Received: 10 December 2020 | Revised: 27 April 2021 | Accepted: 9 May 2021

DOI: 10.1002/wcs.1569

ADVANCED REVIEW

WIREs COGNITIVE SCIENCE WILEY

Anomalies in implicit attitudes research

Edouard Machery

Center for Philosophy of Science, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Correspondence
Edouard Machery, Center for Philosophy of Science, University of Pittsburgh, 1117 CL Pittsburgh PA 15260.
Email: machery@pitt.edu

Edited by: Wayne Wu, Editor

Abstract

In this review, I provide a pessimistic assessment of the indirect measurement of attitudes by highlighting the persisting anomalies in the science of implicit attitudes, focusing on their validity, reliability, predictive power, and causal efficiency, and I draw some conclusions concerning the validity of the implicit bias construct.

This article is categorized under:

Psychology > Reasoning and Decision Making

KEYWORDS

bias, construct validity, implicit attitude, indirect measure

1 | INTRODUCTION

Brownstein et al. (2019) have put forward an insightful review of the theoretical issues raised by the research on implicit attitudes. The premise of their article is that "The advent of implicit measures has given researchers tools to capture unintended biases in precise and empirically tractable ways" (Brownstein et al., 2019, p. 1). The goal of my response is to provide a much more pessimistic assessment of the indirect measurement of attitudes by highlighting the persisting anomalies in the science of implicit attitudes, and to draw some conclusions concerning the validity of the implicit bias construct (see also Mitchell & Tetlock, 2017).[1,2]

Indirect measures of attitudes have been put to different uses in psychology and cognitive neuroscience, and the anomalies discussed in this article do not bear on all these uses equally. I will be mostly concerned with the indirect measurement of attitudes in relation to the prediction of real-life outcomes at the individual level, but other uses of these measures will be discussed at the end of this article.

Section 1 briefly contextualizes the research on indirect measures historically, by looking both at the long-term history of social psychology and at the shorter-term history of implicit attitudes research. This contextualization explains why critics have expressed the concerns they have. Section 2 turns to a first aspect of the validation of the implicit attitude construct—the discriminant validity of indirect measures—and argues that after 30 years of research discriminant validity is astoundingly still up for grabs. Section 3 examines the low reliability of indirect measures. Section 4 discusses their predictive validity. Section 5 turns to the issue of causality. Section 6 brings together the points discussed in Sections 2–5 to assess the validity of the implicit attitude construct as well as psychologists' propensity to advertise at best unproven, at worse false scientific theories to policy makers. Finally, Section 7 briefly examines whether indirect measures could be saved by focusing on other areas of research.

2 | HISTORICAL PREAMBLE

While the exact definition and the ontological status of attitudes have been extensively debated (see, e.g., Machery, 2016), at the very least they manifest themselves in (if they are not identified with) behavioral preferences for or against

the object of these attitudes. This idea is common to the otherwise divergent definitions of attitudes found throughout the decades of implicit attitude research. Thurstone defined them as follows (Thurstone, 1928, p. 531):

> The concept "attitude" will be used here to denote the sum total of a man's inclinations and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions about any specified topic.

In a later work, he defined an attitude more narrowly as "the affect for or against a psychological object" (Thurstone, 1931, p. 261). Allport (1935) defined attitude as follows:

> An attitude is a mental and neural state of readiness, organized through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related.

He further added:

> An attitude characteristically provokes behavior that is acquisitive or avertive, favorable or unfavorable, affirmative or negative toward the object or class of objects with which it is related. This double polarity in the direction of attitudes is often regarded as their most distinctive feature.

More recently, in the introduction of their influential book, Eagly and Chaiken (1993, p. 1) largely concur, defining an attitude as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (see also Eagly & Chaiken, 2007). Petty, Wegener, and Fabrigar (1997, p. 611) concur: "[a]ttitudes have been defined in a variety of ways, but at the core is the notion of evaluation" (see also Banaji & Heiphetz, 2010; Briñol & Petty, 2012). In the present context (e.g., Greenwald & Krieger, 2006, pp. 950–951; for useful discussion, see Holroyd et al., 2017, section 5), biases are just morally objectionable attitudes toward groups: racism, sexism, ageism, and so on.[3]

Measuring attitudes (toward groups, individuals, brands and products, ideas such as democracy, or activities such as professions) was the raison d'être of social psychology in the 1920s soon after its emergence as a distinct discipline (Allport, 1935; Danziger, 1994; McGuire, 1986) and a constant goal since then. Since 1935, every edition of the *Handbook of Social Psychology* has included a chapter on attitudes and attitude change (Banaji & Heiphetz, 2010). Attitude measurement has however been hampered by many difficulties, one of which is the possibility that people may be unwilling or unable to report their attitudes. Racists may not be willing to express their racism, and the reported attitude may poorly reflect their true attitude. People may also not be aware of their attitudes, and thus may not be able to report them accurately.

To address the first difficulty, social psychologists in the 1960s and 1970s turned toward "unobtrusive" measures of attitudes, which attempt to bypass respondents' response bias (Crosby et al., 1980; Webb et al., 1999), and contemporary *indirect* measures of attitudes are the heir of these measures. Contemporary indirect measures also reflect the exploding interest in automatic processes in cognitive psychology in the 1970s and 1980s. When attitudes are indirectly measured, instead of asking people to answer questions about their attitudes (as is done with, e.g., feeling thermometers) or questions about the objects of their attitudes (e.g., whether they believe that "Discrimination against blacks is no longer a problem in the United States," as asked in the Modern Racism Scale), people are asked to engage in tasks that do not involve questions about their attitudes or their objects, but that are supposed to index, to some degree, the automatic expression of their attitudes: As a result, people's performance measures the strength of their attitude.[4] Well-known tasks, such as the Implicit Attitude Test (Greenwald et al., 1998), the Affect Misattribution Procedure (Payne et al., 2005), or the Evaluative Priming Task (Fazio et al., 1986), illustrate this indirect approach to the measurement of attitudes.

A few articles have been turning points in the contemporary development of indirect measures of attitudes. Fazio et al.' (1986) article about what he called the "bona fide pipeline" to attitudes, which was based on the then recent research on priming in cognitive psychology, opened the road for the contemporary research on indirect measures of attitudes (Fazio et al., 1995), and its influence is evidenced by its more than 3000 citations. Fazio's work was kindred to the work on the automatic aspect of prejudice that emerged in the 1980s, particularly with Dovidio and Gaertner's (1986) extremely influential work on aversive racism and Devine's (1989) on the automatic components of prejudice. A decade later, Wittenbrink et al. (1997) appear to be the first to refer to "implicit prejudice" in the journals indexed in the PsycINFO database (Mitchell & Tetlock, 2017). Of crucial importance is the article published in 1998 by Greenwald, McGhee, and Schwartz, which introduced the Implicit Association Test (IAT), the most commonly used indirect

