Measuring Implicit Intergroup Biases

Calvin K. Lai
Washington University in St. Louis

Megan E. Wilson
Washington University in St. Louis

Implicit intergroup biases are automatically activated prejudices and stereotypes that may influence judgments of others on the basis of group membership. We review evidence on the measurement of implicit intergroup biases, finding: implicit intergroup biases reflect the personal and the cultural; implicit measures vary in reliability and validity; and implicit measures vary greatly in their prediction of explicit and behavioral outcomes due to theoretical and methodological moderators. We then discuss three challenges to the application of implicit intergroup biases to real-world problems: 1) a lack of research on social groups of scientific and public interest; 2) developing implicit measures with diagnostic capabilities; and 3) resolving ongoing ambiguities in the relationship between implicit bias and behavior. Making progress on these issues will clarify the role of implicit intergroup biases in perpetuating inequality.

To understand prejudice and stereotyping, one can often ask people what they feel and believe. When it comes to political parties and sports teams, people often have no qualms or difficulties in expressing their biases for one group over another. This enthusiasm and ease of reporting contrasts with questions about social groups in areas where inequalities in status and power are rife. When asked about race and gender stereotypes, most White American undergraduates report little or no racial stereotypes (Madon et al., 2001), and a majority of Americans state men and women are equally competent and intelligent (Eagly et al., 2019).

The absence of self-reported biases in race and gender stands in contrast to the inequalities revealed in everyday behavior. Over 70% of Black Americans report experiencing racial discrimination at least occasionally in daily life, and 42% of women report experiencing gender discrimination at work (Pew Research Center, 2016, 2017). These everyday experiences are linked to systemic inequalities in income (U.S. Bureau of Labor Statistics, 2019), health (Laveist, 2003; Unruh, 1996), and education (Robinson & Lubienski, 2011; U.S. Department of Education, 2016).

The gap between peoples' egalitarian values and inequalities in everyday life has been the subject of much attention across the social sciences, including research on symbolic racism and ambivalent sexism (Glick & Fiske, 1996; McConahay & Hough, 1976). One line of research has examined the role of implicit biases. The origins of implicit bias research are in cognitive psychology (Jacoby & Dallas, 1981; Shiffrin & Schneider, 1977), where evidence for automatic attention and unconscious memories led to the insight that perceptions of people may be similar (Fazio et al., 1986; Greenwald & Banaji, 1995; Payne & Gawronski, 2010). Much of this research within social psychology has attended to *implicit measures* that indirectly assess mental content through performance in ostensibly unrelated tasks.[1] Example tasks include the Implicit Association Test

---

[1] We refer to 'implicit biases' as the construct of interest assessed by implicit measures. We define implicit measures as measurement outcomes where the measured mental content in the task is automatically retrieved (i.e., uncontrolled, unintentional, unconscious, or efficient). In contrast, explicit measures are measurement outcomes where the measured mental content is deliberately retrieved (i.e., controlled, intentional, conscious, or inefficient; De Houwer & Moors, 2007)

which assesses the speed at which people can categorize social stimuli to one category or another (Greenwald et al., 1998); the Affect Misattribution Procedure which assesses the extent to which observing a stimulus influences the evaluation of another stimulus (Payne et al., 2005); and the First-Person Shooter Task which assesses the propensity for people to shoot individuals with or without a gun when those individuals vary in race (Correll et al., 2002).

Research on implicit measures indicates that implicit preferences and stereotypes in favor of socially dominant groups are pervasive and widespread.[2] In a large-scale study of 2.5 million people who visited the website Project Implicit (Nosek et al., 2007), participants exhibited large implicit preferences for White over Black people ($d = .86$ compared to the point of no preference), young over old people ($d = 1.23$), straight over gay people ($d = .74$), and light skin over dark skin ($d = .73$). Large implicit stereotypes were also observed in topics such as gender and career/family ($d = 1.10$) and gender and science/liberal arts ($d = .93$). In contrast, corresponding explicit self-report measures suggested considerably weaker prejudices and stereotypes (ds = .36, .51, .54, .25, .89, .79).

Implicit biases reflect personal favoritism for the ingroup and knowledge of cultural hierarchy. In the study described above, White participants show implicit preferences in favor of White individuals and Black participants show implicit preferences in favor of Black individuals (Nosek et al., 2007). This tendency toward ingroup favoritism can be attributable to having more favorable experiences with one's own group and being motivated to like one's own group (Richeson & Shelton, 2007; Tajfel & Turner, 1979).

Simultaneously, implicit biases reflect knowledge about cultural hierarchies. While White people's implicit preferences for their own racial group are strong ($d = 1.00$), Black people's implicit preferences in favor of their own group are very weak ($d = .05$, Nosek et al., 2007). This pattern can also be observed among Asian and Hispanic participants who showed implicit biases in favor of White people over Black people, but to a lesser degree than White people do (ds = .88, .79). This is suggestive of Asian and Hispanic people being exposed to the same cultural hierarchies but lacking ingroup favoritism for either White or Black people. Similar relationships to ingroup favoritism and cultural knowledge have been observed experimentally and across religion, age, and sexual orientation (Axt et al., 2014; Olson et al., 2009; Westgate et al., 2015).

Overview

The contrast between explicit prejudices and implicit intergroup biases raises questions about the application of implicit measures for understanding the mind and societal inequality. This review summarizes evidence on measures used to assess implicit intergroup bias. The psychometric evidence demonstrates that implicit measures vary in reliability and validity, reflect construct-related and non-construct-related processes, and vary greatly in their prediction of explicit and behavioral outcomes. We then discuss challenges to the application of implicit intergroup biases to real-world problems.

