Lessons From Two Decades of Project Implicit

Kate A. Ratliff and Colin Tucker Smith

Department of Psychology, University of Florida

Citation: Ratliff, K. A., & Smith, C. T. (2021, in press). Lessons from two decades of Project Implicit. In Krosnick, J. A., Stark, T. H., & Scott, A. L. (Eds.). The Cambridge handbook of implicit bias and racism. Cambridge, England: Cambridge University Press.

When we started graduate school in 2003, few people outside of the field of social psychology were talking about implicit bias. We earnestly explained to our friends and family that people have attitudes and stereotypes that influence how they see and interpret the world around them, and they might not even know it is happening. They were skeptical. We told them about tests that help scientists uncover and quantify these biases. They were not convinced. We told them to read *Blink* (Gladwell, 2005). A "real" author wrote that; they started to get it. Now, of course, implicit bias is discussed everywhere – courtrooms, police departments, offices of human resources, corporate boardrooms, elementary schools, and colleges. The idea that even "good people" may harbor unwanted attitudes and stereotypes is commonplace, ordinary, perhaps even a bit insipid. We seem to have forgotten that, just two decades ago, these ideas were quite radical.

There are many reasons why the implicit bias construct took root in everyday conversation, but one of those reasons is that millions of people have experienced the most commonly used measure of implicit bias—the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998)—by visiting the Project Implicit websites. Project Implicit is a non-profit organization and international research collaboration between behavioral scientists interested in implicit social cognition. In this chapter we provide an overview of Project Implicit and reflect on the contributions and challenges of more than 20 years of internet-based data collection on implicit attitudes and stereotypes.

This chapter is intentionally less formal than a typical academic manuscript as we combine scientific knowledge with a bit of organizational (auto)biography. We have been involved with Project Implicit since the year it was incorporated as a non-profit (2003), first as co-founder Brian Nosek's graduate students and then as members of the Scientific Advisory Board. At the time of writing, Colin Smith is the Director of Education and Kate Ratliff just stepped down from a formal role after serving as the Executive Director for five years in addition to two previous years as the Director of Research. Thus, we are writing "from the inside" though we want to be clear that we are not speaking on behalf of the organization or its other members[1].

In this chapter we first provide an overview of Project Implicit as an organization with a focus on history, mission, and structure. We then describe what we have learned from millions of people's data and discuss how changes in scientific knowledge about implicit bias are reflected in Project Implicit's public education efforts.

Project Implicit, Inc.

In 1998, Brian Nosek was a graduate student at Yale University working with Mahzarin Banaji, a former student of Tony Greenwald's. Greenwald and Banaji had been working for several years on a project, funded by the National Science Foundation, developing the theory of implicit social cognition (Greenwald & Banaji, 1995) that gave rise to the IAT. Nosek wanted to use the Implicit Association Test in his research but was only allotted 15 participant hours through the Yale participant pool. He had recently learned about the problems with low-powered

[1] Though we are not speaking on behalf of the organization, we do thank the following Project Implicit affiliates for their thoughtful comments and suggestions on this chapter: Jordan Axt, Yoav Bar-Anan, Carlee Beth Hawkins, Benedek Kurdi, Calvin Lai, Nicole Lofaro, and Brian Nosek, and Brian O'Shea.

research in a course taught by Alan Kazdin, and so he and his fellow graduate students began searching for creative solutions to increase sample sizes. One strategy was to set up a large tent at Hammonasset Beach on Long Island Sound. They divided the tent up into four sections and administered paper-and-pencil IATs, paying participants $1, a lottery ticket, or a cold soda for the 10-minute study. They would get a few hundred participants a weekend, but it was a lot of work and the strategy was not particularly sustainable.

The next idea was to try collecting data from participants on the internet. In an email exchange with the authors (November 21, 2020), Nosek reports his thinking was that "*Non-scientists loved taking the IAT. I took decks of cards home with me that were IATs with the four categories replacing the suits. My friends sat in a coffee shop for hours playing with it and debating about what it all meant. I know that Mahzarin had similar experiences with people early on—it was so obvious that performing the task was fascinating, instructional, and a way to get people to think about evidence, methodology, bias, and (even) psychology as a science.*"

So Nosek pitched his idea to Banaji and Greenwald, promising that he could build the necessary website based on his undergraduate experience in computer engineering. It was a bold (i.e., overconfident) move; he had, in fact, never programmed in any language for a web application. But the team agreed to try it, seeing it as a way to introduce the about-to-be-published IAT paper in an engaging way—and Greenwald expressed his agreement with the plan by scheduling a news conference for six weeks later. Nosek finished the site on the day he got on the plane to Seattle for the news conference. He recalls watching data come in live as the news stories came out: "*In the first version of the website, I set up the application to compute the scores within the app and just send a single line of data to the database -- e.g., block means, errors. I could watch the file grow live with each person completing a test and their result being* *added to the database. It was truly mesmerizing. Watching a new line come in every few seconds compared to how laborious data collection had been before. It was something of a conversion experience to going all-in on on-line data collection."* Thus, the Project Implicit website was born; as of now, late in 2020, more than 50 million study sessions have been launched with more than 28 million IATs completed.

Between 2002 and 2003, Nosek started a faculty position at the University of Virginia, Banaji moved to Harvard, eventually taking the Project Implicit name and infrastructure along with her (and launching the site's current internet home on the Harvard servers), the National Institutes of Mental Health awarded the team $2.5 million in funding to further develop the virtual laboratory on the internet[2], and the organization was formally incorporated in Massachusetts as Project Implicit, Inc., a non-profit with 501(c)(3), tax-exempt status, as it remains today. Project Implicit's mission is to: (1) develop and deliver methods for investigating implicit social cognition, (2) advance the public's awareness and understanding of implicit social cognition and implicit bias, and (3) advance the field of psychological research.

The organization is structured such that the Board of Directors is the "big picture" decision-making body, approving priorities and budget on an annual basis. The Executive Director is responsible for carrying out the day-to-day work of the organization with the help of a small staff (which has ranged from two to four paid employees). The Scientific Advisory Board is a group of volunteer faculty and post-doctoral scientists who work with the Executive Director and staff to carry out the scientific mission of the organization. One of the organization's goals for the immediate future is to expand and diversify the Scientific Advisory Board, which to date has consisted entirely of "academic descendants" of the organization's founders. In the last

[2] GOLDEN GOOSE. https://www.goldengooseaward.org/awardees

decade, funding for Project Implicit has come primarily through clients who pay for the two main services offered: (1) use of the Project Implicit infrastructure for custom research studies programmed by Project Implicit staff[3], and (2) educational presentations on implicit bias. Project Implicit does accept donations, though no specific efforts at fundraising have been initiated and this income represents a tiny fraction of overall income[4].

The Project Implicit Websites

Project Implicit's primary contribution to the field—which was acknowledged with a 2020 Service to the Field Award from the Society for Personality and Social Psychology—is through its public website (https://implicit.harvard.edu). The site is set up on the model of an interactive exhibit at a science museum (Nosek, Banaji, & Greenwald, 2002a), and includes several core components: U.S. and International Demonstration Sites, Research Sites, and Educational Materials.[5,6] We describe these in more detail in the sections that follow.

United States and International Demonstration Sites

Overview of the Demonstration Sites

The U.S. Project Implicit Demonstration Site is the site people often refer to as *the* Project Implicit website. Visitors to the site choose from a list of 15 tasks (see Table 1 for an overview of task names, category labels, and stimuli used throughout this chapter category labels

[3] These are purely business relationships; Project Implicit staff are not granted authorship on papers stemming from these paid programming transactions.

[4] Over the last five years, Project Implicit's average annual budget was around $450,000.