measure of attitudes. Mitchell and Tetlock (2017) make a compelling case that the publication of the IAT in 1998 was a turning point for the research on implicit attitudes from both a scholarly and a scientific popularization perspective. In and outside academia, implicit attitude research is sometimes simply identified with the use of the IAT. Greenwald et al.'s article is one of the most cited of the *Journal of Personality and Social Psychology*, with more than 4000 citations. Banaji, Greenwald, and Nosek held a press conference in 1998, reporting that the IAT had revealed that most people hold unconscious prejudice and expressing hope that this test could help people address their prejudice (Mitchell & Tetlock, 2017). The IAT was immediately discussed in popular media (e.g., Goode, 1998). The scientific and popular impact of Greenwald et al. (1998) was compounded by the creation of a public website in 1998, where people could take IATs, and received feedback about their scores (https://implicit.harvard.edu/implicit/). Millions of people have logged in, contributing to the research and social impact of implicit attitude researchers. Other indirect measures of attitudes, sometimes based on different psychological assumptions, have been developed since the IAT, but they have not received the same scientific and lay attention. Remarkably, while scientists have typically been careful to distinguish these measures, comparing for instance their psychometric properties (e.g., Bar-Anan & Nosek, 2014), they have sometimes been less careful when they theorize about implicit attitudes in general. Similarly, the significant differences between the theories about implicit attitudes or automatic prejudice are sometimes glossed over (Mitchell & Tetlock, 2017). Differences between measures and theories are almost always erased in public discussions.

The construct of implicit attitudes provides a distinct interpretation of the indirect measurement of attitudes: It assumes that direct and indirect measures measure two distinct (but not necessarily entirely unrelated) types of attitudes (e.g., Banaji & Heiphetz, 2010; Buttrick et al., 2020; Charlesworth & Banaji, 2019; Greenwald et al., 1998; Nosek et al., 2007; Rydell & McConnell, 2006; Strack & Deutsch, 2004; Wilson et al., 2000).[5] Early research took implicit attitudes to be both unconscious, in the sense that people are unaware of having them (e.g., Greenwald & Banaji, 1995), and automatic, but, as Brownstein et al. (2019) explain, recent research has cast doubt on this once common characterization (e.g., Gawronski et al., 2006; Hahn et al., 2014; Hahn & Gawronski, 2019). Many social psychologists now only appeal to automaticity to characterize implicit attitudes (e.g., Van Dessel et al., 2020; Vianello & Bar-Anan, 2021). While the dual approach to attitudes is dominant in social psychology, some of the leading psychologists who developed indirect measures reject it (e.g., Fazio, 2007).

Indirect measures are often presented as measuring *individuals'* attitudes: They tell something about the attitudes of each test taker. Greenwald et al.' (1998) article was entitled "Measuring Individual Differences in Implicit Cognitions." More recently, Rae and Olson (2018, p. 309) have observed that "developmental researchers have treated the IAT as an individual difference measure." Popular presentation of implicit attitude research has not shied away from presenting indirect measures as revealing the hidden depths of individual minds. Greenwald and Banaji's influential book, *Blindspot*, described the case of a gay activist whose IAT scores apparently indicated negative associations toward gays (Greenwald and Banaji, 2017, p. 56). Greenwald and Banaji did not indicate that it might be a mistake or at least careless to infer the valence and strength of someone's attitude from an IAT score. Rather, they endorse this inference, writing about Banaji's own indirect measure scores (p. 57):

> This gap between Mahzarin's reflective and automatic reactions to the same thing (Blacks-Whites attitudes) would not have been unveiled without the insight the test inflicted. The powerful message that the test gave her about the force of the unconscious has been among the most significant self-revelations she has experienced, impressing upon her the full import of the observation that the 'self is more distance than any star.'

Similarly, as a guest of an NPR show on "tests that can reveal your hidden bigotry" in 2008, Nosek explained the importance of indirect measures in terms of the capacity to modify our own individual biases (www.npr.org/transcripts/93137786):

> [I]f we have associations in our minds that we don't like. Then what's the implications of that? The first step of changing associations that we don't like is knowing that we possess them. So, from my own practical interest as a person, realizing that I have implicit race biases very much like your guest, Maureen, is starting to think about, well, what is it in my environment that I can change so that I might develop different associations, so that these things aren't dominating the behavior that I have everyday.

Indirect measures are also supposed to be *predictive* at the *individual* level. Rae and Olson (2018, p. 309) noted that "predicting behavior is a key motivation behind the use of implicit measures." Similarly, Greenwald and Banaji (2017) wrote that "It [the IAT] predicts discriminatory behavior even among research participants who earnestly (and, we believe, honestly) espouse egalitarian beliefs." Note that Greenwald and Banaji's claim is not that *sometimes*, or in *some* circumstances, the IAT is predictive.

Implicit attitudes are also taken to be *causally* responsible for various outcomes. Devine et al. (2012) wrote the following (p. 1267; emphasis added):

> [A]ccumulating evidence reveals that implicit biases *are linked to* discriminatory outcomes ranging from the seemingly mundane, such as poorer quality interactions (...), to the undeniably consequential, such as constrained employment opportunities and a decreased likelihood of receiving life-saving emergency medical treatments (...).[...] [Implicit bias] *leads people* to be unwittingly complicit in the perpetuation of discrimination.

There is no doubt that Devine et al. have in mind not a correlation, but a causal relation between indirect measures and discriminatory outcomes. Greenwald and Banaji's *Blindspot* (Greenwald & Banaji, 2017) is replete with causal language, connecting unconscious and automatic processes to behavior. For instance, they ask, "what caused the gap between the professor's reflective intention and his actual behavior?", and they respond "automatic associations" (Greenwald & Banaji, 2017, p. 55). Similar causal language is found in Eberhardt's (2020) *Biased: Uncovering implicit prejudices that shapes what we see, think, and do*. Philosophers who have embraced the research on implicit attitudes are no different. Madva, for instance, writes (Madva, 2017, p. 146; emphasis added) that "Implicit biases *influence* whom we trust and whom we ignore, whom we promote and for whom we vote. They *affect* interactions between teachers and students, doctors and patients, police and civilians, and lawyers and jurors."

Finally, implicit biases are connected to socially significant ills, when they are not simply taken to be their main causes. Implicit racism has often been taken to be a cause of the continuing racial ills that pervade the United States and implicit sexism is regularly blamed for social inequalities among genders.

It may be unfair to paint all psychologists with a single brush: While influential and to some extent the public face of implicit attitude research, Greenwald and Banaji are only two of the many psychologists whose research relies on indirect measures of attitudes. And indeed social psychologists have sometimes attempted to promote a more accurate assessment of the research on implicit attitudes (e.g., Payne et al., 2017; Van Dessel et al., 2020). On the other hand, as the references and quotations above make clear, Greenwald, Banaji, Eberhardt, Devine, Nosek, and others such as law Professor Kang have repeatedly made it clear that for them indirect measures tell something about individuals, measuring a distinct construct that influence behavior in a way that is socially meaningful. It may also be unfair to take scientists' pronouncements in popular presentations of their work as reflecting their considered opinion, but similar claims are made in scientific articles, and in any case scientists should be extra-careful when talking to lay people rather than overlooking complexities for the sake of painting an appealing, simple picture.

