Automaticity and Implicit Measures

Bertram Gawronski
University of Texas at Austin

Various areas in psychology are interested in whether specific processes underlying judgments and behavior operate in an automatic or non-automatic fashion. In social psychology, valuable insights can be gained from evidence on whether and how judgments and behavior under suboptimal processing conditions differ from judgments and behavior under optimal processing conditions. In personality psychology, valuable insights can be gained from individual differences in behavioral tendencies under optimal and suboptimal processing conditions. The current chapter provides a method-focused overview of different features of automaticity (i.e., unintentionality, efficiency, uncontrollability, unconsciousness), how these features can be studied empirically, and pragmatic issues in research on automaticity. Expanding on this overview, the chapter describes the procedures of extant implicit measures and the value of implicit measures for studying automatic processes in judgments and behavior. The chapter concludes with a discussion of pragmatic issues in research using implicit measures.

Keywords: automaticity; cognitive control; cognitive resources; consciousness; implicit measures; intentionality; measurement

The rise of dual-process theories in social and personality psychology has fueled interest in whether the processes underlying judgments and behavior operate in an automatic or non-automatic fashion (Gawronski et al., in press). In addition to providing insights into the mental underpinnings of human behavior, evidence addressing this question offers valuable information on the determinants of judgments and behavior under suboptimal processing conditions, and the extent to which judgments and behavior differ under suboptimal versus optimal processing conditions. To provide a companion for rigorous research on automaticity, the current chapter provides a method-focused overview of different features of automaticity, how these features can be studied empirically, and the value of implicit measures in this endeavor.

Four Features of Automaticity

Although some researchers proposed more fine-grained conceptualizations of automaticity (e.g., Moors & De Houwer, 2006), there is consensus about the centrality of four basic criteria of automatic processing: unintentionality, efficiency, uncontrollability, and unconsciousness (Bargh, 1994).

Unintentionality

The unintentionality criterion refers to the question of whether a given process operates in the absence of a goal to start the process. A common approach to study unintentionality is to compare the emergence of a given effect under different task instructions (e.g., Uleman & Moskowitz, 1994). For example, to investigate whether people form social impressions of others unintentionally, participants may be presented with impression-relevant information about a target individual and the effects of this information may be compared across conditions where participants are explicitly instructed to form an impression and conditions where participants are not explicitly instructed to form an impression. Alternatively, effects of impression-relevant information may be compared across conditions where participants are instructed to form an impression and conditions where participants are instructed to pursue a goal that is different from the goal of forming an impression (e.g., memorizing the information). Because participants who are not given explicit goal instructions may spontaneously adopt the instructed goal of the comparison group, manipulations comparing effects across conditions with different goal instructions are often superior in determining the unintentionality of a given effect compared to manipulations involving control conditions without explicit goal instructions.

Efficiency

The efficiency criterion refers to the question of whether a given process operates when the amount of invested or available processing resources is small. The modal approach to study efficiency is to compare the emergence of a given effect under conditions of different processing resources (e.g., Gilbert et al., 1988). For example, to investigate whether information about situational influences on observed behavior shapes dispositional inferences via an efficient or inefficient process, effects of situational information on dispositional inferences may be compared across conditions where the time available to form an impression is either long or short. Alternatively, effects of situational information may be compared across conditions where participants have to perform a concurrent secondary task that requires either a high or a low amount of mental resources (e.g., rehearsing either a short or a long digit-string while forming an impression). Although manipulations of the latter type often use control conditions without any secondary task, it is worth noting that observed differences across conditions that do versus do not involve a secondary task are conceptually ambiguous. On the one hand, it is possible that such differences are driven by the different amount of residual resources in the two conditions. On

To appear in: H. T. Reis, T. West, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (3rd edition). New York: Cambridge University Press.

the other hand, such differences could be driven by the pursuit of a single goal versus two goals. In the former case, the underlying process would qualify as resource-dependent in terms of the efficiency dimension. However, in the latter case, the process would be more appropriately described as goal-dependent, which is not the same as resource-dependent. For these reasons, a superior approach is to use manipulations with secondary tasks that require either a high or a low amount of mental resources (e.g., Yzerbyt et al., 1999). Another issue is that time pressure and secondary tasks can lead to strategic shifts in the allocation of mental resources, and such effects can occur independent of the intended reduction of available resources (Inzlicht et al., 2014). Strategic shifts in resource allocation can render the meaning of time-pressure and secondary-task effects ambiguous when the intended reduction of available resources is compensated by the investment of greater mental effort (e.g., when participants under high time-pressure allocate more resources to a cognitively demanding primary task compared to participants under low time-pressure). In such cases, manipulations of time pressure and secondary tasks may show null effects, not because the focal effect is driven by an efficient process, but because the intended reduction of overall resources is compensated by greater resource allocation. At this time, there are no effective procedures available to tackle this problem.

Uncontrollability

Similar to the unintentionality criterion, the uncontrollability criterion is concerned with goal-dependence. However, while the unintentionality criterion is concerned with the goal to start a process, the uncontrollability criterion is concerned with goals to alter or stop a process. A common approach to study uncontrollability is to investigate the effectiveness of instructions to alter or prevent a given effect (e.g., Gawronski et al., 2014). For example, to investigate whether repeated pairings of a neutral object with positive or negative stimuli influence evaluative responses to the object via an uncontrollable learning mechanism, participants may be instructed to avoid being influenced by the pairings, and effects of the pairings under such conditions may be compared to the effects under conditions where participants do not receive such instructions. However, an important caveat is that evidence for the ineffectiveness of control instructions merely demonstrates that the underlying process is *uncontrolled*, but such evidence is insufficient to conclude that the process is *uncontrollable*. After all, it is possible that the participants in the study used an ineffective control strategy and that a different control strategy would have been more effective in altering or stopping the underlying process. Thus, to provide more compelling evidence for the uncontrollability of a given process, it can be helpful to test the relative effectiveness of different control strategies by giving participants specific instructions on how they are supposed to control a given effect.

