Reflecting on Twenty-Five Years of Research Using Implicit Measures: Recommendations for their Future Use

Pieter Van Dessel, Jamie Cummins, Sean Hughes, Sarah Kasran, Femke Cathelyn, & Tal Moran

Ghent University, Belgium

Author note: Pieter Van Dessel and Jamie Cummins contributed equally. Correspondence concerning this article should be addressed to Pieter Van Dessel or Jamie Cummins, Department of Experimental-Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium. E-mail: Pieter.VanDessel@UGent.be or Jamie.Cummins@UGent.be. This manuscript is supported by the Scientific Research Foundation, Flanders under Grant FWO19/PDS/041 and by Ghent University under grants BOF16/MET_V/002 and 01P05517. The authors declare that they have no competing interests.

The authors would like to thank all our present and past collaborators for the many discussions that shaped this paper, foremost Jan De Houwer for his continued support and critical insight, and Ian Hussey for valuable comments and contributions on drafts of this manuscript.

This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article as published in Social Cognition special issue on "25 Years of Research Using Implicit Measures". The final article will be available, upon publication, via its DOI.

Abstract

For more than twenty-five years implicit measures have shaped research, theorizing, and intervention in psychological science. During this period, the development and deployment of implicit measures have been predicated on a number of theoretical, methodological, and applied assumptions. Yet these assumptions are frequently violated and rarely met. As a result, the merit of research using implicit measures has increasingly been cast into doubt. In this paper, we argue that future implicit measure research could benefit from adherence to four guidelines based on a functional approach wherein performance on implicit measures is described and analyzed as *behavior* emitted under specific conditions and captured in a specific measurement context. We unpack this approach and highlight recent work illustrating both its theoretical and practical value.

Keywords: implicit measures, behavior, automaticity, levels of analysis

Reflecting on Twenty-Five Years of Research Using Implicit Measures: Recommendations for their Future Use

Implicit measures are widely used in psychological science (see Gawronski & Hahn, 2019). Their popularity has been primarily based on three key assumptions. First, implicit measures are assumed to provide unique insight into mental processes operating under conditions of automaticity (Greenwald et al., 1998). For instance, the Implicit Association Test (IAT; Greenwald et al., 1998), was originally introduced as a measure of 'unconscious associations' between mental concepts (Greenwald & Banaji, 1995). Such mental associations were assumed to drive more unconscious and spontaneous thoughts and behavior (Fazio & Olson, 2003; Smith & DeCoster, 2000). The assumption that implicit measures provide a window into the 'unconscious' mind gave rise to a second key assumption: that implicit measure performance predicts other behavior in a unique manner, either independently, additively, or interactively with self-reports (Dovidio et al., 1997). Finally, at the methodological level, the use of implicit measures was (and still is) predicated on a third assumption: that these measures represent a reliable and valid indicator of the probed construct of interest (e.g., Nosek et al., 2005).

It seems fair to say that implicit measures have sparked an incredible amount of empirical work and led to some useful insights. For instance, implicit measures research has stimulated a massive amount of theory-building, and these theories have been used to generate new research hypotheses (e.g., Dovidio et al., 1997; Greenwald et al., 2003). There is also evidence to suggest a practical value of implicit measures, in the sense that implicit measure responses *sometimes* predict behavior (Friese et al., 2009). Yet, reflecting back on twenty-five years' worth of work, it is also clear that the initial fervor and enthusiasm has not been met. Indeed, together with the recent shift in focus in psychological science towards openness and replicability, it has become clear that research on implicit measures is not without its problems, and that the three aforementioned assumptions are unlikely to be true in the ways previously assumed.

A Critical Analysis of Three Assumptions Underlying Implicit Measures Research

Assumption 1: Implicit measure performance is mediated by specific mental processes. Evidence supporting the assumption that (a) implicit measure responses are mediated by specific mental processes and (b) these processes are distinguishable from those mediating responses on other (explicit) measures is weak. Researchers have typically treated responses on most implicit measures as proxies for mental associations (or associative processes). Yet, claiming that a behavior (e.g., an IAT score) is a proxy for a mental process (e.g., activation of a mental association) builds on untested and questionable assumptions (e.g., that the sole determinant of the behavior is the mental process: De Houwer, 2011). As one example, though IAT performance is often equated with association activation, research repeatedly shows that associative explanations of IAT performance fail to adequately account for empirical findings (see Brownstein, Madva, & Gawronski, 2019; Corneille & Stahl, 2019; De Houwer, 2014) and that observed discrepancies between IAT scores and explicit measure scores can be explained without reference to distinct types of processes (Heycke et al., 2018; Payne et al., 2008). As we discuss later on, this problem of (unverified) behavioral proxies can be solved by defining responses on implicit measures in behavioral rather than mental terms.

In addition to these issues with treating behavior as a proxy for mental processes, there is the long-standing problem that researchers use the same term (‘implicit’) in many different ways (Corneille & Hütter, 2020). Some use the term ‘implicit’ to refer to conditions under which mental processes are assumed to operate (i.e., the mental processes are implicit in the sense of ‘automatic’). This is problematic, as this has led researchers to map behavior in implicit measures onto a whole class of mental processes. Others use the term 'implicit' to refer to a class of procedures (i.e., the procedures are implicit in the sense of 'indirect'). Still others use 'implicit' in reference to the outcome of an indirect procedure (i.e., the IAT effect is 'implicit'). This heterogeneity has long been an issue that has plagued the literature and one that leads to conceptual confusion and wrong-headed debate, slows communication, and impedes scientific progress. Despite being repeatedly highlighted (Brownstein et al., 2019), the issue remains. As we note below, this can also be solved by defining performance on implicit measures in behavioral terms.

Assumption 2: Implicit measures have added value in predicting behavior. After twenty-five years of work, it seems reasonable to say that implicit measures have generally proven far worse at predicting behavior than was initially hoped (for a review see Oswald et al., 2013). Although applied research provides *some* evidence for their predictive utility (e.g., in the prediction of suicidality: Nock et al., 2010; see Tello et al., 2018, for a direct replication), when we step back and consider the field as a whole, there are many more instances where implicit measures have failed to provide any added utility in predicting behavior above and beyond simply asking people what they think, feel, or do (e.g., Larsen et al., 2012; Lindgren et al., 2019).