Americans have been particularly receptive to implicit bias research. Pundits have echoed implicit attitudes researchers' claims. *New York Times* columnist Nicholas Kristof reported on Banaji's work as follows (May 7, 1995): "I encourage you to test yourself at implicit.harvard.edu. It's sobering to discover that whatever you believe intellectually, you're biased about race, gender, age or disability." In a column written nearly a decade before, he confessed, "To my horror, I am a racist" (April 6, 2008), after discussing first-person shooter tasks. Two assumptions are noteworthy: Indirect measures are supposed to measure individuals' attitudes, and, when they are biases, whatever they measure can be identified to, or at the very least connected to, racism, sexism, and other forms of social ills: Whatever it is they measure matters. Of course, scientists are not responsible for the simplifications introduced by journalists, politicians, and pundits, but social psychologists' message to society did nothing but encourage these simplifications.

In addition to pundits, institutions, from the police, to the bench, to school districts, to universities, to companies, as well as politicians and policy makers have often bought into implicit attitude research, particularly when it bears on socially sensitive topics such as racism and sexism. To give a few examples among many, Obama's Task Force on 21st Century Policing recommended implicit-bias training to law enforcement; the FBI started an implicit bias training program for their 28,000 employees in 2016; the New York Police Department started its own in 2018; in 2020, Connecticut Governor Ned Lamont signed a bill requiring the police to undergo implicit bias training, and so did Governor Phil Murphy for state, county, and municipal law enforcement departments in New Jersey. Universities often require or encourage implicit bias training for staff and faculty: The University of California has added a six-course series on

implicit bias as a core requirement to the "existing UC Systemwide People Management Series and Certificate"; the University of Nevada, Reno requires "all persons serving as members of search committees and all persons who participate in one or more of the functions of a search committee" to "complete the Integrated Implicit Bias Search Committee Training"; corporate stores and offices are regularly closed for sometimes costly implicit bias training (e.g., Starbucks: Meyer, 2018). Needless to say, implicit bias training is extremely costly for the relevant institutions and also very profitable for those offering the training: The New York Police Department's program costs $4.5 million. And, important to stress, there is little evidence that implicit bias training has any effect on people's behavior (Section 5). The assumptions behind the trend illustrated by these examples are identical to the ones identified in Kristoff's articles above. They are well illustrated on the website of the Association of American Medical Colleges: "In 2012, each member of The Ohio State University College of Medicine admissions committee took various implicit association tests (IATs). When results showed that many committee members exhibited implicit white race preference, implicit bias against homosexuals, and unconscious association of men with "career" and women with "homemaker," we initiated annual implicit bias mitigation workshops." Policy makers and politicians have followed suit. In 2016, Hillary Clinton asserted that "implicit bias is a problem for everyone" in the first debate with Donald Trump, of all people, and implicit bias appeared again in the vice-presidential debate in 2016. New York City Mayor Bill de Blasio has pushed for implicit bias training since his 2016 State of the City Address.

Criticism of the validity and significance of indirect measures of individuals' implicit biases have been around for years too. Psychologists expressed technical concerns in scientific venues (e.g., Arkes & Tetlock, 2004; Blanton & Jaccard, 2006), which were then echoed in the popular media. As early as 2008, New York Times columnist John Tierney reported on Hart Barton's and others' criticism of the IAT and other indirect measures (November 17, 2008) as well as the already brewing tensions about implicit bias research (Tierney, 2008). Critical reports about indirect measures of biases have also become more common in recent years (e.g., Bartlett, 2017; Singal, 2017a, 2017b), sometimes with a political slant (MacDonald, 2017)

3 | THE DISCRIMINANT VALIDITY OF INDIRECT MEASURES

As noted above, indirect measures can be understood as measuring a distinct type of attitudes, namely implicit attitudes, or as a different way of measuring a single construct, attitudes. The former interpretation, which reifies indirect measures, has received the most attention among academics and outside academia. This dual-attitude interpretation of indirect measures is justified by two main considerations. First, direct and indirect measures of attitudes often, though not always, correlate poorly with one another (e.g., Buttrick et al., 2020; Greenwald et al., 1998; Greenwald & Farnham, 2000; Izuma et al., 2018), a phenomenon illustrated by the frequently described self-conscious liberal who nonetheless displays racist and sexist behavior (e.g., Kristof, 1995; Greenwald & Banaji, 2017, p. 866). Greenwald et al. (1998) put the point as follows (p. 1470): "This conceptual divergence between the implicit and explicit measures is of course expected from theorization about implicit social cognition." Second, indirect measures of attitudes have predictive validity above and beyond direct measures (incremental predictive validity): They still correlate with the criteria once correlations between criteria and direct measures are partialled out (e.g., Greenwald et al., 2009). In addition, Nosek and Smyth (2007) argued that across domains a two-factor model fit data better than single-factor model (see also Bar-Anan & Vianello, 2018).

Surprisingly, however, *30 years* after the implicit revolution, the question is not settled. Schimmack (2021) has recently challenged the dual-attitude interpretation of the IAT. Low correlations between direct and indirect measures provide little evidence that they measure different constructs since such correlations can merely result from either measure's low reliability or low validity, not from them measuring different constructs (see also Dang et al., 2020), or from having low correspondence, that is from being directed at different objects (e.g., Gawronski, 2019; Payne et al., 2018).[6] What's more, Schimmack's reanalysis of existing data sets suggests that indirect and direct measures' factors are very highly correlated—providing no evidence for the reality of two distinct constructs—and that indirect measures have reasonable validity for some objects (such as political preferences), at best modest validity for some objects (e.g., race), and no validity for others (self-concept).

Incremental validity can itself be due to measurement error (Schimmack, 2021), which was not taken into account in the earlier metaanalyses of the predictive validity of indirect measures (Greenwald et al., 2009; Oswald et al., 2013; but see more recently Kurdi et al., 2019). What's more, in contrast to previous studies, Schimmack found no evidence for the incremental validity of indirect measures related to race at the individual level.

Proponents of the dual-attitude interpretation of indirect measures have responded to Schimmack (Kurdi et al., 2021; Vianello & Bar-Anan, 2021), and the issue remains under examination. Vianello and Bar-Anan (2021) show that

Schimmack's pessimistic conclusions depend in part on modeling decisions, but for the reader concerned with the construct validity of implicit attitudes, their response demonstrates that the available evidence concerning the indirect measures of individuals' attitudes is not robust to modeling decisions. For our purposes, the issue is that a basic issue in implicit attitude research—what do indirect measures measure?—is still unanswered. What's more the low validity of some indirect measures, including the most well-known measures such as the IAT, shows that they are poor measures of *individuals'* attitudes, defeating the push for a translational use of indirect measures to deal with, among others, biases at the individual level.