Unconsciousness

The unconsciousness criterion refers to the question of whether a given process operates in the absence of conscious awareness. Different from the relative clarity of the previous three criteria, empirical tests of unconsciousness require further specification of what exactly is claimed to be outside of awareness (Gawronski & Bodenhausen, 2012). Is it the relevant stimulus? Is it the thought or feeling that is elicited by the stimulus? Or is it the effect of a thought or feeling on judgments and behavior?

Effects of unconscious (vs. conscious) stimuli can be studied by comparing their effects across conditions when they are presented supraliminally or subliminally (e.g., Stahl et al., 2016). Subliminal presentations typically involve very short presentations of the focal target stimulus and longer presentations of masking stimuli that appear before and/or after the target stimulus (e.g., 15-millisecond presentation of a focal target word, followed by a 500-millisecond presentation of a meaningless letter string). An alternative technique is continuous flash suppression (CFS), which permits unconscious stimulus presentations for longer durations (e.g., Högden et al., 2018). To avoid premature conclusions, studies using either of these approaches should include objective awareness tests to confirm that participants are indeed unable to identify the presented stimuli.

Unconsciousness of thoughts and feelings is often inferred when (a) participants' judgments or behavior suggest a particular underlying thought or feeling and, at the same time, (b) participants do not report having this thought or feeling when they are directly asked. Inferences of unconsciousness from such evidence crucially depend on the validity of participants' self-reports, in that the employed self-report measure has to be sensitive to the relevant thoughts and, at the same time, captures them in a psychometrically reliable manner (Shanks & St. Johns, 1994). If either of the two conditions is not met (e.g., when participants are not honest about their thoughts and feelings, or the measure has poor reliability), inferences of unconsciousness would be premature and potentially flawed.

Research on unconscious effects of thoughts and feelings typically relies on an indirect inferential strategy derived from extant theories of mental control. A shared assumption of these theories is that effective control of a given influence depends on (a) awareness of the influence, (b) motivation to control for the influence, and (c) ability to control for the influence (e.g., Wegener & Petty, 1997; Wilson & Brekke, 1994). To the extent that the influence of a given thought or

feeling on overt responses remains uncontrolled despite high motivation and high ability to control for that influence, it is inferred that participants are most likely unaware of the influence (e.g., Gawronski et al., 2003).

Pragmatic Issues in Research on Automaticity

A few things are worth keeping in mind when studying features of automaticity. First, different features of automaticity do not overlap, in that empirical evidence for one feature provides no information about whether a given process is also characterized by one or more of the other three features (Bargh, 1994). Empirically, this means that each feature of automaticity has to be tested independently. Conceptually, it suggests that the umbrella term *automatic* remains ambiguous if it is not specified in which particular sense a given process is claimed to be automatic: is it claimed to be unintentional, efficient, uncontrollable, or unconscious? Using terminology referring to specific automaticity features is a simple and effective way to avoid conceptual and empirical ambiguities.

Second, it is important to note that statements about automaticity features specify only the conditions under which a given process operates (i.e., *operating conditions*); they do not specify the mental operations by which the process translates environmental inputs into behavioral outputs (i.e., *operating principles*). Without a clear specification of operating principles, research on automaticity features can lead to circular inferences when evidence for a particular automaticity feature is used to infer the operation of a particular mental process (see Gawronski et al., in press). For example, in research guided by dual-system theories, a common inference is that a given effect is caused by "System 1" when the effect is resource-independent and by "System 2" when it is resource-dependent (e.g., Dhar & Gorlin, 2013). However, in the absence of an empirical criterion that signifies the operation of the two systems independent of their presumed resource-dependence, such inferences can be criticized for providing nothing more than a different pair of labels for the more precise terms *resource-independent* and *resource-dependent* (Gawronski, 2013).

Third, it is important to keep in mind that inferences of automatic features are often based on null effects. For example, unintentionality is inferred when a given effect is not qualified by task instructions; efficiency is inferred when a given effect is not qualified by time pressure or secondary tasks; and uncontrollability is inferred when a given effect is not qualified by instructions to alter or prevent the effect. To overcome the well-known obstacles in interpreting null effects, it can be helpful to compare effects of two kinds of stimuli or effects on two kinds of outcomes, one capturing the effect that is presumed to be automatic and the other capturing an effect that is presumed to be non-automatic. In such designs, the effectiveness of manipulations targeting automaticity features can be confirmed via a significant effect on one outcome, even when the manipulation shows a null effect on the other outcome (e.g., Gawronski et al., 2014). To the extent that the two effects or measures are comparable in terms of their basic psychometric properties, such designs provide a stronger basis for inferences of automaticity features from observed null effects.

Fourth, in many studies on automaticity features, it can be important to distinguish between the processes involved in the formation of mental representations and the processes involved in the behavioral expression of mental representations (Gawronski et al., in press). For example, in studies on whether repeated pairings of a neutral object with positive or negative stimuli influence evaluative responses to the object via an uncontrollable learning mechanism, it is important to distinguish between (a) the processes by which the pairings influence a person's mental representation and (b) the processes involved in the behavioral expression of the mental representation. Evidence for the controllability of such effects on self-reported evaluative judgments does not necessarily mean that participants were able to control the impact of the pairings on their mental representations. After all, participants may simply control the overt expression of their mental representations on the self-report measure. In research dealing with ambiguities of this kind, it can be helpful to use measurement instruments that impose processing constraints on the behavioral expression of mental representations (e.g., Gawronski et al., 2014). The following section provides an overview of a particular class of such instruments, commonly referred to as *implicit measures*.