In part, this may be explained by poor measurement of the to-be-predicted 'behavior' in experimental work. For instance, measurement of the target behavior often involves verbal report (e.g., in self-reported behavioral intention measures) and it is unsurprising that such reports more strongly correspond to responses from other self-report measures (see Payne et al., 2008). Studies that do look at real-life behavior often use poorly-validated measures of behavior that may have little to do with the probed construct (e.g., seating distance from out-group members to validate racial prejudice measures: Amodio & Devine, 2006; see Dang et al., 2020, for a similar argumentation in the context of correspondence between self-report and behavioral measures).

Research that better deals with these measurement issues might show more evidence for predictive validity of implicit measures (e.g., when IAT scores are related to medical records of suicide attempts: Nock et al., 2010).

Furthermore, some recent studies have provided initial evidence that, in specific contexts, performances on implicit measures might explain added variance in other types of behavior after controlling for performances on explicit measures (Kurdi et al., 2019; see also recent research looking at predictive validity of response scores at the aggregate level: Payne et al., 2017). Yet, it remains clear that the field as a whole has fallen short of the broader claim that these measures provide substantive predictive utility when used in tandem with (or instead of) self-report measures (Meissner et al., 2019). Similarly, the promise that implicit measures would translate into interventions that could be used to impact real-life behavior has not come to pass (e.g., Carter et al., in press; Cristea et al., 2015; Forscher et al., 2019; Lai et al., 2014, 2016). This is not to say that intervention research is not at times useful, but rather that much of this research was based on a number of assumptions about these measures that do not concord with the actual trend of evidence. After twenty-five years of work, we still do not know in which real-life contexts (if any) implicit measures have any substantial added value (i.e., domains where they do not merely have a statistically significant impact but a clinical, practical, or meaningful one).

Assumption 3: Implicit measure scores are valid and reliable. Methodological research has long indicated that the psychometric properties of implicit measure scores are poor. First, there are problems with construct validity. These problems are not surprising given that the construct has often been tied to specific mental processes (e.g., the automatic activation of mental associations). However, even when implicit measures are defined at the behavioral level (i.e., as behavior that occurs ‘automatically’, irrespective of the mental processes at play) this leads to issues (Cummins et al. 2019). For instance, automaticity is often considered a multi-dimensional concept with conditions defined at the level of mental processes (e.g., unconsciousness, unawareness, unintentionality; Moors & De Houwer, 2006). As a solution, automaticity conditions can be defined at the behavioral level (e.g., implicit measure performance can be defined as unintentional when instructions to try and modify performance do not lead to congruent changes in performance) but even then, these conditions do not always relate to implicit measures in the way that they were originally assumed (Cummins et al., 2019; Hahn & Gawronski, 2019). Because of these conceptual issues, some authors have reversed or revised their earlier stance on this issue (e.g., Greenwald & Banaji, 2017) which has further complicated testing this kind of validity.

There are also issues with other psychometric properties of implicit measures, such as structural and external validity. In particular, research has often demonstrated poor test-retest reliability and weak correlations among different implicit measures of the same construct (e.g., Bar-Anan & Nosek, 2014; Blanton & Jaccard, 2006). Scores on implicit measures also fail to meet the measurement model they have been assumed to meet (i.e., that they load onto a single factor distinct from that of an explicit measure; see Schimmack, 2019). Interestingly, these issues have been noted for many years and guidelines have been provided to improve the statistical properties of implicit measure scores (e.g., Nosek et al., 2007). While there is published work on (a) the reliability of implicit measures, (b) calls to improve their reliability, and (c) suggestions on how to do so, this body of work is ignored by the majority of studies that use implicit measures. To take a concrete example, the procedural parameters of the IATs used in contemporary published studies are most often identical to those proposed in the original Greenwald et al. (1998) paper with no modifications. In short, implicit measures in general have failed to meet normative criteria at all three levels of validity (i.e., construct, structural, and external validity), but their use has persisted regardless of this.

Conclusion. Several decades of work has eroded our confidence in three key assumptions underpinning research on implicit measures, and scholars have grown increasingly uncertain whether those measures can (a) shed light on specific types of mental processes (Schimmack, 2019), (b) be used for predicting behavior in unique and substantial ways (Jost, 2019), and (c) represent reliable and valid measures of the construct of interest (Mitchell & Tetlock, 2017). This work has in turn led to growing doubt about the nature, role, and utility of implicit measures in general (e.g., De Houwer, 2019; Jost, 2019; Payne et al., 2017). In what follows, we offer recommendations designed to improve research on implicit measures in the years to come.

Four Guidelines for Future Research Using Implicit Measures

We are not the first to raise these issues with implicit measures: many others have offered important perspectives which have too often been disregarded (Flake et al., 2017; Payne et al., 2008; Rothermund & Wentura, 2004). As a result, we face a situation where we collectively make use of measures that are riddled with issues and often continue "business as usual". Clearly something needs to change. With this in mind, we reiterate and expand these prior suggestions, offering four concrete guidelines that researchers should adhere to when using implicit measures, both now and in the future. We will first outline these guidelines and then illustrate their potential theoretical and applied value.

Guideline 1: Define implicit measure responses as functional effects. In line with De Houwer et al. (2013), we believe that scientific progress is facilitated when researchers separate the phenomenon they want to explain (behavior) and the thing they use to explain that phenomenon (mental constructs and environmental variables; see also De Houwer, 2019; Hughes et al., 2016).

Therefore, in the context of research with implicit measures, our first recommendation is that researchers start by describing implicit measures in purely functional terms (i.e., as behavior observed in the context of a specific procedure). Following this guideline requires that implicit measures research begins with an analysis of (a) the contextual properties that characterize the procedure (e.g., the stimuli) and (b) the behavior captured by that procedure. The behavior of interest is “automatic behavior” (i.e., behavior captured under conditions of automaticity) and the set of contextual properties are those that influence the emission or elicitation of automatic behavior (see De Schryver et al., 2018; Gawronski & De Houwer, 2014). We will return to what we mean by “automatic” in Guideline 2.

