preference for Whites over Blacks) and a psychological attribute on the other side of a continuum if the score is lower than zero (e.g., preference for Blacks over Whites), with a value of zero being interpreted as a neutral reference point. Although metric interpretations of this kind are rather common in the literature, we consider them as problematic for at least two reasons. Aside from the fact that the metric of any given measure remains ambiguous without proper calibration (Blanton & Jaccard, 2006), contingent features of the employed stimulus materials have been shown to influence both the size and the direction of implicit measurement scores (e.g., Bluemke & Friese, 2006; Scherer & Lambert, 2009). Because it is virtually impossible to quantify the contribution of 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).

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 absolute interpretations of measurement scores of individual participants or samples.

Multiple Processes Underlying Implicit Measures

A final caveat concerns the lack of process purity of implicit measures. It is commonly assumed that implicit measures provide direct access to mental associations that are activated automatically upon the encounter of a relevant stimulus. However, responses on indirect measurement procedures are the product of multiple distinct processes that jointly influence performance on the task. To overcome this problem, researchers have developed mathematical modeling techniques that provide a more fine-grained analysis of data obtained with indirect measurement procedures (e.g., Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007; Payne, 2008). The main advantage of these modeling techniques is that they allow researchers to quantify the individual contributions of multiple distinct processes to task-performance instead of relying on a single measurement score. Because the mathematical underpinnings of these procedures go beyond the scope of this chapter, we limit our discussion to a brief description of Conrey et al.'s (2005) quad-model to illustrate how responses on indirect measures depend on multiple processes.[4]

[4] For more information about the use of mathematical modeling techniques in research using implicit measures, we recommend the introductory overview by Sherman, Klauer, and Allen (2010). For more detailed information about particular modeling procedures, readers may consult Conrey et al. (2005), Klauer et al. (2007), and Payne (2008).

To illustrate the basic assumptions of Conrey et al.'s (2005) quad-model, consider a race IAT with the target categories *Black* versus *White* and the attribute categories *pleasant* versus *unpleasant*. In the combined blocks of this IAT, a black face may elicit a response tendency to press the *Black* key, and, to the extent that negative associations are activated, another response tendency to press the *unpleasant* key. If *Black* and *negative* responses are mapped onto the same key ("compatible" block), responses will be facilitated. If, however, *Black* and *negative* responses are mapped onto different keys ("incompatible" block), the tendency to press the *negative* key has to be inhibited, so that the accurate tendency to press the *Black* key can be executed. Importantly, because the inhibition of the incorrect response tendency requires executive control processes, the impact of race-related associations is confounded with executive control processes in the traditional calculation of IAT scores.

To address this limitation, Conrey et al.'s (2005) quad-model includes statistical parameters for four qualitatively distinct processes: (a) the likelihood that an association-related response tendency is activated (*Association Activation* or *AC*); (b) the likelihood that the correct response to the stimulus can be determined (*Discriminability* or *D*); (c) the likelihood that an automatic association is successfully overcome in favor of the correct response (*Overcoming Bias* or *OB*); and (d) the likelihood that a general response bias (e.g., right-hand bias) drives the response (*Guessing* or *G*).

The proposed interplay of these processes in the quad-model can be depicted as a processing tree that specifies how their joint operation can lead to correct or incorrect responses on compatible and incompatible trials (see Figure 3).[5] To illustrate the logic of this processing tree, consider the presentation of a Black face in the two combined blocks of the race IAT. If the

[5] Note that the quad-model has been designed for indirect measurement procedures that are based on response interference. It is not applicable to tasks that are based on other mechanisms, such as the AMP (Gawronski, Deutsch, LeBel, & Peters, 2008).