Inventory of Implicit Tasks

There are many implicit tasks that assess mental contents where the influence of the measured mental content on responses is automatic (i.e., uncontrolled, unintentional, unconscious, or efficient; for reviews, see Gawronski & De Houwer, 2014; Nosek et al., 2011). Table 1 shows 18 tasks assessing implicit intergroup biases organized into three families. All tasks we review in this paper are computerized (for a review of paper-and-pencil tasks, see Sekaquaptewa et al., 2010). All of these tasks have been tested on general adult samples and six tasks have been adapted for child or adolescent samples (Phipps et al., 2019; Rae & Olson, 2018).

One family of tasks includes the Implicit Association Test and its variants. In this family, implicit biases are assessed by the speed and accuracy by which people sort stimuli into different categories. The second family focuses on priming. In most priming tasks, subjects are presented with a prime stimulus followed by a target stimulus. Then, subjects are instructed to respond to the target stimulus. Implicit biases are assessed by the ease or likelihood by which responses are facilitated or impaired by the prime stimulus. Finally, a third family of miscellaneous tasks assesses implicit biases through methods such as indirect self-report, role-playing, or mouse movements.

---

[2] In this paper, we operationalize 'preferences' or 'prejudices' as mental links between social groups and valence (i.e., good/bad) and 'stereotypes' as mental links between social groups and concepts/attributes (e.g., family, athletic).

Table 1
Meta-Analysis of the Internal Consistency and Test-Retest Reliability of Measures of Implicit Intergroup Bias

Families of Tasks	Primary Publication	Meta-analytic Internal Consistency		Meta-Analytic Test-Retest Reliability	
		k	α (95% CI)	*k*	*r* (95% CI)
Implicit Association Test variants					
Implicit Association Test (IAT)	Greenwald et al. (1998)	62	.80 (.77, .83)	16	.49 (.38, .59)
Brief IAT (BIAT)	Sriram & Greenwald (2009)	41	.82 (.77, .87)	21	.51 (.26, .70)
Go/No-Go Association Task (GNAT)	Nosek & Banaji (2001)	6	.65 (.39, .80)	2	.54 (-1.00, 1.00)
Single Target IAT (ST-IAT)	Bluemke & Friese (2008)	9	.77 (-1.00, .86)	7	.46 (.15, .68)
Single Category IAT (SC-IAT)	Karpinski & Steinman (2006)	3	.77 (-1.00, 1.00)	0	N/A
Extrinsic Affective Simon Task (EAST)	De Houwer (2003)	2	-.06 (-1.00, .70)	0	N/A
Recoding Free IAT (IAT-RF)	Rothermund et al. (2009)	2	.74 (.58, .84)	0	N/A
Priming tasks					
Affect Misattribution Procedure (AMP)	Payne et al. (2005)	24	.82 (.75, .87)	4	.56 (-.09, .88)
Evaluative Priming Task	Fazio et al. (1986)	7	.53 (.13, .74)	8	.33 (-.16, .69)
First-Person Shooter Task (FPST)	Correll et al. (2002)	0	N/A	0	N/A
Lexical Decision Task (LDT)	Wittenbrink et al. (1997)	0	N/A	0	N/A
Semantic Priming Task	Blair & Banaji (1996)	0	N/A	0	N/A
Weapons Identification Task (WIT)	Payne (2001)	0	N/A	0	N/A
Other tasks					
Implicit Relational Assessment Procedure (IRAP)	Barnes-Holmes et al. (2006)	7	.66 (.48, .78)	0	N/A
Relational Responding Task (RRT)	De Houwer et al. (2015)	3	.58 (-1.00, .92)	0	N/A
Linguistic Intergroup Bias (LIB)	Maass et al. (1989)	0	N/A	0	N/A
Stereotypic Explanatory Bias (SEB)	Sekaquaptewa et al. (2003)	0	N/A	0	N/A
MouseTracker	Freeman & Ambady (2010)	0	N/A	0	N/A

Note. *k* = Number of effect sizes included in the model, α = raw alpha coefficient, *r* = correlation coefficient, CI = confidence interval. Due to a lack of data, four measures yielded highly imprecise internal consistency estimates (ST-IAT, SC-IAT, EAST, RRT) and one measure (GNAT) yielded highly imprecise test-rest reliability estimates.

Implicit Association Test (IAT; Greenwald et al., 1998). The key feature of tasks like the IAT is the categorization of stimuli into different categories under conditions that are either compatible or incompatible with mental contents. In a race attitude IAT, participants are presented with good and bad words and pictures of White and Black faces. In one set of critical blocks, subjects respond to good words and White faces with one key and bad words and Black faces with the other key. In the other set of critical blocks, subjects respond to the reverse: good words and Black faces with one key and bad words and White faces with the other key. Generally, participants are faster and more accurate when the critical block requires them to respond in a manner that is compatible with their preference (e.g., pairing White faces with good words, Black faces with bad words) and slower and less accurate when the critical block requires them to respond in a manner that is incompatible with their preference (e.g., pairing Black faces with good words, White faces with bad words).

However, the IAT has been subject to criticism for potential methodological limitations. As the IAT is structured as a relative measure, it is difficult to assess biases about a single concept. Attempts to decompose response latencies into single-concept biases do not appear to yield valid results (Nosek et al., 2005), although modeling of error rates has been more successful (Sherman et al., 2008). Another limitation is the possibility

that presenting compatible and incompatible trials separately produces systematic measurement error related to recoding processes (Rothermund et al., 2009). In response, IAT variants have been developed which address these concerns. These include the Single-Category IAT and Single-Target IAT for single targets (Bluemke & Friese, 2008; Karpinski & Steinman, 2006); the Brief IAT for pairs of targets (Sriram & Greenwald, 2009); and the Recoding-Free IAT for eliminating the influence of recoding processes (Rothermund et al., 2009).