For now, let us illustrate our first guideline using research on ‘implicit racial bias’. Previous work has long equated the phenomenon that needs to be explained (e.g., race IAT scores) with the phenomenon that is used to explain (e.g., mental associations; Hughes et al., 2011). When IAT scores are found to predict racial bias on other measures, researchers have assumed that it was mental associations which predicted such performances. Problems arise when evidence emerges questioning such an associative account, and because IAT scores and mental associations are conflated, the validity of the race IAT effect is also drawn into question (e.g., Schimmack, 2019). Adopting a functional perspective avoids this issue: by viewing IAT scores as behavior that is emitted in the context of the IAT procedure we separate the phenomenon which needs to be explained (IAT scores) from the phenomenon used to explain it (e.g., mental associations). This has several advantages: (a) uncertainty regarding mental causes does not lead to uncertainty about the observed racial IAT effect, (b) the behavioral effect can now be of interest regardless of its assumed mediators, and (c) researchers can investigate relations between race IAT scores and other (behavioral) phenomena while remaining agnostic to their assumed mental mediators.[1] In sum, defining implicit measure responses as behavioral effects will improve clarity and reduce bias, improving implicit measures research.

Guideline 2: Specify what you mean by 'automatic'. A second guideline for implicit measures research is to perform a precise analysis of what exactly qualifies a behavior as being "automatic" in the context of the procedure being used. Given the multiplicity of ways the term 'implicit measures' is used, and the long history of confusion it has left in its wake, we echo Corneille and Hütter's (2020) suggestion that the term be abandoned. We propose that researchers instead (a) clearly specify which properties of the behavior being captured in a given task qualify as "automatic" and (b) outline how the procedure serves to elicit the "automatic" behavior of interest. We unpack each of these recommendations in turn.

2.1. Specify and test automaticity conditions. We view automaticity as an 'orientating term', a word that serves to highlight that behavior can occur in ways that are uncontrolled, unaware, efficient, or fast (see Moors & De Houwer, 2006). Importantly, however, each of these automaticity conditions is ill-defined at the behavioral level and continuous rather than dichotomous. In addition, recent work suggests that these conditions do not map onto one another (Melnikoff & Bargh, 2018). It is not surprising then that implicit measure performance does not relate to automaticity in the way it was often thought (with all measures having all automaticity conditions; Cummins et al., 2019; Hahn et al., 2018) and it seems of little use to say that automatic behavior *in general* is measured within a given procedure (such over-simplification is often tied to

[1] Note that research designed to address questions about mental mediators of race IAT effects can also benefit from the functional approach outlined here. Defining behavior on implicit measures as automatic behaviors emitted in a specific context ensures theoretical freedom and debate at the mental level, allowing for those effects to be mediated by any number of mental processes, not just associations (for more see Hughes et al., 2011; De Houwer et al., 2020).

very broad theoretical perspectives; Melnikoff & Bargh, 2018). Instead, researchers using automatic behavior measures should always clarify (and possibly test) to what extent procedures give rise to behavior under different automaticity conditions.

In this analysis, we recommend taking into account each of the three issues noted above. First, behavioral definitions of the automaticity conditions of interest should be provided. For instance, intentionality can be functionally defined as the extent to which performance in the measure can be changed in a certain direction when instructed to do so (see De Houwer & Moors, 2007, for such an approach). Second, the continuous nature of automaticity conditions should be taken into account. We propose that researchers describe automaticity conditions in relative terms (e.g., in contrast to behavior in other measures that might be used; see De Schryver et al., 2018). For instance, researchers could test the extent to which IAT performance more strongly adheres to a condition than behavior in self-report evaluation measures. Third, it should be clarified which of the conditions of automaticity have been tested. When only one automaticity condition is met, the probed measure can be described accordingly (e.g., as a measure of fast evaluation).

For example, IAT effects have been considered to be more “unintentional” than responses on self-reported liking scales on the basis that stimulus evaluations within the IAT are not task-relevant (e.g., Banaji, 2001). When one uses the functional definition of intentionality noted above, it can be examined whether IAT scores can be intentionally shifted compared to a baseline (e.g., the baseline of no shift in effects), and the extent to which this intentional shift occurs relative to a self-report measure of the same construct. Some studies have provided evidence for the intentionality of IAT scores in this sense (Fiedler & Bluemke, 2005; Stieger et al., 2011). Similarly, the automaticity condition of awareness can be examined when defined in behavioral terms (e.g., whether people report awareness when probed). Here, research suggests that participants sometimes show a similar level of awareness of their IAT effects as of their self-reported ratings (Hahn & Gawronski, 2019).

Given that implicitness is the raison d'être for the use of implicit measures, it is both surprising and worrying that we have failed to substantively investigate the validity of the claim that these measures are in fact implicit. It therefore seems unwarranted to draw any strong conclusions about automaticity in implicit measures at present. We recommend that researchers refrain from saying that behavior in the IAT or any other task is or is not 'automatic' in general. Instead, when using implicit measures, researchers can highlight for which automaticity conditions there is behavioral evidence compared to a set value and in a given (procedural) context, or they can perform their own measurement of these conditions in the context of their experimentation. Notably, this approach is imperfect, given that there are currently no agreed-upon methods which can be used to test for different automaticity conditions. However, this is likely a by-product of failing to define automaticity conditions in functional terms. By describing automaticity as a function of environmental (i.e., task) conditions, arriving at agreed-upon methods for testing these conditions should be more easily facilitated.

2.2. Specify and test the directness of the measure. As previously mentioned, the term 'implicit measures' has been used to refer to measures that do not 'directly' ask about the behavior of interest (i.e., there are several steps required to infer the targeted behavior from the responses on the measure; see also Corneille & Hütter, 2020). We do not recommend this terminology, simply because there is no task that directly probes a given construct (i.e., a construct is always inferred on the basis of behavioral indicators, even in self-report scales) and it therefore makes little sense to refer to indirect measures as a distinct class of measures. Instead, as noted above, we propose to generally refer to implicit measures in terms of the automaticity conditions behavior is captured under. When doing so, however, we recommend taking "indirectness" into account by clarifying (and possibly testing) in what regard the phenomenon of interest is indirectly inferred from behavior in the task.

