Annual Review of Psychology

Implicit Social Cognition

Anthony G. Greenwald[1] and Calvin K. Lai[2]

[1]Department of Psychology, University of Washington, Seattle, Washington 98195, USA; email: agg@uw.edu

[2]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, Missouri 63130, USA

Annu. Rev. Psychol. 2020. 71:419–45

First published as a Review in Advance on October 22, 2019

The *Annual Review of Psychology* is online at psych.annualreviews.org

https://doi.org/10.1146/annurev-psych-010419-050837

Keywords

indirect measures, implicit bias, implicit–explicit relationships, Implicit Association Test, evaluative priming, implicit social cognition

Abstract

In the last 20 years, research on implicit social cognition has established that social judgments and behavior are guided by attitudes and stereotypes of which the actor may lack awareness. Research using the methods of implicit social cognition has produced the concept of implicit bias, which has generated wide attention not only in social, clinical, and developmental psychology, but also in disciplines outside of psychology, including business, law, criminal justice, medicine, education, and political science. Although this rapidly growing body of research offers prospects of useful societal applications, the theory needed to confidently guide those applications remains insufficiently developed. This article describes the methods that have been developed, the findings that have been obtained, and the theoretical questions that remain to be answered.

Contents

INTRODUCTION: WHAT DOES "IMPLICIT" MEAN?

Indirect measure: inference of a construct without instruction to report it, assuming no introspective awareness of the construct

Attitude: association of positive or negative valence with a social object, a physical object, or an abstract concept

Stereotype: association of a social group or a category of people with a trait

Implicit: indirectly measured

In a review of social and personality psychologists' uses of indirect measures in research on attitudes, stereotypes, and self-esteem in the 1970s and 1980s, Greenwald & Banaji (1995) identified a research domain for which they offered the label "implicit social cognition." They proposed that "the signature of implicit cognition is that traces of past experience affect some performance, even though the influential earlier experience is not remembered in the usual sense—that is, it is unavailable to self-report or introspection" (Greenwald & Banaji 1995, p. 4). Fazio & Olson's (2003) review of implicit social cognition—the predecessor of this review—was occasioned, in part, by the extent to which indirect (i.e., implicit) measures were being increasingly used in the years since Greenwald & Banaji's article. That growth in use has subsequently continued and even accelerated.

At the end of 2018, the PsycNET database identified 1,483 peer-reviewed articles, published since 2003, in which the three words "implicit," "social," and "cognition" all appeared in indexing fields. This number does not include (many) peer-reviewed articles published in scholarly journals of applied disciplines such as medicine, education, business, and law, nor does it include book chapters, among them handbook chapters on implicit social cognition (Hahn & Gawronski 2018, Nosek et al. 2012, Payne & Gawronski 2010) and other chapters appearing in Gawronski & Payne's (2010) *Handbook of Implicit Social Cognition*.

As did Fazio & Olson (2003), the authors of the three subsequent handbook overview chapters organized their reviews around the main research questions that were receiving empirical attention (Hahn & Gawronski 2018, Nosek et al. 2012, Payne & Gawronski 2010). The present review similarly evaluates progress in answering those questions and extends to additional questions in which interest has more recently developed. Necessarily limited to a small minority of what has

been published in the last 17 years, this article gives greatest attention to broad narrative reviews, to quantitative reviews (meta-analyses) of the empirical literature, and to influential theoretical and methodological contributions.

Direct measure: assessment of a construct in response to an instruction to report it, typically assuming introspective awareness of the construct

Graf & Schacter (1985) selected "implicit" as their label for a form of memory that had been demonstrated with indirect measures. Schacter (1987) reviewed the history of developments in psychology that led to the introduction of "implicit" in the field of memory research. Because research on memory in the 1980s using indirect measures regularly produced evidence for memories that were not apparent on direct measures of memory (e.g., free recall and recognition), many cognitive psychologists interpreted implicit memory as designating an unconscious form of memory. Very soon thereafter, others argued forcefully that equating implicit with unconscious was not justified either conceptually or empirically (for the strongest of such arguments, see Jacoby 1991, Reingold & Merikle 1988).

By the time of Fazio & Olson's (2003) review, the use of "implicit" had extended to indirectly measured social-cognitive constructs, including attitudes, stereotypes, identities, and self-esteem. The positions taken subsequently in the major overviews of implicit social cognition about how "implicit" should be understood are that (*a*) implicit constructs are ones measured indirectly, without implication for the conscious or unconscious nature of what those measures reveal (Fazio & Olson 2003); (*b*) implicit attitudes are automatic or unconscious attitudes, an interpretation complicated by the lack of general agreement about the meanings of automatic and unconscious (Payne & Gawronski 2010); (*c*) "implicit" is a descriptive definition (Nosek et al. 2012, p. 32), meaning "not committed to any particular theoretical interpretation of [psychological] mechanisms"; and (*d*) the range of variation of definitions of "implicit" signals a "terminological confusion," such that no preferred definition can be identified (Hahn & Gawronski 2018, p. 396).

In a retrospective piece that described the historical roots of research on implicit social cognition, Greenwald & Banaji (2017) explained the conceptual difficulties of identifying implicit with unconscious and opted for the indirect-measurement definition that had been advocated by Fazio & Olson (2003). Greenwald & Banaji's (2017, pp. 861, 862) summary position is conveyed by two headings in their article: "Implicit = Indirect; Explicit = Direct" and "Indirect ≠ Unconscious; Direct ≠ Conscious."

In the context of research on perception of visually masked stimuli, formal definitions of "direct" and "indirect" were offered by Reingold & Merikle (1988, p. 564):

> Discriminations among a set of alternative stimulus states (e.g., S1, S2, …S*n*) should be considered a direct measure of perception if the discriminative response is part of the task definition, as expressed in the instructions given to the subjects. Conversely, if the discriminative response is not part of the task definition, it should be considered an indirect measure of perception.

Reingold & Merikle's definitions can be extended from perception to social-cognitive constructs such as attitude by changing "discriminations among…alternative stimulus states" in their first sentence to "judgments of levels of evaluation."

A virtue of the terms "direct" and "indirect" is that each describes a large class of operational definitions that do not mutually overlap. These are also operational definitions that can be used equally by researchers who hold varying (even mutually conflicting) theoretical stances on the debated terms (e.g., implicit, unconscious, and automatic). Defining "implicit" as meaning "indirectly measured" is therefore a theory-uncommitted definition, allowing research to proceed without need for debate about the conceptual understanding of "implicit."

Fazio & Olson (2003) commented on the lack of theory accompanying the early surge of empirical research on implicit (i.e., indirect) measures. That lack of theory may explain the magnitude

of the empirical surge. Many of the findings summarized by Greenwald & Banaji (1995) had been obtained using innovative methods that produced unexpected findings. The two main research methods in use by 2003 were evaluative priming (Fazio et al. 1986) and the Implicit Association Test (IAT) (Greenwald et al. 1998). Both methods rested theoretically only on the notion of association, which has been a staple concept of psychology since Aristotle (cf. Greenwald et al. 2005, p. 421). Perhaps because this reliance on such a basic, traditional idea left little basis for theoretical criticism, the research response to the (often) unexpected findings obtained with indirect measures was to explore empirical variations, seeking either to establish boundaries on surprising findings or to develop theoretical interpretations more elaborate than the concept of association. (Examples of such theoretical developments are in Meissner & Rothermund 2013, Mierke & Klauer 2001, and Rothermund & Wentura 2004.)

Evaluative priming: indirect attitude measure inferred from a preceding stimulus' effect on speed of an evaluative classification response to a subsequent stimulus

IAT: Implicit Association Test

At present, the use of "implicit" to designate the (large) class of indirect measures in social cognition appears to have more supporters than do definitions that specify a theorized mental process such as automaticity (e.g., De Houwer et al. 2009). Even so, conceptually stated definitions such as De Houwer and colleagues' merit consideration and do not obstruct the orderly pursuit of research on implicit social cognition. Advocates of different understandings of "implicit" have mostly used priming, IAT, and other indirect measurement procedures in the same forms used by those who equate "implicit" with "indirect." A useful consequence is that readers of the remainder of this review should be unencumbered by a need to choose among definitions of "implicit."

MEASUREMENT AND PSYCHOMETRICS

Frequently Used Indirect Measures

Table 1 lists the most frequently used indirect measures of social-cognitive constructs, organized into three families. Prior inventories of implicit measures have been published by Gawronski & De Houwer (2014) and Nosek et al. (2011).

The first four rows of **Table 1** list variations of priming methods, derived from procedures developed by Meyer & Schvaneveldt (1971) to study associative memory. Priming procedures present a stimulus in the form of a word or an image (the prime) alongside or prior to another stimulus (the target) that the subject is asked to classify, usually as pleasant versus unpleasant in valence. Latency measures capture the differential influence of two categories of primes (e.g., Black versus White faces) on latency and/or accuracy of response to the two target categories.

The next seven rows of **Table 1** list various forms of the IAT. Like priming measures, IAT measures typically (although not always) have four categories, with trials alternating between classification of exemplars of two target categories (e.g., male names versus female names) and of two attribute categories (e.g., family words versus career words). IAT procedures have two distinct combined tasks that differ in the assignment of target categories to a left versus a right response key. Subjects are expected to respond faster and/or more accurately when the target and attribute categories assigned to the same key are more strongly associated. In one combined task the left key may be used for both male and career categories, and the right key for both female and family. In the other combined task, the keys used to classify male and female would be switched. The segregation of trials into these two configurations invokes mental operations that are not involved in priming procedures. Competing theoretical understandings of these additional (or different) mental operations involved in the IAT are treated in an overview of IAT research by A.G. Greenwald (unpublished manuscript).