Priming. Priming was the first method identified for assessing individual differences in implicit social cognition (Fazio et al., 1986; Wentura & Degner, 2010). In priming tasks, subjects are presented with a prime stimulus that is followed by (or presented with) a target stimulus. Subjects are then instructed to classify the target stimulus as positive or negative (e.g., Evaluative Priming Task, Fazio et al., 1986) as a word or not (i.e., lexical decision task, Wittenbrink et al., 1997), or as having some property or not (e.g., semantic priming task, Blair & Banaji, 1996). The general finding is that responses are facilitated when the target stimulus matches the prime stimulus and impaired when the target stimulus mismatches with the prime stimulus. For example, in the Evaluative Priming Task, subjects are faster and more accurate in classifying the target word 'disgusting' as 'bad' rather than 'good' if 'disgusting' is preceded by a Black face rather than a White face (Fazio, et al., 1995).

A unique priming task is the Affect Misattribution Procedure (AMP; Payne et al., 2005) and its variants, the Semantic Misattribution Procedure (Krieglmeyer & Sherman, 2012) and Stereotype Misperception Task (Sava et al., 2012). In the AMP, subjects are presented with a prime stimulus that is quickly followed by a neutral target stimulus (e.g., Chinese pictograph, abstract painting) and a visual mask. Subjects then rate whether they liked or disliked the target stimulus. The general finding is that the target stimulus is liked more if it was preceded by a positively evaluated stimulus (e.g., an ingroup face) than if it was preceded by a negatively evaluated stimulus (e.g., an outgroup face). However, there is some controversy about whether AMP performance is driven by conscious or intentional ratings of the primes (see "*Internal Validity*" section).

Two priming tasks are specialized for understanding implicit biases related to race and criminality: the Weapons Identification Task (WIT; Payne, 2001) and the First-Person Shooter Task (FPST; Correll et al., 2002). In the WIT, participants are primed with a Black or White face for 200 ms followed by a gun or tool for 200 ms. Then, participants indicate whether they thought the object was a gun or a tool. Participants in the WIT are more likely to misidentify tools as guns when tools are preceded by Black faces rather than White faces. In the FPST, participants are shown images of Black and White targets in scenes holding either a gun or another object (e.g., wallet). Participants are required to quickly decide to "shoot" if the person is holding a gun or "don't shoot" if the person is holding another object. Participants in the FPST are more likely to shoot unarmed Black targets than unarmed White targets and are more likely to fail to shoot armed White targets than armed Black targets.

Other tasks. There are other implicit tasks that complement the IAT and priming tasks. First, the Linguistic Intergroup Bias (Maass et al., 1989) and Stereotypic Explanatory Bias (Sekaquaptewa et al., 2003) tasks assess implicit bias indirectly through self-report. For example, studies on the Linguistic Intergroup Bias find that people are more likely to use abstract language to describe stereotype-consistent events and concrete language to describe stereotype-inconsistent events. Second, tasks such as the Implicit Relational Assessment Procedure (Barnes-Holmes et al., 2006) and Relational Responding Task (De Houwer et al., 2015) assess propositions rather than associations. To do so, they instruct participants to role play as if they believed certain statements (e.g., "respond as if Flemish people are dumber than immigrants" vs. "respond as if Flemish people are wiser than immigrants" to assess Flemish/immigrant attitudes). Finally, the MouseTracker can capture the dynamics of how processes unfold over time by recording the trajectory of computer mouse movements while categorizing stimuli (Freeman & Ambady, 2010).

Psychometric Properties of Implicit Tasks

The utility of a psychological measure depends on its precision of measurement (reliability), its ability to measure constructs of interest (internal validity), and its ability to predict outcomes of interest (predictive validity). Next, we summarize evidence on the reliability and validity of implicit tasks.

Reliability

The predictive validity of an implicit task depends on the precision of measurement, which can be operationalized through internal consistency and test-retest reliability. A meta-analysis by Greenwald and Lai

(2020) examined the internal consistency and test-retest reliability of 18 implicit tasks. However, that meta-analysis examined tasks across all domains. Measures of intergroup bias may have different psychometric properties than measures of other topics. We repeated their meta-analytic method using only data examining intergroup biases. We supplemented their dataset with a literature search of two tasks that are unique to implicit intergroup bias that were not included in their meta-analysis (i.e., FPST, WIT) and the relatively new Relational Responding Task. In total, we analyzed 166 internal consistency effect sizes and 58 test-retest reliability effects from 91 independent samples. See Table 1 for a summary of the results. For a full report of our meta-analysis, see https://osf.io/vq2h5.

Internal consistency. Internal consistency varied greatly among the eleven measures for which estimates were available, with most of the frequently-used measures showing acceptable or good reliability.3 Internal consistency was good or acceptable for six measures: the AMP (α = .82), BIAT (α = .82), IAT (α = .80), ST-IAT (α = .77), SC-IAT (α = .77), and IAT-RF (α = .74). Internal consistency was questionable, poor, or non-existent for four measures: the IRAP (α = .66), GNAT (α = .65), RRT (α = .58), Evaluative Priming Task (α = .53), and the EAST (α = -.06). Unlike most implicit tasks, participants in the Evaluative Priming Task and the EAST are not given instructions to attend to the social category that the stimulus belongs to. Instead, they are instructed to categorize ostensibly unrelated features or stimuli. As a result, evaluations of a target stimulus are less likely to be influenced by attitudes toward that particular social category on a given trial. This feature may partially explain why these two tasks are particularly unreliable (De Houwer & De Bruycker, 2007; Gawronski et al., 2010).