[5] In this section we describe what is true of the Project Implicit websites at the time of writing; it may no longer be true at the time you are reading. Later we describe some ways in which the information provided at the site has changed over time.

[6] Project Implicit Health (PIH), a site devoted to understanding the role of implicit social cognition in health and wellness, shares the Project Implicit name, and resides on the Harvard server, but operates independently from the main Project Implicit site.

and stimuli). Of the tasks offered, the Black/White Race Attitudes Task receives by far the most traffic. From 2008-2015, 2.2 million participants completed the Black/White Race Attitudes IAT and corresponding self-report items; the task with the next highest amount of traffic was the Gay/Straight Attitudes task with 933,953 participants completing the IAT and corresponding self-report measures. In total, more than 28 million IATs have been completed on the Demonstration Site[7]. In each year since 2016, more than 3 million IATs have been completed. Based on what they self-report, visitors come to the site for the following reasons: An assignment for school or work (61%), recommendation from someone they know (17%), media link (17%), a planned search (1%), or other reasons (4%). Human-subjects ethics approval for studies at the Demonstration Site are granted through Harvard University.

There are also 19 country- or language-specific International Demonstration Sites*,* where visitors complete tasks created in collaboration with a large network of international collaborators. Each task at the U.S. and International Demonstration Sites includes an IAT, a corresponding self-report measure of attitudes or stereotypes, a standard set of demographic items, and, for most tasks, additional self-report measures that are particularly relevant to that task such as items about policy preferences (e.g., support for affirmative action in the race attitudes task, support for the Americans with Disabilities Act in the disability attitudes task) or questions about relevant group memberships that are not covered in the standard demographics (e.g., gender identity in the transgender task, questions about college major in the gender-science stereotypes task)[8]. The U.S. Demonstration Site is presented entirely in English. Anywhere from

---

[7]This is a very conservative estimate; it is a bit complicated to figure out exactly how many tests were completed prior to the migration to the implicit.harvard.edu site in 2003-2004.

[8] We prefer to separately define measures and outcomes in our scientific writing about implicit bias. *Direct measures* of attitudes/stereotypes involve asking participants to self-report their preferences (if an attitude measure) or the traits they ascribe to prototypical members of social groups (if a stereotype measure). *Indirect measures*, on the other hand, either do not alert participants to what is being measured, or reduce participants' deliberative control

10% (Native American task) to 21% (Gender-Career task) of participants are from outside the United States.

IAT Feedback at the Demonstration Sites

At the end of each study, participants receive feedback about their IAT performance (discussed in more detail later in the chapter). Based on the relative latency with which participants sort stimuli in the congruent and incongruent conditions, the computer calculates an IAT *D*-score (Greenwald, Nosek, & Banaji, 2003; D_3 algorithm). Participants then receive one of seven feedback messages for the IAT, ranging from, for example, *Your responses suggested a strong automatic preference for Straight People compared to Gay People* to *Your responses suggested a strong automatic preference for Gay People compared to Straight People*.

Participants with IAT *D*-scores between -0.149 and 0.149 are given the feedback: *Your responses suggested no automatic preference between Gay People and Straight People.* Automatic preference for those participants whose scores IAT *D*-scores range from 0.15 to 0.349 (or -0.15 to -0.349) is described as *slight;* IAT *D*-scores from 0.35 to 0.649 (or -0.35 to -0.649) are *moderate*, and IAT *D*-scores 0.65 or higher (or less than -0.65) are *strong*. The reasoning for these particular cut-offs is that, given that the standard deviations of IAT *D*-scores are rarely greater than 0.5 (Nosek et al., 2007), these IAT *D*-score cutoffs correspond approximately to Cohen's *d* effect sizes of 0.3 (slight preference), 0.7 (moderate preference), and 1.3 (strong preference). These are above Cohen's conventional cutoffs (i.e., 0.2, 0.5, 0.8), because the confidence interval around the estimate of a single score is likely to be greater than that of the

over their responses, even if they are aware of what is being measured (De Houwer, 2006). In this chapter, we refer to *implicit attitudes or stereotypes* as the outcomes of the indirect measure of attitudes/stereotypes (in almost all examples in this chapter this is the IAT), and *explicit attitudes* as the outcomes of direct measures of attitudes/stereotypes (what participants self-report). For ease of discussing attitudes and stereotypes together, we will refer to these as measures of implicit or explicit bias. We do want to be clear, however, that our belief that people harbor internalized prejudices and stereotypes that influence how they see and interpret the world is not dependent on the existence or fitness of any particular measure.

confidence interval based on a sample mean. In other words, the feedback is somewhat conservative. It is our opinion, which we describe in more detail later, that providing performance feedback has an important educational value. That said, the Scientific Advisory Board continues to discuss ways to improve how the IAT is scored and how that score is translated into meaningful feedback for participants.

Limitations of and Future Directions for the Demonstration Sites

There are some notable limitations (and, thus, future directions) of the U.S. and International Demonstration Sites. First, the stimuli on the sites are due for an update; many of the image sets are low-quality and anachronistic. A simple solution, of course, would be to update them. For example, a modernized set of weapons and harmless objects was recently introduced in the Race-Weapons task on the U.S. Demonstration Site. However, any changes to the core tasks on the site require a cost-benefit analysis that weighs what is gained from an update against what is lost by breaking the continuity of data collected for more than 15 years. Answers to some research questions require consistent data across time. At the task level, there are also many instructors out there who use the tasks in their courses and are unhappy to find their favorite is suddenly different or missing entirely. There are some updates that firmly fall into “beneficial” side of a cost-benefit analysis (e.g., implementing best practices for asking about gender identity). Other possible updates are not so clear when weighed against the limitations of change (e.g., is it necessary to change the images from the race task given that they are perfectly acceptable, but a little grainy?).

And, surely the specific tasks that appear on the Demonstration site communicate something—intentionally or not—about what groups are seen as important and worthy of attention. As one example, a Transgender IAT (Axt, Conway, Westgate, & Buttrick, 2020) was introduced this year—the first new task added to the site in 15 years. A Hispanic IAT will go up soon. Thus, some consideration should be given to the tasks themselves and what groups should be represented that currently are not. The pool of tasks could also be expanded to include implicit measures other than the IAT. Having large, geographically and demographically diverse attitudinal data sets for additional implicit measures would increase the breadth of questions we can ask and answer. Further, the inclusion of additional measures would also allow for a more intersectional approach to understanding implicit bias that does not rely on the binary, mutually exclusive categories that the IAT does.

The International Demonstration Sites are in particular need of attention. Initial funding to build the sites came from the National Institutes of Mental Health, and when the funding ran out, development and maintenance on the sites slowed down dramatically. The technology for most of the sites is out-of-date (using Flash for the sites and Javascript on the server-side) as are the supporting scientific materials. Further, many of the International Demonstration Sites do not take advantage of the ability to generate location-specific questions of interest; instead, they only include translated (where applicable) versions of the most popular of the U.S. Demonstration Site tasks. Though there are a few exceptions (e.g., each individual, country-specific International Demonstration Site includes a task assessing attitudes toward that country compared to the United States; there is a caste attitudes task at the India site), the potential for conducting important cross-cultural work is largely untapped. Relatedly, much of the international data are not publicly available due to the lack of resources available to anonymize and clean those data. In the past year, in an attempt to reduce the oversight required, the Scientific Advisory Board trimmed the number of active International Demonstration Sites from 39 to 20. Further, Project Implicit now includes a prominent disclaimer noting these sites are out-of-date and suggesting participants go instead to the U.S. Demonstration site while the team works toward updating the technology and content.

