18 Increasing the Validity of Implicit Measures: New Solutions for Assessment, Conceptualization, and Action Explanation

Franziska Meissner and Klaus Rothermund

Introduction

Egalitarian values, tolerance, and diversity are popular in modern societies. Still, even today, people are discriminated against because of their race, gender, age, or sexual orientation. In interpersonal or intergroup contexts, people often do not behave as intended or as they would have predicted. But this is not a phenomenon observed exclusively in social behavior. Across the board, we often find a gap between people's self-reported intentions and their actual behavior. For instance, people pick another piece of cake although this interferes with their dietary plans. In romantic relationships, some cheat on their beloved partners. Others accept an offered cigarette although they are convinced that smoking is bad.

Why do people not act in line with their attitudes? Dual process/dual system models postulate that these counter-attitudinal behaviors stem from forces that operate below the threshold of personal control and awareness (e.g., Hofmann et al., 2009; Kahneman, 2011; Strack & Deutsch, 2004). In order to detect and identify these hidden forces of behavior, so-called implicit measures of attitudes, like the Implicit Association Test (IAT, Greenwald et al., 1998), have been introduced, and they evoked enthusiastic hopes regarding their predictive value. Unfortunately, however, implicit measures have fallen short of these expectations. In this chapter, we argue that, in order to fulfill their inherent possibilities, implicit measures have to be improved with regard to construct validity, and we outline specific ways how this can be achieved.

This chapter is structured as follows: First, we describe the attitude–behavior gap, that is, the phenomenon that behavior often deviates from explicit values, goals, and intentions. We discuss explanations for this discrepancy and the mostly unsatisfying results of previous research to close the gap by using implicit measures. In the main part of this chapter, we specify different features of implicit measures that we consider responsible for their weak relationship with behavior (cf. Meissner et al., 2019; Rothermund et al., 2020). We review findings illustrating each of these problems, and we present specific solutions which can, in turn, increase the predictive power of implicit measures. We conclude with further, more general recommendations and implications for future research.

The Attitude–Behavior Gap

Intergroup bias and discrimination is just one special case of the more general phenomenon of an *attitude–behavior gap*: People express attitudes and values that are in conflict with their actual behavior. Indeed, although they were postulated to be strongly linked to cognitive processes, judgments, and behavior

 https://doi.org/10.1017/9781108885492.028

(e.g., Ajzen, 1991; Fazio et al., 1983; Katz, 1960), attitudes measured via self-reports revealed only weak predictive validity (correlations were rarely above $r = .30$, Wicker, 1969; see also Kraus, 1995, who found an average $r = .38$). In an attempt to explain and eventually close this attitude–behavior gap, dual process or dual system models were proposed that traced behavior to controlled and automatic influences (e.g., Hofmann et al., 2009; Kahneman, 2011; Strack & Deutsch, 2004), and researchers began to examine the "sub"-personal level of behavior control. The underlying reasoning was that people might not be able to verbalize their mental processes accurately (e.g., Nisbett & Wilson, 1977), implying that self-reports must be insufficient predictors of behavior. Instead, it was argued that behavior might rather be driven by "introspectively unidentified (or inaccurately identified) traces of past experience" (Greenwald & Banaji, 1995, p. 5).

In the following, new attitude measurement procedures were introduced that should tap into these processes (e.g., the IAT, Greenwald et al., 1998; the Affective Priming Paradigm, Fazio et al., 1986; the Affect Misattribution Procedure, Payne et al., 2005; for overviews, see Gawronski & De Houwer, 2014; Gawronski & Hahn, 2019; Teige-Mocigemba et al., 2010; Wentura & Degner, 2010). The new procedures share one characteristic: They do not require introspection. Instead of asking questions about attitudes directly, they involve computerized tasks that require individuals to quickly execute a specific behavioral response to a set of stimuli, capitalizing on stimulus–response compatibility effects to detect and assess automatic evaluations of these stimuli (De Houwer, 2003a). The scores obtained from the observed performance in those tasks are then interpreted in terms of attitude strength.

The hopes evoked by these new measurement procedures, first and foremost by the IAT (Greenwald et al., 1998) were enormous. On the one hand, these measures were assumed to provide less opportunity to control one's responses as compared to self-report measures. They were assumed to be less affected by deliberate manipulation attempts and self-presentational concerns (e.g., Fazio et al., 1986; Greenwald et al., 1998). On the other hand, it was proposed that these procedures would measure an attitude construct (*implicit attitude*) that is introspectively less accessible and thus conceptually distinct from that captured via self-report (*explicit attitude*; Greenwald & Banaji, 1995; Wilson et al., 2000; but see Fazio, 2007, for a different view). Based on these considerations, the corresponding procedures were often labeled as *implicit measures* and *explicit measures* (for a recent critique of the label "implicit measures," see Corneille & Hütter, 2020; for an alternative suggestion see Rothermund et al., 2020).

Overall, implicit measures came along with the promise that they would detect the hidden forces of behavior; forces that make us act in a way that deviates from our intentions to act. Consequently, it was assumed that implicit measures would predict behavior over and above self-report (for an overview of theoretical models of the relationship between cognitive processes, measurement procedures, and behavior, see Perugini et al., 2010).

Unfortunately, implicit measures have not met these expectations. Over the years, the predictive validity of the most popular implicit measure of attitudes, the IAT (Greenwald et al., 1998), was the subject of several meta-analyses (Greenwald et al., 2009; Kurdi et al., 2019; Oswald et al., 2013). All of them point to the same conclusion: The implicit-criterion correlation (ICC) is unsatisfyingly low (average $r_{ICC} = .27$, Greenwald et al., 2009; average $r_{ICC} = .14$, Oswald et al., 2013; 90-percent prediction interval for ICCs from

$r = -.14$ to $r = .32$; Kurdi et al., 2019). What is more, the incremental predictive validity that is provided by implicit measures over and above self-report is more-or-less negligible (i.e., ranging between one and five percent; Greenwald et al., 2009; Kurdi et al., 2019; Oswald et al., 2013). Altogether, the predictive validity of implicit measures is quite disappointing. This is a frustrating state of affairs, especially because it was the unsatisfying predictive value of self-reported attitudes that gave rise to the development of implicit measures in the first place.

Possible Causes of Low Predictive Validity of Implicit Measures

So, should we now stop using implicit measures? Obviously, such a conclusion would be premature. Instead, we first have to answer the question *why* implicit measures are so weakly related to behavioral criteria. In the past two decades, numerous studies pointed to important shortcomings of the IAT and its derivatives (e.g., for overviews, see Fiedler et al., 2006; Gawronski & Hahn, 2019; Teige-Mocigemba et al., 2010; Wentura & Rothermund, 2007). In order to close the attitude–behavior gap, we must no longer ignore them.

Specifically, recent research points to at least four crucial features of implicit measures that might be responsible for their weak predictive validity. First, implicit measures are not process-pure. They suffer from extraneous influences (for an overview, see Teige-Mocigemba et al., 2010). In order to reliably predict behavior, we have to filter out construct-irrelevant variance. Second, most implicit measures focus on evaluation instead of motivation. However, liking and wanting are not necessarily related (e.g., Tibboel et al., 2015b). Regarding the prediction of behavior, especially in critical situations when behavior deviates from what people value, strive for, and intend, it might be more relevant what people want than what they like. Third, most implicit measures focus on quantifying associations. Associations, however, might be too unspecific to unambiguously relate to and account for a particular behavior in a specific situation. Instead, (implicit) propositional beliefs could be a more plausible precursor of behavior (e.g., Hughes et al., 2011). Finally, most implicit measures aim at assessing global attitudes or stereotypes in a situational vacuum. These global beliefs do not adequately reflect the structure of mental representations of attitudes and stereotypes, and they do not match the situatedness of real-life behavior, which always occurs in specific contexts. Assessing implicit beliefs in a more context-dependent and domain-specific way will help to overcome this lack of specificity and will help to improve their predictive validity for behavior in real-life conditions.

In the remainder of this chapter, we explain these potentially problematic features of existing implicit measures in detail, and we present specific solutions for each of these issues (see Table 18.1, for an overview).

Extraneous Influences on Implicit Measures

Implicit measures (just like explicit ones) are not process-pure. They are contaminated with processes that we do not intend to measure, and this kind of error variance reduces their predictive validity. In order to illustrate this point, we will focus on the IAT (Greenwald et al., 1998), as one of the most popular implicit measures.