Test-retest reliability. We found test-retest reliability for only six measures.4 Four of the five implicit measures yielded test-retest reliability estimates around r = .50: AMP (r = .56), GNAT (r = .54), BIAT (r = .51), IAT (r = .49), and ST-IAT (r = .46). The Evaluative Priming Task had the lowest test-retest reliability (r = .33). These estimates converge with the meta-analysis of test-retest reliability of implicit measures by Gawronski and colleagues (2017), which also found that comparable estimates for test-retest reliability for

explicit measures tended to be considerably higher (r = .75).

The lack of temporal stability has been interpreted as a problem of measurement error or as a feature of the underlying representation. It is plausible that current implicit tasks are noisy measures of psychological traits. Supporting that account, methodological innovations like using images of groups of people as stimuli produce higher test-retest reliability than using images of individuals (Cooley & Payne, 2017). It is also plausible that the construct assessed by implicit measures is inherently unstable and state-like. Rather than being an enduring attitude, Payne and colleagues (2017) have argued that implicit biases are a form of mental accessibility that is more indicative of transient differences in situations than chronic individual differences.

Internal Validity

Like almost any psychological measure, implicit measures are incomplete accounts of the constructs they seek to assess. Any single implicit task can assess only some of the ways in which a construct can be automatic, and implicit measures may be consistent with multiple theories about the underlying representations.

Automaticity of implicit measures. A mental construct's automaticity can be conceived of as the product of overlapping but distinct features: unconscious, unintentional, uncontrollable, or efficient (Bargh, 1994; Melnikoff & Bargh, 2018; Moors & De Houwer, 2006). Any implicit task will assess aspects of automaticity to different degrees. For example, the IAT has been assumed to assess unconscious mental representations (e.g., McConnell et al., 2011). This is true in that people are often unaware or inattentive to the mental representations measured on the IAT (Hahn & Goedderz, 2020). However, studies find that people are capable of accurately introspecting on how they will perform on the IAT when asked to (Hahn et al., 2014), suggesting that the IAT does not reflect purely unconscious representations. This indicates a need to validate the features of automaticity underlying implicit measures. Although some aspects of automaticity can be pre-specified by a task's design (e.g., some tasks require quick responding, which constrains the ability to engage in slow deliberation),

---

3 As a rule of thumb, α = .90 is excellent, α = .80 is good, α = .70 is acceptable, α = .60 is questionable, and α = .50 is poor. However, this rule of thumb should be interpreted cautiously as the labels are arbitrary and α depends on the number of items and the breadth or narrowness of the construct (Peters, 2014; Taber, 2018)

4 There are few standards for acceptable test-retest reliability as test-retest reliability varies greatly based on the construct, sample, and time between tests (Crocker & Algina, 1986).

other aspects of automaticity must be validated through empirical assessment.

Form of representation. An unresolved debate is about whether the representations assessed by implicit tasks are associative or propositional (Brownstein et al., 2019). Associative representations are defined by simple links between concepts (e.g., police are linked to danger), whereas propositional representations are defined by causal relations between concepts (e.g., police *prevent* danger; police *cause* danger). Determining which account is more plausible has been difficult as some theories argue that associative representations can accommodate propositional information (e.g., Gawronski & Bodenhausen, 2011) whereas other theories argue they cannot (De Houwer, 2009; Mandelbaum, 2016). Another debate is about whether the representations are more like states or traits (see "*Reliability*").

Non-construct-related processes in implicit measurement. Like almost any psychological measure, implicit measures are not process-pure assessments of mental constructs (Calanchini & Sherman, 2013). For instance, implicit measures based on response latencies may be related to general processing speed if not properly controlled for with statistical methods (Greenwald et al., 2003) or formally examined with mathematical models (Calanchini & Sherman, 2013). Mathematical models to identify non-construct-related processes that contribute to implicit measures include the process dissociation procedure (Payne, 2001; Stewart et al., 2009), diffusion modeling (Klauer et al., 2007), and the Quadruple Process model (Quad model; Conrey et al., 2005; Sherman et al., 2008).

The most prominent mathematical model for implicit measures is the Quad model (Conrey et al., 2005; Sherman et al., 2008). In the Quad model, the distribution of errors within an implicit task is used to distinguish four processes which contribute to performance: Activation (the activation of associative content within memory), Overcoming Bias (the ability to override biased associative content), Detection (the ability to detect the correct answer), and Guessing (general response biases). Studies using these parameters have discovered that increased implicit racial biases among older adults reflect reduced self-regulatory ability to override biases rather than more biased associative content (Gonsalkorale et al., 2009) and that Black-White participant differences in implicit racial biases reflect differences in associative content rather than differences in overcoming biases, detection, or guessing (Gonsalkorale et al., 2010).

Failure to follow instructions. Concerns have been raised about the possibility that implicit task performance reflects a failure to follow task instructions. One concern is the possibility that participants fake their responses. Existing evidence finds that people do not spontaneously fake the IAT or the Evaluative Priming Task, suggesting that faking is not a parsimonious account of implicit task performance. However, people can fake implicit measures if instructed to do so (Banse et al., 2001; Czellar, 2006; Degner, 2009; Fiedler & Bluemke, 2005; Kim, 2003; Teige-Mocigemba & Klauer, 2008).