Another major area for improvement is in implementing universal design standards at the sites. Universal design principles urge products that are usable by people with the widest possible range of abilities in the widest possible range of situations. For example, the research experience at the sites is likely inaccessible to people who are blind or visually impaired and require assistive technology (e.g., a computer screen reader).

Finally, we turn our discussion to the issue of participant sampling at the Project Implicit websites. We preface this with a clear statement that participants who visit the U.S. Demonstration Site are not representative of the U.S. as a whole, nor are they representative of any particular population within the United States. We try to be very up front about this when we write papers using Project Implicit data, and we hope others will do the same. That said, there are reasons to feel hopeful about the utility of Project Implicit data for drawing (limited) conclusions about attitudes and stereotypes.

In terms of sheer numbers, the accumulated data at Project Implicit (i.e., 28 million people) is higher than the total combined population of Wyoming, Vermont, Alaska, North Dakota, South Dakota, Delaware, Rhode Island, Montana, Maine, New Hampshire, Hawaii, Idaho, West Virginia, Nebraska, New Mexico, Kansas, Mississippi, and Arkansas. It is likely the largest database of implicit and explicit bias in existence.

Comparing the demographics of participants who completed the Black-White Race task at the Project Implicit U.S. Demonstration Site to a representative population, Hehman, Calanchini, Flake, & Leitner (2019), found that the Project Implicit sample is younger, more likely to be female, and more educated than the general public. These differences, combined with the fact that visitors to the Demonstration Sites self-select a topic of choice, make it difficult to know how *average* responses on measures of implicit and explicit bias differ between Project Implicit participants and a nationally representative sample (though based on the patterns of difference between Project Implicit samples and the general public, we would speculate that Project Implicit data provides an underestimation of the degree of bias).

We feel somewhat more confident that data from the Project Implicit Demonstration Sites performs similarly to representative data in correlational studies. Hehman et al. (2019) found that the regional percentages of Black and White participants at the Demonstration Site correlate at $r = .91$ with local racial demographics. And bias among Project Implicit participants analyzed at the state or county level predicts a number of real-world outcomes (see below for more on this), at least some of which have been shown to replicate in nationally representative samples. For example, Ofosu, Chambers, Chen, and Hehman (2019) used Project Implicit data to demonstrate that same-sex marriage legalization was associated with reduced implicit and explicit anti-gay bias; they then replicated the finding using data from the American National Election Studies (ANES). Further research comparing data collected at the Project Implicit sites to nationally representative samples is urgently necessary for increasing the utility of the vast amounts of data collected at the site.

Research Site

Participants who want to have a research experience not available at the Demonstration Site can register using a unique username and password. They complete a set of demographics at the time of registration. Each time a user returns to the site and login, they are randomly assigned to a research study from the pool of studies that are available at the time. There are currently 2,103,710 accounts registered that have completed at least one study.

The pool of available research studies is maintained by members of the Scientific Advisory Board. Those researchers and their lab members get access as compensation for their labor on the sites. Human-subjects ethics approval for studies at the Research Site is granted by the home institutions of the individual researchers running studies, though Project Implicit also maintains a strict set of guidelines for studies at the site in order to create a positive, educational experience for participants. One member of the Scientific Advisory Board is responsible for approving all studies that go into the research pool to ensure they meet Project Implicit's requirements, which are that: (1) studies can be no longer than 15 minutes (~10 is the goal), (2) study text should be no higher than an 8^{th} grade reading level, (3) studies may not include deception, (4) studies must include some kind of measure about which participants receive feedback, and (5) an appropriate debriefing that fulfills the educational mission of the organization must be offered.

With the exception of self-selection into study topics, the limitations of the Demonstration Site (discussed previously) are also applicable to the Research Site. The Research Site has the additional potential for differential drop-out by condition, which may be particularly worrisome if one condition is more onerous than another given that participants are volunteers who can drop out by simply closing their browser window. Despite these drawbacks, studies at the Research Site can generate data more quickly than most other sources and the large number of visitors at the site make it feasible to obtain very large sample sizes. For example, Bar-Anan and Vianello (2018) reported data from more than 24,000 participants who completed subsets of 3 self-report measures and 7 indirect measures in three attitude domains (see also Schimmack, 2020; Vianello & Bar-Anan, 2020; Kurdi, Ratliff, & Cunningham, 2020). The research pool can also be used to collect data from participants at multiple time points (e.g., Friese, Smith, Plischke, Bluemke, & Nosek, 2012; Ratliff & Nosek, 2008; Smith, Ratliff, & Nosek, 2012), although participant retention is a challenge.

Over the years, participants have had the opportunity to experience literally thousands of different studies in the Project Implicit research pool. These have featured nearly every implicit measure in existence across a wide variety of attitude objects and employed a breadth of designs and explicit measures. We feel that has real value. Of course this benefits the researchers who conduct the studies as well. But we believe that being an active participant in social science and, thereby, contributing to the accumulation of human knowledge has tangible benefits for the participant. We realize the responsibility involved in taking 10 minutes' time from millions of human beings and take that seriously.

Educational and Technological Materials

In addition to the tasks available on the Project Implicit sites, a number of educational resources are available to participants including an overview of the IAT, a Frequently Asked Questions page, and a page outlining ethical considerations people should take into advisement if they want to implement an IAT in research or for other use. Although the site provides basic information about the science of bias, there is considerable room for expanding what is offered to include discussion guides, sample curricula, and other resources for educators, journalists, scientists, and others. Project Implicit has also made its most recent technological platform accessible for anyone to use at no cost (https://minnojs.github.io/docsite/minnosuitedashboard/).

What We Have Learned About Bias from the Project Implicit Datasets

Since 2015, all data collected at the Project Implicit Demonstration Site are made publicly available on the Open Science Framework (OSF; https://osf.io/y9hiq/). Each January-February data from the previous year are released. Project Implicit commits substantial resources to cleaning and posting these data. Datasets from some International Demonstration Sites are also posted when possible (https://osf.io/kaqi5/). Members of the Scientific Advisory Board are all strongly committed to open science; thus, although it is not required, published data from the Project Implicit Research Site are almost always available on the OSF as well and are often reused for multiple projects. For example, Lai and colleagues (2016) used the Research Site infrastructure in a project demonstrating that 9 bias-reduction interventions, implemented across 18 university campuses, changed implicit bias in the short-term, but were not effective after a delay. Vuletich and Payne (2019) re-analyzed these public data to show that individual biases did not return to baseline levels; instead, they fluctuated mostly randomly—but *campus* means returned to preexisting campus means.

As noted earlier, the Project Implicit Demonstration Site datasets are probably the most comprehensive documentation of implicit and explicit biases in existence. Although the datasets would benefit greatly from comparison with representative data, as described previously, they have become a highly valuable scientific resource in their own right. As of now there are approximately 75 published papers using the data, an average of about 15 per year since the data were made available (though the linear trend is toward a substantial increase each year). These publications have been authored by more than 150 unique researchers and have appeared in top general science journals (e.g., *Nature Human Behavior, Proceedings of the National Academy of Sciences*), general psychological science journals (e.g., *Psychological Inquiry, Journal of Experimental Psychology: General*, *Psychological Science*), and social psychology journals (e.g., *Journal of Personality and Social Psychology, Personality and Social Psychology Bulletin, Social Psychological and Personality Science*).

The majority of the knowledge accumulated from access to these publicly available resources falls roughly into four general lines of inquiry: (1) Pervasiveness and correlates of bias, (2) bias over time and in response to specific events, (3) geolocating bias, and (4) understanding responses to learning about bias. Project Implicit data are uniquely suited to answering these questions due to the sheer number of subjects, continuous data collection day and night, and diversity of participants across geographic regions. We describe what we have learned from these lines of inquiry, as well as future directions, in more detail in the sections that follow.