The IAT involves two speeded binary classification tasks that are combined in two test blocks with varying response compatibility. More precisely, there is a target task involving the classification of exemplars into one of two target categories representing opposing attitude

Table 18.1 *Improving the predictive validity of implicit measures: Problems and solutions*

Problems	Solutions
1. Extraneous influences unrelated to implicit biases contaminate the assessment of implicit attitudes	- Development of process-pure measures of implicit associations (e.g., IAT-RF, Rothermund et al., 2009; SB-IAT, Teige-Mocigemba et al., 2008) - Statistical modeling to distinguish effects of different processes (e.g., multinomial models: ReAL model, Meissner & Rothermund, 2013; Quad model, Conrey et al., 2005; diffusion model: Klauer et al., 2007)
2. Assessment of evaluations ("liking") does not capture motivational qualities ("wanting")	- Development of measures that capture implicit wanting (e.g., W-IAT, Koranyi et al., 2017)
3. Mere associations are ambiguous and do not capture meaningful beliefs ("propositions")	- Development of measures that capture automatic evaluations of propositional beliefs (e.g., PEP, Müller & Rothermund, 2019; IRAP, Barnes-Holmes et al., 2010; RRT, De Houwer et al., 2015)
4. Global attitudes and stereotypes do not capture the context-dependency of social cognition and behavior	- Development of measures that assess domain-specific attitudes and beliefs (e.g., Casper et al., 2011; Kornadt et al., 2016; Wigboldus et al., 2003)

objects (e.g., deciding whether a given face shows a Black or a White individual) and an attribute task where stimuli have to be categorized with regard to a particular attribute of interest. In attitude IATs, this would be valence, but it could also be stereotypical attributes, or personality traits, depending on the research question. In the IAT, exemplars appear successively on screen, and have to be categorized quickly via pressing one of two response keys. Importantly, the response assignment varies across the different blocks of the IAT. In the *compatible block* of an attitude IAT, the positively evaluated target category (e.g., White) and the positive pole of the attribute dimension (e.g., positive) are assigned to the same response key while the more negative target and attribute categories (e.g., Black and negative) are assigned to the second key. In the *incompatible block*, participants are instructed to press one key for negative targets and positive attributes (Black and positive) and to press the other key for positive targets and negative attributes (White and negative). Typically, participants are faster and produce fewer errors in compatible compared to incompatible IAT blocks. That is, responding is usually easier when associated categories are assigned to the same response key. The performance difference between compatible and incompatible blocks (*compatibility effect*, *IAT effect* or *IAT score*) is then interpreted as a measure for the strength of associations between the respective categories (Greenwald et al., 1998).

In the last two decades, however, the construct validity of the IAT was repeatedly challenged (for an overview, see Teige-Mocigemba et al., 2010). For example, it has been observed that content-unrelated IATs (i.e., IATs comprising no overlap with respect to the involved target concepts) share a substantial amount of variance (so-called *method variance*; e.g., Back et al., 2005; Greenwald et al., 1998; Klauer et al., 2010; McFarland & Crouch, 2002; Mierke & Klauer, 2003). To account for this shared method variance, several attitude-unrelated processes were proposed to influence the IAT, such as general processing speed (Blanton et al., 2006; McFarland & Crouch, 2002) or executive functions like task-switching ability (Ito et al., 2015; Klauer et al., 2010). Besides that, the IAT also suffers from unwanted block order effects: IAT scores are larger when participants completed the compatible block first (e.g., Greenwald et al., 1998; Nosek et al., 2005; for a possible explanation, see Klauer & Mierke, 2005). Furthermore, contrary to initial expectations, the IAT is not only driven by the valence of the target categories but can also be contaminated by unwanted stimulus effects (e.g., Bluemke & Friese, 2006; Gast & Rothermund, 2010; Govan & Williams, 2004; Mitchell et al., 2003; Steffens & Plewe, 2001).

The IAT is thus not a pure measure of associations. In the following, we will outline that there is a common core behind the contaminating processes: *recoding* (e.g., De Houwer, 2003b; Rothermund et al., 2009; Wentura & Rothermund, 2007).

The Problem: Recoding

In the IAT, participants are instructed to perform two binary classification tasks simultaneously. This is demanding, and hard. Recoding constitutes a possibility to simplify the task by combining targets and attributes to superordinate categories based on some feature that targets and attributes share (e.g., salience, familiarity, valence, or even perceptual features like color or shape; Rothermund et al., 2009; see also Chang & Mitchell, 2009; De Houwer et al., 2005; Kinoshita & Peek-O'Leary, 2006; Mierke & Klauer, 2003; Rothermund & Wentura, 2004).[1] For example, in a Black–White IAT, both the target and the attribute dimension is characterized by a salience asymmetry (i.e., the outgroup is more salient than the ingroup, negative words are more salient than positive words; Rothermund & Wentura, 2004; see also Kinoshita & Peak-O'Leary, 2005). In the compatible block of the Black–White IAT, the congruent response assignment allows for a task recoding based on salience. This reduces the IAT's double categorization task to a simple binary classification (i.e., the correct response can be identified for all stimuli by categorizing them as salient vs. non-salient). In the incompatible block, however, the incongruent response assignment of target and attribute categories prevents recoding. Here, participants have no possibility to simplify the task and thus have to perform the double categorization task, and to respond on the basis of the nominal categories of the target and attribute tasks. This difference in task difficulty between compatible and incompatible blocks (i.e., simple vs. double binary classification) accounts for the typical block difference

[1] In this sense, the recoding account combines two related process models for the IAT, the so-called figure-ground account (Rothermund & Wentura, 2001, 2004; Rothermund et al., 2005; see also Chang & Mitchell, 2009; Kinoshita & Peak-O'Leary, 2006; Mitchell, 2004) and the task-switching account (Klauer & Mierke, 2005; Mierke & Klauer, 2001, 2003). These and other process accounts for the IAT have been reviewed by Teige-Mocigemba and colleagues (Teige-Mocigemba & Klauer, 2015; Teige-Mocigemba et al., 2010).

in response times and error rates observed in the IAT (e.g., Rothermund et al., 2009). By this, recoding can even explain IAT effects in the absence of any category-based associations (e.g., De Houwer et al., 2005; Mierke & Klauer, 2003; Rothermund & Wentura, 2004).

Note that recoding triggers responses that are unrelated to the nominal categories and thus are also unrelated to the attitudes toward those categories. That is, if the task is recoded, the stimuli are processed and categorized based on the shared feature rather than according to their nominal category membership (e.g., target faces are no longer processed as Black vs. White but rather as more vs. less salient, Kinoshita & Peek-O'Leary, 2005). Moreover, recoding should not be understood as an irrelevant constant which is added to the IAT score but rather as a further source of variance that potentially distorts the correlation of IAT scores and behavioral criteria. More precisely, there might be inter-individual differences in task recoding (e.g., due to individual differences in familiarity, Greenwald et al., 1998 [Exp. 2], salience, Rothermund & Wentura, 2004 [Exp.'s 2A & 2B], or fluid intelligence, von Stülpnagel & Steffens, 2010) that can be unrelated to the to-be-measured attitudes. In this sense, recoding represents more than random error. It should be understood as a systematic source of variance in IAT scores that can distort the predictive validity of these scores for behavioral criteria. Although individual differences in recoding could in principle reflect evaluations (e.g., recoding that is based on valence), recent research findings do not support the claim that recoding adds anything on top of explicit attitudes in predicting behavior. A more detailed elaboration on this issue is beyond the scope of the present paper. For the interested reader, however, who wants to learn about theoretical ideas and empirical findings showing that recoding is unrelated to the construct of interest, we refer to our pertinent work (Meissner & Rothermund, 2013, 2015a, 2015b, 2015c).

We consider recoding as the most crucial extraneous influence in the IAT because it can account for other extraneous influences that were identified in IAT research. For instance, recoding might be the driving force behind the negative correlation of task-switching ability with IAT scores (e.g., Klauer et al., 2010): Task-switching ability enables fast and effortless switches between target and attribute classification. In the incompatible block of the IAT, people scoring high on switching ability should thus be faster than people with low switching ability. In the compatible block, however, task recoding renders switches between targets and attributes unnecessary: By combining pairs of targets and attributes that are assigned to the same response into superordinate categories, the dual task architecture of target and attribute classifications is reduced to a single binary classification task. Eliminating task switches in the compatible block implies that people with high vs. low switching ability will perform equally well in this block. Thus, due to recoding, IAT scores should decrease with increasing levels of switching ability. Similarly, the negative correlation of IAT scores with general processing speed (e.g., McFarland & Crouch, 2002) can be explained with recoding. Finally, task recoding can also account for unwanted stimulus effects in the IAT because it is due to recoding that participants do not process stimuli in terms of their nominal category membership but rather based on some other, irrelevant feature (e.g., Gast & Rothermund, 2010).

To sum up, the IAT score is a mixture of both relevant influences (i.e., associations) and irrelevant influences, first and foremost, recoding. In order to increase the predictive validity of the IAT, we thus have to control for recoding processes. In the following, we present two approaches that aim to separate effects of associations from the influence of recoding in the IAT. The first approach minimizes recoding processes by modifying the IAT procedure while the second disentangles associations and recoding processes with the help of multinomial modeling.

The First Solution: Dropping the Block Structure

Task recoding can be traced back to the IAT's block structure. Hence, in order to address the problem of recoding, new variants of the IAT were developed that dropped this structure. These procedures, namely, the Single-Block IAT (SB-IAT, Teige-Mocigemba et al., 2008) and the Recoding-Free IAT (IAT-RF, Rothermund et al., 2009) varied response compatibility within one test block. Scores for attitudinal preferences are then obtained by computing performance differences between compatible and incompatible *trials* rather than between compatible and incompatible *blocks*.[2] In other words, response assignment is not constant throughout a whole block but varies randomly from trial to trial. Participants are informed about the response assignment either by simply showing the current response assignment shortly before the current stimulus appears (IAT-RF) or by using stimulus position as a cue (with a compatible response assignment holding in the upper half of the screen, and an incompatible response assignment holding in the lower half of the screen; SB-IAT).

By varying response compatibility randomly from trial to trial, the upcoming category–response assignment is not predictable. Hence, implementing a stable and efficient recoding strategy specifically for the compatible response assignment becomes much harder than in the standard IAT. In fact, Rothermund et al. (2009) showed that dropping the IAT's block structure successfully reduces switch cost asymmetries, a marker of recoding processes.

The new IAT variants did not only successfully reduce the role of task recoding but come with some further advantages. First, due to the omission of the block structure, block order effects (e.g., Greenwald et al., 1998) can no longer influence conclusions. Furthermore, and in line with the assumption that recoding is the core problem behind the IAT's validity issues, the block-free versions of the IAT eliminate method-related variance (Teige-Mocigemba et al., 2008) and stimulus effects (Gast & Rothermund, 2010). Finally, these IAT variants are not only correlated with behavioral criteria (Houben et al., 2009; Teige-Mocigemba et al., 2008), it was also shown that dropping the block structure of the IAT can actually improve its predictive validity (Kraus & Scholderer, 2015).