One could perhaps object to generalizing from the failure of the IAT's discriminant validity to the failure of indirect measures' validity in general, but what other indirect measures happen to measure is also unclear. Cummins et al. (2019) have recently argued that the Affect Misattribution Procedure does not measure anything implicit and is not a valid measure of attitudes because effects are driven by a subset of participants who respond to the allegedly unattended prime.[7]

I turn to other anomalies in implicit attitude research in the remainder of this article. Since it is surprisingly still unclear what indirect measures measure, I will use the term "measurand" to refer to their object, remaining agnostic about its nature.

4 | THE RELIABILITY OF INDIRECT MEASURES

The reliability of a measure is the proportion of the variance in test scores for a given population of test takers that is explained by their true scores. The reliability of a measure yields an upper bound to its validity (Lord & Novick, 1968, p. 72), and unreliable measures only partly measure what they are intended to measure. Unreliability also means that indirect measure scores should not be taken at face value, except when they come with a confidence interval. From the perspective of scientific communication, it is irresponsible not to correct people's disposition to take the scores of unreliable indirect measure at face value.

How reliable are indirect measures of attitudes? Numerous studies have shown that indirect measures are not very reliable, and typically less so than explicit measures. Unreliability can be estimated in various ways, including by examining test–retest reliability, that is the similarity of scores across occasions for a given experimental unit (e.g., a person, a region, a group). Gawronski et al. (2017) examined the test–retest reliability (measurement times separated by several weeks) of the IAT and the Affect Misattribution Procedure (AMT) in three domains—self-concept, race, and political orientation. Both indirect measures showed weak reliability—between $r = 0.38$ (AMT for racial attitudes) and $r = 0.64$ (IAT for the self-concept)—and were less reliable than direct measures for all domains. Reanalyzing more than 30 studies that vary widely in terms of domains, method, and time intervals, they report a mean correlation equal to 0.41. Rae and Olson (2018) report similar test–retest reliabilities for the IAT given to children and teens for the domains of race and gender. In both domains, most explicit measures are more reliable.

Variation in indirect measure scores could either result from the variation in the measured trait or from variation in the host of extraneous factors that influence performance during indirect measurement (i.e., from measurement error). Brownstein et al. favor the former interpretation (Brownstein et al., 2019, section 4), but it is remarkable that in the third decade of research on indirect measures we still do not know whether this is actually the case. What's more, the low validity of many indirect measures (Section 2) suggests that test performance is influenced by many different factors in addition to the measured trait. It is thus likely that the low test–retest reliability is due to variation in one or several of these extraneous factors rather than in the measured trait.

Brownstein et al. (2020) address the low reliability of indirect measures in an insightful manner: They invite us to set our expectations appropriately, by noting that unreliable measures such as heart rate and blood pressure measurements are successfully used outside psychology. This response is however too quick since whether low reliability is an issue depends on what use a measure is put to. For instance, if the point is to measure large deviations from a baseline of repeated measurements, low reliability may not be a serious issue. The situation is different when the point is to measure individual differences so as to predict or explain behavior.

5 | THE PREDICTIVE VALIDITY OF INDIRECT MEASURES

Much of the critical discussion of indirect measures has focused on their predictive validity, following the conflicting metaanalyses of Greenwald et al. (2009) and Oswald et al. (2013). The stakes are high, since some took the predictive validity of some indirect measures to validate them (Banaji & Heiphetz, 2010, p. 360):

> Questions of validation have been addressed most reassuringly though through studies of the relationship between IAT scores and behaviors that satisfy the desire for ecological validity.

However, Oswald et al. found a lower zero-order correlation between the IAT and behavioral measures than Greenwald et al. (Greenwald et al., 2009, $r = 0.27$; Oswald et al., 2013, $r = 0.14$), and concluded that "the IAT provides little insight into who will discriminate against whom, and provides no more insight than explicit measures of bias. (...) social psychology's long search for an unobtrusive measure of prejudice that reliably predicts discrimination must continue" (p. 188). Oswald and colleagues also reported a minuscule incremental predictive validity of the IAT in the domains of race (0.1–2.0% of additional predicted variance) and ethnicity (0.2–5.4%). As we have seen, Schimmack (2021) reports no incremental predictive validity of race indirect measures.

Some psychologists have acknowledged that the predictive power of indirect measures is disappointing (Meissner et al., 2019; Van Dessel et al., 2020), but others have dismissed this concern, focusing on the accumulation of small effects, a point originally made in the discussion of sexism (Valian, 1999; see also Mallon, 2016; Buttrick et al., 2020). The idea is that small effects accumulate, either over time or across people, and end up being socially significant. Greenwald, Banaji, and Nosek (2015, p. 558) have given the following fictional example in their response to Oswald et al. (2013):

> Using OMBJT's $r = 0.148$ value as the IAT–profiling correlation generates the expectation that, if all police officers were at 1 SD below the IAT mean, the city-wide Black–White difference in stops would be reduced by 9976 per year (5.7% of total number of stops) relative to the situation if all police officers were at 1 SD above the mean.

This line of argument is problematic. First, it is unjustified to assume that the real-world impact of implicit attitudes is of the same magnitude as the effect sizes observed in experimental conditions, when the criteria are often not behavioral (e.g., preferences about policies, imagined behavior, intentions to act, or perceptions of others) or consist of behaviors that, while perhaps not harmless, are definitely much less harmful than many real-world biased behaviors (e.g., body posture, emotion display, distance from someone, etc.). A more realistic effect size for real world impact may undercut the social significance of the measurand of indirect measures. Second, while possible, it is at best an unsubstantiated hypothesis that when they combine across times, the effect sizes combine additively rather than in a way that increasingly dampens their effect. These two issues point toward the crucial problem for this first response: While cumulative models of the effects of implicit biases should definitely be considered, they are empirical hypotheses that call for empirical evidence, which is still missing in the third decade of indirect measures.

A second response, discussed in Brownstein et al. (2019, section 6), argues that the correlations between indirect measures and behavioral criteria are in fact higher than reported in Oswald et al. (2013), "when the variables specified by extant theories are considered" (Brownstein et al., 2019). This second line of response often draws on the results reported by Kurdi et al. (2019). On average the IAT-criterion correlation reported by Kurdi et al. is of the same size as the zero-order correlation reported by Oswald et al. ($r = 0.10$; see also $r = 0.09$ in Forscher et al., 2019), but they also identify contexts in which the correlation between IAT and the behavioral criteria appears higher (up to $r = 0.37$ when all contexts are taken into account) such as contexts where these correlations obey Ajzen and Fishbein's (1977) principle of correspondence (viz., roughly, correlations between measures and behavior should be higher when the object of the attitude and the behavior are similar).[8]

While suggestive, these higher correlations are unlikely to convince a skeptic. First, implicit attitude researchers were not qualifying the predictive validity of indirect measures when they toted their new tools in scientific and public contexts, as we saw earlier. Second, contexts are bound to exist where correlations will be *overestimated* and others where they will be underestimated. That higher correlations are found in some contexts is uninformative since we do not know whether this is merely due to an overestimation of a smaller true correlation or an accurate estimation of a genuinely larger true correlation. Kurdi et al. did not preregister their predictions about which contexts would lead to higher correlations and merely examined various contexts to find those with higher correlations, and one should be concerned that they are merely capitalizing on random variation. Following Kurdi et al., Brownstein et al. (2019, 2020) attempt to fend off this concern by arguing that the contexts where higher correlations were found are theoretically meaningful, but one suspects that a theoretical reason would have been found for any context in which the correlations were found (Kerr, 1998). Identifying contexts in which correlations are genuinely higher is an important area of research, but theoretically-driven, *preregistered* studies are called for. What's more, contrary to what Brownstein et al.