For instance, when the IAT is used as a measure of automatic evaluation, researchers should explain not only how the probed behavior is automatic (e.g., it is fast to the extent that it is emitted more quickly than behavior in a self-report measure) but also how the probed measurement index (e.g., the IAT score) relates to automatic evaluation. In an IAT, automatic evaluation is typically inferred on the basis of differences in response times in categorization. Specifically, because the procedure elicits categorization in the context of evaluative categorization of other stimuli with the same response keys, (differences in) fast categorizations of stimuli are thought to reflect automatic evaluation. Importantly, this inferential step is based on many assumptions and these assumptions can depend on procedural aspects of the measure. For instance, in the context of differences in the salience of stimuli used in the IAT, IAT scores may reflect salience asymmetries rather than automatic evaluation (see Rothermund & Wentura, 2004). When using implicit measures, researchers should clarify indirectness, testing or noting how the phenomenon of interest is inferred from behavior in the task and how this relates to procedural aspects of the selected measure.

Summary. In sum, we recommend that researchers describe implicit measures in terms of automatic behavior and conceptualize automaticity conditions from a functional perspective, as well as clarify how the phenomenon of interest relates to the measure.[2]

[2] In principle, researchers could also omit any reference to automaticity (or implicitness and indirectness) and simply refer to a measure by its name (e.g., the race IAT) without framing it as a measure of automatic behavior when they are not interested in making any claims about the construct that is measured. For instance, one could use a race IAT because it has been found to predict certain behavior. However, even in this case, the researcher inherits the theoretical assumptions which underpin IAT use in that context in the first place, and so knowledge of the probed construct is important to facilitate adequate explanation and interpretation of the findings.

Guideline 3: Choose the implicit measure that has the characteristics that you require. A meter stick might help us measure a person's height, but not the size of a subatomic particle. Relatedly, it should be clear that there is little benefit in adopting an 'off-the-shelf' or 'one-size-fits-all' approach to implicit measures. Selection of implicit measures in the absence of consideration of their automaticity conditions is likely to reduce applied (and theoretical) value. It is clear that not all implicit measures have the same predictive utility (e.g., Spruyt et al., 2013, 2015) and the approach of using any implicit measure because it is 'implicit' in a generic sense does not seem to generate the applied value that was initially promised from these measures (Lindgren et al., 2019). Instead, we recommend carefully selecting implicit measures that most closely fulfil or meet the specific conditions that characterize the phenomenon one is trying to assess/predict. In other words, select implicit measures so that there is a 'match' between the measure and the aims of the research. Researchers have already applied this logic to stimulus identity; Irving and Smith (2019) found that donations to build a border wall between Mexico and the USA were better-predicted by IATs which assessed evaluations of border walls (i.e., a good match between stimuli and outcome) compared to IATs which assessed generic evaluations of immigrants. This same matching logic can be applied to automaticity conditions: if researchers wish to predict a behavior that is unintentional, then a measure that meets the automaticity condition of unintentionality should be employed. If, for example, an addictive behavior of interest appears to occur under one particular condition of automaticity based on prior investigations, then subsequent work for predicting this behavior can be well-informed in terms of how to best optimize this prediction (i.e., by focusing on the relevant automaticity condition).

How can this selection be done? Take again the example of a race IAT. Researchers investigating automatic racial bias can select from a range of different procedures (e.g., an IAT, AMP, PEP). When they do so, we recommend they consider which automaticity conditions performances will be emitted under, and to determine if such conditions are consistent with the broader aims of their research agenda. For instance, one might be interested in people's *immediate* racial evaluations (i.e., in 'fast' behavior). Although the IAT may be a tempting candidate procedure, one first needs to ensure that adequate research exists showing that IAT performance qualifies as 'fast' relative to other tasks. Such research might be available for some measures (albeit in certain contexts; Gawronski & De Houwer, 2014), but for other measures (or for adapted measures) researchers will need to test the conditions themselves. In sum, just as improving the match between stimuli in implicit measures and to-be-predicted behavior can improve the utility of implicit measures, so too can specifying and matching the automaticity conditions under which behavior in the measure is captured and the conditions assumed to be present in the to-be-predicted behavior.

As an important note, adherence to this guideline might seem difficult given that there is ample debate on almost every well-known implicit measure in terms of the automaticity conditions that performances on that procedure do or do not exhibit (e.g., Bar-Anan & Nosek, 2014; Fiedler et al., 2006). This is unsurprising given that there are few (if any) agreed-upon methods for testing automaticity conditions (see Cummins et al., 2019). Essentially all methods of testing automaticity conditions rely on bespoke manipulations, which are problematic because the manipulations themselves have unknown measurement properties (see Chester & Lasko, 2019).

We hope that defining automaticity in terms of the properties of measurement procedures can lead to agreed-upon ways of testing these conditions. Indeed, it seems imperative that researchers focus on the development of normative methods for testing each automaticity condition which can - in principle - be applied regardless of the specific implicit measure being investigated.

One potential method to achieve this is by starting from validated nonautomatic measures, modifying them, and then testing for differences between the original and modified measures. For example, suppose a researcher wishes to develop a measure which captures body dissatisfaction under the automaticity condition of *fast*. Researchers might typically opt for an IAT or some other similar measure in this context. However, if valid nonautomatic measures of body dissatisfaction already exist, it might be more useful to start with such a measure, and then manipulate the measure in order to ensure that responding is fast (for example, by requiring responses within 1s). The original and modified measures should be essentially identical and vary only in terms of speed of responding. From here, the researcher can compare the original and manipulated measure to (i) ensure that responding is faster, and (ii) determine whether capturing such fast responding provides any incremental utility beyond the original measure (see Payne et al., 2008).