I. Pervasiveness and Correlates of Bias

The first aggregate report of data collected at the Project Implicit website summarized data from approximately 600,000 completed sessions between October 1998 and April 2000 (Nosek et al., 2002a). The next paper summarized data from 2.5 million sessions completed between July 2000 and May 2006 (Nosek et al., 2007). A new paper is currently under review that summarizes data from more than 7 million completed sessions between 2007-2015. Three broad findings emerge from these papers[9].

Bias Favoring Culturally Dominant or Societally Valued Groups is Strong and Pervasive

The data from the Project Implicit Demonstration Site provide clear and consistent evidence of societal bias favoring higher-status groups over lower-status groups. On average, people hold more positive implicit and explicit attitudes toward and stereotypes about those groups that are more culturally dominant or societally valued (e.g., men, White people, straight people, thin people) compared to those that are less culturally-dominant or societally-valued (e.g., women, Black or Asian people, gay people, fat people). While this might seem obvious, it is nevertheless a highly important demonstration—perhaps *particularly* with regard to explicit

[9] The 2002 and 2007 papers summarized Demonstration Site data before Project Implicit made data publicly available; the 2020 paper makes use of the public datasets.

biases—in a time of increased insistence that prejudice is a thing of the past, and that we are "colorblind" or "post-racial" (Neville, Gallardo, & Sue, 2016). The pervasiveness of bias is further evidenced by the finding that there is little difference in strength or direction of these biases when comparing U.S. citizens to citizens of other countries. And the effect sizes of biases favoring more culturally dominant or valued groups over less dominant or valued groups are not trivial. In the Ratliff et al. (2020) analysis, the meta-analytic effect size for implicit bias was d = 0.80 (a large effect) and all but two tasks have effect sizes greater than d = 0.50 (a medium effect size). The meta-analytic effect size for explicit bias was smaller, though also still substantial, with an average d of 0.51.

A specific sub-category of research falling into this general category of descriptive summary data on the pervasiveness of biases examines the magnitude of implicit and explicit biases among people in a particular profession. These findings are typically used to illustrate that people in a specific profession also have biases, just like people in the general population. For example, implicit pro-White/anti-Black race bias is evidenced by teachers (Starck, Riddle, Sinclair, & Warikoo, 2020) and genetic counselors (Lowe, Beach, & Roter, 2020), and health care providers show implicit pro-straight/anti-gay (Prairie, Wrye, Bowman, Weatherby, & Thareja, 2019; Sabin, Riskind, & Nosek, 2015) and pro-young people/anti-old people biases (Maximiano-Barreto, Fabricio, Luchesi, & Chagas, 2020). Although this type of investigation can be conducted without Project Implicit's data, the size of the samples and their geographic coverage serves to increase the reliability and generalizability of specific findings.

Ingroups Are Evaluated More Positively Than Outgroups

The majority of participants show implicit and explicit biases that are more positive toward the ingroup than the outgroup (Ratliff et al., 2020). For example, straight people have a stronger implicit and explicit preference for straight people (relative to gay people) than do gay people, abled people have a stronger implicit and explicit preference for abled people (relative to disabled people) than do disabled people, etc. This overall pattern of implicit and explicit ingroup preference on evaluative tasks is clear and robust.

An important-but-underutilized contribution of Project Implicit Demonstration Site datasets is the ability to examine attitudes and beliefs among members of groups that are underrepresented in psychological research. For example, data are available for over 1.4 million Hispanic participants, 1.3 million Black participants, and 1 million multiracial participants.[10] Jiang, Vitiello, Axt, Campbell, and Ratliff (2020) used these data to understand attitudes toward ingroups and outgroups among people who belong to multiple stigmatized groups (compared to one stigmatized group) in the domains of gender, race, and sexual orientation. The results showed that: (1) there is considerable variation in the strength of ingroup favoritism across members of stigmatized groups, (2) Black people (particularly Black men) showed the weakest levels of ingroup preference, and (3) White women in particular showed the greatest degree of ingroup preferences. This work highlights how occupying multiple stigmatized identities has unique effects on the formation and expression of ingroup attitudes and contributes to a growing literature focusing on people belonging to underrepresented groups as perceivers rather than as targets.

Demographic Differences in Bias Are Quite Small

As described above, participant demographics are a strong predictor of implicit and explicit bias *when the demographic factor in question is directly pertinent to the task and reflects an ingroup/outgroup distinction* (e.g., Black participants show lower levels of implicit and

[10] Some of these participants could be the same people participating on multiple occasions, though the modal responses to the item asking people how many previous IATs they have completed is “none”.

explicit pro-White/anti-Black bias than do White participants). However, across tasks, demographic variables do not relate strongly to implicit or explicit bias. Nosek et al. (2007) and Ratliff et al. (2020) found that there is a slight tendency for men to have stronger implicit and explicit biases than women, particularly for attitude tasks (compared to stereotyping tasks), but these effects are relatively small (implicit: $d = 0.17$; explicit: $d = 0.26$) There is a similarly small relationship between political orientation and biases, such that political conservatives show slightly larger biases (implicit: $r = .09$; explicit: $r = .12$). The effects of other demographic variables were even smaller. While it is clear that bias favoring culturally dominant or societally valued groups is pervasive and robust, variance in these biases is not particularly well-explained by perceiver characteristics (at least not when they are examined as single, simple predictors).

II. Bias Over Time and In Response to Specific Events

Two general approaches have been used to examine time as a moderator of bias. One is to *assess change in average scores on IATs and self-report measures across a long period of time*. For example, Westgate, Riskind, and Nosek (2015) found that, from 2006 to 2013, self-reported preference for straight people relative to gay people decreased by 26%, and IAT scores indicating preference for straight people relative to gay people decreased by 14% (unpublished data from the first author and Jennifer Howell also show that defensive response to feedback indicating an implicit pro-straight preference went up during this time period). In the most comprehensive publication in this vein, Charlesworth and Banaji (2019) summarized patterns of change among 4.4 million data points collected between 2007 and 2016 in six social-group attitudes: sexual orientation, race, skin tone, age, disability, and body weight. They found that self-report measures moved toward neutrality for all six attitude objects. IAT scores moved toward neutrality for sexual orientation, race, and skin-tone attitudes; IAT scores for age and disability attitudes remained stable over time, and IAT scores for weight attitudes moved away from neutrality (i.e., pro-thin/anti-fat preferences increased).

The second approach to looking at time as a moderator of bias *compares average bias in some time period before and after a particular event.* For example, Schmidt and Axt (2016) found that Black/White IAT scores from approximately 2.3 million participants did not discernably change during the first seven years of President Obama's tenure (c.f., Charlesworth & Banaji, 2019). Other interesting examples of this approach include the finding that implicit and explicit pro-White/anti-Black bias is greater when the economy is worse (Bianchi, Hall, & Lee, 2018), and that implicit (but not explicit) pro-thin/anti-fat bias was higher shortly after 20 different highly publicized "fat-shaming" statements made by celebrities (Ravary, Baldwin, & Bartz, 2019). In addition, implicit and explicit pro-straight/anti-gay bias tracked with state level same-sex marriage legislation (Ofosu, Chambers, Chen, & Hehman, 2019) and implicit (but not explicit) pro-straight/anti-gay bias was *very* slightly higher in the time period after the 2014 Ebola epidemic (Inbar, Westgate, Pizarro, & Nosek, 2016), possibly reflecting a more general phenomenon in which society-wide pathogen threat influences attitudes toward gay people[11].