(2019, 2020) suggest, it is not the case that theoretically predicted moderators are always or even often observed. One of the highlights of Kurdi et al. (2019) is that, contrary to what most theories predict, the correlation between indirect measures and behavior is not moderated by controllability, awareness, or social sensitivity (on the latter, see also Hofmann et al., 2005). What's more evidence that correspondence increases the predictive validity of indirect measures is limited since in Kurdi et al. (2019) correspondence in fact failed to moderate the correlation between indirect measures and behavior when the variables were blind-coded (see also Section 4).

This response also largely undermines the appeal of indirect measures, particularly in a translational context: If indirect measures are only predictive in narrow contexts, then it makes no sense to conclude that one is racist, sexist, and so on, when one receives an apparently damning IAT score since racism and other biases manifest themselves across contexts: Racists think, speak, and act racist in many different contexts, although of course not invariably. It also makes little sense to blame the measurand of indirect measures for social ills.

A third, important response focuses on the kind of effect size we should expect, and argues that the correlations between indirect measures and criteria are as expected (Brownstein et al., 2019, section 6; Brownstein et al., 2020). Appearances to the contrary are merely due to inflated expectations about the correlations between behavior and psychological measures. Brownstein et al. (2019) note that psychologists have long agreed that when contextual mediators are not taken into account, measures of attitudes should only predict behavior weakly, and indirect measures do as expected in this regard: "Not a single meta-analysis of implicit measures has reported nonsignificant correlations close to zero or negative correlations with behavior." Brownstein et al. (2020) also note that the effect size is in line with the typical predictive validity of psychological measures, such as IQ, SES, SAT, and so on. They also remind us that behavior is multiply determined, and that we should thus not expect the measure of a single factor to correlate highly with behavior.

This third response is the most compelling, and previous critical discussions of the low predictive power of indirect measures (e.g., Machery, 2017) may have been in part driven by failing to calibrate expectations. It is however important to calibrate expectations properly and not to choose standards in a self-serving manner. Possible standards range widely, from the correlation between aspirin and the probability of a heart attack (<1% in absolute reduction) to the correlation between IQ and academic performance ($r > 0.5$; Kobrin et al., 2008) and between conscientiousness and academic performance ($r = 0.46$; Poropat, 2009). A difference between a 0.1 and 0.3 correlation might not seem much, but the latter accounts for nine times for variance than the former. Second, indirect measures of attitudes differ from other measures such as IQ in that they were supposed to measure something that existing, direct measures failed to capture, but their incremental predictive validity—what they add to direct measures—is low.

In any case, this response concedes a lot to critics: Saying that indirect measures do not do worse than other measures is not saying they do well. In effect, this response throws indirect measures on the heap of social–psychological measures that cannot be used to predict how people are going to behave as individuals. Banaji, Greenwald, and others could have been clear that indirect measures predict behavior very poorly, but the scientific community and the larger public would not have been fascinated by their research, and social psychologists would not have become public figures and consultants. And let us not forget, the difference between what science tells us and how it is depicted and reported may not be costless since it detracts efforts from other, perhaps more significant ways to addressing social ills such as racism and sexism.

Claims about the incremental predictive validity of indirect measures such as the IAT face an additional difficulty: When measurement error is not controlled for, the risk of Type I error is inflated when scientists claim that indirect measures have incremental predictive validity (Westfall & Yarkoni, 2016). Kurdi et al. (2019) addressed this issue, but as they acknowledged, their conclusions were tentative since most studies did not report the relevant information, compelling them to impute missing data. Buttrick et al. (2020) have examined this question in turn, showing that, controlling for measurement error, the IAT has incremental predictive validity for the outcome measures they focus on. However, while suggestive, this recent analysis says little about the impact of the measurand of indirect measures on behavior since almost none of the outcome measures were behavioral. It also assumes that direct and indirect measures measure different latent variables, following Nosek and Smyth (2007), a problematic assumption in light of Schimmack's (2021) modeling work. Be that as it may, three features emerge from their analysis: The IAT is not predictive for about 30% of outcome variables; it does not have incremental predictive validity for nearly 60% of the outcome variables; and when corrected for measurement error the incremental predictive validity is very small (median semi partial correlation of 0.05). Since Brownstein and colleagues have helpfully highlighted the importance of benchmarks when thinking about effect sizes, the incremental predictive validity of the IAT was much smaller than that of explicit measures.

We have been discussing the predictive validity of the IAT, and one might wonder whether the IAT is a particularly poor predictor among indirect measures. Because the most careful discussion of predictive validity has focused on the IAT, it is difficult to be fully confident, but other measures might be more predictive than the IAT. Sequential priming measures appear to correlate with behavioral criteria to a larger degree in some theoretically predicted contexts ($r = 0.28$), and to have some incremental predictive validity (Cameron et al., 2012). As was noted in the discussion of Kurdi et al. (2019), it is however too easy to explain post hoc why observed large correlations should be theoretically expected. What's more, as we have seen, whether the sources of these effects can be described as implicit has not yet been settled.

Taking stock, there is no sugarcoating it: At the individual level, indirect measures are poorly predictive of behavior, and their incremental validity, while not null, is very limited. Predictive validity could be higher in some contexts, but compelling evidence is lacking. The limitation of the significance of indirect measures to a narrow context undermines their social significance and is definitely at odds with the ambitions of their inventors.

6 | THE CAUSAL CONTRIBUTION OF IMPLICIT ATTITUDES

As we saw in the first section of this article, implicit attitudes are often understood as impacting behavior causally. However, after 30 years of research, there is almost no evidence that indirect measures measure something causally efficient rather than merely epiphenomenal.