The final seven rows of **Table 1** list indirect measures using methods other than priming or IAT. Some have roots, at least indirectly, in the interference method developed by Stroop (1935),

Table 1 Frequently cited measures used in implicit social cognition research[a]

Families of measures	Primary publication	Times cited 2014–2018[b]	Times used 2014–2018[c]	Meta-analytic IC		Meta-analytic TRR		Used to assess			Association type varied within or between trial blocks	Used on the Internet	Used with children	Provides absolute or relative measure	Predictive validity evidence available
				k^d	α	k^d	r	Attitude	Stereo-type	Self-esteem and/or self-concept					
Priming variations															
Evaluative Priming (EPT)	Fazio et al. 1986	359	103[e]	21	0.53	16	0.26	Yes	No	Yes	Within	Yes	Yes	Absolute	Yes
Semantic Priming Task	Blair & Banaji 1996	92	10[e]	NA	NA	NA	NA	No	Yes	No	Within	No	No	Absolute	No
Lexical Decision Priming (LDT)	Wittenbrink et al. 1997	87	20[e]	NA	NA	NA	NA	Yes	Yes	No	Within	No	No	Absolute	Yes
Affect Misattribution Procedure (AMP)	Payne et al. 2005	333	58	73	0.81	7	0.52	Yes	Yes	Yes	Within	Yes	Yes	Absolute	Yes
Implicit Association Test variations															
Implicit Association Test (IAT)	Greenwald et al. 1998	2,116	767	257	0.80	58	0.50	Yes	Yes	Yes	Between	Yes	Yes	Relative	Yes
Go/No-Go Association Test (GNAT)	Nosek & Banaji 2001	204	33	18	0.66	5	0.48	Yes	Yes	Yes	Between	Yes	No	Absolute	Yes
Single-Category IAT (SC-IAT)	Karpinski & Steinman 2006	259	72	33	0.76	7	0.25	Yes	Yes	Yes	Between	Yes	No	Absolute	Yes
Implicit Relational Assessment Procedure (IRAP)	Barnes-Holmes et al. 2006	89	71	23	0.60	5	0.43	Yes	Yes	Yes	Between	Yes	Yes	Relative	Yes
Single-Target IAT (ST-IAT)	Bluemke & Friese 2008[f]	69	31	16	0.78	8	0.43	Yes	Yes	Yes	Between	Yes	Yes	Absolute	Yes
Brief IAT (BIAT)	Sriram & Greenwald 2009	121	36	61	0.79	32	0.43	Yes	Yes	Yes	Between	Yes	No	Relative	Yes
Recoding-Free IAT (IAT-RF)	Rothermund et al. 2009[g]	32	3	4	0.69	NA	NA	Yes	Yes	No	Within	No	No	Relative	Yes

(*Continued*)

Table 1 (*Continued*)

Families of measures	Primary publication	Times cited 2014–2018[b]	Times used 2014–2018[c]	Meta-analytic IC		Meta-analytic TRR		Used to assess			Association type varied within or between trial blocks	Used on the Internet	Used with children	Provides absolute or relative measure	Predictive validity evidence available
				k[d]	α	k[d]	r	Attitude	Stereo-type	Self-esteem and/or self-concept					
Other methods															
Name-Letter Effect (NLE)	Nuttin 1985	85	9	8	0.66	24	0.56	No	No	Yes	NA	Yes	Yes	Absolute	Yes
Linguistic Intergroup Bias (LIB)	Maass et al. 1989	80	14	NA	NA	NA	NA	Yes	No	No	NA	No	Yes	Relative	No
Stimulus Response Compati-bility Task (SRCT)	Mogg et al. 2003	112	22	5	0.81	NA	NA	No	Yes	No	Between	Yes	Yes	Relative	Yes
Extrinsic Affective Simon Task (EAST)	De Houwer 2003	74	7	24	0.38	3	0.24	Yes	Yes	Yes	Within	No	No	Absolute	Yes
Stereotypic Explana-tory Bias (SEB)	Sekaquaptewa et al. 2003	20	2	NA	NA	NA	NA	No	Yes	No	NA	No	No	Absolute	Yes
Approach-Avoidance Task (AAT)	Rinck & Becker 2007[h]	291	152	19	0.62	4	0.10	No	Yes	No	Between	No	Yes	Relative	Yes
MouseTracker	Freeman & Ambady 2010	157	13	NA	NA	NA	NA	No	Yes	No	Within	No	No	Absolute	No

[a]This table does not describe methods that have been cited fewer than 100 times on Google Scholar as of January 9, 2019; that are practically unused in social cognition research (e.g., Emotional Stroop); that have not been used outside of their original publications; that have very limited applications; or that are close relatives of other measures presented here (e.g., slight changes in instructions or stimuli/categories). More information on meta-analytic estimates is provided in the **Supplemental Text**. Abbreviations: IC, internal consistency; NA, not available; TRR, test–retest reliability.

[b]This indicates the number of times the primary publication has been cited from 2014 to 2018 in the Scopus database.

[c]This indicates the number of times a measure has been mentioned in the title, abstract, or keywords of academic articles from 2014 to 2018 in the Scopus database, combined with additional articles identified by original authors who helped in the construction of this table.

[d]k indicates the number of effects associated with meta-analytic effect size.

[e]Estimates of usage for these measures are less precise than for other measures because these measures have been used with multiple names.

[f]This method is a close relative of the SC-IAT and was created by Wigboldus and colleagues (D.H.J. Wigboldus, R.W. Holland, A. van Knippenberg, unpublished manuscript).

[g]This method is similar to the Single-Block IAT that was created by Teige-Mocigemba et al. (2008).

[h]This method is derived from a method introduced by Chen & Bargh (1999).

Supplemental Material >

in free association (Jung 1910), or in projective methods, such as Morgan & Murray's (1935) Thematic Apperception Test (TAT).

Measurement precision. Two well-known psychometric indicators, internal consistency (IC) and test–retest reliability (TRR), assess the precision of psychological measures. **Table 1** lists aggregate estimates of IC and TRR derived from a robust variance estimation meta-analysis method (Hedges et al. 2010). This meta-analysis used articles retrieved from the Scopus database using the names of implicit measures and search terms related to IC or TRR, including prominent reference articles for those methods. Additionally, authors of the primary publications of each method were contacted for assistance in identifying additional articles. These search procedures yielded 166 articles that reported results from 331 independent samples with 562 IC effect sizes and 169 TRR effect sizes. A report of the meta-analysis's methods and details of its results are available in the **Supplemental Text**.

IC: internal consistency

TRR: test–retest reliability

Construct validity: evidence, often correlational, supporting the validity of a measure of a construct such as an attitude or a stereotype

Supplemental Material >

Four overview points about IC and TRR can guide understanding of how these indicators bear on measurement precision and why they should be reported more routinely.[1] First, IC describes the proportion of systematic variance (nonerror variance) available from a single measurement occasion. Second, TRR indicates the proportion of variance that is preserved across measurement occasions. Third, TRR is generally smaller than IC (as is readily observable in **Table 1**); this difference indicates the presence of systematic variance in IC that is unshared among measurement occasions. Fourth, the variance captured by TRR can be apportioned into (*a*) person-associated variance that is unrelated to the construct of interest (i.e., contamination or artifact) and (*b*) construct-relevant variance due to the construct of interest. This last component provides a basis for assessing construct validity. It follows that TRR places an upper limit on correlational tests of construct validity. However, that upper limit can be exceeded in studies using multiple measures of each of a set of variables, allowing structural equation modeling (e.g., MacCallum & Austin 2000).

Just one previously published report on IC and TRR used methods that enabled comparisons among some of the more widely used indirect measures of social-cognitive constructs (Bar-Anan & Nosek 2014). Bar-Anan & Nosek investigated six of **Table 1**'s measures (IAT, BIAT, GNAT, ST-IAT, EPT, and AMP), adhering to procedures for each based on those used by the methods' introducers. Averaged over three attitude domains (race, politics, and self-esteem), Bar-Anan & Nosek found average ICs above $\alpha = 0.70$ for four of the six measures, and average TRRs above $r = 0.40$ for five of the six. For those same six measures, the figures in **Table 1** are close to those reported by Bar-Anan & Nosek.

TRRs of the magnitudes shown in **Table 1** are satisfactory for using most of **Table 1**'s measures in correlational studies. They are, however, problematic for individual diagnostic uses—i.e., uses that seek either to describe traits of individuals or to characterize differences between individuals. There has not yet been any development of high-precision implicit measures. However, need for high-precision measures should increase as clinical psychological use of these measures grows. One example of such clinical use is Nock et al.'s (2010) study of suicide risk. Nock and colleagues found that an IAT measure of association of self with death or suicide predicted odds of a suicide attempt in the six months following the measure. That predictive success in a correlational study notwithstanding, it is not psychometrically justifiable to use Nock and colleagues' measure diagnostically to predict an imminent suicide attempt for an individual unless the measure's TRR can be substantially increased from what is expectable for a single IAT measurement. The only

[1]The four points in this paragraph are necessarily much simplified relative to treatments available in texts on psychometrics (e.g., Nunnally & Bernstein 1994).

IAT that (to the present authors' knowledge) has ever been proposed for individual diagnostic use is the autobiographical IAT that Agosta & Sartori (2013) proposed for use as a means of lie detection.

Scoring methods. Only the IAT and the Brief IAT (BIAT) have been the object of studies seeking to refine scoring of the data obtained by their procedures (Greenwald et al. 2003 for the IAT, Nosek et al. 2014 for the BIAT). The IAT's initial scoring method (Greenwald et al. 1998) used the average of log-transformed response latencies as the speed measure for each of the IAT's two combined tasks. In comparison with the log-transform measure, Greenwald and colleagues' *D* algorithm was found to be superior in IC, in sensitivity to known group differences, and in strength of correlation with parallel self-report measures, and it showed reduced contamination associated with subject differences in performance speed, the order in which the two combined tasks were completed (Greenwald et al. 2003), or individual differences in executive function (Cai et al. 2004, Ito et al. 2015, Mierke & Klauer 2003).

Three further efforts to develop improved scoring of IAT measures have been based on theoretical models of performance in the IAT. These are (*a*) the quadruple-process (Quad) model by Conrey et al. (2005), which separates automatic attitude activation from three other theorized process components of responding to each trial of an IAT measure; (*b*) the diffusion model by Klauer et al. (2007), which includes two nonassociative parameters based on Ratcliff et al.'s (2004) drift-diffusion model of reaction times in two-choice tasks; and (*c*) the ReAL model by Meissner & Rothermund (2013), which also includes two nonassociative parameters. All three approaches share a goal of modeling (and thereby isolating) processes that contribute nonassociative variance to IAT measures.

Interpretation of zero points. Studies that compare treatment group means on an indirect measure or that assess correlations of indirect measures to other variables require only a modest scaling assumption—that the indirect measure is monotonically related to the construct that it is assumed to measure. However, to assess biases involving, for example, race, gender, or age, or to measure self-esteem and identities, the interpretation of a measure's zero point becomes critical. Results obtained with indirect measures of attitude are often described as revealing implicit preferences. For example, with large Internet samples, (*a*) the assumption that the IAT's zero point indicates equal racial preference for White and Black allows the statement that approximately 75% of the sample's respondents demonstrate automatic preference for White relative to Black (based on data reported in Nosek et al. 2007), and (*b*) the assumption that the gender–career IAT's zero point indicates equal association of male and female with career (relative to family) allows the statement that more women (84%) than men (77%) possess this stereotype (also based on Nosek et al. 2007).