In the future, the Project Implicit websites could be updated more rapidly to collect data in response to events as they are happening rather than waiting for retrospective analysis. For example, early on in the COVID-19 pandemic (March of 2020), Project Implicit issued a general call to researchers to solicit suggestions for items to add the Demonstration Site. The team narrowed the responses down to a few scales—including the Perceived Vulnerability to Disease

---

[11] We note that this paper has a particularly thoughtful and nuanced discussion of their findings with an appreciation for the importance of distinguishing between effect sizes and statistical significance when working with datasets as large as these.

Scale (Duncan et al., 2009) and Intolerance of Uncertainty Scale (Carlton et al., 2007)—that received rapid IRB approval and were deployed across all tasks two weeks later. Data will be released to the public earlier than the typicalzJanuary data release. The site has also hosted special data collections around U.S. elections (e.g., Friese et al., 2012, Greenwald et al., 2009; Ratliff et al., 2019) and around major U.S. sporting events like March Madness or the Superbowl. So, although these examples of rapid response do exist, this kind of adaptability to real world events could be expanded to much greater effect.

III. Geolocating Bias

One way in which Project Implicit's data has proved to be particularly powerful in terms of advancing theory while also contributing to an understanding of systemic racism is through the ability to look at geographically-based correlates of bias. Out of concern for participant data privacy, Project Implicit does not make zip code data publicly available but does use zip code to assign a Metropolitan Statistical Area (MSA) to participants reporting U.S. zip codes. MSAs represent geographical regions bigger than zip codes or counties, but smaller than states. MSAs typically have an area of relatively high population density at the core with surrounding areas in close economic ties to the core. Currently there are 392 MSAs in the United States.

Papers using the Demonstration Site data for this purpose typically aggregate IAT and self-report scores at some geographic unit (e.g., state, MSA) and then correlate those scores with another indicator that is also aggregated at the same geographic unit, while controlling for relevant demographic characteristics at the individual and/or geographic level (see Hehman, Calanchini, Flake, & Leitner, 2019 and Hoover & Dehghani, 2020 for recommendations regarding best practices for this type of analysis). For example, region-level Race IAT scores (higher = more pro-White/anti-Black bias) positively correlate with racial disparities in school discipline (Chin, Quinn, Dhaliwal, & Lovison, 2020; Riddle & Sinclair, 2019), health outcomes at birth (Orchard & Price, 2017), use of lethal force by police officers (Hehman, Flake, & Calanchini, 2018), and rates of Google searches for racist terms (Rae, Newheiser, & Olson, 2015).

These kinds of analysis have also been used to investigate questions related to whether intergroup contact increases or reduces inter-group conflict. For example, Rae et al. (2015) showed that states with larger proportions of Black residents also had stronger implicit and explicit ingroup bias among both White and Black respondents. Contradictorily, a higher proportion of gay people living in an MSA is associated with lower pro-straight/anti-gay bias in the area (MacInnis, Page-Gould, & Hodson, 2017), and implicit pro-straight/anti-gay bias decreased in states that passed legislation legalizing same-sex marriage whereas it increased in states that did not pass such legislation (Ofosu et al., 2019).

Project Implicit data, aggregated at the geographic level of analysis, was used to support arguments that Payne, Vuletich, and Lundberg (2017) put forward in their influential *Bias of Crowds* theory, in which they treat measures of implicit bias as reliable measures of situations rather than of persons. Specifically, they show that Black/White Race IAT scores are highly stable across time (comparing 2006 to 2017) when aggregated at the level of a place (i.e., a state), although they commonly show relatively lower test-retest reliability at the individual level (see Gawronski, Morrison, Phills, & Galdi, 2017). Relatedly, and remarkably, IAT scores of White people collected from 2002 to 2016 are more pro-White/anti-Black in counties and states that had higher proportions of enslaved Black people in 1860 (Payne, Vuletich, & Brown-Iannuzzi, 2019). Interestingly, Project Implicit data have also been used to counter these claims. For example, Rae and Greenwald (2017) showed that a substantial portion of the stability attributed to geography is due instead to individual-level racial identity, and Connor and Evers (2020) re-analyzed the data and concluded that the low test-retest reliability of the IAT is primarily due to measurement error.

We see this as one of the most important and exciting areas for future research on implicit bias. It is not yet clear whether variation in responding on the IAT (and implicit bias more generally) primarily reflects individual differences or mostly reflects situations—or, whether it reflects some kind of multiplicative variation in both. (Or, of course, it could reflect something else entirely.) The research that is most necessary to better understand these issues is highly intensive and resource dependent and would examine people's attitude and behavior across multiple domains, in multiple contexts, and with multiple measures.

IV. Understanding Responses to Learning about Bias

As mentioned previously, at the end of each study, participants receive feedback about their performance on the IAT. Following the feedback, participants respond to several items that, when used as a scale, capture a general defensiveness about their feedback: (1) Whether I like my IAT score or not, it captures something important about me (reverse-coded), (2) The IAT reflects something about my automatic thoughts and feelings concerning this topic (reverse-coded), and (3) The IAT does not reflect anything about my thoughts or feelings unconscious or otherwise. These items have been used in a series of investigations of the correlates and outcomes of defensive responding.

Howell, Gaither, and Ratliff (2015) investigated discrepancy between self-report racial bias and implicit racial bias on general defensiveness among Black Americans, White Americans, and Black/White bi-racial Americans, finding that monoracial Black *and* White people are higher in defensiveness when: (1) their implicit bias feedback shows a pro-White bias compared to a pro-Black bias, and (2) their implicit and explicit biases are more discrepant. Biracial individuals are also more defensive when their implicit bias and self-report bias are more discrepant, but their defensiveness is increased in both directions of bias feedback (pro-Black bias and pro-White bias).

In a larger-scale investigation of defensive responding to IAT feedback, Howell, Redford, Pogge, and Ratliff (2017) found that participants responded most defensively to feedback when: (1) their implicit and explicit attitudes were more discrepant than congruent, (2) their implicit attitudes aligned more with societal bias than did their explicit attitudes (e.g., a preference for Straight People relative to Gay People), and (3) they were majority group members (e.g., White participants in a race-relevant task). An experimental follow-up study demonstrated, among those for whom feedback elicits defensiveness, defensiveness predicts lower intentions to engage in egalitarian behavior.

Providing education about implicit bias (personal or general) is tricky. On one hand, defensiveness is a barrier to acceptance of and willingness to change bias, so care must be taken to present feedback in a way that reduces people's desire to derogate the science. On the other hand, we do not see a need to exculpate people from responsibility for their implicit biases (or any potential relationship of those biases to behavior). Redford and Ratliff (2016) found that people hold a hiring manager less accountable for discriminatory behavior if the discrimination is described as stemming from biases of which he is unaware compared to aware. Similarly, Daumeyer et al. (2019) found that attributions of discrimination to implicit bias, compared to explicit bias, resulted in lower judgements of accountability and, in some cases, lower intention to punish.

Future research in this area should focus on resolving this tension between defensiveness and accountability. Daumeyer et al. (2019) provide several excellent suggestions for more nuanced public conversations about implicit bias. These suggestions, which should be empirically tested on a larger scale and in multiple attitude domains, include avoiding presenting implicit bias as unconscious (see more on this below), pointing out that people might be able to override their implicit biases with sufficient motivation and opportunity, and using implicit bias as a starting point for conversations about implementing structural and institutional policies to correct for individual-level biases.

Changes Over Time at Project Implicit in Response to Developments in the Science of Implicit Bias

Greenwald et al.'s (1998) paper introducing the IAT has been cited almost 13,000 times. The body of scientific literature on implicit bias is both enormous and ever-changing. We have been in this line of work long enough to see various ideas drift in and out of fashion (e.g., the IAT is too unrelated to self-report to possibly be measuring something real; no wait, the IAT is actually too *related* to self-report to be interesting). In contrast, Project Implicit has been relatively conservative with regard to changing how the IAT and implicit bias are described. There are a number of reasons for this, including, but not limited to, the fact that keeping up with the latest developments is a Herculean task, getting a group of scientists to agree on updated materials is an even more Herculean task, and the organization's scientific board has a strong commitment to ensuring that any changes reflect solid, replicable, accumulated knowledge and not individual papers or passing fads.