Another concern is the possibility that AMP performance is driven by intentional or conscious ratings of the stimuli. Findings have been mixed. Bar-Anan and Nosek (2012) found that AMP effects were driven by people who reported that they had intentionally rated the prime rather than the target. However, follow-up studies found evidence that intentional rating of primes were the result of people creating post-hoc confabulations about their performance (Gawronski & Ye, 2015; Mann et al., 2019; Payne et al., 2013). People appeared to be misattributing the primes to the targets and assuming they rated the primes intentionally when then did not. Finally, a new working paper using a refined version of prior study designs finds evidence of conscious (but not necessarily intentional) ratings of the primes (Cummins et al., 2019).

Predictive Validity

Implicit intergroup biases are predictive of explicit biases, behavioral outcomes, and regional differences in inequality.

Relationship to explicit prejudice & stereotypes. The relationship between implicit and explicit measures of intergroup bias is consistently positive, but the size of the relationship depends on the topic. In a large-scale study of 57 attitudes (Nosek, 2005), the relationship between IAT scores and explicit intergroup attitudes was as high as $r = .59$ (Democrats vs. Republicans) and as low as $r = .33$ (European Americans vs. African Americans) or $r = .10$ (Thin people vs. Fat people). Generally, implicit-explicit relations are lower in studies on intergroup topics than in other topics (Cameron et al., 2012; Greenwald et al., 2009).
The strength of the relationship between implicit and explicit intergroup biases is moderated by factors which have been documented in one large-scale study and several meta-analyses (Cameron et al., 2012; Greenwald et al., 2009; Hofmann et al., 2005; Nosek,

2005; Oswald et al., 2013). Much of this work has focused on the IAT, finding that implicit-explicit relations are stronger when the attitude is more strongly elaborated, perceived as distinct from other people, has a bipolar structure (i.e., liking for one group implies disliking of the other), and the explicit measure assesses a relative preference rather than an absolute preference (Greenwald et al., 2009; Hofmann et al., 2005; Nosek, 2005).

Relationship to intergroup behavior. Implicit biases predict intergroup behavior over-and-above explicit biases in three of four meta-analyses studying the issue (Cameron et al., 2012; Greenwald et al., 2009; Kurdi et al., 2019; cf. Oswald et al., 2013). The magnitude of the implicit-behavior relation varies greatly depending on the topic, context, and measurement correspondence between the focal implicit and behavioral measures. Meta-analyses consistently find that implicit-behavior relations are stronger when implicit biases are more strongly correlated with explicit biases (Cameron et al., 2012; Greenwald et al., 2009; Kurdi et al., 2019). This supports an account by which low implicit-explicit relations reflect a form of ambivalence that impedes behavioral prediction. However, there are several issues in the research that suggest the implicit-behavior relation is less understood than it appears (see "*Challenges for Applying Implicit Intergroup Bias to Real-World Problems*").

Regional differences in inequality. Emerging evidence finds that implicit intergroup biases have stronger predictive validity for regional differences in group inequalities than individual-level outcomes. For example, county-level differences in implicit racial attitudes are related to county-level racial disparities in the use of lethal force in policing (Hehman et al., 2018), infant health (Orchard & Price, 2017) death rates (Leitner et al., 2016), and disciplinary actions within schools (Riddle & Sinclair, 2019). While Payne and colleagues' (2017) bias-of-crowds model theorizes that this stronger predictive validity for regional inequalities may be because implicit biases are better understood as features of situations rather than of individuals, Connor and Evers (2020) have demonstrated that such patterns may be due to reduced measurement error from statistical aggregation.

Comparing the Psychometric Properties of Implicit Measures

Overall, implicit measures demonstrate evidence for reliability and validity. However, our review shows that specific tasks varied greatly in their reliability and validity. By far, the IAT has the strongest and most robust body of evidence for its reliability and validity. The IAT is high in internal reliability relative to other implicit measures, its strengths and limitations in assessing constructs are well-understood, and almost all meta-analyses of predictive validity have concentrated on the IAT. The Brief IAT seems to have similar properties to the IAT, although there is less research on this measure. The AMP ranked highest in internal reliability and has evidence of predictive validity, but new questions about the internal validity of the task have not yet been resolved (Cummins et al., 2019). A surprisingly poor-performing measure (given its popularity) was the Evaluative Priming Task, which had much lower reliability than most other implicit measures. Finally, many of the other implicit tasks reviewed were relatively lacking in psychometric evaluation, with some tasks having zero relevant studies in our meta-analysis of reliability.

Challenges for Applying Implicit Intergroup Bias to Real-World Problems

The past several decades have shown the value of implicit measures for understanding intergroup biases. However, the continued application of implicit intergroup biases to real-world problems depend on continual improvements to measurement and theory. We review three challenges to future researchers that would improve the utility of implicit biases for real-world problems if addressed: 1) lack of research on groups of scientific and public interest; 2) developing implicit measures with diagnostic capabilities; and 3) resolving ambiguities in the relationship between implicit bias and behavior.

A Lack of Research on Groups of Scientific and Public Interest

The majority of research in this article describes work on a small number of groups that are not representative of broader scientific and public interest (Henrich et al., 2010). In our meta-analysis, 85% of the effects were about race/ethnicity, gender, sexual orientation, or political party. In Kurdi and colleagues' (2019) meta-analysis, 61% of the effects were about race/ethnicity, gender, or sexual orientation (political party was not examined). Within these categories, race/ethnicity was primarily about U.S. White/Black race relations; gender was primarily about men vs. women; sexual orientation was primarily about gay or lesbian individuals; and political party was primarily about U.S. parties.