Use of reaction times in indirect measures rests on a tradition [dating to Donders 1969 (1868)] of interpreting latency of an instructed response to a stimulus as a measure of the mental processes activated by the stimulus and leading to the response. It is a leap from that statement to conclude that equal latencies of classification responses to two different stimuli indicate equal strengths of association between the two stimuli and their responses, let alone to conclude that equal mean latencies of a collection of responses to a collection of stimuli represent equal strengths of association between the categories represented by the collections. (The latter is the assumed understanding of the zero point of the IAT.) However plausible that assumption may appear, empirical support for the validity of a zero point as indicating equal association strengths is desirable. This evidence was recently provided for the IAT in a meta-analysis by Cvencek and colleagues, who showed that predictions from a theory about interrelations among social cognitive constructs (balanced identity theory) (see Greenwald et al. 2002) were more strongly confirmed when the IAT *D* measure

was used than when the same predictions were tested with *D* measures for which the zero point had been displaced by adding or subtracting a constant (D. Cvencek, A.N. Meltzoff, C.D. Maddox, B.A. Nosek, L.A. Rudman, et al., unpublished manuscript).

Explicit: directly measured

Construct validity. Cronbach & Meehl (1955, p. 290) described construct validity of psychological traits as resting on a nomological network, which they defined as "the interlocking system of laws which constitute a theory." They further wrote, "Construct validation is possible only when some of the statements in the network lead to predicted relations among observables" (Cronbach & Meehl 1955, p. 300). The evidence for construct validity of implicit social-cognitive constructs therefore consists of correlations with measures of other constructs that should, by theory, be related to them. Such correlational evidence for validity is considered in the section titled Theory and the section titled Findings.

THEORY

Fazio & Olson's (2003) treatment of theory in implicit social cognition was limited to describing (*a*) the associative basis for knowledge underlying implicit measures and (*b*) Fazio's (1990) dual-process theory of dissociations between indirect and direct measures of social attitudes (MODE model). Subsequent findings with indirect measures have prompted substantial further theory development. This section starts with the largest piece of that development, which consists of alternative conceptions based on the duality theme that was central to Fazio's MODE model.

How Does Implicit Social Cognition Connect to Social Judgment and Behavior?

The relation of social cognition to behavior has been addressed in multiple dual-construct theories. The duality has been formulated in terms of mental representations (e.g., associations versus propositions), mental processes (e.g., automatic versus controlled), mental systems (e.g., impulsive versus reflective), research operations (e.g., implicit versus explicit), and abstract categories (e.g., Type 1 and Type 2). The large number of these formulations makes it impractical to focus on more than a few in this article. Stanovich et al. (2014) have provided a broad overview that describes similarities and differences among dual-construct conceptions.

Several dual-construct conceptions were influenced by Bargh's (1994) description of four component properties of automaticity: awareness, intention, efficiency, and control. Bargh (1994, p. 31) elaborated: "The outcomes of social cognitive processes are very different, depending on whether one is aware of influences, whether one has specific intentions or goals within the situation, whether attentional resources are in ample or short supply, and whether one is motivated to take control over one's decisions and behavior." Fifteen years later, De Houwer et al. (2009) advocated a view that treated "automatic" and "implicit" as synonyms, the definitions of which would need to be developed in future programmatic experimental research. Describing the current status of that research, they referred their readers to an eight-component conception of automaticity offered by Moors & De Houwer (2006). At present, many scholars agree that automaticity plays a role in indirect (implicit) measures, but little effort has been invested in validating multicomponent definitions of automaticity. Contemporary researchers often use "automaticity" primarily as it is understood in everyday speech.

Most dual-construct conceptions are stated broadly enough to make it difficult to find empirical confrontations among them. This article therefore makes no attempt to choose among them. In the following discussion, the duality will be distinguished generically as associative versus rule-based conceptions (borrowed from Sloman 1996 and Smith & DeCoster 2000). As used here,

these two terms should not be read as implying a preference among the various modes of dual-construct theory—representations, processes, systems, operations, and abstract categories.

Among the various dual-construct conceptions, Strack & Deutsch (2004) most directly considered both (*a*) how associative knowledge might, by itself, produce behavior and (*b*) how it might cooperate with rule-based knowledge in producing behavior. They suggested two possibilities for direct causation of behavior by associative knowledge. One is ideomotor action (see James 1890, chapter 10), by which the thought or perception of an action may elicit performance of that action. The second was the hypothesis that "semantic concepts can be directly connected to motor programs" (Strack & Deutsch 2004, p. 224). For this, Strack & Deutsch offered the example of Bargh et al.'s (1996) conception of stereotype activation.

Other than a direct path from associative knowledge to behavior, there are two forms of explanations for findings of correlations between indirect measures and behavior. One is that association strengths measured by priming or IAT measures and the behaviors with which they correlate are shaped by some (perhaps many) of the same influences. The other is that associative and propositional processes—to use Gawronski & Bodenhausen's (2006) designations—may cooperate in influencing behavior. Pursuing this cooperative causation idea, Greenwald & Banaji (2017) proposed that IAT-measured associations (automatically) shape the content of conscious thoughts but do not produce behavior directly. Rather, the automatically provoked conscious thoughts may guide judgments and decisions. There is presently no empirical basis to choose among the three proposed forms of explanation for observed correlations between indirect measures and behavior: (*a*) automatic effects of associations on behavior, (*b*) overlapping influences that (independently but relatedly) produce both associative knowledge and related behavior, and (*c*) cooperative causation in which automatically activated associations produce conscious judgments that play at least a partial role in guiding judgments and behavior.

The Roles of Person, Culture, and Situation

Fazio & Olson (2003, p. 316) described a distinction between personal and extrapersonal influences on implicit measures. They viewed the standard IAT as an extrapersonally influenced measure that captures "associations that do not contribute to one's evaluation of an attitude object and thus do not become activated when one encounters the object but that are nevertheless available in memory" (Olson & Fazio 2004, p. 653). To measure the knowledge that they theorized to be more directly contributing to the evaluation of attitude objects, Olson & Fazio (2004) created a personalized IAT procedure that replaced the standard attitude IAT's "pleasant" and "unpleasant" classification labels with "I like" and "I don't like." Olson & Fazio theorized that classifying into these categories would oblige IAT respondents to base their responses more on personal liking for stimuli representing the target concepts than on (presumed) extrapersonal categorical associations with those stimuli.

Payne et al. (2017, p. 233) offered a different conception of extrapersonal influences, proposing that "most of the systematic variance in implicit bias is situational." They defined situational as "refer[ring] to an immediate social situation such as interacting with a Black experimenter during a lab study, or broader situations such as living in a particular place and time" (Payne et al. 2017, p. 236). Several of the commentaries that accompanied the publication of their article suggested that Payne and colleagues had not sufficiently credited evidence that IAT measures capture reliable individual differences. At the same time, Payne and colleagues had not yet developed methods to identify sources of situation-induced effects that would explain substantial variance in IAT or other indirect measures. Their main relevant empirical finding was a demonstration of "very high levels of stability [across time] in the average level" of IAT-measured racial attitudes, using mean IATs of

White respondents in each US state, compared across a period of 10 years (see Payne et al. 2017, table 1).

To test Olson & Fazio's (2004) hypothesis [and, without being aware of it at the time, also Payne et al.'s (2017) later hypothesis] that extrapersonal knowledge provided the basis for indirect (especially IAT) measures, Nosek & Hansen (2008, table 3) asked subjects to provide judgments of attitudes of "the average person," "most people," and "the culture you live in" for 95 attitude topics ($N > 100{,}000$). Contrary to expectations of the extrapersonal knowledge hypothesis, these measures of the expected attitudes of others did not predict IAT scores. Rather, subjects' IAT scores were predicted more strongly by their self-report (explicit) attitude measures. Nosek & Hansen nevertheless concluded—along with virtually all others who have commented on their findings—that the experience of one's own culture is the most plausible source of associative knowledge revealed by IAT measures. As they stated, "We suggest that people do not have introspective access to the associations formed via experience in a culture" (Nosek & Hansen 2008, p. 553). Relatedly, Banaji (2001, p. 139) had previously observed that "implicit attitudes [reflect] experiences within a culture that have become so integral a part of the individual's own mental and social makeup that it is artificial…to separate such attitudes into 'culture' versus 'self' parts."

BIT: balanced identity theory

The Roles of Self and Identity in Implicit Social Cognition

An unexpected finding from an early IAT study of gender stereotypes prompted a theoretical development that is still unfolding. Rudman et al. (2001) tested whether men and women held similar gender–potency (male = strong, female = weak) and gender–warmth (female = warm, male = cold) stereotypes. On the basis of evidence obtained from Project Implicit's online data archive, showing similar gender–science and gender–career stereotypes in men and women (see Nosek et al. 2007), Rudman and colleagues expected that their implicitly measured stereotypes would be similar for men and women. To their surprise, findings showed that men (but not women) associated female more with weak than with strong, and women (but not men) associated male more with cold than with warm. They concluded that men and women possessed "implicitly self-favorable" gender stereotypes (Rudman et al. 2001, p. 1176).

Rudman and colleagues' findings prompted the formulation of a "unified theory" of implicit social cognition (Greenwald et al. 2002), later renamed balanced identity theory (BIT). The theory borrowed (and elaborated) consistency principles from Heider's (1958) balance theory to derive predictions of consistency relations among attitudes, stereotypes, identities, and self-esteem.

Tests of BIT require trios of implicit or self-report measures, including (*a*) a group–self association (an identity), (*b*) a group–attribute association (often an attitude toward the group or a stereotype of the group), and (*c*) a self–attribute association (self-esteem or self-concept). As an example, BIT's balance-congruity principle (Greenwald et al. 2002, pp. 5–6) predicts that men who have (*a*) a strong male gender identity and (*b*) a gender stereotype that associates male with career and female with family should also have (*c*) a strong career self-concept that associates self more with career than with family. In contrast, women who have a strong female identity along with the same gender stereotype (associating female more with family than with career) should associate self more with family than with career.

Combining principles of cognitive consistency developed in the 1940s and 1950s (e.g., Heider 1946, 1958) with concepts of social identity developed in the 1970s and 1980s (e.g., Tajfel et al. 1971, Turner et al. 1987), BIT generated predictions that could not be made by either of these substantial bodies of prior theory separately. In a meta-analysis of 36 studies, BIT's predictions were more consistently and strongly confirmed in tests with IAT measures than with self-report measures (D. Cvencek, A.N. Meltzoff, C.D. Maddox, B.A. Nosek, L.A.

Rudman, et al., unpublished manuscript). This confirmation of BIT provided what amounted to nomological validation (Cronbach & Meehl 1955) of the IAT as a method for measuring associatively conceived social-cognitive constructs.