We recognize that the organization has faced a great deal of scrutiny, and some criticism, over the years for the ways in which its educational efforts describe implicit bias and the IAT.

The most substantial of these concerns center around conceptualizations of the IAT as a measure of unconscious cognitions, the relationship between the IAT and behavior, and the ethics of giving feedback about IAT performance. In the section that follows we describe each of these concerns, the relevant evidence, how Project Implicit's presentation of implicit bias—on the websites and in educational presentations—has changed in response, and possible future directions.

Consciousness, Awareness, and Implicit Bias

The history of implicit social cognition research stems from two roots (Hahn & Gawronski, 2018; Payne & Gawronski, 2010). The *automaticity root* is focused on the automatic activation of attitudes, demonstrating that attitudes can influence responses without intention, particularly when motivation and opportunity to control responses is limited (Fazio, Jackson, Dunton, & Williams, 1995; Fazio, 2007).

The second root, the *unconsciousness root*, was initially favored by the founders of Project Implicit. Greenwald and Banaji (1995) defined implicit attitudes/stereotypes as "introspectively unidentified (or inaccurately identified) traces of past experience that mediate responses" (p. 5). The interpretation of this definition (though perhaps not the intention) is that implicit biases are outside of conscious awareness and inaccessible to introspection. "Unconscious bias" became, for many, synonymous with "implicit bias". In line with that, for many years the Project Implicit website defined implicit attitudes and stereotypes as those that people are "unwilling or unable to report".

It is impossible to address the question of awareness of implicit bias without going down a rabbit hole of definitions and teasing apart constructs (e.g., implicit bias) and measures (e.g., the IAT). Taking on this complexity with the care it deserves is beyond the scope of this chapter, though we appreciate the scholars who have done so (e.g., Corneille & Hütter, 2020; Moors & De Houwer, 2007; Gawronski, Hofmann, & Wilbur, 2006). For example, Gawronski et al. (2006) describe three ways in which one might be (un)aware of a particular bias: (1) its source (where did the bias come from?), (2) its content (what is the bias?), and (3) its influence on behavior (see also Hahn & Goedderz, 2020). Most research focuses on the second type of awareness, asking whether people are able to report the valence and/or strength of their biases. In this chapter we are largely concerned with Project Implicit, and Project Implicit has been largely concerned with the IAT, so we will (mostly) limit our discussion to that particular measure. This is also for practical reason – the majority of data available on this topic are about the IAT.

If we take the hard line that an attitude or stereotype (i.e., bias) is only implicit if its content is entirely outside of conscious awareness (i.e., inaccessible to introspection), we then have to grapple with what is meant by "outside of awareness". Most research in this area attempts to falsify the claim that people are unaware of their implicit biases by showing that IAT scores relate to self-report at some level above zero. And clearly that is the case. Across tasks at the Demonstration site, the correlation between IAT score and self-report ranges from $r = .13$ (Age Task) to $r = .43$ (Sexuality Task; Ratliff et al., 2020). For the Black/White Race Task the correlation is $r = .32$. The correlation between IAT scores and self-report is certainly influenced by topic (correlations for 57 attitude objects range from $r = -.05$ and $r = .70$ [Nosek, 2005]), and almost certainly influenced by choice of stimuli and self-report measure, but the relationship is clearly not non-existent. And there are ways to increase the correlation even further; for example, by asking participants to self-report their "gut reactions" rather than their "actual feelings" (Ratliff, Smith, & Nosek, 2008).

Another way of showing that people are aware of their implicit biases is by asking for a prediction. Hahn and colleagues (Hahn & Gawronski, 2019; Hahn, Judd, Hirsh, & Blair, 2014) demonstrate that people have some ability to predict their IAT scores (at least within a given set of IATs). Using this line of evidence, combined with the generally moderate relationship between the IAT and self-report measures, it is easy to conclude that people have at least some awareness of their implicit biases (either because they are truly introspecting based on some internal cues or because they are able to predict their own biases based on external knowledge of what biases they are most likely to have).

Thus, we certainly accept that the IAT does not capture attitudes and stereotypes that are fully inaccessible to introspection. Project Implicit no longer describes implicit attitudes and stereotypes as those that people are unwilling or unable to report. There is no mention of the unconscious in our talks or in our educational materials. That said, we disagree with arguments that moderate correlations between IAT scores and self-report suggest that the constructs are redundant (Schimmack, 2020), and thus implicit bias is uninteresting. These and similar arguments are difficult to reconcile with many people's surprise and even resistance when confronted with evidence of their own bias.

As scientists, we are obviously wary of using our personal experience as evidence, but this is an area in which we have a considerable amount of personal experience. Combined, we have presented *to tens of thousands* of people about implicit bias. They are surprised at their performance on the IAT. This is true in live demonstrations where no feedback is provided (where, in our experience, it is always clear that the audience slows down on the incongruent pairing compared to the congruent pairing), and in the data showing defensiveness in response to IAT feedback that diverges from self-report (Howell et al., 2015; Howell et al., 2017; Howell &

Ratliff, 2017). Gawronski (2019) argues that this surprise might occur because the metric used to convert an IAT score into verbal feedback (e.g., "Your responses suggested a strong automatic preference…) does not match people's lay theories of how they would quantify their own bias. This is certainly possible and is likely to be a fruitful avenue for future research. However, ongoing research from the Ratliff lab has examined many different ways of providing feedback to participants without diminishing the effect of implicit-explicit discrepancy on surprise.

Because there seems to be some element of unawareness of implicit bias that prompts reactions in people, for several years now Project Implicit has used the term *active awareness* to reflect the fact that unawareness of implicit bias might be because one *has not* reflected deeply about their biases rather than because one *cannot*. This is similar to the argument that implicit biases might reflect state-unconsciousness rather than trait-unconsciousness (Hahn & Goedderz, 2020). There are entire programs of research in this area left to undertake, including examining whether some people are better at introspecting than others, how we might promote accurate introspection, what aspects of situations lead to more accurate introspection (e.g., focusing on emotional reactions; Smith & Nosek, 2011), and whether there are particular attitude objects about which people are better able to introspect. We should also examine people's ability to accurately report on and predict their implicit biases measured by tasks other than the IAT. And, of course, all of these questions are specifically about people's ability to introspect on the *content* of their implicit biases – we have barely even begun to address these questions with regard to people's awareness of the source or impact of their implicit biases.

The Relationship between Implicit Bias and Behavior

It surely seems odd to suggest that an area of inquiry that has produced hundreds of papers is in its infancy, but our understanding of the relationship between the IAT and behavior has progressed shockingly little in the last two decades. The literature is full of individual papers that use an IAT in some attitude domain (e.g., Black/White race attitudes), with some stimuli (e.g., stereotypical names of Black and White people) to predict some behavior (e.g., seating distance from a Black confederate). These idiosyncratic papers have been collected into several influential meta-analyses (see Greenwald & Lai, 2020 for summary). Some authors (e.g., Kurdi et al., 2019) have been careful to say that we should not draw conclusions about **the** relationship between the IAT and behavior on the basis of meta-analyses that combine papers across many domains, with different original research questions, many scholars attempt to do so. For example, Greenwald et al. (2009) conclude that the average correlation between IAT scores and behavioral outcomes is $r = .27$; Oswald et al. (2013) narrow down the inclusion criteria to those focusing on intergroups relations and conclude that the average correlation is $r = .14$ (which is also the average estimate provided by Kurdi et al., 2019, though we are warned against overinterpretation of this single number).