Black face activates a prejudicial response tendency (*AC*) and participants are able to identify the correct response (*D*), whether or not the prejudicial response tendency will drive the final response depends on whether participants are able to inhibit the prejudicial response tendency. If they are able to inhibit the prejudicial response tendency (*OB*), they will show the correct response on both compatible and incompatible trials and regardless of whether the required response is on the left or on the right (first row in Figure 3). However, if they are unable to inhibit the prejudicial response tendency (*1 – OB*), they will show the correct response on compatible trials, but an incorrect response on incompatible trials (second row in Figure 3). Moreover, if a prejudicial response tendency is activated (*AC*) and, at the same time, participants are not able to identify the correct response (*1 – D*), the quad-model assumes that the prejudicial response tendency will drive the final response in the task. In this case, participants will show the correct response on compatible trials, but an incorrect response on incompatible trials (third row in Figure 3). If no prejudicial response tendency is activated (*1 – AC*) and participants are able to identify the correct response (*D*), they will show the correct response on both compatible and incompatible trials and regardless of whether the required response is on the left or on the right (fourth row in Figure 3). Finally, if no prejudicial response tendency is activated (*1 – AC*) and participants are unable to identify the correct response (*D*), a guessing bias is assumed to drive the final response. For example, if participants show a bias toward responding with the right key (G), they will show the correct response on both compatible and incompatible trials when the correct response is on the right but not when it is on the left (fifth row in Figure 3). Conversely, if participants show a bias toward responding with the left key (*1 – G*), they will show the correct response when the correct response is on the left but not when it is on the right (sixth row in Figure 3).

The contribution of multiple processes to responses on indirect measurement procedures has important implications for the interpretation of empirical findings. First, when using traditional measurement scores as independent variables (e.g., in studies on the prediction of behavior), the obtained relations to a criterion measure could be driven by an overlap in construct-unrelated processes. A potential example might be the correlation between an implicit measure of food attitudes and impulsive eating behavior, which could be driven by individual differences in the ability to inhibit unwanted response tendencies instead of genuine differences in food attitudes. Second, when using traditional measurement scores as dependent variables (e.g., in studies on attitude change), the measurement scores may be influenced by experimentally induced changes in construct-unrelated processes. For example, increased levels of prejudice on the race IAT after alcohol consumption have been shown to be the result of impaired inhibitory control rather than genuine changes in prejudice levels (Sherman et al., 2008). Such ambiguities can be resolved by means of mathematical modeling techniques, such as the quad-model (Conrey et al., 2005) and other kinds of modeling procedures (e.g., Klauer et al., 2007; Payne, 2008).

Where Are We Going?

Up to now, method-focused research on implicit measures has primarily focused on the development of new measurement procedures and attempts to improve existing paradigms (Payne & Gawronski, 2010). For the decade to come, we believe that the field would benefit from a stronger focus on underlying mechanisms with regard to the measures themselves as well as their capability to predict behavior (see also Nosek, Hawkins, & Frazier, 2011). The groundwork for this focus has already been set by the development of mathematical modeling techniques (e.g., Conrey et al., 2005; Klauer et al., 2007; Payne, 2008), in which measurement outcomes are treated as behaviors that are themselves in need of a psychological explanation rather than as direct reflections of mental constructs (e.g., automatic associations) that can be used to explain behavior. As we will outline in the final sections of this chapter, this perspective has several important implications.

Mechanisms Underlying Behavior Prediction

If the outcomes of psychological measurements are treated as behaviors rather than as direct reflections of mental constructs, one could argue that direct and indirect measurement procedures differ with regard to the processing constraints that they impose during the measurement of behavior. For example, traditional self-report measures of attitudes ask participants to intentionally evaluate the relevant attitude object and the time for this evaluation is typically unlimited. In contrast, there is no requirement to intentionally evaluate the primes in an evaluative priming task and participants are asked to respond as quickly as possible. Yet, when the similarity between the processing constraints of direct and indirect measures is increased (e.g., by imposing a time limit in the self-report measure), the correspondence of their measurement outcomes increases accordingly (e.g., Ranganath, Smith, & Nosek, 2008).