Cohen's *d*: a measure quantified in standard deviation units

Conscious Awareness of Associative Knowledge

Even though implicitly measured attitudes are not defined here as unconscious attitudes, it is of interest to know the extent to which those who provide implicit measures are aware of the knowledge that those measures reveal. To pursue that interest, Hahn et al. (2014) asked subjects to predict the relative preferences that would be revealed by five IAT attitude measures that those subjects had not yet taken. The subjects had received a description of what each IAT measured and how the IAT worked, after which they completed two practice IATs unrelated to the five whose results they would later predict. Each of the five IATs contrasted two social categories—Black versus White, Latino versus White, Asian versus White, celebrity versus regular person, and child versus adult. Before making each prediction, the subjects received descriptions of the pair of concepts involved in each IAT. Hahn and colleagues concluded that their subjects' level of accuracy in predictions revealed some level of conscious access to association strengths measured by the IAT. Hahn & Gawronski (2019) later found that asking subjects to predict their IAT scores can increase acknowledgment of bias, a finding that might eventually contribute to designing methods to ameliorate the discriminatory effects of implicit biases.

A related question about the role of conscious process in performance on indirect measures is, Can respondents to implicit measures consciously control what those measures will reveal? This was first investigated for the IAT, in several studies of faking by Banse et al. (2001) and Kim (2003). Even in the presence of automatic processes that affect IAT performance, subjects can control their IAT scores by deliberately slowing their responding in one or the other combined task. Although few subjects spontaneously succeed when asked to fake an IAT result, most can easily follow the (effective) faking instruction to give slow responses for one of the IAT's two combined tasks. There is no evidence, however, that subjects can control their IAT scores by trying to increase speed (relative to performance following standard instructions) in either combined task. Cvencek et al. (2010) reviewed faking studies and evaluated statistical methods to detect faking. At present, there is little reason to believe that more than a small percentage of subjects attempt to fake their scores in research or educational uses of the IAT.

FINDINGS

Pervasiveness of Implicit Biases

The first large-sample study of Internet-administered attitude and stereotype IAT measures (Nosek et al. 2007) documented a striking difference between IAT and parallel self-report measures: IAT measures were generally more polarized (further from neutral) than were parallel self-report measures. Measuring in standard deviation units (Cohen's *d*), Nosek and colleagues found that polarization of the mean for the self-report race attitude measure was $d = 0.31$, indicating a level of explicit racial preference for White that is conventionally between small and medium.[2] The race attitude IAT for the same respondents ($N = 732{,}881$) had a mean *d* of 0.86, revealing a substantially more polarized average level of implicit (compared to explicit) preference for White.

[2]The established conventions for Cohen's *d* interpret values of 0.2, 0.5, and 0.8, respectively, as small, medium, and large effect sizes.

Applying statistics of the normal distribution, a mean d of 0.31 corresponds to 54.4% of a population having at least a conventionally small ($d = 0.2$) level of self-report-measured preference for White; $d = 0.86$ corresponds to 74.5% of a population having a more than conventionally small IAT-measured preference for White. Across 14 social attitude and stereotype measures for which Nosek et al. (2007) reported findings (with sample sizes ranging from 28,816 to 732,881), the weighted mean absolute level of polarization was $d = 0.51$ for self-report measures, compared to 0.87 for IAT measures. Mean IAT was more polarized than mean self-report for 12 of the attitude or stereotype topics. In contrast, for three political (i.e., non-intergroup) attitude topics, mean IAT scores were less polarized than were mean self-report scores.

IEC: implicit–explicit correlations

A necessary consequence of the greater polarization of IAT-measured compared to self-report-measured attitudes and stereotypes is that more respondents meet criteria for having scores that deviate from neutrality in a biased direction; implicitly measured biases are therefore often found to be more pervasive than are explicitly measured biases. However, there are occasional important exceptions to this generalization, including the observation that, for African Americans, ingroup-favorable explicit attitudes are substantially stronger than are their implicitly measured ingroup-favoring attitudes.

Moderators of Implicit–Explicit Correlations and Implicit–Criterion Correlations

Table 2 summarizes a very large amount of data that were quantitatively reviewed in one meta-analysis of priming measures (Cameron et al. 2012), four meta-analyses of IAT findings (Greenwald et al. 2009, Hofmann et al. 2005a, Kurdi et al. 2018, Oswald et al. 2013), and one experimental study of 57 IAT measures (Nosek 2005).[3] The results will be summarized briefly, focusing on conclusions that can be stated confidently about magnitudes of implicit–explicit correlations (IECs) and implicit–criterion correlations (ICCs) (see aggregate r column in **Table 2**) and variables tested as moderators of those magnitudes.

Both implicit–explicit and implicit–criterion correlations are consistently positive and small to moderate in magnitude. The consistent positivity of IECs (apparent in the upper portion of **Table 2**) is a distressing finding to those who assume that implicit measures capture unconscious attitudes and stereotypes that should often be dissociated from (i.e., uncorrelated with) parallel explicit measures.

IECs and ICCs are shown in **Table 2** to be smaller in magnitude in the more recently reported quantitative reviews. There are three plausible explanations: (*a*) Researchers may have been more reluctant to publish nonsignificant findings in the earlier years of research using implicit measures; (*b*) meta-analyses in more recent years have not included content domains that have relatively large IECs and ICCs; and (*c*) the more recent meta-analysis authors may have searched more diligently to include broad variation on potential moderators. It would be inappropriate to conclude that IECs or ICCs are diminishing with the passage of time.

Implicit–explicit and implicit–criterion correlations are lower when social sensitivity of topic is high. Social sensitivity is the extent to which self-reporting an attitude or stereotype "might activate concerns about the impression that the response would make on others" (Greenwald et al. 2009, p. 20). Social sensitivity is high for attitudes or stereotypes regarding

[3]Another large study (95 topics) of IECs for IAT measures is not included in **Table 2** because its data for moderators, although recorded in its archived data set, were not analyzed for the published report (Nosek & Hansen 2008).

Table 2 Summarized findings of six meta-analyses of correlational validity

Reference	Journal	IAT or priming	Number of samples	Number of subjects	Aggregate *r*	Tests of moderator variables[a]					
						Social sensitivity[b]	Controllability[c]	Correspondence	Affectivity[d]	Relative scoring[e]	I–E correlation[f]
Analyses of implicit–explicit correlations											
Hofmann et al. 2005a	PSPB	IAT	126	12,289	0.240	ns	Negative	No data	Positive[g]	Positive[g]	NA
Nosek 2005	JEPG	IAT	57	6,836	0.360	Negative	No data	No data	Positive	Positive	NA
Greenwald et al. 2009	JPSP	IAT	155	13,121	0.214	Negative	ns[i]	ns[i]	ns	Positive	NA
Cameron et al. 2012	PSPR	Priming	116	11,236[g]	0.200	ns	ns	No data	No data	No data	NA
Oswald et al. 2013	JPSP	IAT	51	12,078	0.140	No data	No data	No data	No data	No data	NA
Kurdi et al. 2018	AmPsy	IAT	160	10,218	0.116	ns[h]	ns[g,h,i]	ns[g,h,i]	ns	ns[g]	NA
Analyses of implicit–criterion correlations											
Greenwald et al. 2009	JPSP	IAT	184	14,900	0.274	Negative	ns	Positive	ns	Positive	Positive
Cameron et al. 2012	PSPR	Priming	86	10,949[g]	0.280	ns	ns	No data	No data	No data	Positive
Oswald et al. 2013	JPSP	IAT	86	17,470	0.140	No data	No data	No data	Positive	Positive	?[j]
Kurdi et al. 2018	AmPsy	IAT	253	36,071	0.097	ns[h]	ns[h]	Positive[h]	ns	Positive	Positive

[a]Moderators that were tested in both univariate and simultaneous regressions are reported as showing a positive or negative relation only if their moderating effect was associated with $p \leq 0.05$ in both tests. "No data" indicates that no data on the moderator were obtained in the meta-analysis. Abbreviations: AmPsy, *American Psychologist*; IAT, Implicit Association Test; JEPG, *Journal of Experimental Psychology: General*; JPSP, *Journal of Personality and Social Psychology*; NA, not applicable; ns, not statistically significant; PSPB, *Personality and Social Psychology Bulletin*; PSPR, *Personality and Social Psychology Review*.

[b]This moderator was called "social desirability" by Hofmann et al. (2005a) and "self-presentation" by Nosek (2005).

[c]This moderator was called "spontaneity" (reversed in direction) by Hofmann et al. (2005a).

[d]This moderator was called "predictor type" by Greenwald et al. (2009) and "evaluative strength" by Nosek (2005).

[e]This includes moderators called "dimensionality" by Nosek (2005) and "complementarity" by Greenwald et al. (2009).

[f]I–E (impicit–explicit) correlation is a potential moderator only in analysis of implicit–criterion correlations.

[g]Information was provided or confirmed by personal communication with authors.

[h]This result is averaged over two tests reported in the meta-analysis.

[i]Moderator was coded in terms of response to the criterion measure rather than response to the explicit measure.

[j]Data for this moderator are available in the data set, but the effect was not tested in the published meta-analysis.

stigmatized groups and low for attitudes toward political parties and for consumer attitudes. Social sensitivity was a significant negative moderator of IEC or ICC magnitudes in three of the eight moderation tests summarized in **Table 2**. Two of the five nonsignificant findings were from the Kurdi et al. (2018) meta-analysis, which was limited to intergroup behavior involving stigmatized groups. Kurdi and colleagues noted that social sensitivity could not be strongly tested in their study because of this restriction in range of the social sensitivity moderator in their data set.

Implicit–explicit and implicit–criterion correlations are unaffected by controllability of responses to explicit measures. Controllability has been defined as "the extent to which [required] responses…were judged easy to consciously control" (Greenwald et al. 2009, p. 20). In Fazio's (1990) MODE model, controllability of responses is expected to weaken both IECs and ICCs, which are predicted to be greater when measures represent spontaneous (i.e., not controllable) processes. The expected negative moderating effect of controllability on IECs and ICCs was found in only one of the eight tests represented in **Table 2**. Kurdi et al. (2018) observed that their meta-analysis found "a sizable number of large ICCs [for] highly controllable behaviors." The absence of a moderating effect of controllability is problematic for the MODE model.

Implicit–criterion, but not implicit–explicit, correlations are higher when there is higher correspondence between implicit and explicit measures. Correspondence was first introduced as a theorized moderator of attitude–behavior correlations by Ajzen & Fishbein (1977). They found that correlations between self-report measures of attitude and behavior were greater when these measures involved the same target (attitude object) and the same action toward the object. They offered this example: "When the behavioral criterion is [attendance at] next Sunday's worship service in his church [a high correspondence] attitudinal predictor would be a measure of the person's evaluation of 'attending my church's worship service next Sunday'" (Ajzen & Fishbein 1977, p. 890). The two nonsignificant effects of correspondence on IECs should be interpreted cautiously; in those two meta-analyses, correspondence was scored in terms of correspondence between IAT and a criterion behavior measure rather than in relation to a parallel explicit measure.