How do we make sense of these simple correlations? One response would be that these correlations are too small to be meaningful, and thus we should abandon the idea that the IAT predicts behavior (Oswald et al., 2013). Another argues that even small effects can be meaningful when you take into consideration the number of decisions a given person makes and the number of people making decisions (Greenwald, Banaji, & Nosek, 2015). But we find these responses a bit empty, because the question that prompts them—whether or not implicit bias predicts behavior—fails to capture the nuance of human experience. In other words, there is no such thing as *the* relationship between implicit measures and behavior.

Nearly 40 years ago, in 1981, Fazio and Zanna wrote: "*We have to ask, under what conditions do what kinds of attitudes held by what kinds of individuals predict what kinds of* *behavior?" (p. 165).* They were, of course, responding to concern that attitudes—explicit, self-reported attitudes—do not predict behavior. But a sustained, programmatic, theoretically-driven approach to this work yielded a different conclusion, that attitudes *do* predict behavior under particular circumstances. As one example, general attitudes predict general behaviors better than specific behaviors, and specific attitudes predict specific behaviors better than general behaviors (Ajzen & Fishbein, 1977; Fishbein & Ajzen, 1977; Kraus, 1995); some parallel evidence has accrued for a similar pattern of results when examining the relationship between *implicit* attitudes and behavior (Irving & Smith, 2020).

Why have we not used the literature on explicit attitudes to guide theorizing on the conditions under which implicit attitudes will predict behavior? We speculate that this is due in part to early theoretical stances that implicit and explicit attitudes were wholly different from one another. Theoretically driven dual-attitudes models like those proposed by Smith and DeCoster (2000) and Wilson, Lindsey, and Schooler (2000) were enormously important for initiating the field of implicit social cognition, but, in hindsight, perhaps made too strong of claims about the distinction between implicit and explicit attitudes. (Or perhaps the nuance of the arguments was lost in translation, distilling complex models into a simplistic distinction that was not what the original authors intended.) Gawronski and Bodenhausen's (2006, 2011) claims that implicit and explicit evaluations are the outcome of associative processes and propositional processes, respectively, and Cunningham et al.'s (2001) conceptualization of implicit and explicit attitudes as "separate but distinct" were used to support the argument that implicit and explicit attitudes draw on different kinds of evaluative knowledge and experiences, forming and changing based on entirely different information (Rydell & McConnell, 2006; Rydell, McConnell, Mackie, & Strain, 2006).

The strict dual-process account of implicit and explicit attitudes has fallen out of favor over time, in part due to strong evidence that implicit attitudes can be impacted by propositions (i.e., information that specifies the relation between concepts; De Houwer, 2014). In other words, implicit measures may not reflect simple associations (e.g., object + good), but can incorporate additional logical information. For example, in one experiment (Zanon, De Houwer, Gast, & Smith, 2014), participants viewed one set of non-words paired with positive stimuli and another set paired with negative stimuli. Before viewing the pairings, participants were told that the meaning of the nonwords was the opposite of the paired stimuli. For a purely associative account, this should not matter – an implicit measure should reflect the observed pairings. However, participants showed an implicit preference for the nonwords paired with negative over those paired with positive, thereby reflecting the instructed pairing (i.e., a proposition) rather than the observed pairing (i.e., an association).

Relatedly, mere instruction about a learning procedure leads to IAT scores that are at least as large as experiencing the actual learning procedure (e.g., Kurdi & Banaji, 2017; Smith, Calanchini, Hughes, Van Dessel, & De Houwer, 2019). Breaking down this strict duality has led to several other demonstrations that implicit attitudes can be formed in many of the same ways that explicit attitudes can – demonstrations that we would have been unlikely to pursue previously. For example, just like for explicit attitudes, implicit attitudes form based on source cues such as credibility (Smith, De Houwer, & Nosek, 2013) and attractiveness and likeability (Smith & De Houwer, 2014).

In sum, the road map for determining how effectively we can predict behavior from implicit attitudes is straightforward in principle, but complex and burdensome to carry out. We cannot rely on secondary data analysis alone; we will need to design studies specifically for this purpose, focusing on questions about the *conditions* under which implicit attitudes predict behavior, rather than *whether* implicit attitudes predict behavior (e.g., Friese, Hofmann, & Schmitt, 2008). This was the approach in a flurry of papers in the early 2000s outlining moderators of the relationship between implicit attitudes and explicit attitudes (e.g., Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Hoffman, Gschwendner, Nosek, & Schmitt, 2005; Karpinski, Steinman, & Hilton, 2005; Nosek, 2005), and is likely to be relevant to the relationship between implicit attitudes and behavior.

Understanding the relationship between implicit attitudes and behavior will require experimental designs where the experiments are high-powered and feature multiple measures in multiple domains. Further, they should incorporate individual differences in people *and* differences across situations. Finally, the designs should simultaneously take explicit attitudes into account. In other words, we should follow the advice of Fazio and Zanna (1981) to: "*Treat the strength of the attitude-behavior relation as we would treat virtually any other dependent variable and determine what factors affect it"* (p. 165).

The preceding discussion about the relationship between the IAT and behavior has an implicit assumption that the IAT is a predictor and behavior is an outcome. De Houwer (2019) provides what is, to us, a very compelling argument for rejecting the framing of implicit bias as a hidden mental bias that resides inside of minds, and instead to conceptualizes performance on measures like the IAT as an instance of implicitly biased *behavior*[12].

Accuracy and Ethics in Providing IAT Feedback

The Project Implicit Demonstration Site has included an "Ethical Considerations" page since it launched in 1998. The current ethical considerations site raises three main issues. First,

[12] For the sake of giving credit where credit is due, and maintaining some semblance of intellectual humility, we note that Brian Nosek tried to get his lab to adopt this perspective at least 15 years ago and we resisted mightily.

there is a section stating that studies at Project Implicit are scientific research, and scientific research should always be voluntary. Instructors and employers are encouraged to direct their students and staff to the site for education, but they should not make study participation mandatory and should not ask that people share their feedback. There is also a section recommending that the IAT not be used for research outside of the safeguards of a research institution.

Since 1998 site lunch there has been a clear statement on the ethics page that the IAT (like the majority of psychological tests commonly in use for research and education purposes), should not be used for diagnostic purposes. The most relevant line of text reads: *Research shows the IAT is an effective educational tool for raising awareness about implicit bias, but the IAT cannot and should not be used for diagnostic or selection purposes (e.g., hiring or qualification decisions). For example, using the IAT to choose jurors is not justifiable, but it is appropriate to use the IAT to teach jurors about implicit bias.*

This brings up an important question on which Project Implicit's Scientific Advisory Board reflects frequently—is it ethical to provide participants feedback on their IAT performance? Thus far, the team has answered this question in the affirmative (a point to which we will return at the end of this section), but the team closely follows the literatures on IAT reliability and malleability to make this decision and are open to reconsidering should the evidence suggest it is prudent to do so.

IAT Reliability

One of the most long-standing criticisms of implicit measures is that they have low reliability. The meta-analytic internal consistency on the Black/White Race IAT is $\alpha = .80$ (Greenwald & Lai, 2020). Though there may be some variability in internal reliability based on stimulus and category labels and other features of a given IAT, on the whole internal reliability of the IAT is not seen as being particularly problematic given that it is reasonable and higher than most other implicit measures (e.g., Bar-Anan & Nosek, 2014; Cummins, Lindgren, & De Houwer, 2020).