If indirect measures measure something causally relevant to behavior, then it should be possible to intervene on it to modify behavior. Finding the proper intervention might not be easy and, for the reasons mentioned in the discussion of the predictive validity of indirect measures, interventions might not dramatically transform behavior, but lack of evidence of our capacity to manipulate indirect measures' measurand would raise questions about their causal efficiency. As it turns out, psychologists have developed and assessed many interventions to reduce implicit biases (Forscher et al., 2019): Evidence shows that indirect measure scores can be manipulated (e.g., Forscher et al., 2019; Lai et al., 2014; Madva, 2017), at least in the short term (Forscher et al., 2017; Lai et al., 2016), but this body of evidence says nothing about whether manipulating the measurand of indirect measures has any behavioral impact. In fact, metaanalytic evidence suggests instead that interventions to reduce the measurand of indirect measures have little effect on behavior (Forscher et al., 2019): Procedures "generally produced trivial changes in behavior. (...) [C]hanges in implicit measures did not mediate changes in explicit measures or behavior. Our findings suggest that changes in implicit measures are possible, but those changes do not necessarily translate into changes in explicit measures or behavior" (p. 522). Of the 492 studies Forscher et al. examined, 94 reported a behavioral task. Some of the interventions had an impact on behavior, but no impact on implicit measure scores; other interventions had no impact on behavior; and no intervention had anything but a trivial impact on behavior. The mediation analysis, involving a smaller set of studies ($n = 63$), failed to find any evidence that the trivial changes in behavior that were observed were due to a change in indirect measure scores. These results are particularly striking since Forscher et al. included not only actual, but also intended behavior in their category of behavioral study: While manipulating the measurand of indirect measures might understandably have only a small, hard to measure effect on actual behavior, why would the impact be small for behavior that is just intended? Also noteworthy is the fact that correspondence did not influence the effect sizes of the manipulations on behavioral measures.

To assess whether one can change behavior by manipulating the measurand of indirect measures one can also look at the efficacy of implicit bias training. This approach has the advantage of examining real-life, consequential behavior instead of imagined, intended, or inconsequential behavior, as is often the case in laboratory studies of interventions on bias. As we saw, the NYPD has been requiring implicit bias training since 2018, like many law enforcement agencies, but this training has had no measurable effect on various behavioral indicators, including ethnic disparities among people arrested (Worden et al., 2020), despite "the training (...) being credited with elevating officers' comprehension of what implicit bias is" (Worden, quoted in Kaste, 2020). Perhaps unsurprisingly given the financial stakes, the trainers working with the NYPD are "undeterred" and believe that [their] training reduces biased behavior on the streets of the jurisdictions where [they] train" (criminology Professor Fridell, quoted in Kaste, 2020). Research on diversity training, including implicit bias training, more broadly shows that they are largely inefficient (Carter et al., 2020).

In contrast to the study of interventions, Charlesworth and Banaji (2019) examined the relation between changes in direct and indirect measures of attitudes for six different targets: sexual orientation, race, skin tone, age, disability, and body weight, using a data set of more than 4 million participants over a 13-year period. They were interested in examining whether changes in the measurand of their direct measure (a simple question about preferences) cause changes in

the measurand of their indirect measure (the IAT), whether the opposite is the case, or whether no causal relation can be detected. To answer this question, they used Granger causality: They computed (1) whether direct measure scores at $t - 1$ incrementally predicts IAT scores at t, but not the opposite, (2) whether IAT scores at $t - 1$ incrementally predicts direct scores at t, but not the opposite, (3) whether both fail to be incrementally predictive, or (4) whether they are both incrementally predictive. If (1) is true, changes in the measurand of direct measures cause changes in the measurand of the IAT; vice-versa if (2) is true. They argue in particular that "change in race attitudes likely flows from implicit to explicit attitudes" (p. 182); similar findings were made for skin color.

Charlesworth and Banaji's (2019) article is remarkable, but their causal analysis is unconvincing.[9] First, as we saw in Section 1, it is unclear whether indirect and direct measures measure different things, and if they are not, the asymmetric predictive power should not be interpreted in terms of causation among distinct constructs. Charlesworth and Banaji moves too quickly from measures (e.g., preference scales and the IAT) to constructs (explicit and implicit biases), although it might be that confusingly they just mean "indirectly measured" by "implicit" (Greenwald & Banaji, 2017).[10] Second, for most targets, the indirect and direct measures are not related (table 3, p. 186), casting doubts on general claims about the causal efficiency of the measurand of indirect measures. Third, and perhaps most crucial for my purposes, their article still provides no evidence for the behavioral efficacy of the measurand of indirect measures since it only examines the relation between the measures of bias.

Why isn't there clearer evidence of the causal efficiency of the measurand of indirect measures? Could it really be epiphenomenal ("the scar interpretation" discussed by Forscher et al., 2019)? The reason, I propose, is that few indirect measures are valid to a substantial degree and a change in an indirect measure is unlikely to be a change in the intended measurand of this measure, and if it is not the case, then it is not very surprising that there is no causal relaiton between changes in indirect measure scores and changes in behavior.

7 | VALIDITY AND HYPE

The construct validity of indirect measures as measures of a distinct type of attitudes—namely implicit attitudes—is weak: At the individual level, their reliability is low; their discriminant validity remains unclear; their predictive validity limited; and there is still no evidence for the causal influence of their measurand on behavior. Measures whose construct validity has not been established are of course not unusual in psychology (Flake et al., 2017; Hussey & Hughes, 2020), and construct validation cannot be expected for a recently introduced measure: Construct validation takes time. But indirect measures have been around for several decades and fundamental issues such as discriminant validity and causal influence remain not only unsolved, they have not been tackled at all until recently.

The contrast between the hype around indirect measures, particularly the IAT, and the anomalies discussed in this article is striking. The IAT was heralded by Banaji at the 2001 convention of the American Psychology Society as a third revolution, after the Copernican and Darwinian revolutions (Kester, 2001; cited in in Mitchell & Tetlock, 2017), and psychologists have been proselytizing their research. Banaji took her research to justify reframing the law away from a focus on intentionality (Potier, 2004; as discussed in Mitchell & Tetlock, 2017): "The central idea is to use the energy generated by research on unconscious forms of prejudice to understand and challenge the notion of intentionality in the law." Scientists are of course incentivized to hype their research (Machery & Doris, 2017), and most are savvy enough to discount such hype, but outsiders, from scientists in other fields to policy makers to lay people, may be misled by unjustified hype.

8 | OTHER USES OF INDIRECT MEASURES

In this article, I have focused on indirect measures of attitudes at the individual level, but indirect measures are also used in other contexts. They have for instance recently been used to measure attitudes at the aggregate level, as discussed by Brownstein et al. (2019), section 5; see also Payne et al., 2017; Machery, 2017). Indirect measure scores at the group level appear to correlate with various outcomes: For instance, racial prejudice IAT scores at "the level of core-based statistical areas (CBSAs), a geographic area defined by the U.S. Office of Management and Budget of at least 10,000 people" were correlated with the use of lethal force in policing (Hehman et al., 2018; see also Hehman et al., 2019).[11] Perhaps more important for psychological research, indirect measures are used in experimental contexts to measure the impact of experimental manipulations on their measurand in order to test theories of automatic psychological processing (Kurdi et al., 2021). How are these uses impacted by the discussion in this article?

Many of the problems discussed in this article might not threaten such uses. The inability of indirect measures to measure individual differences in attitudes might not be an issue for their experimental use; indeed, it might be a strength since it reduces the proportion of the variance in measurement that is not related to the manipulation, as is for instance the case of the Stroop task (Dang et al., 2020). Similarly, this inability is not a problem for the measurement of group-level attitudes since unreliable and poorly predictive measures at the individual level may be reliable and predictive at the group level (Payne et al., 2017).