Affectivity as moderator of implicit–explicit and implicit–criterion correlations. The affectivity moderator was operationalized either as an indicator of the strength/polarity of an attitude measure or as the distinction between attitude (higher in affectivity) and stereotype measures (lower in affectivity). This dual definition was applicable both to studies that included only (or mostly) attitude measures and to studies that had substantial numbers of stereotype measures. Affectivity was a positive moderator in three of its seven tests in **Table 2**, indicating that IECs or ICCs are greater in magnitude in attitude studies in which the sample is more polarized on the measure and are also greater in magnitude in studies of attitudes compared to studies of stereotypes.

Relative scoring as a moderator of implicit–explicit and implicit–criterion correlations. This moderator category applies only to studies with IAT measures. In studies of IECs it indicates that the two categories (in both the IAT and the parallel explicit measure) are clearly contrasted, such that liking for one implies disliking of the other. In studies of ICCs it indicates that the criterion behavior measure involves a comparison of the same two categories contrasted in the IAT measure. This moderator yielded positive evidence in six of its seven tests in **Table 2**. The strong conclusion is that correlations will be stronger to the extent that explicit measures or behavioral criterion measures capture the contrast between the two categories used in the IAT that is used to predict them.

ICC: implicit–criterion correlations

Implicit–explicit correlations are potent moderators of implicit–criterion correlations. Correlations indicating that ICCs were significantly predicted by IECs were found in the only three meta-analyses that could test this moderation effect.[4] One plausible explanation for the finding is that higher IECs may occur when the implicit measure and its parallel explicit measure have shared influences. A consequence of these shared influences may be that the two measures reinforce one another in influencing judgments and behavior.

Malleability and Change of Implicit Measures

Prior to 2000, there was widespread belief that implicitly measured attitudes, stereotypes, and self-concepts were quite stable, meaning that they were difficult to change. Devine (1989) and Wilson et al. (2000) considered implicit attitudes to be habitual, requiring sustained motivation and effort to be shifted. Bargh (1999, p. 378) expressed this expectation of stability bluntly: "Once a stereotype is so entrenched that it becomes activated automatically, there is really little that can be done to control its influence." Implicit measures were thought to capture attitudes and stereotypes that could be activated automatically (Devine 1989) and that were difficult to suppress once activated (Macrae et al. 1994).

It did not take long for this dominant view of stability to evolve to one of assumed malleability of implicitly measured attitudes and stereotypes. This change was prompted by studies reporting shifts in implicit attitudes, stereotypes, and self-concepts in response to brief interventions, first reviewed by Blair (2002). In a later review, Gawronski & Sritharan (2010, pp. 224, 233) concluded that "the associations measured by indirect procedures can be formed rather quickly and with relatively little effort" and that "researchers should be cautious in quickly interpreting experimentally induced variations in measurement scores as direct evidence for variations in the underlying associations." Expressing caution as well, Lai et al. (2013) noted that almost no research had examined whether changes in implicit measures corresponded with downstream changes in behavioral outcomes.

More recent research has established that short-term implicit change can coexist with long-term stability. In studies with large sample sizes ($N = 6{,}321$), Lai et al. (2016) showed that eight brief interventions that had previously been shown to reduce implicit race preferences (Lai et al. 2014) had nondurable effects: That is, none of the eight effective interventions produced an effect that persisted after a delay of one or a few days. This lack of persistence was not previously known because more than 90% of prior intervention studies had considered changes only within a single experimental session (Lai et al. 2013).

The few studies reporting evidence for long-term change in implicit attitudes suggest that changing implicit attitudes or stereotypes requires extensive experience. Even these results, however, may be considered uncertain in light of the small number of them produced during a time when the topic of effectiveness of interventions has been of such great interest. In one experiment employing evaluative conditioning (McNulty et al. 2017), subjects completed 13 brief conditioning sessions distributed over 6 weeks, each lasting 6–7 minutes. This intervention produced a change in implicitly measured attitude that persisted for two weeks after the last conditioning session. In another study testing an effect of intergroup contact, college freshmen were randomly assigned to a White or Black roommate. White freshmen who had been assigned to live with a Black roommate showed reduced implicit preference for White after one semester of room sharing (Shook & Fazio 2008).

[4]Oswald et al. (2013) did not report this test, but the archive of their meta-analysis includes the data needed for the test.

PRACTICAL APPLICATIONS

Programs labeled as diversity training and as implicit bias training are now offered in corporations, nonprofit organizations, hospitals, public welfare organizations, public schools, universities, medical schools, law schools, court systems, and police departments. These programs reflect much public interest in understanding and remedying discrimination that could be due at least in part to implicit bias. This section evaluates how scientific work has contributed or might yet contribute to such training efforts. The focus is on findings that have been established by research involving consequential decision making in natural settings.

Diversity or Implicit Bias Training

The stated goals of these training programs are generally in part educational: They typically provide explanations of implicit bias and describe its likely consequences. The functioning of private-sector diversity trainers as educators cannot be evaluated on the basis of published research, mostly because scientists lack opportunities to observe in detail how these efforts are being done and what consequences they have. Nevertheless, it is possible to find in public media various examples of such education, which indicate that at least a portion of the available science is often being properly used in trainers' explanations of how implicit biases can contribute to occurrences of unintended discrimination.

A frequent second goal of diversity training is therapeutic: to reduce audience members' implicit biases or to reduce the likelihood that audience members will be transmitters of unintended discrimination. For this second goal, there is no reason to expect that diversity trainers can reduce implicit biases when researchers cannot reliably produce such effects empirically (see the treatment of malleability of implicit biases in the section titled Findings). Only two studies have tested whether implicit bias education could reduce the likelihood that audience members will become transmitters of unintended discrimination, and the evidence is unclear. Carnes and colleagues (2015) found that gender-bias diversity training increased motivation and self-reported action to promote gender equity among faculty in university departments. However, a follow-up study found that this training did not significantly increase hiring of female faculty (Devine et al. 2017).

Title VII of the US Civil Rights Act of 1964 identified protected classes, which include most prominently race, religion, national origin, age, sex, and disability status. Kalev et al. (2006) and Dobbin et al. (2015) used Equal Employment Opportunity Commission data to evaluate effects of corporate diversity activities on hiring of women and minorities, from which they concluded that most corporate diversity training efforts are ineffective.

In their review of 985 research studies of diversity training, Paluck & Green (2009, p. 339) concluded that "a small fraction [of the studies] speak convincingly to the questions of whether, why, and under what conditions a given type of [prejudice reduction] intervention works.... The causal effects of many widespread prejudice-reduction interventions, such as workplace diversity training and media campaigns, remain unknown." A recent meta-analysis by Bezrukova et al. (2016) found that many of the 260 studies they reviewed reported significant findings. However, they did not offer conclusions about how to construct a successful diversity training program in a corporate setting.

Remedying Unintended Discrimination

Computer software programs that control ongoing physical processes are said to operate in real time. In this sense, implicit biases operate in real time, influencing ongoing social interactions.

An important application question is, Does psychological research identify remedies that (in real time) can disrupt or counteract implicit biases that might produce unintended discrimination? This resolves to two questions: First, while making judgments that could produce disparities, can decision makers sense that implicit biases are operating? Second, when (or if) they are aware that implicit biases are operating, can decision makers avoid unintended discrimination?

The associative-propositional evaluation (APE) theory of Gawronski & Bodenhausen (2006) assumes the existence of this conscious-override possibility. They wrote that "if…the propositional implication of an automatic affective reaction is inconsistent with other relevant propositions, it may be considered invalid" (p. 694). As described in the section titled Theory, Hahn et al. (2014, p. 1387) found noteworthy accuracy of subjects in predicting scores on their own IAT measures and interpreted this as "suggest[ing] that people can sense their internal spontaneous reactions." At the same time, there remain alternative interpretations for the accuracy of self-predictions of IAT results that do not presume conscious awareness of one's associative knowledge and that will be difficult to rule out in further research. There is not yet a research test of the hypothesis that people in responsible positions (e.g., managers, judges, doctors, police) can sense that implicit bias is active in ways that might influence their decision making.

Accounts of diversity training in popular media often suggest relying on one's own resources to intercept implicit biases—perhaps by pausing to think deliberately or by meditating before making decisions that might adversely affect others. Convincing evidence for the effectiveness of these strategies is not yet available in peer-reviewed publications.[5]

As described in the section titled Findings, three meta-analyses (Cameron et al. 2012, Greenwald et al. 2009, Kurdi et al. 2018) found that conscious controllability of performances on criterion measures of discriminatory judgment or behavior did not have its expected effect of reducing the correlation between implicit biases and discriminatory judgments or behaviors. Kurdi et al. (2018) additionally coded subjects' awareness that the criterion behavior involved discrimination, finding (contrary to the conscious-override hypothesis) that both the controllability and the awareness moderator were unrelated to correlations of IAT measures with discriminatory judgment and behavior. Evidence supporting the conscious-override process is, at present, insufficient.

Likening the operation of implicit attitudes and stereotypes to visual illusions, Greenwald & Banaji (2017) proposed that implicit-stereotype-influenced judgments are social illusions that cannot immediately be identified as errors, nor can they be corrected by introspective efforts. This view is less supportive than is Gawronski & Bodenhausen's (2006) APE model of the possibility of online detection and conscious override of implicit biases.

Fortunately, there are remedies for unintended discrimination that require neither awareness that one has potentially biasing associative knowledge nor conscious effort to suppress bias. Because of its simplicity and effectiveness, a preferred strategy is blinding, such as in orchestral blind auditions in which candidates for instrumental positions perform behind a screen (Goldin & Rouse 2000). With effective blinding, bias based on demographic characteristics is not possible.

When blinding is not possible, strategies of discretion elimination may be available. In US court decisions involving employment discrimination, decision-maker discretion in evaluating job applicants and employees has been identified as a policy that enables discriminatory personnel decisions (e.g., Hart 2005, Heilman & Haynes 2008). Discretion can be sharply reduced when decision makers precommit to valid decision criteria before they conduct evaluations, as required

[5]Greenwald & Banaji (1995, p. 17) were early supporters of the now questionable idea that concentrated thought can interrupt implicit bias, as discussed in their section titled "Attention as a Moderator of Implicit Cognition."

in the hiring procedure called structured interviewing (Campion et al. 1988; cf. Uhlmann & Cohen 2005). When decisions are conscientiously based on valid criteria that decision makers will not revise as they are deliberating, implicit biases should have less chance of influencing decisions. A limitation of the discretion-elimination strategy is that there are many circumstances in which the needed valid criteria to which decision makers can be precommitted have not been established.