Still, although confounding factors like recoding processes do not represent the construct that researchers typically attempt to measure when employing the IAT (i.e., evaluative associations), it should nevertheless be considered that they might represent variance that is worth being additionally measured. For example, it has been proposed that task recoding occasionally reflects explicit attitudes (Rothermund et al., 2009) and that it might be related to criteria of interest (e.g., behavior;

[2] Another measurement procedure that dropped the block structure and thus taps into the problem of recoding is the Extrinsic Affective Simon Task (EAST, De Houwer, 2003b; see also its derivative, the Identification EAST, De Houwer & De Bruycker, 2007). However, the EAST does not contain classification responses based on the target categories. It is thus less sensitive to category evaluations but is strongly susceptible to stimulus effects (Gast & Rothermund, 2010). The EAST also suffers from low reliability (De Houwer, 2003b), and thus represents a less recommendable approach to account for the problem of recoding.

Rothermund et al., 2005; Teige-Mocigemba et al., 2008). The SB-IAT and the IAT-RF pursue the strategy to prevent recoding. Employing these variants thus bears the risk to miss potentially interesting relationships between some underlying processes of indirect measures and individual judgment and behavior.

In the following, we will present another approach that follows a different rationale. Accepting that neither the IAT nor its variants will ever be process-pure, this approach aims to measure all relevant processes within the same procedure. By that, the predictive power of both construct-related and method-related variance due to task recoding can be examined separately.

The Second Solution: Multinomial Modeling

The ReAL model (Meissner & Rothermund, 2013) is a multinomial processing tree model that successfully disentangles evaluative associations from the distorting influence of task recoding within a single IAT. Besides addressing the problem of recoding, the ReAL model comes with another important advantage: It overcomes the limitation that IAT scores can only be interpreted as measures for *relative* preferences (which is potentially problematic; for an overview, see Teige-Mocigemba et al., 2010). Instead, the ReAL model measures associations separately for each of the two target categories. Thus, the model can distinguish between equally strong positive, negative, or neutral attitudes for both attitude objects while the conventional IAT score would only yield a null effect in all of these cases. Altogether, the ReAL model enables a remarkably fine-grained analysis of the IAT (note that neither recoding nor relative preferences were addressed in previous mathematical models for the IAT; e.g., the quad model, Conrey et al., 2005; or the diffusion model, Klauer et al., 2007). In the following, we will give an overview of the ReAL model's basic idea, and we will review findings showing that the model successfully addresses several problems of the IAT, including predictive validity.

As a multinomial processing tree model, the ReAL model is based on categorical data, that is, on the observed pattern of correct and incorrect responses in the IAT. The strength of multinomial processing tree models is that they are able to disentangle multiple cognitive processes accounting for the same observable response. The probabilities of these observed responses are predicted by a set of model parameters that are assumed to reflect particular cognitive processes. The assumed interplay of these cognitive processes is displayed in a tree architecture, the so-called *multinomial processing tree*. The tree structure follows theoretical considerations and needs empirical validation. Based on observed response patterns, algorithms estimate values for the model parameters which are then interpreted as measures for the respective cognitive processes (for mathematical details on multinomial processing tree models, see Batchelder & Riefer, 1999; Hu & Batchelder, 1994; Riefer & Batchelder, 1988; for reviews of applications, see Erdfelder et al., 2009; Klauer et al., 2012).

The ReAL model assumes that three kinds of processes drive responding in the IAT, and it measures them in separate model parameters: recoding (*Re*), evaluative associations (*A*), and the resource-consuming label-based identification of the correct response (*L*). In order to disentangle those processes, the ReAL model incorporates two central assumptions. First, task recoding determines responding for both targets and attributes but only in one of the IAT test blocks (i.e., in the compatible block). Second, evaluative associations influence responding in both compatible and incompatible test blocks, and they are triggered only in target trials, not in attribute trials (reflecting the conventional understanding of attitudes as evaluative associations triggered by an attitude object, not vice versa; Fazio et al., 1986; see also Anderson, 1983). Importantly, the ReAL model does not require *a priori* assumptions concerning the question whether an attitude object would be associated with positive or negative valence, and it can even handle situations where both target concepts are associated with the same valence (e.g., Meissner & Rothermund, 2015b).

Various studies showed that the model parameters are valid measures of the respective processes (Meissner & Rothermund, 2013; Meissner & Rothermund, 2015a, 2015b, 2015c; see also Jin, 2016; Koranyi & Meissner, 2015). Most importantly, it was shown that the ReAL model's association parameters reflect the direction and the strength of evaluative associations for each of the two target concepts (Meissner & Rothermund, 2013; Meissner & Rothermund, 2015b), even in applications where task recoding pushed the overall IAT score in the opposite direction (Meissner & Rothermund, 2015a). Studies revealed that the association parameters are sensitive to manipulations of evaluation (Meissner & Rothermund, 2013), and that they are immune against artificial, non-evaluative influences (i.e., salience asymmetries and modality match effects; Meissner & Rothermund, 2015a, 2015c). Furthermore, the association parameters, but not recoding or the overall IAT score, show convergent validity with another indirect measure of attitudes (Meissner & Rothermund, 2015b). Additionally, in line with theoretical considerations (e.g., Fazio & Towles-Schwen, 1999), evaluative associations as measured by the ReAL model correlate with self-reported attitudes in non-sensitive attitude domains (consumer preferences; Meissner & Rothermund, 2013), but not in attitude domains where individuals typically harbor more self-presentational concerns (self-esteem; Jusepeitis & Rothermund, 2022; Meissner & Rothermund, 2015b).

Finally, the ReAL model's association parameters also revealed predictive validity (Meissner & Rothermund, 2013). In a study applying a fruit–chocolate IAT, the amount of chocolate participants consumed while watching a movie clip was successfully predicted by the evaluative associations regarding chocolate as measured by the model's association parameter. Neither the recoding parameter nor the global IAT score was predictive of the observed behavior. These data nicely illustrate the potential of the ReAL model to increase the IAT's predictive validity. It also shows that recoding is not just a theoretical problem. By acknowledging its role and controlling for its influence, the IAT's power to predict behavior can be increased considerably.

The performance of the ReAL model might be improved even further by incorporating recent developments in the field of multinomial models (i.e., allowing the incorporation of response time data, Heck & Erdfelder, 2016; Klauer & Kellen, 2018; and a sophisticated treatment of possible parameter heterogeneity, e.g., Klauer, 2010; Matzke et al., 2015). Still, even without these advances, the ReAL model's association parameters outperformed the global IAT score in terms of construct validity in a number of studies (e.g., Meissner & Rothermund, 2013, 2015a, 2015b, 2015c). We therefore recommend researchers to consider an application of the ReAL model as an alternative, or at least as an additional analysis tool for the IAT in studies on attitudes. Note that the ReAL model has been validated only for attitude IATs so far. While it is technically possible to apply it to other IATs (e.g., stereotype IATs), future research still has to examine the validity of the model parameters in these applications (for a discussion and ideas on adjustments of the ReAL model for stereotype IATs, see Meissner & Rothermund, 2013).

To sum up, there are undeniable extraneous influences on implicit measures. However, there are also promising approaches that address this problem. Hence, just as researchers should try their best to reduce the influence of response biases in questionnaire-based studies by choosing the right question and response format (cf. Pasek & Krosnick, 2010), researchers applying the IAT as a measure of attitudes should not take its procedure and algorithms as a given. By controlling for task recoding in implicit measures of attitudes, either via procedural modifications or via mathematical separation, researchers can measure more validly what people really like. But what if it is not relevant what people like but rather what people *want*?

Measuring Liking versus Wanting

Implicit measures often focus on assessing evaluations (i.e., attitudes, or *liking*). Motivation (i.e., *wanting*), however, is much more closely linked with behavior. Certainly, in many cases, liking and wanting overlap: You like what you want, and you want what you like (Berridge, 1996; Berridge & Robinson, 2003). However, the two can dissociate (e.g., Dai et al., 2010; Dai et al., 2014; Epstein et al., 2003; Hobbs et al., 2005; Litt et al., 2010; Robinson & Berridge, 1993). Consider the moment after finishing a lovely meal. You still like the delicious food you just had, but you are sated, so you do not want more of it. Similarly, despite a highly positive evaluation, procrastinators show a lack of desire for certain objects. Addicts, on the other hand, might show a strong wanting for a substance despite low liking (e.g., Robinson & Berridge, 1993; Wiers et al., 2002). Finally, dissociations like these can also be observed in interpersonal behaviors (e.g., liking vs. wanting of attractive same-sex persons among heterosexual people; Koranyi et al., 2017), and is thus potentially relevant for intergroup behavior and discrimination.

To summarize, liking and wanting are not necessarily related (for an overview, see also Tibboel et al., 2015b). And if wanting is more relevant in determining behavior, especially in critical situations where liking and wanting dissociate, (implicit and explicit) liking measures must have limited predictive validity. In order to address the attitude–behavior gap, implicit measures should therefore be expanded to the assessment of motivation. So, how do we measure what people want?

The Problem: How to Assess Wanting?

Again, self-reports are not a plausible solution. Not only because they are potentially influenced by self-presentational concerns but also because it is not easy to disentangle wanting and liking on a semantic level. Asking participants to do so is not recommendable. Instead, we need an implicit measure of wanting.