Guideline 4: Thoroughly examine and improve the psychometric properties of implicit measures. We believe that research using implicit measures needs to give more priority to work assessing and improving the psychometric properties of those measures. The assumption that implicit measures are both reliable and valid is the foundation upon which all other research, theory, and intervention proceeds. If that foundation is weak then so too are the pillars that stand upon it. For example, the extent to which theory-oriented researchers can clarify the automaticity conditions of probed behavior critically depends on progress that is made by researchers investigating the psychometric properties of specific implicit measures. Likewise, applying implicit measures requires adequate information be available about the specific psychometric properties of different measures in order to select an optimal measure. We see two general issues here.

First, questions regarding measurement generally begin and end with a report of the most common metrics of internal consistency and test-retest reliability (see Flake et al., 2017; Hussey & Hughes, 2019). Little measurement work has assessed other key assumptions in implicit measure research (e.g., that effects conform to specific assumed measurement models, etc.) despite an abundance of freely available and large scale datasets (although see Schimmack, 2019, for a recent exception). Second, when psychometric issues are highlighted, such as poor test-retest reliability, these issues are likely to either be defended as desirable in some way (e.g., implicit attitudes being seen as unstable rather than implicit measures being seen as having poor measurement properties; see Gawronski, 2019) or to be ignored. For example, despite how well-known the IAT's relatively poor test-retest reliability is, to our knowledge, no research has identified ways to remediate this (perhaps except for recommendations to aggregate responses at the group level: Payne et al., 2017). Elsewhere, suggestions to alter the procedural properties of tasks are frequently not taken up[3], and in the event that criticisms of an implicit measure are incorporated, the response of the field tends towards the development of a new measurement procedure, rather than the refinement of the extant measure (De Houwer et al., 2015; Müller & Rothermund, 2019). Such new procedures may themselves be poorly or improperly validated, which would only serve to propagate the cycle of implicit measures with undesirable measurement properties.

A much greater body of psychometric validation research is required in this regard. A reader may ask: why has this work not been done yet? The answer likely resides in the incentives provided to researchers to date. For example, studies are more likely to be published if measures are shown

[3] For example, in their methodological review of the IAT, Nosek et al. (2007) demonstrate that increasing the number of trials in Block 5 from 20 to 40 reduces the magnitude of the block-order effect between participants. However, subsequent research frequently includes only 20 trials in Block 5 on the basis that this is what was included in the original description of the IAT by Greenwald et al. (1998).

to perform well rather than poorly. As such, studies where measures perform poorly psychometrically will be less likely to be made public. For the same reason, psychometric tests that are more stringent tend to be less likely to be conducted: a good margin of literature might report reliability statistics (which measures tend to perform relatively well in) but few if any studies will report tests of measurement invariance or confirmatory factor analyses (which measures tend to fare much more poorly with; see Hussey & Hughes, 2019). Of course, in principle there is little stopping researchers from continuing to ignore or neglect psychometrics of implicit measures. However, if researchers truly wish to improve their utility, then they would do well to pay more heed to these basic psychometric barriers.

Advantages of Following these Guidelines

Theoretical value. Following our guidelines has theoretical value. Describing implicit measure performance as behavior in a specific context can help those interested in developing new or refining their existing theories. To illustrate, imagine that one wants to use an implicit measure to test their novel theory of addiction. If one follows the guidelines mentioned above, then one might start by systematically examining behavioral predictions (e.g., that specific addictive behavior depends on automatic evaluation of specific addiction stimuli) using measures that probe different types of (evaluative or addictive) behavior under different automaticity conditions (e.g., Thush et al., 2007). Once such behavioral evidence is obtained, it can facilitate further development of this theory or other theories. For instance, the study results can help scrutinize the relation between the probed behaviors and mental processes as well as the extent to which these mental processes depend on specific automaticity conditions.

Separately, following the outlined approach to implicit measures research promotes theoretical debate and avoids theoretical hegemony. Approaching implicit measure performance as an 'act-in-context' frees one up to consider how well any one theory accommodates the data which can allow not only dominant theories (e.g., that relate automatic evaluation to mental processes such as the automatic activation of associations that are acquired via stimulus pairings; Gawronski & Bodenhausen, 2006) but also other perspectives to flourish and thrive (Cone et al., 2017; Van Dessel et al., 2017). For example, taking this approach has led to new theories that explain responding on implicit (e.g., automatic evaluation) measures in terms of propositions or automatic inferences (De Houwer, 2014; Van Dessel, Hughes et al., 2019).

Another value of our approach can be found in the domain of dissociations between implicit and explicit measures (Gawronski & Brannon, 2019). The idea that inherently stable mental associations underlie implicit evaluations is often directly mapped onto empirical dissociations between implicit and explicit evaluation (e.g., in research on impression formation: Okten, 2018; racial prejudice: James, 2018; addiction: Wiers et al., 2017). Yet the approach we outline here suggests that evidence for dissociations should not to be interpreted as strong or direct evidence for distinct types of mental constructs or cognitive processes (Van Dessel, Gawronski et al., 2019). Instead, dissociations should be investigated at the functional level to allow adequate conclusions about the environmental conditions that produce these dissociations, such as the operating conditions of the measures (see also Gawronski, 2019). First, observed dissociations are often due to issues of fit between implicit and explicit measures (e.g., when the procedures are structurally – or quantitatively – dissimilar). This issue has already been raised by others (e.g., Payne et al., 2008) but is infrequently taken into account when designing research (although see Cummins & De Houwer, 2019; Moran & Bar-Anan, 2020; Van Dessel et al., 2020). Alternatively, dissociations can also reflect differences in the automaticity conditions under which behavior is emitted or elicited within a given procedure as well as differences in psychometric properties of the measures.

To examine this, studies are needed that probe effects of manipulations on multiple measures that differ only in terms of their automaticity conditions, thus providing behavioral evidence which can be used to test different theoretical explanations (e.g., associative vs. inferential processes) in a bias-free manner.