This idea can also be applied to the assessment of behavior. Specifically, one could argue that the predictive validity of implicit and explicit measures of the same construct should depend on the match versus mismatch of the processing constraints that are imposed by the measurement procedure and the processing constraints in the assessment of the to-be-predicted behavior (Fazio, 2007). Importantly, because indirect measurement procedures may differ with regard to the processing constraints in a given task, the same idea applies to the prediction of behavior by means of implicit measures. For example, when an indirect measurement procedure captures responses that are unintentional yet resource-dependent, these responses might be a better predictor of behavior that is unintentional and resource-dependent. The same responses may be less suitable to predict behavior that is intentional, but resource-independent.

Another implication is that predictive relations between psychological measures and observed behavior do not reflect the causal impact of a directly measured mental construct (e.g., automatic association) on the observed behavior, but covariations between two instances of behavior that are presumably driven by the same combination of processes and representations. Hence, successful prediction of behavior depends not only the correspondence of the processing constraints in the measurement procedure and the to-be-predicted behavior, but on the entire set of processes that are involved in the production of the relevant responses. From this perspective, prediction of behavior by means of implicit measures might be improved by considering the conglomerate of processes that influence responses on the measurement procedure as well as the conglomerate of processes that underlie the to-be-predicted behavior. To the extent that indirect measurement procedures can be designed to match the combination of processes that are relevant in real-life situations, behavior prediction by means of implicit measures should be significantly improved.

To illustrate these arguments, consider the four processes proposed by Conrey et al.'s (2005) quad-model: the activation of an association-related response tendency (*AC*), the discrimination of the target stimulus (*D*), the success at overcoming association-related response tendencies in favor of the correct response (*OB*), and the impact of a general response bias (*G*). As we outlined above, all of these processes play a significant role in the IAT (and other measurement procedures based on responses interference; Gawronski et al., 2008). Although this lack of process-purity may be regarded as a methodological flaw because of the implied confounds, it might in fact be functional for the prediction of behavior that is driven by the same combination of processes. For example, when a police officer has to make a split-second decision whether or not to shoot at a Black suspect holding either a gun or a harmless object (Correll, Park, Judd, Wittenbrink, Sadler, & Keesee, 2007), the officer's decision may be influenced by race-related associations between Black people and guns (*AC*), the officer's ability to identify the object held by the suspect (*D*), the officer's success at overcoming an association-related tendency to pull the trigger (*OB*), and a general response tendency to shoot or not to shoot (*G*). Thus, to the extent that performance on an indirect measurement procedure involves all of these processes, its success in predicting decisions to shoot may be higher than when it involves only a subset. Moreover, because the involved processes may be influenced by different situational affordances, the processing constraints in the indirect measure should be designed to match the ones in the to-be-predicted behavior. For example, the discriminability of the object held by the suspect may depend on visual conditions (e.g., daytime vs. nighttime), whereas success at overcoming an association-related tendency to pull the trigger may be reduced under time pressure. Ideally, both processing constraints should be equivalent in the measurement procedure and the to-be predicted behavior. The bottom-line is that any behavioral response is the product of multiple different processes, and this idea applies to both responses on indirect measurement procedures and to-be-predicted behaviors. Hence, the predictive validity of indirect measures should be higher if their underlying processes and processing constraints match those of the to-be-predicted behavior.

Convergence vs. Divergence between Implicit Measures

These considerations may also help to clarify why different kinds of implicit measures sometimes show diverging effects. For example, a number of studies showed different effects of the same experimental manipulation on Fazio et al.'s (1986) evaluative priming task and Payne et al.'s (2005) AMP (e.g., Deutsch & Gawronski, 2009; Gawronski, Cunningham, LeBel, & Deutsch, 2010). From a traditional measurement perspective, these findings might be attributed to the mechanisms underlying different kinds of priming tasks, and these mechanisms may be distinguished from the to-be-measured psychological construct (e.g., automatic associations influence measurement outcomes by means of different task-specific mechanisms; Gawronski et al., 2008). However, if the outcomes of indirect measurement procedures are treated as behavioral responses, the mechanisms underlying a given measurement procedure become essential for understanding the production of the behavioral responses themselves.