Due to the fact that the IAT typically outperforms other implicit measures in terms of internal consistency and test–retest reliability (e.g., Gawronski & De Houwer, 2014; Nosek et al., 2007), it is not surprising that multiple researchers relied on the general structure of the IAT in order come up with an implicit measure of wanting (for an overview of IAT-based as well as other implicit measures of wanting, see Tibboel et al., 2015b). By now, several different IAT variants have been introduced that aim to measure implicit wanting for a given target dimension of interest (e.g. alcohol vs. no alcohol, smoking vs. no smoking, attractive vs. unattractive persons). They differ with respect to the attribute dimension, that is, in how they transformed the IAT from measuring liking into a measure of wanting. Palfai and Ostafin (2003) for instance, introduced an IAT that employs approach- and avoidance-related words as attributes (e.g., advance, withdraw) which should be classified as "approach" or "avoidance" (see Kraus & Scholderer, 2015, for a similar approach using the IAT-RF). Tibboel et al. (2011, 2015a), on the other hand, used "I want" vs. "I do not want" as attribute categories in their IAT. The exemplars were either positive vs. negative words (e.g., holiday, pain; Tibboel et al., 2011), just as those used in a traditional attitude IAT, or motivational words with a clear positive vs. negative connotation (e.g., gain vs. deprivation; Tibboel et al., 2015a).

Unfortunately, there is only little evidence that these IAT variants actually measure implicit wanting (for an overview, see Tibboel et al., 2015b). A high overlap between those IATs and traditional liking IATs raised strong doubts regarding the validity of these measures (Tibboel et al., 2011, 2015b). Obviously, a simple replacement of the evaluative attribute categories with motivational concepts alone cannot transform the IAT into an implicit measure of wanting. If anything, these IATs will only reflect semantic associations, or a "cognitive form of wanting" (Tibboel et al., 2015a, p. 189). Recently, however, another IAT variant was introduced that follows a totally different, more promising rationale.

The Solution: Endowing Measurement Procedures with Motivational Quality

Our wanting IAT (W-IAT, Koranyi et al., 2017) differs from previous attempts to measure implicit wanting: Instead of incorporating wanting on a purely semantic level, the W-IAT establishes an actual desire for a particular object and maps this wanting onto one of the attribute categories and its associated response. This is done by (1) inducing a strong need (e.g., thirst), and then (2) endowing one response of the attribute task with a consummatory motivational quality (i.e., *I want drinks*). This wanting response of the attribute classification can then be used to assess the strength of wanting for the two target categories of the task.

In detail, this is how the procedure of the W-IAT works: In a first step, a strong need for water is induced by asking participants to quickly consume many salty crackers. As a result, participants get quite thirsty. They want to drink. This wanting is transferred to one of the attributes of the W-IAT. The attribute exemplars in the W-IAT are pictures of drinks versus no drinks (i.e., household items) that have to be classified according to the categories "I want" and "I do not want." Since participants are thirsty, there is an inherent wanting for drinks. By using drinks (vs. no drinks) as attribute exemplars, the response of the "I want" category is endowed with wanting. We further increased the motivational value of this response by having participants earn small amounts of water with this response key if they respond correctly and fast enough to pictures of a drink. This motivational consummatory quality of the response is further emphasized by providing correct and fast responses with visual and auditory action effects (i.e., a glass of water gets filled on screen, a cork plopping or some gurgling sounds).

All of the above (i.e., inducing of a need for drinks, using drinks vs. no drinks as attribute exemplars, and enabling participants to earn water via keypress) serves the same purpose. It secures that one of the responses of the W-IAT is endowed with an actual wanting quality. Having established this, implicit wanting can be measured in the W-IAT by assigning target categories of interest (e.g., smoking, attractive faces) to either the wanting or to the no wanting response in the two blocks of this task.

Koranyi et al. (2017) illustrated the potential of the W-IAT in a study on interpersonal attraction. Heterosexual male participants were assessed for their implicit wanting and liking of very attractive versus less attractive faces. Importantly, half of them were female faces. In this study, a regular liking IAT revealed a preference for attractive over less attractive faces, irrespective of the gender of those faces. Heterosexual men thus like attractiveness in both men and women. The W-IAT, however, revealed a different pattern. That is, heterosexual men showed a wanting only for attractive female – but not male – faces. Thus, as expected, wanting was observed only for attractive faces of the opposite sex. Note that a version of the wanting IAT that only uses "I want" and "I do not want" as attribute categories without endowing the responses with a consummatory character (cf. Tibboel et al., 2011, 2015a) could not uncover this other-sex directed wanting (Koranyi et al., 2017). The findings of Koranyi et al. (2017) thus suggest that an implicit measure of wanting should incorporate a motivational quality in the responses.

We further corroborated the validity of the W-IAT in a recent study that compared smokers and nonsmokers for their wanting of smoking (Grigutsch et al., 2019). In this study, a liking IAT revealed that both smokers and nonsmokers showed similar and slightly negative implicit evaluations of the category smoking, while the W-IAT successfully dissociated between the two groups: Only smokers showed positive wanting for smoking whereas non-smokers showed negative wanting for this category.

To sum up, the W-IAT measures actual wanting instead of purely semantic associations (cf. Palfai & Ostafin, 2003, Tibboel et al., 2011, 2015a). Furthermore, our data suggest that the W-IAT is sensitive to variations in wanting when liking is high (Koranyi et al., 2017) and also when liking is low (Grigutsch et al., 2019; Koranyi et al., 2020). There are instances where liking and wanting dissociate, and behavior is not always in line with attitudes or values. Implicit measures of wanting, like the W-IAT, are thus a promising alternative to existing measures of implicit liking when it comes to closing the attitude–behavior gap with regard to interpersonal and other forms of behavior. What remains an open question is how the W-IAT could be applied in other settings (e.g., online studies). It might be unrealistic that the cracker procedure would work outside the lab. Inducing thirst would be difficult in such a setting. Nevertheless, the logic of the W-IAT should hold not only for (induced) thirst but also for other needs. For example, researchers could make use of something that most of their participants chronically want: money. The attribute task will then have to be operationalized in a way that one response satisfies this need. In this sense, participants would get extra monetary reward for fast and accurate responding to stimuli of one of the attribute categories. Such a procedure could be easily applied in an online study. We established the validity of a money-based variant of the W-IAT in a recent study (Koranyi et al., 2020), showing that the W-IAT (but not the standard liking IAT) could discriminate between heavy coffee drinkers and low/non-consumers of coffee. However, further research is needed to more systematically examine whether a W-IAT operationalized in this or a similar way reliably captures (differences in) implicit wanting.

Measuring Associations versus Beliefs

When shifting the focus from self-report measures to indirect measurement procedures in order to address the attitude–behavior gap, researchers also switched the to-be-measured construct. Instead of personal beliefs that can be expressed as a statement in propositional form, most implicit measures (i.e., the IAT, Greenwald et al., 1998, Affective Priming, Fazio et al., 1986, and their derivatives) aim at measuring associations, that is, the mental link between a given object and some kind of attribute (e.g., positive or negative valence). Such an associative connection, however, is necessarily unspecific and no longer has a clear meaning.

The Problem: Associations Are Ambiguous

Associations do not contain qualitative relational information. One and the same association between two concepts can thus reflect different, sometimes even opposite beliefs (cf. Arkes & Tetlock, 2004). For example, the concepts "I" and "good" can be associated either because I believe that I am good, or because I believe that I am no good, or because I would desperately like to be good, or because I know that others would like me to be good (see also De Houwer, 2014; De Houwer et al., 2015). Similar issues have been discussed in the literature on evaluative learning: Several studies have shown that valence transfer from an unconditioned stimulus (US) to a conditioned stimulus (CS) is not determined by mere associative co-occurrence alone, but that it is moderated by relational qualifiers (Bading et al., 2020; Moran & Bar-Anan, 2013). For example, presenting a neutral person (CS) together with a positively evaluated person (US) will lead to positive evaluations of the CS if the relation between the two persons is described as friendship, but it will lead to a negative evaluation of the CS, if the relation between the two is described as being antagonistic (Fiedler & Unkelbach, 2011; see also Förderer & Unkelbach, 2012; Peters & Gawronski, 2011; Van Dessel et al., 2018; Zanon et al., 2014).

Together with findings of weak predictive validity (e.g., Greenwald et al., 2009; Oswald et al., 2013), these issues raise the question whether implicit measures of associations might be too unspecific to predict behavior. Even more, it has been argued that all information stored in memory is inherently propositional, and that therefore, the attempt to explain behavior with associations is bound to fail (e.g., De Houwer, 2014; Hughes et al., 2011). Whether we need the concept of associations (in addition to propositions) at all, however, is an ongoing debate, and we will not address it in detail in this chapter. Still, what remains is that (implicit measures of) associations are ambiguous with regard to the qualitative relation between the concepts that are associated, and that this could be responsible for the weak predictive validity of implicit measures. We therefore propose that the attitude–behavior gap can be addressed more convincingly with implicit measures of propositional beliefs instead of associations.

The Solution: Implicit Measures of Beliefs

In recent years, different implicit measures of beliefs have been introduced (Barnes-Holmes et al., 2010; De Houwer et al., 2015; Müller & Rothermund, 2019). They, however, do share some main characteristics. For example, all of them ground on the assumption that complex personal beliefs regarding the truth of certain propositions become activated automatically when processing the content of these statements (e.g., Wiswede et al., 2013). Furthermore, while traditional implicit measures of attitudes are blind regarding the semantic relation between the associated concepts, implicit measures of beliefs assess the automatic evaluation of propositions, that is, of more complex, relational content. By implication, while implicit measures of associations use single words or pictures as stimuli, implicit measures of beliefs present combinations of stimuli, or even whole sentences that specify the relation between those concepts. The procedural details of these new implicit measures of propositional attitudes and beliefs, however, differ. In the following, we provide a brief overview.