Consider the finding that the recency and diagnosticity of information play a key role in producing implicit-explicit dissociations. Specifically, diagnostic counter-attitudinal information can sometimes lead to changes in explicit but not implicit evaluation (e.g., Gregg et al., 2006; Rydell et al., 2007). Importantly, however, these studies typically use measures with very different procedures to establish these dissociations (e.g., self-reported liking scales and IAT procedures). In contrast, a recent study manipulated information diagnosticity and primacy and investigated distinct effects on evaluative responding in two AMPs that differed only in their instructions. Specifically, they instructed participants to either evaluate a target person prime or a Chinese ideograph that followed the prime and were therefore assumed to differ only in terms of intentionality, and not in terms of structural fit or any of the other automaticity conditions (Van Dessel et al., 2020; see also Payne et al., 2008). Whereas information diagnosticity had strong effects on scores on both measures and influenced the former AMP scores more strongly, information primacy did not influence scores on either of these measures, hinting at the importance of only the diagnosticity manipulation in relation to the intentionality of evaluation. This provided information unbiased by specific explanatory mental theories which could be used to update certain assumptions of dual-process and inferential theories of evaluation, facilitating theoretical precision and improving the value of these theories (see Moran & Bar-Anan, 2020, for an example of this approach in the context of testing differential effects of pairings versus relational information).

Applied value. Adopting our four guidelines can also facilitate progress in applied research. The question of whether implicit measures have something to offer in the prediction of real-life behavior can be addressed systematically, in research that directly compares the 'match' between behavior that is measured under different automaticity conditions and real-life behavior. This research can be closely tailored to those specific conditions of implicit measures that are of interest in the relevant context (e.g., probing unintentional responding). Adhering to our guidelines means that, as well as examining and improving the measurement properties of implicit measures, we should also be strategic about the contexts in which we attempt to find utility for those measures. First, it might have little use to build measures for behavior that applied fields are not immediately asking for simply because such measures are not likely to be of any use. Second, researchers should also be more selective and explicit about choosing domains and contexts in which there may be reason that an implicit measure has (added) utility. For example, rather than a broad appeal to their supposed ability to tap unconscious or unaware processes, applied researchers should attempt to define whether their behavioral phenomenon of interest is evident under a given condition of automaticity and, if so, then select an implicit measure that demonstrably captures behavior under that same automaticity condition(s). We recommend that researchers give greater thought to the fit or congruence between the behavior they are attempting to predict and understand and what they are capable of capturing within an implicit measure (see also Irving & Smith, 2019). Doing so could help specify which implicit measures are most likely to have utility in which contexts and help highlight those contexts in which implicit measures are less likely to be useful.

To illustrate, consider work on lie detection. In this applied domain there is a specific need to measure behavior that people have good reasons not to be truthful about, and by implication, for measures that can capture behavior under conditions that reduce intentionality. Rather than asking people to confirm that they recognize a stimulus, they might be asked to *reject* recognizing that stimulus by pressing a certain button. The time it takes them to do so (relative to their responses to other stimuli which they actually do not recognize) might then be used as an index of recognition. This general approach has utility for detecting previous (illegal) behavior and is regularly utilized by police forces in Japan (Matsuda et al., 2012; Verschuere et al., 2011). Here researchers have specified the automaticity condition of interest (unintentionality), and have utilized a measurement procedure which focuses on capturing unintentional responding, demonstrating the importance of considering automaticity overlap between to-be-predicted behaviors and the measurement procedure to be used.

Applied implicit measures research might particularly benefit from adhering to two specific recommendations. First, one should use those measures that have appropriate psychometric properties for the question of interest. For instance, when lasting individual differences in racial bias are of interest, it might be of little use to sample racial prejudice IAT scores which have low test-retest reliability at the individual level (Lai et al., 2014, 2016; Payne et al., 2017). Second, one should stay as close as possible to the behavior that one wants to predict. That is, the stimuli used within the procedure should relate as closely as possible to the behavior to be predicted, such as in the previously mentioned study where trying to predict border wall donations was better achieved by using an IAT that examined border wall evaluations rather than immigrant evaluations (Irving & Smith, 2019).

As a more detailed example, consider driving under the influence of alcohol (DUIA). People might not always deliberately report on their DUIA behaviour given the negative consequences of doing so. Hence there is an urgent need for a way to measure whether people have driven under the influence of alcohol that is less intentional than the measures that are currently used. In a recent study, we sought to develop a measure of unintentional evaluation of statements that people had driven drunk (Cathelyn et al., 2019). We based our research on theoretical models about beliefs underlying evaluation, but we defined the measure in behavioral terms (Guideline 1). Next, we looked at research providing evidence for measures that evoke behavior under this automaticity condition (defined as behavior being more difficult to change when instructed to do so compared to other measures) and selected a measure with this desired condition in the current context (Guideline 2.1, 3). We therefore decided to develop a variant of the autobiographical IAT (Sartori et al., 2008) that required participants to categorize sentences related to drunk driving behavior as well as other, unrelated autobiographical statements of a known truth value (e.g., “I am doing a computer task”). We assured good correspondence between the probed behavior and the measure as the aIAT’s drunk-driving sentences were related directly to having committed such an act in the past (e.g., the sentence “I have driven while drunk”) (Guideline 3). After initial testing of psychometric properties such as reliability (Guideline 4), further relevant properties of the aIAT were optimized for the context within which it was to be employed. For example, this aIAT was optimized to have good predictive utility for this specific context by calibrating score thresholds to have a low false positive rate. Though more research is needed to establish the utility of this specific measure, this example provides an initial illustration of the potential utility which keeping the noted guidelines in mind may have in developing new and optimizing existing implicit measures for applied goals.

Conclusions

The use of implicit measures has brought about many opportunities in several fields of research, but the measures have not always lived up to their original claims or promise. After 25 years, there is still a lack of clarity as to what they measure, their mediating mental mechanisms, their applied predictive utility, and their psychometric properties. While some may view this uncertainty positively (i.e., as justification for further debate and work on implicit measures), this uncertainty also begs the question: how much more time and resources should we collectively spend creating and defending these measures?