To illustrate this argument, consider the task demands in Fazio et al.'s (1986) evaluative priming task and Payne et al.'s (2005) AMP. In the evaluative priming task, participants have to identify the correct response to the target stimulus, and the execution of this response might be facilitated or impaired by a valence-related response tendency that is elicited by the preceding prime (e.g., a response tendency to press the negative key elicited by a negative prime stimulus). From this perspective, priming effects are due to synergistic versus antagonistic effects of the response tendencies that are elicited by the primes and the targets (Gawronski et al., 2008). This situation is quite different in the AMP, in which participants have to disambiguate the evaluative connotation of a neutral target stimulus. There is no correct or incorrect response in the AMP. In other words, whereas the evaluative priming task involves a situation of response conflict, the AMP involves a situation of evaluative disambiguation.

These considerations have important implications for the relation between the two tasks and their capacity in predicting behavior. For example, whereas the evaluative priming task might be a better predictor of behavior that involves the resolution of response conflicts (e.g., inhibition of an association-related tendency to pull the trigger of a gun in response to a Black man holding an object that is identified as harmless), the AMP might be a better predictor of behavior that involves evaluative disambiguation (e.g., tendency to pull the trigger of a gun in response to a Black man holding an ambiguous object). Moreover, the respective processes that are involved in the two kinds of responses may be differentially affected by the same factor, thereby leading to different outcomes of the same experimental manipulation. For example, attention to particular features of an attitude object may eliminate response conflicts resulting from evaluative connotations of irrelevant stimulus features. However, attention to particular features of an attitude object may be less effective in reducing the impact of irrelevant stimulus features on the processes that are involved in evaluative disambiguation. Consistent with these assumptions, Gawronski et al. (2010) found that attention to the category membership of face primes (i.e., age vs. race) moderated priming effects in Fazio et al.'s (1986) evaluative priming task, but not in Payne et al.'s (2005) AMP.

The bottom-line is that responses on indirect measurement procedures are driven by different underlying mechanisms, and these mechanisms play an essential role in the production of the behavioral responses that are assessed by these procedures. Thus, to the extent that the involved mechanisms respond differently to the same situational influence, different measurement procedures may show diverging outcomes even when they are designed to assess the same psychological construct. Moreover, behavior prediction should be enhanced to the extent that the mechanisms underlying a given measurement procedure match the mechanisms underlying the to-be-predicted behavior.

Final Remarks

The validity of self-report measures is often challenged when people are unwilling or unable to provide accurate reports of their own psychological attributes. This concern has been a driving force in the development of indirect measurement procedures. However, the evidence that has been gathered so far suggests a more complex relation between the two types of measures. Although social desirability and introspective limits may play a role for dissociations between explicit and implicit measures, researchers should be careful to avoid the fallacy of reverse inference by interpreting any dissociation in these terms. To avoid premature conclusions, we recommend that theoretical interpretations of measurement dissociations should be supported with relevant empirical data. Such data will not only provide deeper insights into why implicit and explicit measures show different antecedents and correlates; they may also advance the development of new measurement procedures and ultimately the prediction of behavior.

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 explicit and implicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83,* 380-393.

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

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

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

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.

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic activation with a pronunciation task. *Journal of Personality and Social Psychology, 32,* 104-128.

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.

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

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

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.

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.

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.

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology, 92,* 1006-1023.

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., & De Bruycker, E. (2007). 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., & 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: 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.

Deutsch, R., & Gawronski, B. (2009). When the method makes a difference: Antagonistic effects on “automatic evaluations” as a function of task characteristics of the measure. *Journal of Experimental Social Psychology, 45*, 101-114.

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.