Implicit Relational Assessment Procedure

On each trial of the Implicit Relational Assessment Procedure (IRAP, Barnes-Holmes et al., 2010; see also Remue et al., 2013; Remue et al., 2014), two kinds of information are presented simultaneously, one below the other (e.g., in one trial this could be "I am" and "nice"; in another trial it could be "I am not" and "worthless," Remue et al., 2013, 2014). In contrast to associative measures, the presented information in the IRAP contains relational information between the target ("I") and the attributes ("nice" or "worthless"): The quality of the target–attribute relation is specified as, for example, "am" versus "am not." Thus, in combination, the stimuli that are presented in the IRAP represent a specific propositional belief. By changing the relational qualifier (e.g., from "I am" to "I want to be," Remue et al., 2013, 2014), different kinds of beliefs can be assessed. In the IRAP, participants are instructed to respond to the combined information via pressing one of two response keys (e.g., labelled with "true" vs. "false"). Importantly, however, the specific response rule differs across the IRAP. In one block, whenever the current information conforms with a pre-specified belief (e.g., the belief that "I am good") the correct response would be "true," and whenever the information is inconsistent with this belief, the required response would be "false." In another block of the task, the belief that serves as a comparison standard for the true/false distinction, is reversed (i.e., information consistent with the belief that "I am no good" would have to be responded with "true"). If implicit beliefs drive responding in the IRAP, task performance should differ between these two blocks. In line with that assumption, responding in the IRAP is faster and more accurate if the response rule fits with personal beliefs (Barnes-Holmes et al., 2010). To sum up, the IRAP is a useful tool to dissociate different beliefs (e.g., uncovering differences between actual and ideal self, Remue et al., 2013, 2014), that are not assessable with traditional implicit measures like the IAT. However, there are also some shortcomings: Each IRAP is limited to assessing implicit attitudes toward one single pair of beliefs at a time. In addition, it has been shown that the IRAP is not immune to faking attempts (Hughes et al., 2016), and often reveals only moderate reliability (e.g., Remue et al., 2013, 2014; see also Gawronski & De Houwer, 2014).

Relational Responding Task

In the so-called Relational Responding Task (RRT, De Houwer et al., 2015), there are two kinds of trials. In inducer trials, synonyms of the concepts "true" and "false" are presented, and they have to be classified via keypress as either "true" or "not true." The inducer trials are used to provide the two response keys with a specified meaning, and to prevent a reinterpretation in terms of, for example, position (De Houwer et al., 2015). In the second kind of trials, the target trials, whole sentences are presented. Those are statements reflecting certain kinds of beliefs (e.g., regarding immigrants, De Houwer et al., 2015; or smoking, Tibboel et al., 2017), and they also have to be classified as "true" or "not true." As in the IRAP, the specific response rule in the target trials differs across the RRT: In one block, participants are asked to respond to the statements "as if" they held a specific belief (e.g., as if they believed that immigrants were smarter than natives) while in the second block, they should respond "as if" they hold the opposite belief (e.g., as if they believe that natives are smarter than immigrants). That is, the correct response to a particular target sentence is "true" in one block, and "not true" in the other one. If implicit beliefs drive responding in the RRT, task performance should differ between the two critical blocks. More precisely, responding should be faster and more accurate if the response rule fits with personal beliefs. As expected, De Houwer et al. (2015) found that implicit beliefs of Flemish participants reflect ingroup preferences: On average, they showed better performance if they should respond as if they held pro-Flemish beliefs. In a recent study, Dentale et al. (2020) applied the RRT to the assessment of self-related beliefs and found that this measure predicted depression over and above an explicit measure of self-esteem. The RRT is structurally similar to the IAT (see also De Houwer et al., 2015). Like the IAT, the RRT comprises two intermixed binary classification tasks that are mapped onto two response keys. Furthermore, like the IAT, the RRT consists of two critical blocks that differ with regard to the specific response rules, and the global score relies on the performance difference between these blocks. Finally, like IAT, the RRT is reliable and easy to apply (De Houwer et al., 2015, Tibboel et al., 2017). However, the structural similarity with the IAT also involves the risk that the RRT suffers from similar flaws as the IAT (e.g., recoding).

Propositional Evaluation Paradigm

The most recent implicit measure of beliefs follows a different rationale. While the previously introduced measurement procedures more or less resembled the basic structure of the IAT, the so-called Propositional Evaluation Paradigm (PEP, Müller & Rothermund, 2019; see also Wiswede et al., 2013) shares more characteristics with priming measures. In the PEP, each trial begins with a simple sentence that is presented word by word (e.g., "Milk is red"). Participants are instructed to read the sentence. After a short blank, a target word (either the word "true" or the word "false") has to be categorized as "true" or "false." Importantly, the prime sentence is completely irrelevant for the decision that has to be made, the task is simply to classify the target word. Still, the PEP will reveal compatibility effects: The sentence "Milk is red" is false, hence, "false" is automatically activated. This facilitates the response if "false" is presented as the target word. However, it interferes with correct responding if "true" is presented as a target. Similarly, if the prime is a sentence that is consistent with participants' beliefs, we would observe faster and more accurate responding to "true" as compared to "false."

Of course, we are usually not interested in implicit beliefs regarding factual knowledge. The potential of an implicit measure of beliefs like the PEP rather lies in more sensitive domains, like beliefs regarding social groups. These questions were addressed in a study by Müller and Rothermund (2019). Here, the PEP prime sentences were extracted from established self-report measures of classic and modern racism. The to-be-classified targets were again the words "true" and "false." In a student sample, the PEP scores indicated an endorsement of a tolerant and open stance toward minorities, and a rejection of racist attitudes. More precisely, responding with "true" was facilitated when sentences like "A multicultural Germany would be good." were shown as primes. On the other hand, participants responded faster with "false" when sentences like "Racist groups are no longer a threat toward immigrants." were presented. Again, the sentences themselves were completely irrelevant for the task. Still, they were evaluated as true or false. Unintentionally, beliefs affected responding. In this sense, the PEP is an implicit measure of beliefs. Importantly, there is variance in the PEP response patterns across participants, and this variance is predictive of self-reported beliefs and behavioral intentions. More precisely, stronger endorsement of racist attitudes on the PEP predicts (a) explicit endorsement of these statements in the Classical Racism Scale as well as the Modern Racism Scale, (b) behavioral intentions to donate money to a charity organization for refugees, and (c) political orientation (Müller & Rothermund, 2019). As an implicit measure of beliefs, the PEP overcomes the ambiguity of implicit measures of associations.

Recent studies applied the PEP to measure implicit beliefs regarding age stereotypes (Huang & Rothermund, 2023a), age norms (de Paula Couto & Rothermund, 2022; de Paula Couto et al., 2022), and self-esteem (Jusepeitis & Rothermund, 2023). These studies demonstrate that the PEP is able to reliably assess implicit beliefs that are independent of their explicit counterparts, and even have incremental predictive validity in predicting outcome variables over and above explicit measures of beliefs (Jusepeitis & Rothermund, 2023).

To sum up, this section illustrated that switching the focus from complex beliefs to simple associations when introducing implicit measures, might have obstructed the success of these measures in predicting spontaneous behavior. Apparently, processing of complex propositional content can also occur in a rapid, and automatic (i.e., implicit) fashion. Studying these implicit beliefs, as well as contextual and individual differences in the automatic activation of these beliefs might help us to overcome some of the limits that are faced by implicit measures of association in predicting behavior. Implicit measures of beliefs are valid, and they provide a much more differentiated view than implicit measures of associations. This implies that implicit measures of beliefs have the potential to predict behavior over and above traditional implicit measures – and also that they have the potential to go beyond explicit attitudes and beliefs in predicting behavior. In order to explain the attitude–behavior gap, we should thus not rely solely on associations. We have to get beliefs back on board.

Increasing the Fit Between Predictor and Criterion: The Context-Dependency of Implicit Beliefs and Their Relation to Real-Life Behavior

In the previous paragraphs, we addressed the attitude–behavior gap by focusing on the validity of the predictor, that is, implicit measures. However, this does not mean that (the measurement of the) criterion should be taken as a given. Dependent variables in relevant studies often do not represent actual behavior (Carlsson & Agerström, 2016; Oswald et al., 2013). For example, many studies use self-reports as a criterion. In principle, there is nothing wrong with investigating and predicting self-report data. Given that indirect measures were introduced with the ambition to overcome self-presentational concerns that typically affect self-report measures and/or to measure introspectively less accessible traces of experience, however, it seems somewhat odd to rely on self-reports as the major criterion for predictive validity. Certainly, assessing real behavior is difficult, especially when it comes to interpersonal or intergroup behavior that figures as a case of discrimination.

This type of criticism of previous studies potentially cuts in two ways: On the one hand, some of the reported positive evidence for the incremental validity of implicit measures in predicting critical behaviors can be discounted on the grounds that these critical behaviors were simply not assessed. Some studies did not assess *any* criterion variables and interpreted the mere existence of an effect on a so-called implicit bias measure as sufficient evidence to claim discrimination, which is definitely a misunderstanding: An IAT effect is not evidence for racial bias or discriminatory behavior – it is just a response time difference in a categorization task that is done on a computer (e.g., Arkes & Tetlock, 2004). Other studies used self-report measures or other indirect indicators as criteria (e.g., hypothetical intentions to do or not do something, or activation of certain brain regions) instead of assessing the behavior of interest that reflects real discrimination or impulsivity. In these cases, alternative and more trivial explanations might be brought up to account for these positive relations, without having to assume that there is a relation to real bias and discrimination (e.g., Oswald et al., 2013).