Implicit Measures

Although there is considerable confusion surrounding the conceptual meaning of the term *implicit* (see Corneille & Hütter, 2020), it seems sufficient for the purpose of this chapter to provide a theoretically agnostic list of instruments that are conventionally referred to as implicit measures (see Gawronski et al., 2020). Table 1 provides such a list. A shared feature of the listed instruments is that they capture automatic responses that may differ from the non-automatic responses captured by traditional self-report measures (De Houwer et al., 2009). Based on these considerations, the instruments listed in Table 1 are often referred to as implicit measures, whereas traditional self-report measures are referred to as explicit measures. The following sections provide brief descriptions of the instruments listed in Table 1, followed by a discussion of their range and limits for research on automaticity features. While some features

of automaticity are well captured by implicit measures, other features require alternative approaches. These differences are discussed in more detail after the overview of implicit measures. I will also discuss pragmatic issues for the interpretation of findings obtained with implicit measures.

Measurement Instruments

Implicit Association Test. The most frequently used task among the instruments listed in Table 1 is the Implicit Association Test (IAT; Greenwald et al., 1998). The IAT consists of two binary categorization tasks that are combined in a manner that is either compatible or incompatible with a to-be-measured psychological attribute. For example, in an IAT to assess racial bias in favor of 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 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 rationale 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 idea, the difference in participants' speed and accuracy 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 data treatment and the calculation of IAT scores, see Greenwald et al., 2003). Although the IAT is most prominent for its application to measure racial bias, its range of applicability is extremely broad. For example, by using evaluative attribute dimensions (e.g., pleasant vs. unpleasant) the IAT can be used to assess relative preferences between any pairs of objects or categories (e.g., White vs. Black; men vs. women; Coke vs. Pepsi). Alternatively, the evaluative attribute dimension may be replaced with a specific semantic dimension to assess relative semantic responses (e.g., stereotypical responses linking men and women to the concepts *career* versus *household*). Another advantage of the IAT is that it typically shows estimates of internal consistency that are comparable to the ones of traditional self-report measures (see Table 1).

IAT Variants. Although the IAT is the most frequently used task among the instruments listed in Table 1, it has also been the target of methodological criticism (for a detailed discussion, see Teige-Mocigemba et al., 2010). A common concern about the IAT is that its task structure is inherently comparative, which undermines its suitability to address questions about individual target concepts or individual attributes. For example, the race IAT can be used to assess relative preferences for White over Black people (or the other way round), but it is not possible to calculate separate indices for evaluations of Black people and evaluations of White people (see Nosek et al., 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 (see 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 inferences about individual target concepts (Single Category IAT; Karpinski & Steinman, 2006) or individual attributes (Single Attribute IAT; Penke et al., 2006), variants that avoid blocked presentations of compatible and incompatible trials by combining them in a single block (Recoding Free IAT; Rothermund et al., 2009; Single Block IAT; Teige-Mocigemba et al., 2008), and an abbreviated variant that is considerably shorter than the standard IAT (Brief IAT; Sriram & Greenwald, 2009). Although these modifications address concerns about several suboptimal features of the standard IAT, a downside of the new IAT variants is that they tend to undercut desirable characteristics of the standard IAT (e.g., most IAT variants show lower internal consistencies than the standard IAT; see Table 1). 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., 1995). 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 elicit a positive response. However, if the prime stimulus facilitates responses to negative words (compared to a neutral baseline prime), it is assumed to elicit a negative response (for details regarding data treatment and the calculation of priming scores, see Koppehele-Gossel et al., 2020). 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 supraliminal and subliminal prime presentations. 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 et al., 1996) and the naming of positive and negative pictures as target stimuli (Spruyt et al., 2007). 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 of the task is its low internal consistency, which rarely exceeds estimates of .50 (see Table 1).

Semantic Priming with Lexical-Decision Task. A somewhat less common, though very similar paradigm, is semantic priming with a lexical-decision task (Wittenbrink et al., 1997). The basic procedure of this measure is analogous to the 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. 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 measure 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 from the measurement of evaluative responses in the evaluative priming task, semantic priming with a lexical-decision task is primarily concerned with semantic responses (e.g., responses linking *self* and *extraverted*) rather than evaluative responses (e.g., responses linking *self* and *positive*).

Semantic Priming with Semantic-Decision Task. Another variant of semantic priming that is procedurally closer to the 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. 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 versions of semantic priming 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 using lexical decision tasks.

Affect Misattribution Procedure. Another frequently used instrument is the affect misattribution procedure (AMP; Payne et al., 2005). 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. As with the 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 larger effect sizes and estimates of internal consistency that are comparable to the ones of traditional self-report measures (see Table 1). Combined with the procedural advantages of sequential priming (e.g., no need for blocked presentations of compatible and incompatible trials), these features make the AMP one of the most valuable alternatives to the IAT. However, 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, 2012; but see Payne et al., 2013).

Semantic Misattribution Procedure. Although the AMP has originally been designed to capture evaluative responses, some studies have used a modified version of the task that is amenable for the measurement of semantic responses (e.g., Imhoff et al., 2011). This modified version is commonly referred to as the semantic misattribution procedure (SMP). The procedure of the SMP can be illustrated with a study by Ye and Gawronski (2018), who used the task to measure gender stereotypes. Participants were asked to guess whether the Chinese ideographs referred to a

male or a female name. The primes in the task were 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 a stereotypically male occupation than when they were primed with a stereotypically female occupation. Beyond gender stereotypes, examples of SMP applications include the measurement of sexual preferences (Imhoff et al., 2011) and personality self-concepts (Sava et al., 2012). An extension of SMP is the stereotype misperception task, which has been designed to disentangle stereotype activation and stereotype application within a single task (Krieglmeyer & Sherman, 2012). Overall, the psychometric properties of the SMP are somewhat weaker compared to the AMP, but still in a range that makes the task a valuable addition to the toolbox of available instruments.