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*, 262-281.

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.

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.

Field, M., Caren, R., Fernie, G., & De Houwer, J. (in press). Alcohol approach tendencies in heavy drinkers: Comparison of effects in a Relevant Stimulus-Response Compatibility Task and an approach / avoidance Simon task. *Psychology of Addictive Behaviors*.

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

Gaertner, S. L., & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46,* 23-30.

Galdi, S., Arcuri, L., & Gawronski, B. (2008). Automatic mental associations predict future choices of undecided decision makers. *Science, 321,* 1100-1102.

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

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., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorization influence spontaneous evaluations of multiply categorizable objects? *Cognition and Emotion, 24,* 1008-1025.

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: Guilford Press.

Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008). Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment*, *24*, 218-225.

Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (2003). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology, 39,* 16-30.

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

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

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

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

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: 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. (2011). *What drives priming effects in the affect misattribution procedure? Underlying mechanisms and new applications.* Manuscript submitted for publication.

Glick, P. & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70,* 491-512.

Gray, N. S., Brown, A. S., MacCulloch, M. J., Smith, J., & Snowden, R. J. (2005). An implicit test of the associations between children and sex in pedophiles. *Journal of Abnormal Psychology, 114,* 304-308.

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

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. 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.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*, 555-588.

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 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.

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

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorising in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record, 61*, 465-498.

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., & Teige-Mocigemba, S. (2007). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology, 43,* 648-655.

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.

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.

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: 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.

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

Nicholson, E., & Barnes-Holmes, D. (in press). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear. *The Psychological Record*.

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

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166-180.

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences, 15*, 152-159.

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

Payne, B. K. (2008). What mistakes disclose: A process dissociation approach to automatic and controlled processes in social psychology. *Social and Personality Psychology Compass, 2,* 1073-1092.

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

Payne, B. K., Krosnick, J. A., Pasek, J., Lelkes, Y., Akhtar, O., & Tompson, T. (2010). Implicit and explicit prejudice in the 2008 American presidential election. *Journal of Experimental Social Psychology, 46,* 367-374.

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

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: Psychology Press.

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 measures. *European Journal of Personality, 21,* 529-546.

Roddy, S., Stewart, I., & Barnes-Holmes, D. (2011). Facial reactions reveal that slim is good but fat is not bad: Implicit and explicit measures of body-size bias. *European Journal of Social Psychology, 41,* 688-694.

Rothermund, K., Teige-Mocigemba, S., Gast, A., & Wentura, D. (2009). Minimizing the influence of recoding in the IAT: The Recoding-Free Implicit Association Test (IAT-RF). *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., Dohn, M. C., & Fairchild, K. (2007). Implicit self-esteem compensation: Automatic threat defense. *Journal of Personality and Social Psychology, 93,* 798-813.

Rydell, R. J., & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context dependent automatic attitudes. *Cognition and Emotion, 23*, 1118-1152

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 Psychgology, 44,* 1526-1532.

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

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to accurately assess autobiographical events. *Psychological Science, 19,* 772-780.

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

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. B. (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.

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition, 25*, 638-656.

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: Guilford Press.

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.

Spruyt, A., Hermans, D., De Houwer, J., Vandekerckhove, J., & Eelen, P. (2007). On the predictive validity of indirect attitude measures: Prediction of consumer choice behavior on the basis of affective priming in the picture-picture naming task. *Journal of Experimental Social Psychology, 43*, 599-610.

Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology, 56*, 283-294.

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.

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: The 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 the Implicit Association Test and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 117-139). New York: 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: Guilford Press.

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

Williams, B. J., & Kaufmann, L. M. (in press). Reliability of the go/no-go association task. *Journal of Experimental Social Psychology*.

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

Wittenbrink, B. (2007). Measuring attitudes through priming. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 17-58). New York: Guilford Press.

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.