The dominance of a few topics means a lack of knowledge on major dimensions of social stratification and inequality, with only a few studies on topics such as social class (Horwitz & Dovidio, 2017), age (Lin et al., 2010; Nosek et al., 2007), caste (Dunham et al., 2014), Hispanic people (Bean et al., 2013), transgender people (Rae et al., 2019), disability (Ferrara et al., 2015; Nosek et al., 2007), people with mental illness (Teachman et al., 2006), body image (Marini et al., 2013; Sabin et al., 2012), religion (Axt et al., 2014),), and race/ethnic relations outside of the U.S. (Gündemir et al., 2014). This literature has also failed to consider social categories that do not cleanly fit within social dichotomies, such as multi-racial people (cf. Chen & Ratliff, 2015), bisexual people, gender non-binary people, and intersectional social categories. Future research on understudied groups would expand the generalizability of research on implicit biases.

Developing Implicit Measures with Diagnostic Capabilities

A common misconception is that implicit measures can or should reliably diagnose an individual's chronic level of bias (e.g., Singal, 2017). Implicit measures were not designed to do so and are currently inappropriate for diagnostic use on individuals. This lack of diagnosticity may be a core feature of the implicit bias construct if implicit biases are features of situations rather than individuals (Payne et al., 2017; cf. Connor & Evers, 2020). However, if temporal stability is a measurement error problem, the most important constraint is the lack of test-retest reliability. With a maximum test-retest reliability of $r = .56$, any single assessment is unlikely to say much about an individual over time.

We are aware of two potential solutions to measurement error problems. One solution is to have participants take the same implicit task multiple times over an extended period of time within the same context. This is a common approach for self-assessments of blood pressure, which also fluctuate considerably over time (Parati et al., 2010; Zawadzki et al., 2017). Another possibility is to develop single-administration measures which have higher levels of reliability, such as by increasing the number of trials or by aggregating the results of multiple implicit measures within one session.

Resolving Ambiguities in the Relationship between Implicit Bias and Behavior

Although the relationship between implicit bias and intergroup behavior has been well-validated in hundreds of studies (Kurdi et al., 2019), there are ambiguities in what that relationship means. First, the vast majority of intergroup behaviors studied in this literature are not necessarily indicative of discrimination (Carlsson & Agerström, 2016). For example, many behavioral outcomes looked at behavior directed toward only one group without a comparison to another group. The few discriminatory outcomes that have been studied are also often unreliable or invalid. More research is needed to examine the implicit-discrimination relation with higher-fidelity measures.

Second, implicit measures often do not match the behavioral outcomes they seek to predict. This is important as correspondence is a key factor in the implicit-behavior relation (Gawronski, 2019; Irving & Smith, 2020; Kurdi et al., 2019) and in attitude-behavior relations broadly (Ajzen & Fishbein, 2005; Fazio & Towles-Schwen, 1999). For example, the same IAT assessing general White/Black race attitudes has been used to predict preferences for a White vs. Black work partner (Ashburn-Nardo et al., 2003), medical recommendations for White vs. Black patients (Sabin et al., 2008), and cognitive depletion following an interracial interaction (Richeson & Shelton, 2007). It is difficult to think of any single measure that could strongly predict outcomes that differ in so many dimensions. As a result, many existing studies may not have had the fidelity to study the topics they sought to investigate. Researchers should employ implicit measures that are closely linked to the behaviors they seek to predict.

Finally, changes in implicit biases do not appear to explain corresponding changes in behavior. In a meta-analysis of 492 studies, Forscher, Lai, and colleagues (2019) found that experimental manipulations could change implicit and behavioral measures. However, changes in implicit measures did not statistically mediate corresponding changes in behavior in a subset of 63 studies. The reasons for this are unclear: it may be due to a lack of measurement correspondence, failures to measure implicit biases or behavior reliably, methodological issues with the manipulations, or a lack of a causal relationship between automatically retrieved associations and behavior. Future studies must assess the strength of these explanations.

Conclusion

Implicit intergroup biases reflect the personal and the cultural. Decades of research on measures of implicit intergroup biases have documented their reliability and validity. This work has also established the conditions under which implicit biases are related to explicit biases and intergroup behavior. Future research should investigate social groups that are historically understudied, develop diagnostic implicit measures, and resolve ambiguities in the relationship between implicit bias and behavior. Progress on these issues will clarify the role of implicit intergroup biases in perpetuating inequality.

References

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.). *The handbook of attitudes* (pp. 173–221). Mahwah, NJ: Erlbaum.

Ashburn-Nardo, L., Knowles, M. L., & Monteith, M. J. (2003). Black Americans' implicit racial associations and their implications for intergroup judgment. *Social Cognition*, *21*, 61–87.

Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2014). The rules of implicit evaluation by race, religion, and age. *Psychological Science*, *25*, 1804–1815.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift Fur Experimentelle Psychologie: Organ Der Deutschen Gesellschaft Fur Psychologie*, *48*, 145–160.

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality and Social Psychology Bulletin*, *38*, 1194–1208.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. W. Jr & T. K. Srull (Eds.), *Handbook of Social Cognition* (2nd ed.). Hillsdale NJ: Lawrence Erlbaum.

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. *The Irish Psychologist*, *32*, 169–177.

Bean, M. G., Stone, J., Badger, T. A., Focella, E. S., & Moskowitz, G. B. (2013). Evidence of nonconscious stereotyping of Hispanic patients by nursing and medical students. *Nursing Research*, *62*, 362–367.

Blair, I., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, *70*, 1142–1163.

Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology*, *38*, 977–997.