EXPECTABLE FUTURE DEVELOPMENTS

Since Fazio & Olson's (2003) review of implicit social cognition, research on the topic has been growing. This large body of research has achieved many accomplishments, but important questions remain to be answered.

Measurement: Improvement in Psychometrics of Implicit Measures

Measures introduced 33 years ago (evaluative priming) and 21 years ago (IAT) remain the most used measures in implicit social cognition research. The success of those methods is due to their adaptability in assessing multiple social-cognitive constructs—attitudes, stereotypes, identities, and self-esteem. This usefulness notwithstanding, measures derived from priming and IAT methods have (at best) modest TRR. The trial-to-trial variability of reaction-time methods (14 of the 18 methods in **Table 1**) poses a general challenge to measurement reliability. Indirect measures based on self-report (of which there are three in the third section of **Table 1**) often have good TRR, but they remain at a disadvantage in establishing correlational evidence for construct validity. Although it remains appropriate to seek new measures that offer superior combinations of TRR and correlational construct validity, there is presently no indication that such new measures will emerge soon.

Two established effective methods of coping with limitations due to measurement unreliability can be increasingly used in future implicit social cognition research. The TRR of measures on individuals is managed in other research domains by averaging separated repetitions of measures. The repetition strategy is easily applicable to IAT and priming measures, the main cost being increased subject participation time. In tests of correlational hypotheses, structural equation modeling is an established method of dis-attenuating correlations that are reduced in magnitude due to unreliability of measures. The first such use of structural equation modeling was with IAT measures by Cunningham et al. (2001).

Theory: Dual-Construct Theories Need Sharpening

The section titled Theory noted that the numerous dual-construct theories are often flexible enough to avoid mutual empirical disagreements. Findings of the meta-analyses summarized in **Table 2** show the need for refinement of dual-construct theories by questioning two theoretical ideas that had previously received both scholarly and public acceptance. One is that implicit measures of bias should often be unrelated to parallel explicit measures. Wide acceptance of this expectation of dissociation is likely based in part on experiences with the IAT. Many avowed racial egalitarians have received a race-attitude IAT result crediting them with moderate or strong automatic preference for White. One way to explain away this often undesired discrepancy is to assume that there is no relation between implicit and explicit attitudes. However, meta-analyses consistently find that implicit and explicit attitudes are positively correlated (see **Table 2**). These positive correlations occur for almost every IAT measure ever tested (see the significant positive implicit–explicit correlations found for 95 of 95 IATs in Nosek & Hansen 2008).

Predictive validity: a form of construct validity assessed by finding theoretically expected correlations of the construct's measure with judgment or behavior

The second dual-construct idea to have come under question recently is that implicit measures predict spontaneous or impulsive behavior, while explicit measures predict deliberate or thoughtful behavior. This duality hypothesis, which stems from Fazio's (1990) MODE model, has been a mainstay of diversity training practitioners' armories, prompting their frequent (but scientifically unjustified) advice that one can avoid the influence of implicit biases by slowing oneself and/or actively deliberating when making decisions. Questioning of this duality hypothesis comes from the three predictive validity meta-analyses that tested controllability as a moderator (see lower portion of **Table 2**). If the duality hypothesis is correct, these meta-analyses should have found that behavioral criterion measures are more positively predicted by IAT measures when controllability is low (i.e., spontaneity is high). However, none of the three meta-analyses found the theorized negative moderating effect of controllability.

Theory needed to understand effects of implicit–explicit discrepancies. The case of self-esteem is illustrative. Consequences of discrepancies between implicit and explicit self-esteem have been observed in multiple North American studies since the one by Jordan et al. (2003). The combination of high implicit self-esteem and low explicit self-esteem has been labeled damaged self-esteem (Schröder-Abé et al. 2007) and has been found to be associated with depression (e.g., Smeijers et al. 2017). Interestingly, this discrepancy is the dominant (and certainly non-pathological) pattern among Japanese adults (e.g., Yamaguchi et al. 2007). Another example presents a theoretical puzzle: In the United States, approximately 30% of African Americans show strong Black (i.e., own-group) preference on explicit measures but implicit White (i.e., outgroup) preference on IATs [based on data reported by Nosek et al. (2007, table 3)]. The consequences of this widely occurring discrepancy—e.g., for personality functioning of African Americans—have received neither empirical investigation nor theoretical explanation.

Possibilities for development of theory concerning self-concept and self-esteem. The section titled Theory briefly described some successes of BIT (Greenwald et al. 2002) in predicting correlational findings involving self-esteem. Confirmations of these predictions with implicit measures provide some of the best available evidence of construct validity for implicit measures of self-esteem, also supporting the hypothesis that self-esteem serves an identity-maintenance function (D. Cvencek, A.N. Meltzoff, C.D. Maddox, B.A. Nosek, L.A. Rudman, et al., unpublished manuscript). Cvencek et al. (2016, p. 51) concluded (metaphorically) that self-esteem is "the central gear of an affective–cognitive system" that connects identities (associations of self with a social category) to attitudes (associations of that social category with valence). Although identity maintenance is plausibly an important function of self-esteem, other theories suggest that self-esteem functions more importantly to protect/defend the self (e.g., Greenberg et al. 1992) or to promote the self (e.g., Rogers 1959). There should be future value in studies of these theorized functions using both implicit and explicit measures of self-esteem.

BIT also credits cultural stereotypes with playing critical roles in the formation of self-concepts that associate one's self with the stereotyped trait. An example served as the title of an article by Nosek et al. (2002): "Math = Male, Me = Female, Therefore Math ≠ Me." Nosek and colleagues reported correlational evidence consistent with this prediction, but without empirically pursuing two related questions. First, does this gender–math stereotype function as a self-fulfilling prophecy that may improve men's math performance and impair women's? Second, does this stereotype deter young girls and women from starting or continuing on paths that lead to careers involving math? The explicit form of this stereotype is already understood as a contributor to anxiety (stereotype threat) for women in situations that require math ability (e.g., Spencer et al. 1999).

More questions. The following short list suggests some further possibilities for future implicit social cognition research. The empirical answers to these questions have potential use both in refining theory and in developing useful applications.

- What are the most important childhood experiences that create implicit attitudes, implicit stereotypes, implicit identities, and implicit self-esteem?
- In what order do implicit attitudes, identities, stereotypes, and self-esteem develop?
- What are the long-term trajectories of implicit attitudes and stereotypes through the life span?
- Can implicit measures (perhaps especially of self-esteem) be predictively useful in longitudinal studies?
- How can implicit biases be durably altered?

CONCLUSION

Greenwald & Banaji (2017) recently reflected on the role of their 1995 article as a step in a not-then-apparent revolution in understanding the relation between conscious and unconscious mental processes. They interpreted the subsequent progress of this revolution as undermining both (*a*) the widely shared intuition that conscious sensory perceptions necessarily capture valid external reality and (*b*) the belief that introspection can accurately perceive mental process or content.

Greenwald & Banaji (2017, p. 868) elaborated this view only with a metaphor: "A mass of associative knowledge acts as a cultural filter that elaborates perception and judgment, in ways that can vary across persons when cultural environments have constructed the associative mass idiosyncratically." This metaphor falls well short of being the type of elaboration of theory that is needed for future development of research in implicit social cognition. It does, however, suggest the difficulty of the challenge.

SUMMARY POINTS

1. The currently dominant understanding of "implicit" among social cognition researchers is "indirectly measured." The labels "indirectly measured attitude" and "implicit attitude" are used interchangeably in this review.
2. Some interpret "implicit attitude" as meaning "unconscious attitude." This needlessly commits to a theoretical interpretation that is not established and seems unlikely to become established in the foreseeable future.
3. The social-cognitive constructs most studied in implicit social cognition research are attitudes, stereotypes, identities, self-concepts, and self-esteem.
4. Multiple dual-construct theories have interpreted the implicit–explicit distinction in terms of the relationship between conscious and unconscious determinants of judgment and behavior.
5. Conceptual disagreements concerning the difference between "implicit" and "explicit" have not prevented accumulation of many reproducible research findings using the two most common methods of implicit social cognition research: evaluative priming and the Implicit Association Test (IAT).

6. Six meta-analyses have established that there are generally positive correlations between implicit and explicit measures of social-cognitive constructs and that predictive validity correlations for these constructs are generally positive, although varying widely in magnitude.
7. Several indirect measures of social-cognitive constructs have sufficient internal consistency and test–retest reliability to be effectively useful in correlational studies. Nevertheless, these measures are not yet used in research with the precision required for diagnostic use at the individual level.
8. There is wide interest in developing practical remedies for unintended discrimination, whether due to implicit bias or to other causes. The readiness of practitioners to offer remedies that lack empirical support poses a challenge to researchers.

FUTURE ISSUES

1. Researchers should invest effort in developing novel indirect measures, especially for children aged 2–5.
2. Researchers should refine existing dual-construct (and other forms of) theory to explain effects of implicit–explicit discrepancies and to predict not-yet-observed phenomena.
3. Developmental researchers should incorporate implicit measures in longitudinal studies that use both children and adults as subjects.
4. Indirect and direct measures of self-esteem should be used in combination to increase understanding of the role of self-esteem in personality and development.
5. Innovative methods that might durably modify implicit biases warrant substantial effort at development.
6. In the absence of established implicit-bias-reduction methods, methods to eliminate discretion in personnel decision making should receive much greater use than they have at present.

DISCLOSURE STATEMENT

A.G.G. and C.K.L. are unpaid officers of Project Implicit, Inc., a nonprofit organization that includes in its mission "to develop and deliver methods for investigating and applying phenomena of implicit social cognition, including especially phenomena of implicit bias based on age, race, gender or other factors." As a paid consultant for Project Implicit, C.K.L. provides educational presentations and consulting on research and policy. A.G.G. receives pay for occasional expert testimony on implicit bias in employment discrimination and criminal cases.

ACKNOWLEDGMENTS

The authors wish to thank for their help Mahzarin Banaji, Matthias Bluemke, Daryl Cameron, Russell Fazio, Jon Freeman, Bertram Gawronski, Wilhelm Hofmann, Vera Hoorens, Benedek Kurdi, Brian Nosek, Michael Olson, Fred Oswald, Keith Payne, Mike Rinck, Klaus Rothermund, Denise Sekaquaptewa, Gün Semin, Daniël Wigboldus, and Bernd Wittenbrink.