Table 1. *Task structure of an Implicit Association Test (Greenwald et al., 1998) designed to assess preferences for Whites over Blacks (Race-IAT).*

	Key Assignment			
	Compatible-Incompatible Block Order		Incompatible-Compatible Block Order	
Block	Left Key	Right Key	Left Key	Right Key
1	Negative	Positive	Negative	Positive
2	Black	White	White	Black
3	Negative/Black	Positive/White	Negative/White	Positive/Black
4	White	Black	Black	White
5	Negative/White	Positive/Black	Negative/Black	Positive/White

Table 2. *Overview of measurement procedures, flexibility of applications, and approximate range of reliability estimates.*

Task	Reference	Applications	Targets	Attributes	Reliability
Action Interference Paradigm	Banse et al. (2010)	(content-specific) [a]	pairs	pairs	.30 - .50
Affect Misattribution Procedure	Payne et al. (2005)	evaluative, semantic	individual	pairs	.70 - .90
Approach-Avoidance Task	Chen & Bargh (1999)	evaluative	individual	individual	.00 - .90 [b]
Brief Implicit Association Test	Sriram & Greenwald (2009)	evaluative, semantic	pairs	pairs	.55 - .95
Evaluative Movement Assessment	Brendl et al. (2005)	evaluative	individual	individual	.30 - .80 [c]
Evaluative Priming Task	Fazio et al. (1986)	evaluative	individual	individual	.00 - .55
Extrinsic Affective Simon Task	De Houwer (2003)	evaluative, semantic	individual	individual	.15 - .65
Go/No-go Association Task	Nosek & Banaji (2001)	evaluative, semantic	individual	pairs	.45 - .75
Identification Extrinsic Affective Simon Task	De Houwer & De Bruycker (2007)	evaluative, semantic	individual	pairs	.60 - .70
Implicit Association Procedure	Schnabel et al. (2006)	self-related	individual	pairs	.75 - .85
Implicit Association Test	Greenwald et al. (1998)	evaluative, semantic	pairs	pairs	.70 - .90 [d]
Implicit Relational Assessment Procedure	Barnes-Holmes et al. (2010)	evaluative, semantic	individual	individual	.20 - .80
Recoding Free Implicit Association Test	Rothermund et al. (2009)	evaluative, semantic	pairs	pairs	.55 - .65
Semantic Priming (Lexical Decision Task)	Wittenbrink et al. (1997)	semantic	individual	individual	n/a
Semantic Priming (Semantic Decision Task)	Banaji & Hardin (1996)	semantic	individual	individual	n/a
Single Attribute Implicit Association Test	Penke et al. (2006)	evaluative, semantic	pairs	individual	.70 - .80

Single Block Implicit Association Test	Teige-Mocigemba et al. (2008)	evaluative, semantic	pairs	pairs	.60 - .90
Single Category Implicit Association Test	Karpinski & Hilton (2006)	evaluative, semantic	individual	pairs	.70 - .90
Sorting Paired Features Task	Bar-Anan et al. (2009)	evaluative, semantic	individual	individual	.40-.70

[a] Previous applications are limited to gender-stereotyping, although alternative applications seem possible.

[b] Reliability estimates differ depending on whether approach-avoidance responses involve valence-relevant or valence-irrelevant categorizations, with valence-irrelevant categorizations showing lower reliability estimates (.00-.35) compared to valence-relevant categorizations (.70-.90).

[c] Reliability estimates differ depending on whether the scores involve within-participant comparisons of preferences for different objects or between-participant comparisons of evaluations of the same object, with between-participant comparisons showing lower reliability estimates (.30-.75) compared to within-participant comparisons (~.80).

[d] Reliability estimates tend to be lower (.40 - .60) for second and subsequent IATs if more than one IAT is administered in the same session.

Figure 1. Patterns of behavior prediction by implicit measures. Figure adapted from Perugini, Richetin, and Zogmaister (2010). Reprinted with permission.