Of course, the aforementioned is beyond the scope of the manuscript. We recognize that implicit measures research will continue, and if it will continue it is better that it be done well. We have here outlined four guidelines to aid future implicit measures research in moving beyond the issues which have held the field stationary for many years. Some readers might argue that they are already aware of these guidelines. But awareness clearly does not equate to action: we have not collectively adhered to these guidelines despite knowing about them. Thus, in addition to providing these guidelines, we have tried to provide concrete ways to follow them. As such, our first recommendation is to approach performance on implicit measures as an ‘act-in-context’ (i.e., as behavior emitted within a measurement context and under certain environmental conditions; Guideline 1). We suggest that the use of mental-level theories should only be introduced after the measures are well-described at the functional level and even then, one should always avoid defining the measure in terms of mental processes. We also recommend that researchers better specify and test the key features (e.g., automaticity, indirectness) that behavior in a given implicit measure is assumed to have and provide a label that better fits these features than the label of ‘implicit measures’ (Guideline 2). Researchers should also test the match between behavior in the implicit measure and the behavior of interest when selecting measures for their research (Guideline 3). Our final recommendation is that researchers focus more acutely on assessing and improving the psychometric properties of the measurement procedures that they are using (Guideline 4).

Taking these guidelines into account has numerous benefits for theory and practice and avoids certain pitfalls. It avoids theoretical hegemony and helps to advance theory, predictions, and possibilities. It might also inform researchers about the contexts in which implicit measures are more likely to provide utility and help improve existing implicit measures. Of course, it may not be possible for every researcher to fully adhere to all the proposed guidelines. However, taking these guidelines into account when doing implicit measures research might already help the field progress and avoid continuing to succumb to the issues which have been present in the last twenty-five years of work.

References

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *Journal of personality and social psychology, 91*, 652.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), The nature of remembering: Essays in honor of Robert G. Crowder (pp. 117–150). Washington, DC: American Psychological Association.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods, 46(3)*, 668-688.

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

Brownstein, M., Madva, A., & Gawronski, B. (2019). What do implicit measures measure? *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(5):e1501.

Carter, E., Onyeador, I., & Lewis, N. A., Jr. (in press). Developing and Delivering Effective Anti-Bias Training: Challenges and Recommendations. *Behavioral Science and Policy.*

Cathelyn, F., Van Dessel, P., Cummins, J., & De Houwer, J. (2019). Driving Under the Influence of Alcohol. Retrieved from osf.io/b7wur

Chester, D., & Lasko, E. (2019). Construct Validation of Experimental Manipulations in Social Psychology: Current Practices and Recommendations for the Future. https://doi.org/10.31234/osf.io/t7ev9

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In *Advances in Experimental Social Psychology* (Vol. 56, pp. 131–199). Academic Press.

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*. 1088868320911325. https://doi.org/10.1177/1088868320911325

Corneille, O., & Stahl, C. (2019). Associative Attitude Learning: A Closer Look at Evidence and How It Relates to Attitude Models. *Personality and Social Psychology Review*, *23*(2), 161–189.

Cristea, I. A., Kok, R. N., & Cuijpers, P. (2015). Efficacy of cognitive bias modification interventions in anxiety and depression: Meta-analysis. *British Journal of Psychiatry, 206*, 7-16.

Cummins, J., & Houwer, J. D. (2019). An inkblot for beliefs: The Truth Misattribution Procedure. *PLOS ONE*, *14*(6), e0218661.

Cummins, J., Hussey, I., & Hughes, S. (2019). The AMPeror's New Clothes: Performance on the Affect Misattribution Procedure is Mainly Driven by Awareness of Influence of the Primes. https://doi.org/10.31234/osf.io/d5zn8

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are Self-Report and Behavioral Measures Weakly Correlated? *Trends in Cognitive Science, 24*(4), 267-269. 10.1016/j.tics.2020.01.007

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6,* 202–209.

De Houwer, J. (2014). A Propositional Model of Implicit Evaluation. *Social and Personality Psychology Compass*, *8*, 342–353.

De Houwer, J. (2019). Implicit Bias Is Behavior: A Functional-Cognitive Perspective on Implicit Bias. *Perspectives on Psychological Science, 14*, 835-840.

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*(1), 252–287.

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

De Houwer, J., Van Dessel, P., & Moran, T. (2020). Attitudes Beyond Associations: On the Role of Propositional Representations in Stimulus Evaluation. *Advances in Experimental Social Psychology*, *61,* 127-83.

De Schryver, M., Hughes, S., De Houwer, J., & Rosseel, Y. (2018). On the Reliability of Implicit Measures: Current Practices and Novel Recommendations. https://doi.org/10.31234/osf.io/w7j86

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510-540.

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

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the implicit association tests. *Basic and Applied Social Psychology, 27,* 307–316.

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*, 74–147.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378.

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology, 117*(3), 522-559.

Friese, M., Hofmann, W., & Schmitt, M. (2009). When and why do implicit measures predict behaviour? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology*, *19*, 285-338.

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

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

Gawronski, B., & Brannon, S. M. (2019). Attitudes and the implicit-explicit dualism. In D. Albarracín & B. T. Johnson (Eds.), *The handbook of attitudes. Volume 1: Basic principles* (2nd edition, pp. 158-196). New York, NY: Routledge.

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 edition, pp. 283–310). New York: 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* (pp. 29-55). New York, NY: Taylor & Francis.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27.

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist, 72*, 861–871. https://doi.org/0.1037/amp0000238

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90,* 1-20. doi: 10.1037/0022-3514.90.1.1

Hahn, A., & Gawronski, B. (2019). Facing one's implicit biases: From awareness to acknowledgement. *Journal of Personality and Social Psychology, 116*, 769–794.

Heycke, T., Gehrmann, S. M., Haaf, J., & Stahl, C. (2018). Of two minds or one? A registered replication of Rydell et al. (2006). *Emotion and Cognition, 32*, 1-20.

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

Hughes, S., De Houwer, J., & Perugini, M. (2016). Expanding the boundaries of evaluative learning research: How intersecting regularities shape our likes and dislikes. *Journal of Experimental Psychology: General, 145*(6), 731-754.

Hussey, I., & Hughes, S. (2020). Hidden Invalidity Among 15 Commonly Used Measures in Social and Personality Psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 166–184.

Irving, L. H., & Smith, C. T. (2019). Measure what you are trying to predict: Applying the correspondence principle to the Implicit Association Test. *Journal of Experimental Social Psychology.*

James, L. (2018), The stability of implicit racial bias in police officers, *Police Quarterly, 21*, 30-52.