On the other hand, null findings regarding the predictive validity of implicit measures might similarly be discounted on the grounds that if the critical criterion variable was not – or not properly – assessed, then not finding a relation does not speak against the predictive validity of the predictor for the criterion in question either. In this regard, a crucial issue regards the fit between predictor and criterion (Payne et al., 2008). Implicit measures usually do not provide contextual information and capture attitudes, stereotypes or beliefs in only a very general fashion. This type of global assessment of "the" attitude toward, for instance, Blacks, women, gays, or older people ignores the fact that more or less all attitudes, beliefs, and stereotypes are context-dependent: they become activated and influence behavior in a highly situation-specific way (Blair, 2002; Casper, Rothermund & Wentura, 2010, 2011; Gawronski & Cesario, 2013; Huang & Rothermund, 2022, 2023b; Kornadt & Rothermund, 2011, 2015; Müller & Rothermund, 2012; Wigboldus et al., 2003). Assessing implicit biases in a global way – that is, in a situational vacuum – will thus often be not specific enough to predict a particular behavior toward a specific attitude object in a certain situation (Blanton & Jaccard, 2015).

One possibility to increase correspondence between implicit measures and behavioral criteria is to measure and to aggregate several behavioral outcomes over different situations, time points, and target objects in order to get a more general behavioral criterion (e.g., of discriminatory behavior; Ajzen 1991). Another and maybe more economic possibility to increase correspondence would be to contextualize attitude assessment with the IAT by incorporating context information in the procedure (e.g., Blanton & Jaccard, 2015) and to use these contextualized measures to predict behaviors in corresponding real-life situations. In this regard, we recommend using measures that use dual primes (Casper et al., 2010, 2011; Huang & Rothermund, 2023b), combining context and category information, in the assessment of automatic evaluations, or to specify context-dependent evaluative meanings when choosing attribute categories in the IAT (Kornadt et al., 2016).

To sum up, in order to investigate the potential of implicit measures for explaining and closing the attitude–behavior gap, both predictors (implicit attitudes and desires) and criterion variables (e.g., discriminatory behaviors) have to be assessed in a reliable, valid, and contextualized fashion. Future research has to face the challenge to improve both the assessment of predictor and criterion variables, and to come up with assessment procedures that adequately reflect the situational embeddedness of attitudes as well as behavior.

Closing Thoughts

In this chapter, we presented an overview of possible reasons for the weak relationship between implicit measures like the IAT and behavioral criteria like discriminatory behavior. We outlined that the weak predictive value of implicit measures is due (1) to extraneous influences like recoding, (2) to the measurement of liking instead of wanting, (3) to the measurement of associations instead of complex beliefs, and (4) to the lack of context-dependency and domain-specificity in the assessment of implicit biases (cf. Meissner et al., 2019). For each problem, we presented precise solutions (see Table 18.1 for an overview). That is, we suggested to switch to procedural variations that minimize extraneous influences (i.e., the SB-IAT, Teige-Mocigemba et al., 2008; and the IAT-RF; Rothermund et al., 2009), and to apply sophisticated analysis tools (i.e., the ReAL model, Meissner & Rothermund, 2013) in order to separate relevant processes from extraneous influences. Furthermore, we presented an implicit measure that goes beyond purely evaluative associations and quantifies actual wanting in an implicit fashion (i.e., the W-IAT, Koranyi et al., 2017). Third, we pointed to implicit measures of beliefs (e.g., the PEP, Müller & Rothermund, 2019) that allow a more differentiated view on individual attitudes and values than measures of associations. Finally, we sketched that contextualized measures might be more adequate in assessing the structure of implicit beliefs and their implications for behavior (Casper et al., 2011; Kornadt et al., 2016). In sum, each of the presented advances has the potential to increase the predictive power of implicit measures substantially. Future research also has to clarify whether a combination of these approaches may lead to further improvement.

So, what are the implications of these reflections for our understanding of the attitude–behavior gap and of the failure of existing research on implicit biases to close this gap? Apparently, early measurement attempts suffer from important shortcomings, and for decades, research did not fully take into account the complexity of those measures. The reason for this reluctance to identify, accept, and creatively address these issues may stem from a tendency to cling to those measures that are familiar and easy to apply. This somewhat conservative attitude, however, has brought research on implicit biases into a troublesome situation. However, by now, our toolbox has been enriched with promising, more sophisticated methods and procedures that researchers should eagerly adopt and adapt to their specific research questions. Future research should thus switch its focus. We have tried to improve our understanding of those measures, we have to consider the problems that they face, and we have to optimize them. Hence, if recent meta-analytical findings tempted some researchers to drop implicit measures from their agenda because they have not fulfilled their expectations, we would interfere with a definite: Don't! We are not even close to the point of making a definitive decision on this issue. We definitely don't know enough, and we have not yet exploited the full potential that is inherent in the "implicit" approach to social cognition and behavior.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295. https://doi.org/10.1016/S0022-5371(83)90201-3

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the Implicit Association

Test?” *Psychological Inquiry*, *15*(4), 257–278. https://doi.org/10.1207/s15327965pli1504_01

Back, M. D., Schmukle, S. C., & Egloff, B. (2005). Measuring task-switching ability in the Implicit Association Test. *Experimental Psychology*, *52*(3), 167–179. https://doi.org/10.1027/1618-3169.52.3.167

Bading, K., Stahl, C., & Rothermund, K. (2020). Why a standard IAT effect cannot provide evidence for association formation: The role of similarity construction. *Cognition and Emotion*, *34*(1), 128–143. https://doi.org/10.1080/02699931.2019.1604322

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., et al. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record*, *60*, 527–542. https://doi.org/10.1007/BF03395726

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. https://doi.org/10.3758/BF03210812

Berridge, K. C. (1996). Food reward: Brain substrates of wanting and liking. *Neuroscience and Biobehavioral Reviews*, *20*, 1–25. https://doi.org/10.1016/0149-7634(95)00033-B

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences*, *26*, 507–513. https://doi.org/10.1016/S0166-2236(03)00233-9

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*(3), 242–261. https://doi.org/10.1207/S15327957PSPR0603_8

Blanton, H., & Jaccard, J. (2015). Not so fast: Ten challenges to importing implicit attitude measures to media psychology. *Media Psychology*, *18*(3), 338–369. https://doi.org/10.1080/15213269.2015.1008102

Blanton, H., Jaccard, J., Gonzales, P. M., et al. (2006). Decoding the Implicit Association Test: Implications for criterion prediction. *Journal of Experimental Social Psychology*, *42*(2), 192–212. https://doi.org/10.1016/j.jesp.2005.07.003

Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, *42*(2), 163–176. https://doi.org/10.1016/j.jesp.2005.03.004

Carlsson, R., & Agerström, J. (2016). A closer look at the discrimination outcomes in the IAT literature. *Scandinavian Journal of Psychology*, *57*(4), 278–287. https://doi.org/10.1111/sjop.12288

Casper, C., Rothermund, K., & Wentura, D. (2010). Automatic stereotype activation is context dependent. *Social Psychology*, *41*(3), 131–136. https://doi.org/10.1027/1864-9335/a000019

Casper, C., Rothermund, K., & Wentura, D. (2011). The activation of specific facets of age stereotypes depends on individuating information. *Social Cognition*, *29*(4), 393–414. https://doi.org/10.1521/soco.2011.29.4.393

Chang, B. P. I., & Mitchell, C. J. (2009). Processing fluency as a predictor of salience asymmetries in the Implicit Association Test. *The Quarterly Journal of Experimental Psychology*, *62*(10), 2030–2054. https://doi.org/10.1080/17470210802651737

Conrey, F. R., Sherman, J. W., Gawronski, B., et al. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*(4), 469–487. https://doi.org/10.1037/0022-3514.89.4.469

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*(3), 212–232. https://doi.org/10.1177/1088868320911325

Dai, X. C., Brendl, C. M., & Ariely, D. (2010). Wanting, liking, and preference construction. *Emotion*, *10*(3), 324–334. https://doi.org/10.1037/a0017987

Dai, X. C., Dong, P., & Jia, J. S. (2014). When does playing hard to get increase romantic attraction? *Journal of Experimental Psychology: General*, *143*(2), 521–526. https://doi.org/10.1037/a0032989

De Houwer, J. (2003a). 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 Associates.

De Houwer, J. (2003b). The extrinsic affective Simon task. *Experimental Psychology*, *50*(2), 77–85. https://doi.org/10.1026//1618-3169.50.2.77

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social Psychology and Personality Compass*, *8*(7), 342–353. https://doi.org/10.1111/spc3.12111

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*(2), 133–143. https://doi.org/10.1016/j.jbtep.2006.10.004

De Houwer, J., Geldof, T., & De Bruycker, E. (2005). The Implicit Association Test as a general measure of similarity. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *59*(4), 228–239. https://doi.org/10.1037/h0087478

De Houwer, J., Heider, N., Spruyt, A., et al. (2015). The relational responding task: Toward a new implicit measure of beliefs. *Frontiers in Psychology*, *6*(319), 1–9. https://doi.org/10.3389/fpsyg.2015.00319

de Paula Couto, M. C. P., Huang, T., & Rothermund, K. (2022). Age specificity in explicit and implicit endorsement of prescriptive age stereotypes. *Frontiers in Psychology, 13,* 820739. https://doi.org/10.3389/fpsyg.2021.820739

de Paula Couto, M. C. P., & Rothermund, K. (2022). Prescriptive views of aging: Disengagement, activation, wisdom, and dignity as normative expectations for older people. In Y. Palgi, A. Shrira, & M. Diehl (Eds.), *Subjective Views of Aging*. Cham, Switzerland: Springer Nature, pp. 59–75.