Go/No-go Association Task. The go/no-go association task (GNAT; Nosek & Banaji, 2001) was inspired by the basic structure of the IAT with an attempt to make the task amenable for the assessment of responses toward 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 milliseconds). 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 responses to the target concept of interest in terms of 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 of individuals, groups, and non-social objects (e.g., partner evaluations, self-concept, racial prejudice, consumer preferences). Estimates of internal consistency reported for the GNAT are lower compared to the Single Category IAT and the AMP, but still higher compared to the evaluative priming task (see Table 1). 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, 2003a). 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. Participants are instructed to categorize the presented words in terms of their valence when they are shown in white, and to categorize them in terms of their color when they are colored. For example, in an EAST designed to measure evaluative responses to alcoholic beverages, participants may be presented with positive and negative words in white (e.g., spider, sunrise) and with names of alcoholic and non-alcoholic beverages (e.g., beer, soda) that are presented in yellow on some trials and in blue 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 and to press a right-hand key when they see a white word of positive valence or a word printed in yellow. 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. A variant of the EAST is the Identification-EAST (ID-EAST), which includes presentations of target and attribute words in upper and lower cases instead of different colors (De Houwer & De Bruycker, 2007). Participants' task to is categorize positive and negative attribute words 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. This procedural modification helped to increase the relatively low internal consistency of the original EAST, although estimates obtained for the ID-EAST are still lower than the average estimates for the IAT and the AMP (see Table 1). Although the EAST was originally designed as a measure of evaluative responses, some studies have demonstrated its applicability to other domains, such as the assessment of semantic responses to self-related stimuli (e.g., Teige et al., 2004).

Approach-Avoidance Tasks. Another group of instruments can be subsumed under the label *approach-avoidance tasks*. The rationale 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 regardless of the word's valence). 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 et al., 2005), and the Implicit Association Procedure (IAP), in which motor movements are used to assess responses to self-related stimuli (Schnabel et al., 2006). An important caveat regarding the use of approach-avoidance tasks is that their internal consistency varies substantially as a function of specific task characteristics (see Table 1). For example, estimates of internal consistency are lower for tasks in which stimulus valence is response-irrelevant compared with tasks in which stimulus valence is response-relevant (Krieglmeyer & Deutsch, 2010). Moreover, estimates of internal consistency 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 1).

Sorting Paired Features Task. A major advantage of the sorting paired features (SPF) task is that it can capture four separate response dimensions in a single response block (Bar-Anan et al., 2009). By using combinations of two simultaneously presented stimuli and four (instead of two) response options, the SPF task breaks the four response dimensions that are confounded in the standard IAT (e.g., Black-positive, Black-negative, White-positive, White-negative) into separate indices. For example, in an application of the SPF task to measure racial bias, 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). Responses are analyzed by subtracting a participant's mean response latency on all trials with a 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. The SPF has been successfully applied to assess evaluative responses to various targets, including racial and

political groups (e.g., Democrats vs. Republicans). However, estimates of internal consistencies reported in these studies tend to be lower compared to the estimates obtained for the IAT and the AMP (see Table 1).

Action Interference Paradigm. The action interference paradigm (AIP) has been developed for research with very young children, who might get overwhelmed by the complex task requirements of other instruments. In one application of the AIP to study the development of gender stereotypes, Banse et al. (2010) told young children that Santa Claus needs their help in delivering Christmas gifts 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. Although the AIP has been specifically designed for the assessment of gender-stereotypes, it seems possible to modify the task for the assessment of other constructs. For example, to assess evaluative responses to racial groups, the assignment task may involve the distribution of desirable and undesirable objects to Black and White children. However, it is important to note that applications of the AIP to other domains would 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 overall (see Table 1).

Implicit Relational Assessment Procedure. A unique characteristic of the IRAP is that it has been designed to capture relational rather than associative responses. Whereas associative responses link two concepts without specifying the particular way in which these concepts are related (e.g., Aspirin-headache), relational responses are sensitive to the way in which concepts are related (e.g., Aspirin relieves headaches; see Hughes et al., 2011). For example, while one person might hold the belief *I am good,* another person might hold the belief *I want to be good*. An implicit measure that captures mere associations between *self* and *good* would not be able to differentiate between the two cases. In the IRAP, the two cases can be separated 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*). To this end, participants are presented with two stimuli on the screen 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). Although the IRAP has been primarily used to measure evaluative beliefs, it is also amenable to the assessment of semantic beliefs. Estimates of internal consistency reported for the IRAP differ substantially across studies (see Table 1). Although little is known about procedural factors that are responsible for the wide range of estimates, some studies suggest that the internal consistency of the IRAP is higher with shorter response deadlines (Barnes-Holmes et al., 2010).