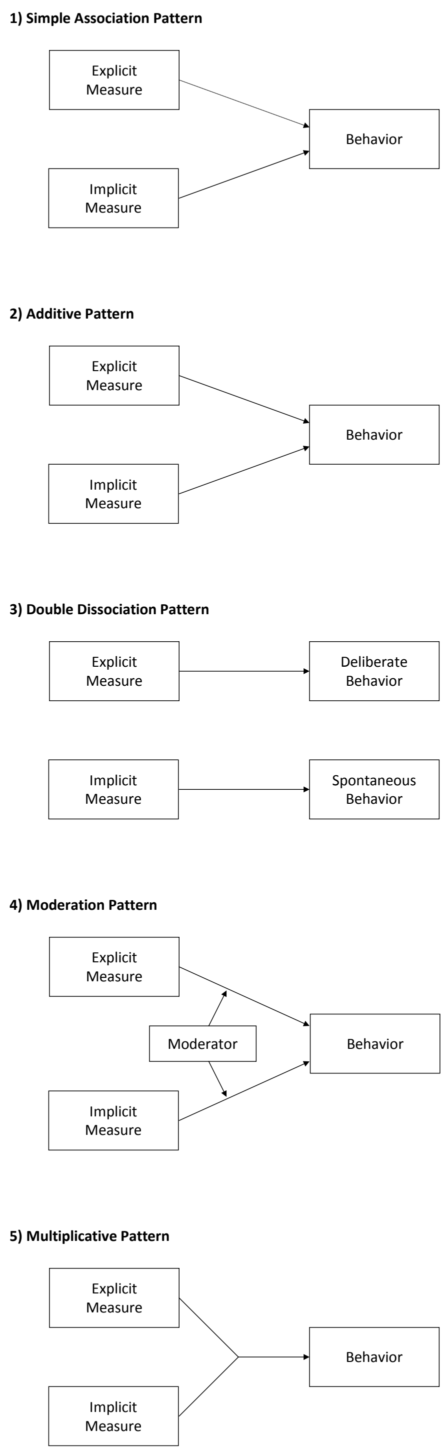

Figure 2. Automatic versus controlled effects of an experimental manipulation on a psychological attribute (Path A) and automatic versus controlled effects of a psychological attribute on measurement outcomes (Path B). Empirical evidence for the automatic nature of Path B does not speak to the automatic nature of Path A.

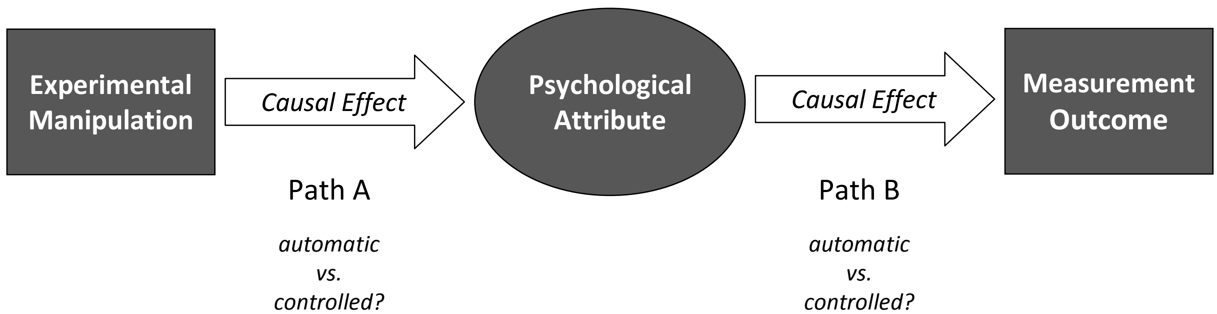

Figure 3. The quad-model of processes underlying correct (+) and incorrect (-) responses on indirect measurement procedures that are based on response interference. Figure adapted from Conrey, Sherman, Gawronski, Hugenberg, and Groom (2005). Reprinted with permission.