Brownstein, M., Madva, A., & Gawronski, B. (2019). What do implicit measures measure? *WIREs Cognitive Science*, *10*, e1501.

Calanchini, J., & Sherman, J. W. (2013). Implicit attitudes reflect associative, non-associative, and non-attitudinal processes. *Social and Personality Psychology Compass*, *7*, 654–667.

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

Carlsson, R., & Agerström, J. (2016). A closer look at the discrimination outcomes in the IAT literature. *Scandinavian Journal of Psychology*, *57*, 278–287.

Chen, J. M., & Ratliff, K. A. (2015). Implicit attitude generalization from Black to Black–White biracial group members. *Social Psychological and Personality Science*, *6*, 544–550.

Connor, P., & Evers, E. R. K. (2020). The bias of individuals (in crowds): Why implicit bias is probably a noisily measured individual-level construct. *Perspectives on Psychological Science*.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The Quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–487.

Cooley, E., & Payne, B. K. (2017). Using groups to measure intergroup prejudice. *Personality and Social Psychology Bulletin, 43*, 46–59.

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

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace Jovanovich.

Cummins, J., Hussey, I., & Hughes, S. (2019). *The AMPeror's new clothes: Performance on the Affect Misattribution Procedure is mainly driven by awareness of influence of the primes.* https://doi.org/10.31234/osf.io/d5zn8

Czellar, S. (2006). Self-presentational effects in the Implicit Association Test. *Journal of Consumer Psychology*, *16*, 92–100.

De Houwer, J. (2003). The Extrinsic Affective Simon Task. *Experimental Psychology*, *50*, 77–85.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37*, 1–20.

De Houwer, J., & De Bruycker, E. (2007). The implicit association test outperforms the extrinsic affective Simon task as an implicit measure of inter-individual differences in attitudes. *British Journal of Social Psychology, 46*, 401–421.

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

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes*. Guilford Press.

Degner, J. (2009). On the (un-)controllability of affective priming: Strategic manipulation is feasible but can possibly be prevented. *Cognition and Emotion*, *23*, 327–354.

Dunham, Y., Srinivasan, M., Dotsch, R., & Barner, D. (2014). Religion insulates ingroup evaluations: The development of intergroup attitudes in India. *Developmental Science*, *17*, 311–319.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2019). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*. https://doi.org/10.1037/amp0000494

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

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 10.

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 97–116). Guilford Press.

Ferrara, K., Burns, J., & Mills, H. (2015). Public attitudes toward people with intellectual disabilities after viewing Olympic or Paralympic performance. *Adapted Physical Activity Quarterly: APAQ, 32*, 19–33.

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

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

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*, 226–241.

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

Gawronski, B., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorisation influence spontaneous evaluations of multiply categorisable objects? *Cognition and Emotion, 24*, 1008–1025.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 44, pp. 59–127). Amsterdam, The Netherlands: Elsevier.

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

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

Gawronski, B., & Ye, Y. (2015). Prevention of intention invention in the affect misattribution procedure. *Social Psychological and Personality Science, 6*, 101–108.

Glick, P., & Fiske, S. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*, 491–512.

Gonsalkorale, K., Allen, T. J., Sherman, J. W., & Klauer, K. C. (2010). Mechanisms of group membership and exemplar exposure effects on implicit attitudes. *Social Psychology, 41*, 158–168.

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology, 45*, 410–414.

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

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology, 20*, 419-445.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 17.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm. *Journal of Personality and Social Psychology, 85*, 197–216.

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

Gündemir, S., Homan, A. C., de Dreu, C. K. W., & van Vugt, M. (2014). Think leader, think White? Capturing and weakening an implicit pro-White leadership bias. *PloS One, 9*, e83915.

Hahn, A., & Goedderz, A. (2020). Trait-unconsciousness, state-unconsciousness, preconsciousness, and social miscalibration in the context of implicit evaluations. *Social Cognition.*

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

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

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–83.

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

Horwitz, S. R., & Dovidio, J. F. (2017). The rich—love them or hate them? Divergent implicit and explicit attitudes toward the wealthy. *Group Processes & Intergroup Relations, 20*, 3–31.

Irving, L. H., & Smith, C. T. (2020). Measure what you are trying to predict: Applying the correspondence principle to the Implicit Association Test. *Journal of Experimental Social Psychology, 86*, 103898.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*, 16–32.

Kim, D. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly; Washington, 66*, 83–96.

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality and Social Psychology, 93*, 353–368.

Krieglmeyer, R., & Sherman, J. W. (2012). Disentangling stereotype activation and stereotype application in the stereotype misperception task. *Journal of Personality and Social Psychology, 103*, 205–224.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., … Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist, 74*, 589-586.

Laveist, T. A. (2003). Racial segregation and longevity among African Americans: An individual-level analysis. *Health Services Research, 38*, 1719–1734.

Leitner, J. B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016). Racial bias is associated with ingroup death rate for Blacks and Whites: Insights from Project Implicit. *Social Science & Medicine, 170*, 220–227.

Lin, X., Bryant, C., & Boldero, J. (2010). Measures for assessing student attitudes toward older people. *Educational Gerontology*, *37*, 12–26.

Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, *57*, 981–993.

Madon, S., Guyll, M., Aboufadel, K., Montiel, E., Smith, A., Palumbo, P., & Jussim, L. (2001). Ethnic and national stereotypes: The Princeton Trilogy revisited and revised. *Personality and Social Psychology Bulletin*, *27*, 996–1010.

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, *50*, 629–658.