Relational Responding Task. Although the IRAP has the advantage of capturing the nature of perceived relations between objects (e.g., *I am good* vs. *I want to be good*), two notable limitations of the task are the high complexity of the instructions and the overall length of the task. To address these limitations, De Houwer et al. (2015) developed the Relational Responding Task (RRT), which captures relational responses in a simpler and more efficient way. In the RRT, participants are presented with statements about the relation between objects (e.g., *women are smarter than men*) and asked to indicate if the statement is true or false. In one block of the task, participants are asked to respond to the statements as if they held a particular belief (e.g., respond as if they believed that women are smarter than men). In another block of the task, participants are asked to respond to the statements as if they held the opposite belief (e.g., respond as if they believed that men are smarter than women). The difference in the speed and accuracy of responses in the two blocks is interpreted as an implicit measure of the extent to which participants hold the focal belief. This conclusion is based on the finding that responses are faster and more accurate when the required response aligns with participants' personal beliefs. Although the RRT is extremely flexible in terms of its application, estimates of internal consistency reported for the task tend to be only moderate (see Table 1).

Weapon Identification and Shooter Tasks. Several of instruments reviewed thus far are amenable for the measurement of both evaluative and semantic responses. Other measures can be used to measure one type of response, but not the other. Nevertheless, all of them are relatively flexible in that they can be used to measure a broad range of responses to various kinds of stimuli. Some implicit measures are less flexible in their range of applications, in that they have been designed to measure responses that are highly content-specific. Two examples are the weapon identification task (Payne 2001) and the shooter task (Correll et al., 2002), which have been designed to measure racial bias in weapon identification and decisions to shoot (for a review, see Payne & Correll, 2020). In the weapon identification task, participants are briefly presented with either a Black or a White face prime, which is immediately followed by a target picture showing either a gun or a harmless object. The target picture is quickly replaced by a black-and-white pattern mask, and participants' task is to indicate whether the target picture showed a gun or a harmless object. The common result is that harmless objects are more frequently misidentified as guns when the face prime was Black than when it was White, whereas guns are more frequently misidentified as harmless objects when the face prime was White than when it was Black. In the shooter task, participants are presented with images of various scenes (e.g., inner-city street corner) in which either a White or Black person is holding either a gun or a harmless object. Participants' task is to press a "shoot" button when the target person holds a gun and a "no-shoot" button whenever the target person holds a harmless object. The common finding is that participants are more likely to respond "shoot" for unarmed Black targets than unarmed White targets. Because estimates of internal consistency have not been reported is studies using the two tasks, it is difficult to gauge their overall reliability.

Automaticity Features

A central assumption in research using implicit measures is that they capture automatic responses, whereas explicit measures capture non-automatic responses. However, the lack of overlap between automaticity features suggests that a more nuanced analysis is warranted for each individual feature (see De Houwer et al., 2009). Moreover, because different implicit measures are based on different underlying mechanisms, the degree to which a given automaticity feature is captured by implicit measures can vary across tasks (see De Houwer et al., 2009).

Regarding the unintentionality criterion, a central characteristic of implicit measures is that their unobtrusive task structure permits the measurement of unintentional responses. For example, implicit measures of evaluation can be said to capture evaluative responses that are elicited by a given object in the absence of a goal to evaluate the object. Explicit measures of evaluation are different in this regard, because they depend on respondents' goal to evaluate the focal object. Although the capacity to capture unintentional responses is shared by all instruments listed in Table 1, it is worth noting that this capacity does not permit the reverse inference that responses on implicit measures are unaffected by intentional processes. There is an abundance of research showing that intentional processes (e.g., intentional retrieval of specific memories) can influence responses on implicit measures (e.g., Blair et al., 2001; Peters & Gawronski, 2011), which poses a challenge to the idea that responses on implicit measures can be interpreted as uncontaminated indicators of unintentional processes.

Regarding the efficiency criterion, most of the instruments in Table 1 require fast responses, which is different from the typical lack of time constraints on traditional explicit measures. Two exceptions are the AMP and the SMP, which are based on judgments of ambiguous target stimuli rather than response latencies. A valuable feature of these two instruments is that they permit direct investigations of how time influences responses in the task. Because longer delays between the onset of the prime stimulus and the onset of the target stimulus (i.e., stimulus onset asynchrony, or SOA) do not affect the sensitivity of the AMP and the SMP, it is possible to experimentally manipulate SOAs to investigate how time influences responses in the two tasks (e.g., Hofmann et al., 2009). Note that such manipulations would not provide meaningful effects for sequential priming tasks based on response latencies (e.g., evaluative priming task), because long SOAs generally eliminate priming effects in these tasks (e.g., Hermans et al., 2001).

Regarding the uncontrollability criterion, it is commonly assumed that implicit measures are less susceptible to influences of strategic control than explicit measures. Although this assumption is generally correct in a relative sense, it is not the case that responses on implicit measures are immune to influences of strategic control. Of the instruments listed in Table 1, the ones that seem most susceptible to strategic influences are the AMP and the SMP (e.g., Teige-Mocigemba et al., 2016), followed by the IAT and its variants (e.g., Röhner et al., 2013). The least susceptible tasks are the EPT and the two variants of semantic priming, but even those have been found to be affected by strategic influences under certain conditions (e.g., Klauer & Teige-Mocigemba, 2007).

Although implicit measures are often described as capturing thoughts and feelings that people are not aware of, the available evidence does not support such claims. For example, research using the IAT has found that people can predict their IAT scores prior to

completing the task with a high degree of accuracy (e.g., Hahn et al., 2014), which is difficult to reconcile with the idea that the IAT captures thoughts and feelings that people are not aware of. Moreover, surprise reactions in response to IAT feedback can be explained by the fact that participants and IAT researchers use different arbitrary metrics to label IAT outcomes, which poses further challenges to strong claims of unconsciousness (Gawronski, 2019). Although some instruments capture responses to stimuli without participants being aware of what is being measured and how (e.g., semantic priming with subliminal prime presentations; see Wittenbrink et al., 1997), unawareness of the measurement process should not be confused with unawareness of the thoughts and feelings underlying responses on implicit measures (see Gawronski & Bodenhausen, 2012).