On the other hand, the lack of clarity of the nature of the measurand of indirect measures would seem to be an issue for these two uses of indirect measures: What do indirect measures measure at the group-level? How are these measurements related to whatever it is direct measures measure? And what is measured in experimental contexts? What do the experimental manipulations influence? Until we have clearer, evidentially supported (i.e., not merely based on theory) answer to these questions, the use of these measures is deeply problematic. It will not do to reply that indirect measures measure automatic processes: As even proponents of indirect measures of attitudes acknowledge, evidence for this claim remains incomplete (Vianello & Bar-Anan, 2021).

9 | CONCLUSION

We do not know what indirect measures measure; indirect measures are unreliable at the individual level, and people's scores vary from occasion to occasion; indirect measures predict behavior poorly, and we do not know in which contexts they could be more predictive; in any case, the hope of measuring broad traits is not fulfilled by the development of indirect measures; and there is still no reason to believe that they measure anything that makes a causal difference. These issues would not be too concerning for a budding science; they are anomalies for a 30-year-old research tradition that has been extremely successful at selling itself to policy makers and the public at large. So, should social psychologists pack up and move to other research topics or should they stubbornly try to address the anomalies pointed out in this article? It is unwise to predict the future of science, and the issues presented here could well be resolved by the many psychologists working on indirect measures, but it would also be unwise to dismiss them as mere challenges to be addressed in the course of normal science.

ENDNOTES

[1] For defenses of this research tradition (see Brownstein et al., 2020; Gawronski, 2019; Jost, 2019; Jost et al., 2009; Payne et al., 2018).

[2] An anomaly is an empirical result that threatens an existing scientific paradigm and typically leads to its demise (Kuhn, 1970).

[3] "Bias" is used in various ways in psychology: For instance, it is sometimes merely synonymous with "preference" (as when psychologists refer to people's in-group bias); in the literature on testing, "bias" typically refers to a systematic error in the estimation of a given value. In other contexts, a more general characterization of the notion of bias as an unjustified or irrational attitude toward groups or individuals might be preferable (for discussion see Holroyd et al., 2017).

[4] I use "direct" and "indirect" to characterize measures and "implicit" versus "explicit" to characterize the constructs measures are intended to measure. This approach is at odds with the recent push for using "implicit" to characterize the measures and to define "implicit" operationally (Brownstein et al., 2019; Greenwald & Banaji, 2017). In addition to a problematic commitment to operationalism, this proposal introduces yet further confusions about what "implicit" means. For instance, Brownstein et al.'s discussion of my claim that attitudes are traits (Brownstein et al., 2019, p. 7) confuses properties of measures with properties of traits.

[5] At times, however, some psychologists engage in some sort of bait-and-switch: For instance, after referring to the "mind's two modes of operation," Banaji and Heiphetz (2010) assert that this is just a "way of speaking about data that points to dissociations in the attitudes that emerge based on variations in methods used to measure attitudes" (p. 356). The carefully entertained equivocation between measures and constructs, and the constant back and forth between realism about constructs and a reductionist operationalism, are an unfortunately common characteristic of this literature.

[6] The same point applies to the absence of correlation between changes in direct and indirect measures (Charlesworth & Banaji, 2019).

[7] It is also worth noting that many indirect measures correlate very poorly with one another (Bar-Anan & Nosek, 2014), raising question about what this implicit attitude is that they are all supposed to be measuring (Machery, 2016).

[8] Kurdi et al. also show that the correlation remains significant when measurement error is controlled for.

[9] One could also question whether Granger causality is genuinely about causality rather than merely forecasting.

[10] For instance, they refer to "population-level implicit attitudes" (p. 175) instead of IAT scores aggregated at some group level. Furthermore, they do not consider the possibility that direct and indirect measures might measure the same thing and that it might make little sense to analyze whether the measurand of the latter causally influence the measurand of the former (an instance of the Jangle fallacy, Weidman et al., 2017; Machery, 2021).

[11] One worries, however, about researchers' degrees of freedom in defining the level of aggregation and in choosing the indirect measures and behavioral criteria.

CONFLICT OF INTEREST

The author has declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

RELATED WIREs ARTICLE

What do implicit measures meausre?

Further Reading

Banaji, M. R., & Greenwald, A. G. (2016). *Blindspot: Hidden biases of good people*. Bantam.

Kristof, N. (2008). Our racist, sexist selves. The New York Times. Retrieved from https://www.nytimes.com/2008/04/06/opinion/06kristof.html

REFERENCES

Allport, G. W. (1935). Attitudes. In C. A. Murchinson (Ed.), *A handbook of social psychology* (Vol. 2, pp. 798–844). Clark University Press.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "would Jesse Jackson 'fail' the implicit association test?". *Psychological Inquiry*, *15*, 257–278.

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), 888–918.

Banaji, M. R., & Heiphetz, L. (2010). Attitudes. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 353–393). John Wiley & Sons.

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

Bar-Anan, Y., & Vianello, M. (2018). A multi-method multitrait test of the dual-attitude perspective. *Journal of Experimental Psychology: General*, *147*, 1264–1272.

Bartlett, T. (2017). Can we really measure implicit bias? Maybe not. The Chronicle of Higher Education. Retrieved from https://www.chronicle.com/article/Can-We-Really-Measure-Implicit/238807

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*(1), 27–41.

Briñol, P., & Petty, R. E. (2012). The history of attitudes and persuasion research. In A. Kruglanski & W. Stroebe (Eds.), *Handbook of the history of social psychology* (pp. 285–320). Psychology Press.

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

Brownstein, M., Madva, A., & Gawronski, B. (2020). Understanding implicit bias: Putting the criticism into perspective. *Pacific Philosophical Quarterly*, *102*, 276–307.

Buttrick, N., Axt, J., Ebersole, C. R., & Huband, J. (2020). Re-assessing the incremental predictive validity of implicit association tests. *Journal of Experimental Social Psychology*, *88*, 103941.

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

Carter, E. R., Onyeador, I. N., & Lewis, N. A., Jr. (2020). Developing & delivering effective anti-bias training: Challenges & recommendations. *Behavioral Science & Policy*, *6*(1), 57–70.

Charlesworth, T. E., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. long-term change and stability from 2007 to 2016. *Psychological Science*, *30*(2), 174–192.

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of black and white discrimination and prejudice: A literature review. *Psychological Bulletin*, *87*(3), 546–563.

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. *PsyArXiv*. 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 Sciences*, *24* (4), 267–269.

Danziger, K. (1994). *Constructing the subject: Historical origins of psychological research*. Cambridge University Press.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56* (1), 5–18.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, *48*(6), 1267–1278.

Dovidio, J. F., & Gaertner, S. L. (1986). *Prejudice, discrimination, and racism*. Academic Press.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers.

Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition*, *25*(5), 582–602.

Eberhardt, J. L. (2020). *Biased: Uncovering the hidden prejudice that shapes what we see, think, and do*. Viking.

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

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229–238.

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

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., Mitamura, C., Dix, E. L., Cox, W. T., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of Experimental Social Psychology*, *72*, 133–146.

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.

Gawronski, B. (2019). Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science*, *14*(4), 574–595.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition*, *15*(3), 485–499.