Dentale, F., Vecchione, M., Ghezzi, V., et al. (2020). Beyond an associative conception of automatic self-evaluations: Applying the Relational Responding Task to measure self-esteem. *The Psychological Record*, *70*(2), 227–242. https://doi.org/10.1007/s40732-020-00392-4

Epstein, L. H., Truesdale, R., Wojcik, A., et al. (2003). Effects of deprivation on hedonics and reinforcing value of food. *Physiology and Behavior*, *78*(2), 221–227. https://doi.org/10.1016/S0031-9384(02)00978-2

Erdfelder, E., Auer, T. S., Hilbig, B. E., et al. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*(5), 603–637. https://doi.org/10.1521/soco.2007.25.5.603

Fazio, R. H., Powell, M. C., & Herr, P. M. (1983). Toward a process model of the attitude-behavior relation: Accessing one's attitude upon mere observation of the attitude object. *Journal of Personality and Social Psychology*, *44*(4), 723–735. https://doi.org/10.1037/0022-3514.44.4.723

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., et al. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229–238. https://doi.org/10.1037/0022-3514.50.2.229

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken, & Y. Trope (Eds.), *Dual-Process Theories in Social Psychology* (pp. 97–116). New York, NY: Guilford Press.

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the "I," the "A," and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology*, *17*(1), 74–147. https://doi.org/10.1080/10463280600681248

Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition and Emotion*, *25*(4), 639–656. https://doi.org/10.1080/02699931.2010.513497

Förderer, S., & Unkelbach, C. (2012). Hating the cute kitten or loving the aggressive pit-bull: EC effects depend on CS-US relations. *Cognition & Emotion*, *26*(3), 534–540. https://doi.org/10.1080/02699931.2011.588687

Gast, A., & Rothermund, K. (2010). When old and frail is not the same. Dissociating category-based and stimulus-based influences on compatibility effects in four implicit measurement methods. *The Quarterly Journal of Experimental Psychology*, *63*(3), 479–498. https://doi.org/10.1080/17470210903049963

Gawronski, B., & Cesario, J. (2013). Of mice and men: What animal research can tell us about context effects on automatic responses in humans. *Personality and Social Psychology Review*, *17*(2), 187–215. https://doi.org/10.1177/1088868313480096

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

Gawronski, B., & Hahn, A. (2019). Implicit measures: Procedures, use, and interpretation. In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.), *Measurement in Social Psychology*. New York, NY: Taylor & Francis, pp. 29–55.

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology*, *40*(3), 357–365. https://doi.org/10.1016/j.jesp.2003.07.002

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., et al. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. https://doi.org/10.1037/a0015575

Grigutsch, L. A., Lewe, G., Rothermund, K., et al. (2019). Implicit 'wanting' without implicit 'liking': A test of incentive-sensitization theory in the context of smoking addiction using the Wanting-Implicit-Association-Test (W-IAT). *Journal of Behavior Therapy and Experimental Psychiatry*, *64*, 9–14. https://doi.org/10.1016/j.jbtep.2019.01.002.

Heck, D. W., & Erdfelder E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review*, *23*(5), 1440–1465. https://doi.org/10.3758/s13423-016-1025-6

Hobbs, M., Remington, B., & Glautier, S. (2005). Dissociation of wanting and liking for alcohol in humans: A test of the incentive-sensitisation theory. *Psychopharmacology*, *178*(4), 493–499. https://doi.org/10.1007/s00213-004-2026-0

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science*, *4*(2), 162–176. https://doi.org/10.1111/j.1745-6924.2009.01116.x

Houben, K., Rothermund, K., & Wiers, R. W. (2009). Predicting alcohol use with a recoding-free variant of the Implicit Association Test. *Addictive Behaviors*, *34*(5), 487–489. https://doi.org/10.1016/j.addbeh.2008.12.012

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*(1), 21–47. https://doi.org/10.1007/BF02294263

Huang, T., & Rothermund, K. (2022). Automatic activation of age stereotypes: Is attention to category information sufficient for stereotype priming? *Social Psychology*, *53*(5), 303–314. https://doi.org/10.1027/1864-9335/a000503

Huang, T., & Rothermund, K. (2023a). Endorsement and embodiment of cautiousness-related age stereotypes. *Frontiers in Psychology*, *14*, 1091763. https://doi.org/10.3389/fpsyg.2023.1091763

Huang, T., & Rothermund, K. (2023b). Implicit and explicit age stereotypes assessed in the same contexts are still independent. *Experimental Aging Research*, *49*(1), 41–57. https://doi.org/10.1080/0361073X.2022.2039507

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record*, *61*(3), 465–496. https://doi.org/10.1007/BF03395772

Hughes, S., Hussey, I., Corrigan, B., et al. (2016). Faking revisited: Exerting strategic control over performance on the Implicit Relational Assessment Procedure. *European Journal of Social Psychology*, *46*(5), 632–648. https://doi.org/10.1002/ejsp.2207

Ito, T. A., Friedman, N. P., Bartholow, B. D., et al. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology*, *108*(2), 187–218. https://doi.org/10.1037/a0038557

Jin, Z. (2016). Disentangling recoding processes and evaluative associations in a gender attitude implicit association test among adult males. *The Quarterly Journal of Experimental Psychology*, *69*(11), 2276–2284. https://doi.org/10.1080/17470218.2015.1126290

Jusepeitis, A., & Rothermund, K. (2022). No elephant in the room: The incremental validity of implicit self-esteem measures. *Journal of Personality*, *90*(6), 916–936. https://doi.org/10.1111/jopy.12705

Jusepeitis, A., & Rothermund, K. (2023). Is there something you don't tell me? Evidence for the incremental predictive validity of indirectly assessed self-esteem using the Propositional Evaluation Paradigm. *Manuscript submitted for publication*.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus, and Giroux.

Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly*, *24*(2), 163–204. https://doi.org/10.1086/266945

Kinoshita, S., & Peek-O'Leary, M. (2005). Does the compatibility effect in the race Implicit Association Test (IAT) reflect familiarity or affect? *Psychonomic Bulletin & Review*, *12*(3), 442–452. https://doi.org/10.3758/BF03193786

Kinoshita, S., & Peek-O'Leary, M. (2006). Two bases of the compatibility effect in the Implicit Association Test (IAT). *The Quarterly Journal of Experimental Psychology*, *59*(12), 2102–2120. https://doi.org/10.1080/17470210500451141

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Klauer, K. C., & Kellen, D. (2018). RT-MPTs: Process models for response-time distributions based on multinomial processing trees with applications to recognition memory. *Journal of Mathematical Psychology*, *82*, 111–130. https://doi.org/10.1016/j.jmp.2017.12.003

Klauer, K. C., & Mierke, J. (2005). Task-set inertia, attitude accessibility, and compatibility-order effects: New evidence for a task-set switching account of the Implicit Association Test effect. *Personality and Social Psychology Bulletin*, *31*(2), 208–217. https://doi.org/10.1177/0146167204271416

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., et al. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *The Quarterly Journal of Experimental Psychology*, *63*(3), 595–619. https://doi.org/10.1080/17470210903076826

Klauer, K. C., Stahl, C., & Voss, A. (2012). Multinomial models and diffusion models. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive Methods in Social Psychology. Abridged edition* (pp. 331–354). New York, NY: Guilford Press.

Klauer, K. C., Voss, A., Schmitz, F., et al. (2007). Process components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality and Social Psychology*, *93*(3), 353–368. https://doi.org/10.1037/0022-3514.93.3.353

Koranyi, N., Brückner, E., Jäckel, A., et al. (2020). Dissociation between wanting and liking for coffee in heavy drinkers. *Journal of Psychopharmacology*, *34*(12), 1350–1356. https://doi.org/10.1177/0269881120922960

Koranyi, N., Grigutsch, L. A., Algermissen, J., et al. (2017). Dissociating implicit wanting from implicit liking: Development and validation of the Wanting-Implicit-Association-

Test (W-IAT). *Journal of Behavior Therapy and Experimental Psychiatry*, *54*, 165–169. https://doi.org/10.1016/j.jbtep.2016.08.008

Koranyi, N., & Meissner, F. (2015). Handing over the reins: Neutralizing negative attitudes toward dependence in response to reciprocal romantic liking. *Social Psychological and Personality Science*, *6*(6), 685–691. https://doi.org/10.1177/1948550615580169

Kornadt, A. E., Meissner, F., & Rothermund, K. (2016). Implicit and explicit age stereotypes for specific life domains across the life span: Distinct patterns and age group differences. *Experimental Aging Research*, *42*(2), 195–211. https://doi.org/10.1080/0361073X.2016.1132899

Kornadt, A. E., & Rothermund, K. (2011). Contexts of aging: Assessing evaluative age stereotypes in different life domains. *Journals of Gerontology: Psychological Sciences*, *66*(5), 547–556. https://doi.org/10.1093/geronb/gbr036

Kornadt, A. E., & Rothermund, K. (2015). Views on aging: Domain-specific approaches and implications for developmental regulation. *Annual Review of Gerontology and Geriatrics*, *35*(1), 121–144. https://doi.org/10.1891/0198-8794.35.121

Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin*, *21*(1), 58–75. https://doi.org/10.1177/0146167295211007

Kraus, A. A., & Scholderer, J. (2015). Indirect measurement of motivation: Developing and testing a motivational recoding-free Implicit Association Test (m-IAT-RF). *Social Psychology*, *46*(3), 142–156. https://doi.org/10.1027/1864-9335/a000234

Kurdi, B., Seitchik, A. E., Axt, J. R., et al. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist*, *74*(5), 569–586. https://doi.org/10.1037/amp0000364

Litt, A., Khan, U., & Shiv, B. (2010). Lusting while loathing: Parallel counterdriving of wanting and liking. *Psychological Science*, *21*(1), 118–125. https://doi.org/10.1177/0956797609355633

Matzke, D., Dolan, C. V, Batchelder, W. H., et al. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*(1), 205–235. https://doi.org/10.1007/s11336-013-9374-9

McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition*, *20*(6), 483–510. https://doi.org/10.1521/soco.20.6.483.22977

Meissner, F., Grigutsch, L. A., Koranyi, N., et al. (2019). Predicting behavior with implicit measures: Disillusioning findings, reasonable explanations, and sophisticated solutions. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02483

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology*, *104*(1), 45–69. https://doi.org/10.1037/a0030734

Meissner, F., & Rothermund, K. (2015a). The insect-nonword IAT revisited: Dissociating between evaluative associations and recoding. *Social Psychology*, *46*(1), 46–54. https://doi.org/10.1027/1864-9335/a000220

Meissner, F., & Rothermund, K. (2015b). *A needle in a haystack: Isolating implicit self-esteem in the self-esteem IAT*. Unpublished manuscript.