Pragmatic Issues in Research Using Implicit Measures

Several issues are worth keeping in mind when using implicit measures to study automaticity features in the behavioral expression of mental representations. First, the currently available instruments differ considerably in terms of their internal consistency (see Table 1). The only tasks that consistently show high internal consistency are the IAT and the AMP. Some tasks have shown moderate estimates of internal consistency that may be deemed acceptable, yet suboptimal from a psychometric view. Others have shown internal consistencies that are clearly unsatisfactory.

Second, even instruments with high internal consistency have shown comparatively low test-retest stabilities with correlations in the range of .40 to .50 (see Gawronski et al., 2017; Greenwald & Lai, 2020). The combination of high internal consistency and low test-retest stability suggests that a considerable portion of variance in responses on implicit measures reflects transient states rather than stable traits. This conclusion is consistent with the findings of several studies that have used latent state-trait analyses to decompose the roles of situation-related and person-related factors in implicit measures (see Klauer & Becker, in press).

Third, when comparing responses on implicit and explicit measures, it is important to avoid confounds between type of measure and the specific materials in the two kinds of measures (Gawronski, 2019). For example, in studies using the IAT to measure self-concepts of personality, researchers have typically been very careful to avoid such confounds by using identical stimuli in the IAT and the self-report measure (e.g., Asendorpf et al., 2002; Peters & Gawronski, 2011). In contrast, confounds between type of measure and stimulus materials are very common in research on prejudice and stereotyping, where participants are often presented with faces of group members in the implicit measure but not in the explicit measure (e.g., Dovidio et al., 2002; Fazio et al., 1995). Such confounds render interpretations of dissociations ambiguous, because they could be driven either by the type of measure or by differences in the stimulus materials. The most stringent way to avoid any such confounds has been proposed by Payne et al. (2008) who used two variants of the AMP: one in which participants were asked to rate the Chinese ideographs and ignore the primes, and one in which participants were asked to rate the primes and ignore the Chinese ideographs. Differences in responses captured by the two AMP variants unambiguously reflect the difference between intentional and unintentional responses, because there are no confounds in terms of stimulus materials or procedural aspects.

Fourth, it is important to keep in mind that there are no process-pure measures. Even responses on implicit measures reflect a mixture of multiple distinct processes, which prohibits direct inferences of underlying mental processes from observed responses. To overcome these issues, researchers have developed various computational models that disentangle the contributions of multiple distinct processes to responses on implicit measures (for reviews, see Calanchini, 2020; Sherman et al., 2010). Precursors of this approach are the use of process dissociation (Jacoby, 1991) to analyze responses in the weapon identification task (Payne, 2001) and the use of signal detection theory (Green & Swets, 1966) to analyze responses in the GNAT (Nosek & Banaji, 2001) and the shooter task (Correll et al., 2002). A prominent extension of these data analytic procedures is the quad-model (Conrey et al., 2005), a multinomial processing tree (MPT) model that quantifies the contributions of four qualitatively distinct processes to IAT performance: activation of an association (*AC*), detection of the correct response required by the task (*D*), success at overcoming associative bias (*OB*), and guessing (*G*). By permitting more fine-grained analyses of the processes underlying responses on implicit measures, computational models are valuable tools to avoid incorrect conclusions from findings obtained with implicit measures (for details on such modeling procedures, see Klauer, this volume)

Fifth, implicit measures constrain processing conditions only during the expression of mental representations. Hence, although dissociations between implicit and explicit measures can provide valuable information about the role of automatic processes during the expression of mental representations, such dissociations do not have any direct implications for the role of automatic processes in the formation of mental representations. Although research on the latter question can sometimes benefit from comparisons of implicit and explicit measures, such research requires direct manipulations of processing conditions during

the formation of mental representations, such as the ones described in the first part of this chapter.

Sixth, there are ongoing debates about the extent to which implicit measures are valuable for the prediction of behavior (for meta-analyses, see Cameron et al., 2012; Greenwald et al., 2009; Kurdi et al., 2019; Oswald et al., 2013). From a purely pragmatic view, three issues are important to consider in research using implicit measures as predictors of behavior. First, because the internal consistency of a given measure sets an upper limit for its relation with another measure, it seems unrealistic to expect strong predictive relations for implicit measures with low estimates of internal consistency (see Table 1). Second, because even implicit measures with high internal consistencies have shown relatively low stability over time (see Gawronski et al., 2017; Greenwald & Lai, 2020), it seems unrealistic to expect strong predictive relations when there is a delay between the completion of the implicit measure and the measurement of the to-be-predicted behavior. Third, the principle of measurement correspondence suggests that implicit measures should show stronger predictive relations when the processing conditions of the to-be-predicted behavior converge to processing conditions imposed by implicit measures (e.g., unintentional behavior under conditions of low elaboration). These issues should be taken into account when planning and evaluating studies that use implicit measures to predict behavior.

Seventh, it seems important to consider that different implicit measures are based on different underlying mechanisms (for a detailed analysis, see De Houwer, 2003b). Although many instruments rely on response compatibility as a mechanism, some instruments are based on other mechanisms including misattribution or stimulus compatibility (see Table 1). These differences are important for two reasons. First, in research using implicit measures as predictors of behavior, stronger predictive relations can be expected when the mechanisms underlying responses on implicit measures are similar to the processes underlying the to-be-predicted behavior (Gawronski et al., 2020). Second, in research using responses on implicit measures as dependent variables in experimental studies, the observed outcomes can be distorted when the experimental manipulation influences aspects of the mechanism underlying the measurement process instead of the to-be-measured psychological construct (Gawronski et al., 2008). In the most extreme cases, influences on the measurement process can lead to opposite effects on implicit measures that are supposed to capture the same psychological construct (e.g., Deutsch & Gawronski, 2009; Gawronski & Bodenhausen, 2005). To avoid potential misinterpretations arising from these issues, it seems prudent to replicate effects observed on one implicit measure with another implicit measure that is based on a different underlying mechanism.