LITERATURE CITED

Agosta S, Sartori G. 2013. The autobiographical IAT: a review. *Front. Psychol.* 4:519

Ajzen I, Fishbein M. 1977. Attitude-behavior relations: a theoretical analysis and review of empirical research. *Psychol. Bull.* 84(5):888–918

Banaji MR. 2001. Implicit attitudes can be measured. In *The Nature of Remembering: Essays in Honor of Robert G. Crowder*, ed. HL Roediger, JS Nairne, pp. 117–50. Washington, DC: Am. Psychol. Assoc.

Banse R, Seise J, Zerbes N. 2001. Implicit attitudes towards homosexuality: reliability, validity, and controllability of the IAT. *Z. Exp. Psychol.* 48(2):145–60

Bar-Anan Y, Nosek BA. 2014. A comparative investigation of seven indirect attitude measures. *Behav. Res. Methods* 46:668–88

Bargh JA. 1994. The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition. In *Handbook of Social Cognition*, Vol. 1: *Basic Processes*, ed. RS Wyer Jr., TK Srull, pp. 1–40. Hillsdale, NJ: Lawrence Erlbaum Assoc.

Bargh JA. 1999. The cognitive monster: the case against the controllability of automatic stereotype effects. In *Dual-Process Theories in Social Psychology*, ed. S Chaiken, Y Trope, pp. 361–82. New York: Guilford Press

Bargh JA, Chen M, Burrows L. 1996. Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. *J. Pers. Soc. Psychol.* 71(2):230–44

Barnes-Holmes D, Barnes-Holmes Y, Power P, Hayden E, Milne R, Stewart I. 2006. Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *Irish Psychol.* 32(7):169–77

Bezrukova K, Spell CS, Perry JL, Jehn KA. 2016. A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychol. Bull.* 142(11):1227–74

Blair IV. 2002. The malleability of automatic stereotypes and prejudice. *Pers. Soc. Psychol. Rev.* 6(3):242–61

Blair IV, Banaji MR. 1996. Automatic and controlled processes in stereotype priming. *J. Pers. Soc. Psychol.* 70(6):1142–63

Bluemke M, Friese M. 2008. Reliability and validity of the Single-Target IAT (ST-IAT): assessing automatic affect towards multiple attitude objects. *Eur. J. Soc. Psychol.* 38(6):977–97

Cai H, Sriram N, Greenwald AG, McFarland SG. 2004. The Implicit Association Test's D measure can minimize a cognitive skill confound: comment on McFarland and Crouch 2002. *Soc. Cogn.* 22(6):673–84

Cameron CD, Brown-Iannuzzi JL, Payne BK. 2012. Sequential priming measures of implicit social cognition: a meta-analysis of associations with behavior and explicit attitudes. *Pers. Soc. Psychol. Rev.* 16(4):330–50

Campion MA, Pursell ED, Brown BK. 1988. Structured interviewing: raising the psychometric properties of the employment interview. *Pers. Psychol.* 41(1):25–42

Carnes M, Devine PG, Manwell LB, Byars-Winston A, Fine E, et al. 2015. The effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial. *Acad. Med.* 90(2):221–30

Chen M, Bargh JA. 1999. Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. *Pers. Soc. Psychol. Bull.* 25:215–24

Conrey FR, Sherman JW, Gawronski B, Hugenberg K, Groom CJ. 2005. Separating multiple processes in implicit social cognition: the Quad model of implicit task performance. *J. Pers. Soc. Psychol.* 89:469–87

Cronbach LJ, Meehl PE. 1955. Construct validity in psychological tests. *Psychol. Bull.* 52(4):281–302

Cunningham WA, Preacher KJ, Banaji MR. 2001. Implicit attitude measures: consistency, stability, and convergent validity. *Psychol. Sci.* 12:163–70

Cvencek D, Greenwald AG, Brown AS, Gray NS, Snowden RJ. 2010. Faking of the Implicit Association Test is statistically detectable and partly correctable. *Basic Appl. Soc. Psychol.* 32:302–14

Cvencek D, Greenwald AG, Meltzoff AN. 2016. Implicit measures for preschool children confirm self-esteem's role in maintaining a balanced identity. *J. Exp. Soc. Psychol.* 62:50–57

De Houwer J. 2003. The extrinsic affective Simon task. *Exp. Psychol.* 50(2):77–85

De Houwer J, Teige-Mocigemba S, Spruyt A, Moors A. 2009. Implicit measures: a normative analysis and review. *Psychol. Bull.* 135:347–68

Devine PG. 1989. Stereotypes and prejudice: their automatic and controlled components. *J. Pers. Soc. Psychol.* 56(1):5–18

Devine PG, Forscher PS, Cox WTL, Kaatz A, Sheridan J, Carnes M. 2017. A gender bias habit-breaking intervention led to increased hiring of female faculty in STEMM departments. *J. Exp. Soc. Psychol.* 73:211–15

Dobbin F, Schrage D, Kalev A. 2015. Rage against the iron cage: the varied effects of bureaucratic personnel reforms on diversity. *Am. Sociol. Rev.* 80(5):1014–44

Donders FC. 1969 (1868). Over de snelheid van psychische processen [On the speed of mental processes], transl. WG Koster. In *Attention and Performance II*, ed. WG Koster, pp. 412–31. Amsterdam, Neth.: North Holland

Fazio RH. 1990. Multiple processes by which attitudes guide behavior: the MODE model as an integrative framework. In *Advances in Experimental Social Psychology*, ed. MP Zanna, pp. 265–343. San Diego, CA: Academic

Fazio RH, Olson MA. 2003. Implicit measures in social cognition research: their meaning and uses. *Annu. Rev. Psychol.* 54:297–327

Fazio RH, Sanbonmatsu DM, Powell MC, Kardes FR. 1986. On the automatic activation of attitudes. *J. Pers. Soc. Psychol.* 50:229–38

Freeman JB, Ambady N. 2010. MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behav. Res. Methods* 42(1):226–41

Gawronski B, Bodenhausen GV. 2006. Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychol. Bull.* 132:692–731

Gawronski B, De Houwer J. 2014. Implicit measures in social and personality psychology. In *Handbook of Research Methods in Social and Personality Psychology*, ed. HT Reis, CM Judd, pp. 283–310. Cambridge, UK: Cambridge Univ. Press

Comprehensive edited volume reviewing findings and theories across the domain of implicit social cognition.

Gawronski B, Payne BK, eds. 2010. *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*. New York: Guilford Press

Gawronski B, Sritharan R. 2010. Formation, change, and contextualization of mental associations. See Gawronski & Payne 2010, pp. 216–40

Goldin C, Rouse C. 2000. Orchestrating impartiality: the impact of "blind" auditions on female musicians. *Am. Econ. Rev.* 90:715–41

Describes the changing understanding of the relation between conscious and unconscious cognition in the past half century.

Graf P, Schacter DL. 1985. Implicit and explicit memory for new associations in normal and amnesic subjects. *J. Exp. Psychol. Learn. Mem. Cogn.* 11:501–18

Greenberg J, Solomon S, Pyszczynski T, Rosenblatt A, Burling J, et al. 1992. Why do people need self-esteem? Converging evidence that self-esteem serves an anxiety-buffering function. *J. Pers. Soc. Psychol.* 63:913–22

Greenwald AG, Banaji MR. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* 102:4–27

Greenwald AG, Banaji MR. 2017. The implicit revolution: reconceiving the relation between conscious and unconscious. *Am. Psychol.* 72:861–71

Theory that predicts observed affective-cognitive consistency among attitudes, stereotypes, self-esteem, and identities.

Greenwald AG, Banaji MR, Rudman LA, Farnham SD, Nosek BA, Mellott DS. 2002. A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychol. Rev.* 109:3–25

Greenwald AG, McGhee DE, Schwartz JLK. 1998. Measuring individual differences in implicit cognition: the Implicit Association Test. *J. Pers. Soc. Psychol.* 74:1464–80

Greenwald AG, Nosek BA, Banaji MR. 2003. Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* 85:197–216

Greenwald AG, Nosek BA, Banaji MR, Klauer KC. 2005. Validity of the salience asymmetry interpretation of the IAT: comment on Rothermund and Wentura 2004. *J. Exp. Psychol. Gen.* 134:420–25

Greenwald AG, Poehlman TA, Uhlmann EL, Banaji MR. 2009. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* 97:17–41

Hahn A, Gawronski B. 2018. Implicit social cognition. In *The Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, Vol. 4, ed. JT Wixted, pp. 395–427. New York: Wiley. 4th ed.

Hahn A, Gawronski B. 2019. Facing one's implicit biases: from awareness to acknowledgment. *J. Pers. Soc. Psychol.* 116:769–94

Hahn A, Judd CM, Hirsh HK, Blair IV. 2014. Awareness of implicit attitudes. *J. Exp. Psychol. Gen.* 143(3):1369–92

Hart M. 2005. Subjective decisionmaking and unconscious discrimination. *Ala. Law Rev.* 56:741–91

Hedges LV, Tipton E, Johnson MC. 2010. Robust variance estimation in meta-regression with dependent effect size estimates. *Res. Synth. Methods* 1(1):39–65

Heider F. 1946. Attitudes and cognitive organization. *J. Psychol.* 21:107–12

Heider F. 1958. *The Psychology of Interpersonal Relations*. New York: Wiley

Heilman ME, Haynes MC. 2008. Subjectivity in the appraisal process: a facilitator of gender bias in work settings. In *Beyond Common Sense: Psychological Science in the Courtroom*, ed. E Borgida, ST Fiske, pp. 127–56. Oxford, UK: Blackwell Publ.

Hofmann W, Gawronski B, Gschwendner T, Le H, Schmitt M. 2005a. A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Pers. Soc. Psychol. Bull.* 31:1369–85

Ito TA, Friedman NP, Bartholow BD, Correll J, Loersch C, et al. 2015. Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *J. Pers. Soc. Psychol.* 108(2):187–218

Jacoby LL. 1991. A process dissociation framework: separating automatic from intentional uses of memory. *J. Mem. Lang.* 30:513–41

James W. 1890. *The Principles of Psychology*, Vol. 2. New York: Henry Holt & Co.