Jost, J. T. (2019). The IAT is dead, long live the IAT: Context-sensitive measures of implicit attitudes are indispensable to social and political psychology. *Current Directions in Psychological Science, 28*, 10-19.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the Implicit Association

Test and intergroup behavior: A meta-analysis. *American Psychologist, 74(5),* 569–586. http://doi.org/10.1037/amp0000364

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., et al. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, 143, 1765-1785.

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., et al. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General, 145*, 1001–1016.

Larsen, H., Engels, R., Wiers, R., Granic, I., & Spijkerman, R. (2012). Implicit and explicit alcohol cognitions and observed alcohol consumption: three studies in (semi)naturalistic drinking settings. *Addiction*, *107*(8), 1420-1428.

Lindgren, K. P., Baldwin, S. A., Ramirez, J. J., Olin, C. C., Peterson, K. P., Wiers, R. W., ... & Neighbors, C. (2019). Self-control, implicit alcohol associations, and the (lack of) prediction of consumption in an alcohol taste test with college student heavy episodic drinkers. *PloS one, 14*, e0209940.

Matsuda, I., Nittono, H., & Allen, J. J. B. (2012). The current and future status of the concealed information test for field use. *Frontiers in Psychology*, *3*, 532.

Meissner, F., Grigutsch, L.A., Koranyi, N. Müller, F., & Rothermund, K. (2019). Predicting behavior with implicit measures: Disillusioning findings, reasonable explanations, and sophisticated solutions. *Frontiers in Psychology.*

Melnikoff, D. E., & Bargh, J. A. (2018). The Mythical Number Two. *Trends in Cognitive Sciences, 22*, 280-293

Mitchell, G., & Tetlock, P. E. (2017). Popularity as a poor proxy for utility: The case of implicit prejudice. In S. Lilienfeld & I. Waldman (Eds.), *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 164–195). West Sussex, England: Wiley.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz, *Implicit measures of attitudes: Procedures and controversies* (pp. 179–194). Guilford Press.

Moran, T., & Bar-Anan, Y. (2020). The effect of co-occurrence and relational information on speeded evaluation. *Cognition and Emotion, 34*(1), 144-155.

Müller, F., & Rothermund, K. (2019). The Propositional Evaluation Paradigm (PEP): Indirect Assessment of Personal Beliefs and Attitudes. *Frontiers in Psychology.*

Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the "suicidal mind:" Implicit cognition predicts suicidal behavior. *Psychological Science, 21*, 511-517.

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

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In JA. Bargh (Ed.), *Automatic processes in social thinking and behavior* (pp. 265–292). New York, NY: Psychology Press.

Okten, I. O. (2018). Studying First Impressions: What to Consider? *APS Observer, 31*.

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

Payne, B. K., 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., Vuletich, H. A., & Lundberg, K. B. (2017). The Bias of Crowds: How Implicit Bias Bridges Personal and Systemic Prejudice. *Psychological Inquiry, 28*, 233-248.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*, 139–165.

Rydell, R., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit attitudes respond differently to increasing amounts of counterattitudinal information. *European Journal of Social Psychology, 37*, 867–878.

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., & Castiello, U. (2008). How to accurately detect autobiographical events. *Psychological Science*, *19*(8), 772–780.

Schimmack, U. (2019). The Implicit Association Test: A Method in Search of a Construct. *Perspectives on Psychological Science.*

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108–131.

Spruyt, A., De Houwer, J., Tibboel, H., Verschuere, B., Crombez, G., Verbanck, P., Hanak, C., Brevers, D., & Noël, X. (2013). On the predictive validity of automatically activated approach/avoidance tendencies in abstaining alcohol-dependent patients. *Drug and Alcohol Dependence, 127*, 81-86.

Spruyt, A., Lemaigre, V., Salhi, B., Van Gucht, D., Tibboel, H., Van Bockstaele, B., De Houwer, J., Van Meerbeeck, J., & Nackaerts, K. (2015). Implicit attitudes towards smoking predict long-term relapse in abstinent smokers. *Psychopharmacology, 232*, 2551-2561.

Stieger, S., Goritz, A. S., Hergovich, A. & Voracek, M. (2011). Intentional faking of the single category Implicit Association Test and the Implicit Association Test. Psychological Reports, 109, 219-230

Tello, N., Harika-Germaneau, G., Serra, W., Jaafari, N., & Chatard, A. (2018). Forecasting a Fatal Decision: Direct Replication of the Predictive Validity of the Suicide-Implicit Association Test.

Thush, C., Wiers, R. W., Ames, S. L., Grenard, J. L., Sussman, S., & Stacy, A. W. (2007). Apples and oranges? Comparing indirect measures of alcohol-related cognition predicting alcohol use in at-risk adolescents. *Psychology of Addictive Behaviors, 21*(4), 587-591.

Van Dessel, P., Cone, J., Gast, A., & De Houwer, J. (2020). The Impact of Valenced Verbal Information on Implicit and Explicit Evaluation: The Role of Information Diagnosticity, Primacy, and Memory Cueing. *Cognition & Emotion.*, *34*, 74-85.

Van Dessel, P., Gawronski, B., & De Houwer, J. (2019). Does explaining social behavior require multiple memory systems. *Trends in cognitive sciences, 23(5)*, 368-369.

Van Dessel, P., Hughes, S., & De Houwer, J. (2019). How Do Actions Influence Attitudes? An Inferential Account of the Impact of Action Performance on Stimulus Evaluation. *Personality and Social Psychology Review, 23*, 267-284.

Van Dessel, P., Gawronski, B., Smith, C. T., & De Houwer, J. (2017). Mechanisms underlying approach-avoidance instruction effects on implicit evaluation: Results of a preregistered adversarial collaboration. *Journal of Experimental Social Psychology, 69*, 23-32.

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (2011). Memory detection: Theory and application of the concealed information test. Cambridge University Press.

Wiers, C. E., Gladwin, T. E., Ludwig, V. U., Gröpper, S., Stuke, H., et al. (2017) Comparing three cognitive biases for alcohol cues in alcohol dependence. *Alcohol, 52*, 242-248.