Mann, T. C., Cone, J., Heggeseth, B., & Ferguson, M. J. (2019). Updating implicit impressions: New evidence on intentionality and the affect misattribution procedure. *Journal of Personality and Social Psychology*, *116*, 349–374.

Marini, M., Sriram, N., Schnabel, K., Maliszewski, N., Devos, T., Ekehammar, B., … Nosek, B. A. (2013). Overweight people have low levels of implicit weight bias, but overweight nations have high levels of implicit weight bias. *PLOS ONE*, *8*, e83543.

McConahay, J. B., & Hough, J. C. (1976). Symbolic racism. *Journal of Social Issues, 32*, 23–45.

McConnell, A. R., Dunn, E. W., Austin, S. N., & Rawn, C. D. (2011). Blind spots in the search for happiness: Implicit attitudes and nonverbal leakage predict affective forecasting errors. *Journal of Experimental Social Psychology*, *47*, 628–634.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*, 280–293.

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

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology. General*, *134*, 565–584.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition*, *19*, 625-666.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 3*1, 166–180.

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, *15*, 152–159.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., … Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88.

Olson, M. A., Fazio, R. H., & Han, H. A. (2009). Conceptualizing personal and extrapersonal associations. *Social and Personality Psychology Compass, 3*, 152–170.

Orchard, J., & Price, J. (2017). County-level racial prejudice and the black-white gap in infant health outcomes. *Social Science & Medicine*, *181*, 191–198.

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

Parati, G., Stergiou, G. S., Asmar, R., Bilo, G., de Leeuw, P., Imai, Y., … Mancia, G. (2010). European Society of Hypertension Practice Guidelines for home blood pressure monitoring. *Journal of Human Hypertension*, *24*, 779–785.

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

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

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

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. K. Payne & B. Gawronski (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1–15). The Guilford Press.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*, 233–248.

Peters, G. J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha. *European Health Psychologist, 16*, 56–69.

Pew Research Center. (2016). On views of race and inequality, Blacks and Whites are worlds apart. https://www.pewsocial-trends.org/2016/06/27/5-personal-experiences-with-discrimination/

Pew Research Center. (2017). 42% of US working women have faced gender discrimination on the job. https://www.pewresearch.org/fact-tank/2017/12/14/gender-discrimination-comes-in-many-forms-for-todays-working-women/

Phipps, D. J., Hagger, M. S., & Hamilton, K. (2019). A meta-analysis of implicit and explicit attitudes in children and adolescents. Unpublished manuscript. Available at https://psyarxiv.com/52jrs/.

Rae, J. R., Gülgöz, S., Durwood, L., DeMeules, M., Lowe, R., Lindquist, G., & Olson, K. R. (2019). Predicting early-childhood gender transitions. *Psychological Science*, *30*, 669–681.

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

Richeson, J. A., & Shelton, J. N. (2007). Negotiating interracial interactions: Costs, consequences, and possibilities. *Current Directions in Psychological Science*, *16*, 316–320.

Riddle, T., & Sinclair, S. (2019). Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Proceedings of the National Academy of Sciences*, *116*, 8255–8260.

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, *48*, 268–302.

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). *The Quarterly Journal of Experimental Psychology, 62*, 84–98.

Sabin, J. A., Marini, M., & Nosek, B. A. (2012). Implicit and explicit anti-fat bias among a large sample of medical doctors by bmi, race/ethnicity and gender. *PLOS ONE*, *7*, e48448.

Sabin, J. A., Rivara, F. P., & Greenwald, A. G. (2008). Physician implicit attitudes and stereotypes about race and quality of medical care. *Medical Care*, *46*, 678–685.

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

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127–190.

Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology*, *39*, 75–82.

Sekaquaptewa, D., Vargas, P., & von Hippel, W. (2010). A practical guide to paper-and-pencil implicit measures of attitudes. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 140–155). The Guilford Press.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*, 314–335.

Singal, J. (2017). Psychology's favorite tool for measuring racism isn't up to the job. *New York Magazine*. Retrieved from http://nymag.com/scienceofus/2017/01/psychologys-racism-measuring-tool-isnt-up-to-the-job.html

Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology, 56*, 283–294.

Stewart, B. D., von Hippel, W., & Radvansky, G. A. (2009). Age, race, and implicit prejudice: Using process dissociation to separate the underlying components. *Psychological Science*, *20*, 164–168.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*, 1273–1296.

Tajfel, H., & Turner, C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks/Cole.

Teachman, B. A., Wilson, J. G., & Komarovskaya, I. (2006). Implicit and explicit stigma of mental illness in diagnosed and healthy samples. *Journal of Social and Clinical Psychology*, *25*, 75–95.

Teige-Mocigemba, S., & Klauer, K. C. (2008). 'Automatic' evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology*, *44*, 1414–1417.

Unruh, A. M. (1996). Gender variations in clinical pain experience. *Pain*, *65*, 123–167.

U.S. Bureau of Labor Statistics. (2019). Labor force statistics from the current population survey. Retrieved August 26, 2019, from https://www.bls.gov/cps/demographics.htm#race

U.S. Department of Education. (2016, May 26). The Condition of Education 2016. Retrieved August 26, 2019, from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2016144

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). New York, NY: Guilford Press.

Westgate, E., Riskind, R., & Nosek, B. (2015). Implicit preferences for straight people over lesbian women and gay men weakened from 2006 to 2013. *Collabra: Psychology, 1*, Art. 1.

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

Zawadzki, M. J., Small, A. K., & Gerin, W. (2017). Ambulatory blood pressure variability: A conceptual review. *Blood Pressure Monitoring*, *22*, 53–58.