Meissner, F., & Rothermund, K. (2015c). A thousand words are worth more than a picture? The effects of stimulus modality on the Implicit Association Test. *Social Psychological and Personality Science*, *6*(7), 740–748. https://doi.org/10.1177/1948550615580381

Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift für Experimentelle Psychologie*, *48*(2), 107–122. https://doi.org/10.1026//0949-3946.48.2.107

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology*, *85*(6), 1180–1192. https://doi.org/10.1037/0022-3514.85.6.1180

Mitchell, C. J. (2004). Mere acceptance produces apparent attitude in the Implicit Association Test. *Journal of Experimental Social Psychology*, *40*(3), 366–373. https://doi.org/10.1016/j.jesp.2003.07.003

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*(3), 455–469. https://doi.org/10.1037/0096-3445.132.3.455

Moran, T., & Bar-Anan, Y. (2013). The effect of object-valence relations on automatic evaluation. *Cognition and Emotion*, *27*(4), 743–752.

Müller, F., & Rothermund, K. (2012). Talking loudly but lazing at work – Behavioral effects of stereotypes are context dependent. *European Journal of Social Psychology*, *42*(5), 557–563. https://doi.org/10.1002/ejsp.1869

Müller, F., & Rothermund, K. (2019). The Propositional Evaluation Paradigm: Indirect assessment of personal beliefs and attitudes. *Frontiers in Psychology*, *10*, 2385. https://doi.org/10.3389/fpsyg.2019.02385

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

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*(2), 166–180. https://doi.org/10.1177/0146167204271418

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Frontiers of Social Psychology: Social Psychology and the Unconscious. The Automaticity of Higher Mental Processes* (pp. 265–292). New York, NY: Psychology Press.

Oswald, F. L., Mitchell, G., Blanton, H., et al. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. https://doi.org/10.1037/a0032734

Palfai, T. P., & Ostafin, B. D. (2003). Alcohol-related motivational tendencies in hazardous drinkers: Assessing implicit response tendencies using the modified-IAT. *Behaviour Research and Therapy*, *41*(10), 1149–1162. https://doi.org/10.1016/S0005-7967(03)00018-4

Pasek, J., & Krosnick, J. A. (2010). Optimizing survey questionnaire design in political science: Insights from psychology. In J. Leighley (Ed.), *Oxford Handbook of American Elections and Political Behavior*. Oxford: Oxford University Press, pp. 27–33. https://doi.org/10.1093/oxfordhb/9780199235476.003.0003

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

Payne, B. K., Cheng, C. M., Govorun, O., et al. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. https://doi.org/10.1037/0022-3514.89.3.277

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, NY: Guilford Press.

Peters, K. R., & Gawronski, B. (2011). Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *37*(4), 557–569. https://doi.org/10.1177/0146167211400423

Remue, J., De Houwer, J., Barnes-Holmes, D., et al. (2013). Self-esteem revisited: Performance on the implicit relational assessment procedure as a measure of self-versus ideal self-related cognitions in dysphoria. *Cognition & Emotion*, *27*(8), 1441–1449. https://doi.org/10.1080/02699931.2013.786681

Remue, J., Hughes, S., De Houwer, J., et al. (2014). To be or want to be: Disentangling the role of actual versus ideal self in implicit self-esteem. *PloS One*, *9*(9), e108837. https://doi.org/10.1371/journal.pone.0108837

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318–339. https://doi.org/10.1037/0033-295X.95.3.318

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving – An incentive-sensitization theory of addiction. *Brain Research Reviews*, *18*(3), 247–291. https://doi.org/10.1016/0165-0173(93)90013-P

Rothermund, K., Grigutsch, L. A., Jusepeitis, A., et al. (2020). Research with implicit measures: Suggestions for a new agenda of sub-personal psychology. *Social Cognition*, *38*, S243–S263. https://doi.org/10.1521/soco.2020.38.supp.s243

Rothermund, K., Teige-Mocigemba, S., Gast, A., et al. (2009). Minimizing the influence of recoding in the Implicit Association Test: The recoding-free Implicit Association Test (IAT-RF). *The Quarterly Journal of Experimental Psychology*, *62*(1), 84–98. https://doi.org/10.1080/17470210701822975

Rothermund, K., & Wentura, D. (2001). Figure-ground asymmetries in the Implicit Association Test (IAT). *Zeitschrift für Experimentelle Psychologie*, *48*(2), 94–106. https://doi.org/10.1026//0949-3946.48.2.94

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*(2), 139–165. https://doi.org/10.1037/0096-3445.133.2.139

Rothermund, K., Wentura, D., & De Houwer, J. (2005). Validity of the salience asymmetry account of the Implicit Association Test: Reply to Greenwald, Nosek, Banaji, and Klauer (2005). *Journal of Experimental Psychology: General*, *134*(3), 426–430. https://doi.org/10.1037/0096-3445.134.3.426

Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie*, *48*(2), 123–134. https://doi.org/10.1026//0949-3946.48.2.123

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1

Teige-Mocigemba, S. & Klauer, K. C. (2015). Implicit Association Test. In J. D. Wright (Ed.), *International Encyclopedia of the Social and Behavioral Sciences* (2nd ed., vol. 11, pp. 703–708). Oxford: Elsevier.

Teige-Mocigemba, S., Klauer, K. C., & Rothermund, K. (2008). Minimizing method-specific variance in the IAT: A single block IAT. *European Journal of Psychological Assessment*, *24*(4), 237–245. https://doi.org/10.1027/1015-5759.24.4.237

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). Practical guide to Implicit Association Tests and related tasks. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications* (pp. 117–139). New York, NY: Guilford Press.

Tibboel, H., De Houwer, J., Dirix, N., et al. (2017). Beyond associations: Do implicit beliefs play a role in smoking addiction? *Journal of Psychopharmacology*, *31*(1), 43–53. https://doi.org/10.1177/0269881116665327

Tibboel, H., De Houwer, J., Spruyt, A., et al. (2011). Testing the validity of implicit measures of wanting and liking. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*, 284–292. https://doi.org/10.1016/j.jbtep.2011.01.002

Tibboel, H., De Houwer, J., Spruyt, A., et al. (2015a). Heavy social drinkers score higher on implicit wanting and liking for alcohol than alcohol-dependent patients and light social drinkers. *Journal of Behavior Therapy and Experimental Psychiatry*, *48*, 185–191. https://doi.org/10.1016/j.jbtep.2015.04.003

Tibboel, H., De Houwer, J., & Van Bockstaele, B. (2015b). Implicit measures of "wanting" and "liking" in humans. *Neuroscience & Biobehavioral Reviews*, *57*, 350–364. https://doi.org/10.1016/j.neubiorev.2015.09.015

Van Dessel, P., De Houwer, J., & Smith, C. T. (2018). Relational information moderates approach-avoidance instruction effects on implicit evaluation. *Acta Psychologica*, *184*, 137–143. https://doi.org/10.1016/j.actpsy.2017.03.016

von Stülpnagel, R., & Steffens, M. C. (2010). Prejudiced or just smart? Intelligence as a confounding factor in the IAT effect. *Zeitschrift für Psychologie/Journal of Psychology*, *218*(1), 51–53. https://doi.org/10.1027/0044-3409/a000008

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications* (pp. 95–116). New York, NY: Guilford Press.

Wentura, D., & Rothermund, K. (2007). Paradigms we live by: A plea for more basic research on the Implicit Association Test. In B. Wittenbrink, & N. S. Schwarz (Eds.), *Implicit Measures of Attitudes: Procedures and Controversies* (pp. 195–215). New York, NY: Guilford Press.

Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, *25*(4), 41–78. https://doi.org/10.1111/j.1540-4560.1969.tb00619.x

Wiers, R. W., Van Woerden, N., Smulders, F. T. Y., et al. (2002). Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology*, *111*(4), 648–658. https://doi.org/10.1037//0021-843X.111.4.648

Wigboldus, D. H. J., Dijksterhuis, A., & van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484. https://doi.org/10.1037/0022-3514.84.3.470

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126. https://doi.org/10.1037/0033-295X.107.1.101

Wiswede, D., Koranyi, N., Müller, F., et al. (2013). Validating the truth of propositions: Behavioral and ERP indicators of truth evaluation processes. *Social Cognitive and Affective Neuroscience*, *8*(6), 647–653. https://doi.org/10.1093/scan/nss042

Zanon, R., De Houwer, J., Gast, A., et al. (2014). When does relational information influence evaluative conditioning? *Quarterly Journal of Experimental Psychology*, *67*(11), 2105–2122. https://doi.org/10.1080/17470218.2014.907324