Jordan CH, Spencer SJ, Zanna MP, Hoshino-Browne E, Correll J. 2003. Secure and defensive high self-esteem. *J. Pers. Soc. Psychol.* 85:969–78

Jung CG. 1910. The association method. *Am. J. Psychol.* 21:219–69

Kalev A, Dobbin F, Kelly E. 2006. Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *Am. Sociol. Rev.* 71:589–617

Karpinski A, Steinman RB. 2006. The single category implicit association test as a measure of implicit social cognition. *J. Pers. Soc. Psychol.* 91(1):16–32

Kim D-Y. 2003. Voluntary controllability of the Implicit Association Test (IAT). *Soc. Psychol. Q.* 66:83–96

Klauer KC, Voss A, Schmitz F, Teige-Mocigemba S. 2007. Process components of the Implicit Association Test: a diffusion-model analysis. *J. Pers. Soc. Psychol.* 93(3):353–68

Kurdi B, Seitchik AE, Axt JR, Carroll TJ, Karapetyan A, et al. 2018. Relationship between the Implicit Association Test and intergroup behavior: a meta-analysis. *Am. Psychol.* 74(5):569–86

Lai CK, Hoffman KM, Nosek BA. 2013. Reducing implicit prejudice. *Soc. Pers. Psychol. Compass* 7(5):315–30

Lai CK, Marini M, Lehr SA, Cerruti C, Shin J-EL, et al. 2014. Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *J. Exp. Psychol. Gen.* 143:1765–85

Lai CK, Skinner AL, Cooley E, Murrar S, Brauer M, et al. 2016. Reducing implicit racial preferences: II. Intervention effectiveness across time. *J. Exp. Psychol. Gen.* 145(8):1001–16

Maass A, Salvi D, Arcuri L, Semin GR. 1989. Language use in intergroup contexts: the linguistic intergroup bias. *J. Pers. Soc. Psychol.* 57(6):981–93

MacCallum RC, Austin JT. 2000. Applications of structural equation modeling in psychological research. *Annu. Rev. Psychol.* 51:201–26

Macrae CN, Bodenhausen GV, Milne AB, Jetten J. 1994. Out of mind but back in sight: stereotypes on the rebound. *J. Pers. Soc. Psychol.* 67(5):808–17

McNulty JK, Olson MA, Jones RE, Acosta LM. 2017. Automatic associations between one's partner and one's affect as the proximal mechanism of change in relationship satisfaction: evidence from evaluative conditioning. *Psychol. Sci.* 28(8):1031–40

Meissner F, Rothermund K. 2013. Estimating the contributions of associations and recoding in the Implicit Association Test: the ReAL model for the IAT. *J. Pers. Soc. Psychol.* 104:45–69

Meyer DE, Schvaneveldt RW. 1971. Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *J. Exp. Psychol.* 90(2):227–34

Mierke J, Klauer KC. 2001. Implicit association measurement with the IAT: evidence for effects of executive control processes. *Z. Exp. Psychol.* 48(2):107–22

Review using Equal Employment Opportunity Commission data to appraise the effectiveness of corporate strategies for improving diversity in hiring.

Series of large-scale experiments comparing 17 interventions to reduce implicit racial preferences.

Mierke J, Klauer KC. 2003. Method-specific variance in the Implicit Association Test. *J. Pers. Soc. Psychol.* 85(6):1180–92

Mogg K, Bradley BP, Field M, De Houwer J. 2003. Eye movements to smoking-related pictures in smokers: relationship between attentional biases and implicit and explicit measures of stimulus valence. *Addiction* 98:825–36

Moors A, De Houwer J. 2006. Automaticity: a theoretical and conceptual analysis. *Psychol. Bull.* 132(2):297–326

Morgan CD, Murray HA. 1935. A method for investigating fantasies: the thematic apperception test. *Arch. Neurol. Psychiatry* 34:289–306

Validation of an indirect (IAT) measure in predicting the likelihood of a repeat suicide attempt.

Nock MK, Park JM, Finn CT, Deliberto TL, Dour HJ, Banaji MR. 2010. Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. *Psychol. Sci.* 21(4):511–17

Nosek BA. 2005. Moderators of the relationship between implicit and explicit evaluation. *J. Exp. Psychol. Gen.* 134(4):565–84

Nosek BA, Banaji MR. 2001. The go/no-go association task. *Soc. Cogn.* 19(6):625–66

Demonstration of math–gender stereotypes' role in reducing women's association of math with self.

Nosek BA, Banaji MR, Greenwald AG. 2002. Math = male, me = female, therefore math ≠ me. *J. Pers. Soc. Psychol.* 83:44–59

Nosek BA, Bar-Anan Y, Sriram N, Axt J, Greenwald AG. 2014. Understanding and using the Brief Implicit Association Test: recommended scoring procedures. *PLOS ONE* 9(12):e110938

Nosek BA, Hansen JJ. 2008. The associations in our heads belong to us: searching for attitudes and knowledge in implicit evaluation. *Cogn. Emot.* 22:553–94

Nosek BA, Hawkins CB, Frazier RS. 2011. Implicit social cognition: from measures to mechanisms. *Trends Cogn. Sci.* 15(4):152–59

Nosek BA, Hawkins CB, Frazier RS. 2012. Implicit social cognition. In *Handbook of Social Cognition*, ed. S Fiske, CN Macrae, pp. 31–53. New York: Sage

Study that summarizes data from 2.5 million completed IATs involving 17 topics.

Nosek BA, Smyth FL, Hansen JJ, Devos T, Lindner NM, et al. 2007. Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur. Rev. Soc. Psychol.* 18:36–88

Nunnally J, Bernstein I. 1994. *Psychometric Theory*. New York: McGraw-Hill. 3rd ed.

Nuttin JM Jr. 1985. Narcissism beyond Gestalt and awareness: the name letter effect. *Eur. J. Soc. Psychol.* 15(3):353–61

Olson MA, Fazio RH. 2004. Reducing the influence of extrapersonal associations on the Implicit Association Test: personalizing the IAT. *J. Pers. Soc. Psychol.* 86:653–67

Oswald FL, Mitchell G, Blanton H, Jaccard J, Tetlock PE. 2013. Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies. *J. Pers. Soc. Psychol.* 105(2):171–92

Paluck EL, Green DP. 2009. Prejudice reduction: What works? A review and assessment of research and practice. *Annu. Rev. Psychol.* 60:339–67

Payne BK, Cheng CM, Govorun O, Stewart BD. 2005. An inkblot for attitudes: affect misattribution as implicit measurement. *J. Pers. Soc. Psychol.* 89(3):277–93

Payne BK, Gawronski B. 2010. A history of implicit social cognition. See Gawronski & Payne 2010, pp. 1–15

Payne BK, Vuletich HA, Lundberg KB. 2017. The bias of crowds: how implicit bias bridges personal and systemic prejudice. *Psychol. Inq.* 28(4):233–48

Ratcliff R, Gomez P, McKoon G. 2004. A diffusion model account of the lexical decision task. *Psychol. Rev.* 111(1):159–82

Reingold EM, Merikle PM. 1988. Using direct and indirect measures to study perception without awareness. *Percept. Psychophys.* 44:563–75

Rinck M, Becker ES. 2007. Approach and avoidance in fear of spiders. *J. Behav. Ther. Exp. Psychiatry* 38(2):105–20

Rogers CR. 1959. A theory of therapy, personality, and interpersonal relationships, as developed in the client-centered framework. In *Psychology: A Study of a Science*, Vol. 3, ed. S Koch, pp. 184–256. New York: McGraw-Hill

Rothermund K, Teige-Mocigemba S, Gast A, Wentura D. 2009. Minimizing the influence of recoding in the implicit association test: the Recoding-Free Implicit Association Test (IAT-RF). *Q. J. Exp. Psychol.* 62:84–98

Rothermund K, Wentura D. 2004. Underlying processes in the Implicit Association Test: dissociating salience from associations. *J. Exp. Psychol. Gen.* 133:139–65

Rudman LA, Greenwald AG, McGhee DE. 2001. Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Pers. Soc. Psychol. Bull.* 27:1164–78

Schacter DL. 1987. Implicit memory: history and current status. *J. Exp. Psychol. Learn. Mem. Cogn.* 13:501–18

Schröder-Abé M, Rudolph A, Schütz A. 2007. High implicit self-esteem is not necessarily advantageous: discrepancies between explicit and implicit self-esteem and their relationship with anger expression and psychological health. *Eur. J. Pers.* 21:319–39

Sekaquaptewa D, Espinoza P, Thompson M, Vargas P, von Hippel W. 2003. Stereotypic explanatory bias: implicit stereotyping as a predictor of discrimination. *J. Exp. Soc. Psychol.* 39(1):75–82

Shook NJ, Fazio RH. 2008. Interracial roommate relationships: an experimental field test of the contact hypothesis. *Psychol. Sci.* 19(7):717–23

Sloman SA. 1996. The empirical case for two systems of reasoning. *Psychol. Bull.* 119(1):3–22

Smeijers D, Vrijsen JN, van Oostrom I, Isaac L, Speckens A, et al. 2017. Implicit and explicit self-esteem in remitted depressed patients. *J. Behav. Ther. Exp. Psychiatry* 54:301–6

Smith ER, DeCoster J. 2000. Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems. *Pers. Soc. Psychol. Rev.* 4(2):108–31

Spencer SJ, Steele CM, Quinn DM. 1999. Stereotype threat and women's math performance. *J. Exp. Soc. Psychol.* 35(1):4–28

Sriram N, Greenwald AG. 2009. The brief implicit association test. *Exp. Psychol.* 56(4):283–94

Stanovich KE, West RF, Toplak ME. 2014. Rationality, intelligence, and the defining features of Type 1 and Type 2 processing. In *Dual-Process Theories of the Social Mind*, ed. JW Sherman, B Gawronski, Y Trope, pp. 80–91. New York: Guilford Press

Review of the wide variety of dual-construct theories in social and cognitive psychology.

Strack F, Deutsch R. 2004. Reflective and impulsive determinants of social behavior. *Pers. Soc. Psychol. Rev.* 8(3):220–47

Stroop JR. 1935. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18(6):643–62

Tajfel H, Billig MG, Bundy RF, Flament C. 1971. Social categorization and intergroup behaviour. *Eur. J. Psychol.* 1:149–77

Teige-Mocigemba S, Klauer KC, Rothermund K. 2008. Minimizing method-specific variance in the IAT: a Single Block IAT. *Eur. J. Psychol. Assess.* 24(4):237–45

Turner JC, Hogg MA, Oakes PJ, Reicher SD, Wetherell MS. 1987. *Rediscovering the Social Group: A Self-Categorization Theory*. Cambridge, MA: Basil Blackwell

Uhlmann EL, Cohen GL. 2005. Constructed criteria: redefining merit to justify discrimination. *Psychol. Sci.* 16:474–80

Wilson TD, Lindsey S, Schooler TY. 2000. A model of dual attitudes. *Psychol. Rev.* 107(1):101–26

Wittenbrink B, Judd CM, Park B. 1997. Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *J. Pers. Soc. Psychol.* 72(2):262–74

Yamaguchi S, Greenwald AG, Banaji MR, Murakami F, Chen D, et al. 2007. Apparent universality of positive implicit self-esteem. *Psychol. Sci.* 18:498–500