A final issue concerns the metric of implicit measurement scores. The scores obtained with implicit measures are often used to draw diagnostic inferences about individuals (e.g., participant X shows a strong preference for Whites over Blacks) or populations (e.g., 70% of the sample showed a strong preference for Whites over Blacks). Although such inferences are very common, they are problematic for at least two reasons. First, both the size and the direction of implicit measurement scores are affected by incidental features of the stimuli (e.g., Bluemke & Friese, 2006; Scherer & Lambert, 2009). Second, random differences between the stimuli within a given task tend to inflate the size of implicit measurement scores when random effects of stimulus sampling are not statistically controlled (Wolsiefer et al., 2017). Both issues render absolute interpretations of implicit measurement scores problematic, which are required for diagnostic inferences of the kind described above. Yet, it is worth noting that most research questions in social and personality psychology do not require absolute interpretations, but instead are based on relative differences between measurement scores. The latter applies to 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 designs in which measurement scores are compared across different individuals (e.g., participants with higher scores on an implicit measure are more likely to show a particular behavior). Because stimulus-related effects are relevant only for diagnostic inferences but not for relative differences between measurement scores, stimulus-related effects do not undermine the usefulness of implicit measures for many of the questions addressed by social and personality psychologists.

Conclusion

Although the interest in automaticity is closely linked to the popularity of dual-process theories (Gawronski et al., in press), questions about the contribution of automatic and non-automatic processes go far beyond the realm of dual-process theories. In social psychology, valuable insights can be gained from evidence on whether and how judgments and behavior under suboptimal processing conditions differ from judgments and behavior under optimal processing conditions. Similarly, in personality psychology, valuable insights can be gained from individual differences in behavioral tendencies under optimal and suboptimal processing conditions. The current chapter provides a method-oriented overview of extant approaches to studying automatic aspects of the

processes underlying judgments and behavior, and the value of implicit measures in this endeavor. I hope that this chapter serves as a valuable starting point for anyone who is interested in studying the automatic underpinnings of human behavior.

References

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. (2012). Reporting intentional rating of the primes predicts priming effects in the Affective Misattribution Procedure. *Personality and Social Psychology Bulletin, 38,* 1194-1208.

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

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

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.

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

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.

Calanchini, J. (2020). How multinomial processing trees have advanced, and can continue to advance, research using implicit measures. *Social Cognition, 38,* s165-s186

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

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

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.

Corneille, O., & Hütter, M. (2020). Implicit? What do you mean? A comprehensive review of the delusive implicitness construct in attitude research. *Personality and Social Psychology Review, 24,* 212-232.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate threatening individuals. *Journal of Personality and Social Psychology, 83,* 1314-1329.

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

De Houwer, J. (2003b). A structural analysis of indirect measures of attitudes. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219-244). Mahwah, NJ: Erlbaum.

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

De Houwer, J., 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.

Dhar, R., & Gorlin, M. (2013). A dual-system framework to understand preference construction

processes in choice. *Journal of Consumer Psychology, 23,* 528-542.

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

Eder, A. B., & Rothermund, K. (2008). When do motor behaviors (mis)match affective stimuli? An evaluative coding view of approach and avoidance reactions. *Journal of Experimental Psychology: General, 137,* 262-281.

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.

Gawronski, B. (2013). What should we expect from a dual-process theory of preference construction in choice? *Journal of Consumer Psychology, 23,* 556-560.

Gawronski, B. (2019). Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science, 14,* 574-595.

Gawronski, B., Balas, R., & Creighton, L. A. (2014). Can the formation of conditioned attitudes be intentionally controlled? *Personality and Social Psychology Bulletin, 40,* 419-432.

Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation versus retrieval experiences. *Journal of Personality and Social Psychology, 89,* 672-685.

Gawronski, B., & Bodenhausen, G. V. (2012). Self-insight from a dual-process perspective. In S. Vazire & T. D. Wilson (Eds.), *Handbook of self-knowledge* (pp. 22-38). New York: Guilford Press.

Gawronski, B., & Creighton, L. A. (2013). Dual-process theories. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition* (pp. 282-312). New York: Oxford University Press.

Gawronski, B., De Houwer, J., & Sherman, J. W. (2020). Twenty-five years of research using implicit measures. *Social Cognition, 38,* s1-s25.

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., 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., Luke, D. M., & Creighton, L. A. (in press). Dual-process theories. In D. E. Carlston, K. Johnson, & K. Hugenberg (Eds.), *The Oxford handbook of social cognition* (2nd edition). New York: Oxford University Press.

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

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology, 54,* 733-740.

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

Greenwald, A G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology, 71,* 419-445.

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.

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

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General, 143,* 1369-1392.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition and Emotion, 15*, 143-165.

Högden, F., Hütter, M., & Unkelbach, C. (2018). Does evaluative conditioning depend on awareness? Evidence from a continuous flash suppression paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 1641-1657.

Hofmann, W., Friese, M., & Roefs, A. (2009). Three ways to resist temptation: The independent contributions of executive attention, inhibitory control, and affect regulation to the impulse control

of eating behavior. *Journal of Experimental Social Psychology, 45,* 431-435.

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.

Imhoff, R., Schmidt, A. F., Bernhardt. J., Dierksmeier, A., & Banse, R. (2011). An inkblot for sexual preference: A semantic variant of the affect misattribution procedure. *Cognition and Emotion, 25,* 676-690.

Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why self-control seems (but may not be) limited. *Trends in Cognitive Sciences, 18,* 127-133.

Jacoby, L. L. (1991). A process-dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, *30*, 513-541.

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., & Becker, M. (in press). Latent state-trait analyses for process models of implicit measures. In J. A. Krosnick, T. H. Stark, & A. L. Scott (Eds.), *The Cambridge Handbook of implicit bias and racism*. Cambridge, UK: Cambridge University Press.

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

Koppehele-Gossel, J., Hoffmann, L., Banse, R., & Gawronski, B. (2020). Evaluative priming as an implicit measure of evaluation: An examination of outlier-treatments for evaluative priming scores. *Journal of Experimental Social Psychology, 87:103905*.

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

Krieglmeyer, R. & Sherman, J. W. (2012). Disentangling stereotype activation and stereotype application in the Stereotype Misperception Task. *Journal of Personality and Social Psychology, 103,* 205-224.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist, 74*, 569-586.

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

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.

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

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81,* 181-192.

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

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

Payne, B. K., Cheng, 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., & Correll, J. (2020). Race, weapons, and the perception of threat. *Advances in Experimental Social Psychology, 62,* 1-50.

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.

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.

Röhner, J., Schröder-Abé, M., & Schütz, A. (2013). What do fakers actually do to fake the IAT? An investiagtion of faking strategies under different faking conditions. *Journal of Research in Personality, 47*, 330-338.

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.

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

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

Schnabel, K., Banse, R., & Asendorpf, J. 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.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17,* 367-447.

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.

Stahl, C., Haaf, J., & Corneille, O. (2016). Subliminal evaluative conditioning? Above-chance CS identification may be necessary and insufficient for attitude learning. *Journal of Experimental Psychology: General, 145*, 1107-1131.

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.

Teige-Mocigemba, S., Penzl, B., Becker, M., Henn, L., & Klauer, K. C. (2016). Controlling the "uncontrollable": Faking effects on the affect misattribution procedure. *Cognition and Emotion, 30,* 1470-1484.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology, 66*, 490-501.

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

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

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.

Wolsiefer, K., Westfall, J., & Judd, C. M. (2017). Modeling stimulus variation in three common implicit attitude tasks. *Behavior Research Methods, 49*, 1193-1209.

Ye, Y., & Gawronski, B. (2018). Validating the semantic misattribution procedure as an implicit measure of gender stereotyping. *European Journal of Social Psychology, 48,* 348-364.

Yzerbyt, V. Y., Coull, A., & Rocher, S. J. (1999). Fencing off the deviant: The role of cognitive resources in the maintenance of stereotypes. *Journal of Personality and Social Psychology, 77,* 449-462.

Acknowledgements

Preparation of this chapter was supported by National Science Foundation Grant BCS-1941440. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Table 1. Overview of measurement instruments, their underlying mechanisms, and approximate range of internal consistency estimates.

Task	Reference	Underlying Mechanism	Internal Consistency
Action Interference Paradigm	Banse et al. (2010)	Response Compatibility	.30 - .50
Affect Misattribution Procedure	Payne et al. (2005)	Misattribution	.70 - .90
Approach-Avoidance Task	Chen & Bargh (1999)	Response Compatibility	.00 - .90 [a]
Brief Implicit Association Test	Sriram & Greenwald (2009)	Response Compatibility	.55 - .95
Evaluative Movement Assessment	Brendl et al. (2005)	Response Compatibility	.30 - .80 [b]
Evaluative Priming with Evaluative Decision Task	Fazio et al. (1995)	Response Compatibility	.00 - .55
Evaluative Priming with Picture Naming Task	Spruyt et al. (2007)	Stimulus Compatibility	n/a
Evaluative Priming with Pronunciation Task	Bargh et al. (1996)	Stimulus Compatibility	n/a
Extrinsic Affective Simon Task	De Houwer (2003a)	Response Compatibility	.15 - .65
Go/No-go Association Task	Nosek & Banaji (2001)	Response Compatibility	.45 - .75
Identification Extrinsic Affective Simon Task	De Houwer & De Bruycker (2007)	Response Compatibility	.60 - .70
Implicit Association Procedure	Schnabel et al. (2006)	Response Compatibility	.75 - .85
Implicit Association Test	Greenwald et al. (1998)	Response Compatibility	.70 - .90 [c]
Implicit Relational Assessment Procedure	Barnes-Holmes et al. (2010)	Response Compatibility	.20 - .80
Recoding Free Implicit Association Test	Rothermund et al. (2009)	Response Compatibility	.55 - .65
Relational Responding Task	De Houwer et al. (2015)	Response Compatibility	.45 - .75
Semantic Misattribution Procedure	Imhoff et al. (2011)	Misattribution	.50 - .85
Semantic Priming with Lexical Decision Task	Wittenbrink et al. (1997)	Stimulus Compatibility	n/a
Semantic Priming with Semantic Decision Task	Banaji & Hardin (1996)	Response Compatibility	n/a
Shooter Task	Correll et al. (2002)	Response Compatibility	n/a
Single Attribute Implicit Association Test	Penke et al. (2006)	Response Compatibility	.70 - .80
Single Block Implicit Association Test	Teige-Mocigemba et al. (2008)	Response Compatibility	.60 - .90
Single Category Implicit Association Test	Karpinski & Hilton (2006)	Response Compatibility	.70 - .90
Sorting Paired Features Task	Bar-Anan et al. (2009)	Response Compatibility	.40-.70
Weapon Identification Task	Payne (2001)	Response Compatibility	n/a

[a] 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).

[b] 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).

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