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, *43*(3), 300–312.

Goode, E. (1998). A computer diagnosis of prejudice. New York Times. Retrieved from https://www.nytimes.com/1998/10/13/health/a-computer-diagnosis-of-prejudice.html

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. Journal of personality and social psychology *Consciousness and Cognition*, 79(6), 1022–1038.

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*(9), 861–871.

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, *94*(4), 945–967.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the implicit association test can have societally large effects. *Journal of Personality and Social Psychology*, *108*, 553–561.

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

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

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

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

Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science*, *9*(4), 393–401.

Hehman, E., Calanchini, J., Flake, J. K., & Leitner, J. B. (2019). Establishing construct validity evidence for regional measures of explicit and implicit racial bias. *Journal of Experimental Psychology: General*, *148*(6), 1022–1040.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385.

Holroyd, J., Scaife, R., & Stafford, T. (2017). What is implicit bias? *Philosophy Compass*, *12*(10), e12437.

Hussey, I., & Hughes, S. (2020). Hidden invalidity among fifteen commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*, *3*(2), 166–184.

Izuma, K., Kennedy, K., Fitzjohn, A., Sedikides, C., & Shibata, K. (2018). Neural activity in the reward-related brain regions predicts implicit self-esteem: A novel validity test of psychological measures using neuroimaging. *Journal of personality and social psychology*, *114*(3), 343–357

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*(1), 10–19.

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior*, *29*, 39–69.

Kaste M. (2020). NYPD study: Implicit bias training changes minds, not necessarily behavior. NPR. Retrieved from https://www.npr.org/2020/09/10/909380525/nypd-study-implicit-bias-training-changes-minds-not-necessarily-behavior

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217.

Kester, J. D. (2001) A revolution in social psychology. APS Observer Online, 14.

Kristof, N. (1995). Our biased brains. The New York Times. Retrieved from https://www.nytimes.com/2015/05/07/opinion/nicholas-kristof-our-biased-brains.html

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for predicting first-year college grade point average*. The College Board.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press.

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.

Kurdi, B., Ratliff, K. A., & Cunningham, W. A. (2021). Can the implicit association test serve as a valid measure of automatic cognition? A response to Schimmack (2021). *Perspectives on Psychological Science*, *16*(2), 422–434.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., ... Nosek, B. A. (2014). A comparative investigation of 17 interventions to reduce implicit racial preferences. *Journal of Experimental Psychology: General*, *143*, 1765–1785.

Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y. J., Pedram, C., Marshburn, C. K., Simon, S., Blanchar, J. C., Joy-Gaba, J. A., Conway, J., Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M. H., Burns, M., ... Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 1001–1016.

Lord, F. M., & Novick, M. R. (1968). *(1968) Statistical theories of mental test score*. Addison-Wesley.

MacDonald, H. (2017). The false "science" of implicit bias. *The Wall Street Journal*. Retrieved from. https://www.wsj.com/articles/the-false-science-ofimplicit-bias-1507590908

Machery, E. (2016). De-Freuding implicit attitudes. In M. Brownstein & J. Saul (Eds.), *Implicit bias and philosophy* (Vol. 1, pp. 104–129). Oxford University Press.

Machery, E. (2017). Do indirect measures of biases measure traits or situations? *Psychological Inquiry*, *28*(4), 288–291.

Machery, E. (2021). A mistaken confidence in data. *European Journal for Philosophy of Science*, *11*(2), 1–17.

Machery, E., & Doris, J. M. (2017). An open letter to our students: Doing interdisciplinary moral psychology. In B. Voyer & T. Tarantola (Eds.), *Moral psychology* (pp. 119–143). Springer.

Madva, A. (2017). Biased against debiasing: On the role of (institutionally sponsored) self-transformation in the struggle against prejudice. *Ergo, An Open Access Journal of Philosophy*, *4*(6), 145–179.

Mallon, R. (2016). *The construction of human kinds*. Oxford University Press.

McGuire, W. J. (1986). The vicissitudes of attitudes and similar representational constructs in twentieth century psychology. *European Journal of Social Psychology*, *16*(2), 89–130.

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*, *10*, 2483.

Meyer, Z. (2018). Starbucks' racial-bias training will be costly, but could pay off in the long run. Retrieved from https://www.usatoday.com/story/money/2018/05/26/starbucks-racial-bias-training-costly/642844002/

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). Wiley.

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

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.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). Psychology Press.

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., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293.

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*(4), 233–248.

Payne, B. K., Niemi, L., & Doris, J. M. (2018). How to think about "implicit bias". *Scientific American*, *27*. Retrieved from https://www.scientificamerican.com/article/how-to-think-about-implicit-bias/.

Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology*, *48*(1), 609–647.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338.

Potier, B. (2004). Making case for concept of "implicit prejudice": Extending the legal definition of discrimination. Harvard University Gazette.

Rae, J. R., & Olson, K. R. (2018). Test-retest reliability and predictive validity of the implicit association test in children. *Developmental Psychology*, *54*, 308–330.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008.

Schimmack, U. (2021). The implicit association test: A method in search of a construct. *Perspectives on Psychological Science*, *16*(2), 396–414.

Singal, J. (2017a). Psychology's favorite tool for measuring racism isn't up to the job. Retrieved from https://www.thecut.com/2017/01/psychologys-racism-measuring-tool-isnt-up-to-the-job.html

Singal, J. (2017b). The creators of the Implicit Association Test should get their story straight. Intelligencer. Retrieved from https://nymag.com/intelligencer/2017/12/iat-behavior-problem.html

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

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*(4), 529–554.

Thurstone, L. L. (1931). The measurement of social attitudes. *The Journal of Abnormal and Social Psychology*, *26*, 249–269.

Tierney, J. (2008). In bias test, shades of grey. The New York Times. Retrieved from https://www.nytimes.com/2008/11/18/science/18tier.html

Valian, V. (1999). *Why so slow? The advancement of women*. MIT Press.

Van Dessel, P., Cummins, J., Hughes, S., Kasran, S., Cathelyn, F., & Moran, T. (2020). Reflecting on twenty-five years of research using implicit measures: Recommendations for their future use. *Social Cognition*, *38*, s223–s242.

Vianello, M., & Bar-Anan, Y. (2021). Can the implicit association test measure automatic judgment? The validation continues. *Perspectives on Psychological Science*, *16*(2), 415–421.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1999). *Unobtrusive measures*. Sage Publications.

Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion*, *17*(2), 267–295.

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS One*, *11*(3), e0152719.

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

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262–274.

Worden, R. E., et al. (2020). The impacts of implicit bias awareness training in the NYPD. IACP/UC Center for Police Research and Policy & John F. Finn Institute for Public Safety. Retrieved from https://www1.nyc.gov/assets/nypd/downloads/pdf/analysis_and_planning/impacts-of-implicit-bias-awareness-training-in-%20the-nypd.pdf

How to cite this article: Machery, E. (2021). Anomalies in implicit attitudes research. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1569. https://doi.org/10.1002/wcs.1569