A more damning criticism of the IAT's reliability focuses on test-retest reliability; that is, the correlation between an IAT score at one time and an IAT score on the same measure at a different time. Test-retest is particularly relevant for IAT feedback as low test-retest reliability would lead to people getting different feedback from the same IAT. Commonly referenced estimates are currently .44 (Gawronski et al., 2017) or .50 (Greenwald & Lai, 2020), and .45 across race, politics, and self-esteem IATs (Bar-Anan & Nosek, 2014).

As with many other issues we have covered in this chapter, it is important to recognize that different versions of the IAT have different characteristics. For example, the Smith lab has data from 591 participants who completed an Obama-McCain/Good-Bad IAT at three different time points (September, October, and November of 2008). The correlations between the IAT scores were: T_1T_2 $r = .59$, T_2T_3 $r = .60$, T_1T_3 $r = .56$. Cronbach's alpha for the three time points was .82, which would be considered acceptable reliability for a scale. Bar-Anan and Nosek (2014) showed test-retest reliability for seven indirect measures to range from a low of .18 for a Race attitudes Evaluative Priming task to a high of .78 for a Democrats-Republicans Brief IAT. This suggests that there is considerable variation in the reliability of indirect measures and, again, that focusing on *the* IAT is not a nuanced enough approach.

We also note that newer approaches to calculating measure reliability might yield different estimates. Haines et al. (2020) used a hierarchical Bayesian approach to re-analyze data from Gawronski et al. (2017), estimating the parameters of a lognormal distribution for each participant to model how performance changed across conditions and across sessions. The central tendency parameters of this model yielded a test-retest reliability of .75, compared to the .55 the original authors reported using IAT *D*-scores. This work is in its early stages, but suggests that introducing "generative" models, which tie observed performance to underlying cognitive processes, will be useful for quantifying performance on the IAT. Developing and applying these models offers the opportunity to characterize behavior in terms of psychological variables that are more reliable than IAT *D*-scores.

More generally, concern with the IATs stability over time stems, in part, from the idea that implicit measures pick up on something that is long-standing and relatively immutable within a person (i.e., a trait). However, if IAT scores tap into something more state-like, then a lack of consistency from one measurement occasion to another should not be taken as a deficit. As early as 2002, Blair argued that implicit measures are sensitive to: (a) self- and social motives, (b) specific strategies, (c) the perceiver's focus of attention, and (d) stimulus cues in the focal task. Around the same time, Banaji (2004) wrote that implicit attitudes will differ depending on the particular construal of an attitude object in a particular situation. Even still, for the next decade, evidence of malleability was treated more as an aberration than as the rule.

Recently, however, the “bias of crowds” theory (Payne et al., 2017) has put forward a different idea—that implicit bias within people reflects the contexts in which they are embedded and, as such, measure-to-measure consistency is evidence of consistency of one’s environment or situations (Payne et al., 2017). Partially in response to this, Jost (2019) makes a compelling case that the IAT is a “context-sensitive measure” that reflects personal *and* cultural forces.

Given that implicit biases are, at the very least, influenced by situations, it is difficult to know what we would expect the test-retest reliability of a given IAT *should* be. We do not mean to just shrug our shoulders here, but we do wish to point out that two people could look at a test-retest correlation of .40 and conclude either that the test is very broken or that it works exactly as it should, depending on whether you expect the measure to capture something stable and trait-like or something situational and state-like. To push the field forward, experimental methods of triangulating on an estimate of the state- vs. trait-like qualities of implicit measures would be an important and needed contribution. Some attempts in this direction have begun, but are so far restricted to the IAT, in the domain of the self, and have not yet gained widespread attention or follow-up (Dentale, Vecchione, Ghezzi, & Barbaranelli, 2020; Schmukle & Egloff, 2005).

Given this complexity, let us return to the question of giving IAT performance feedback to participants. The rationale for giving feedback remains the same now as it was when the sites were launched. In short, Project Implicit gives feedback to participants about their IAT performance because of the perceived educational value in doing so. Further, we know that feedback is a large part of what engages participants about this work. When given a choice, as is the case for the Project Implicit Health site, the overwhelming majority of people want feedback to help understand their performance on the IAT. In the early days of the website, there was a period when, due a programming error, some participants did not get feedback, and that was the source of the most frequent and vitriolic complaints at that time. Project Implicit's mission is education and providing IAT feedback provides a highly engaging mechanism for enhancing education. Participants are told "here's what just happened on the test, let's explore what it might mean." IAT feedback is also presented with a clear statement that the results are not intended to be a definitive assessment of implicit bias. Because concerns about the IAT are primarily about reliability of scores from one time to another, rather than about the internal consistency on any given occasion, the site emphasizes the situational nature of IAT scores (though we do believe that the team should revisit this language and consider making this point even more strongly, particularly as new evidence comes out about the relative nature of situations versus individual differences). Conveying uncertainty—e.g., unreliability, the possible influence of situational factors—is an integral part of helping people learn about implicit bias *and* is an essential part of promoting scientific literacy and education about the scientific process.

Conclusions and Future Directions

Throughout this chapter we have laid out a number of areas in which there are unanswered questions, entrenched puzzles, and need for future research. For Project Implicit, and for the field more generally, future directions should include further refinement of IAT scoring and feedback, taking into account our current understanding of the IAT reflecting both stable and malleable attitudes, updating outdated stimuli and tasks on the U.S. and International Demonstration Sites, using implicit measures other than the IAT, and implementing universal design standards to make the site more accessible.

Putting the science directly into the hands of the participants is important and exciting, but also challenging. In this chapter we have laid out some of the challenges and opportunities with which we are confronted as we do this very public work. After more than 20 years of collecting data at Project Implicit, the goal remains the same – to educate people about implicit bias.

Some people argue that researchers should do away with the term “implicit” altogether due to its ambiguity and lack of precision (Corneille & Hütter, 2020; c.f. Greenwald & Lai, 2020). We agree that precise terminology is essential so that scientists can communicate clearly with one another with certainty that they are speaking the same “language.” To that end, we appreciate and amplify the calls for scientists to clearly define what they mean by “implicit” in their manuscripts (the Corneille & Hütter, 2020 “compromise” position). But, at the same time, implicit bias has taken on a life of its own as a descriptive term in the public domain, and it could be counterproductive to scientific communication to introduce a new set of terminology for public use, especially if that terminology will seem obscure and pretentious to people outside of the academic research world. What we really want people to know is a variation of what we started with: that people have attitudes and stereotypes that influence how they see and interpret the world. In our view, implicit bias is ordinary, it is rooted in culture, and it is pervasive. Those features do not absolve us of the responsibility to confront it in ourselves and in others—in fact, in our opinion, they are exactly why it is our responsibility to change the culture within which it thrives.

References

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

Axt, J. R., Conway, M. A., Westgate, E. C., & Buttrick, N. R. (2020). Implicit transgender attitudes independently predict beliefs about gender and transgender people. In press at *Personality and Social Psychology Bulletin.*

Banaji, M.R. (2004). The opposite of a great truth is also true: Homage to koan #7. In J. T. Jost, M. R. Banaji, & D. A. Prentice (Eds.), *Perspectivism in social psychology: The yin and yang of scientific progress*. (pp. 127-140). American Psychological Association.

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

Bar-Anan, Y., & Vianello, M. (2018). A multi-method multi-trait test of the dual-attitude perspective. *Journal of Experimental Psychology: General*, *147*(8), 1264-1272.

Bianchi, E.C., Hall, E.V., & Lee, S.* (2018). Reexamining the link between economic downturns and racial antipathy: Evidence that prejudice against Blacks rises in recessions. *Psychological Science, 29*(10), 1584-1597.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*(3), 242-261.

Carleton, R. N., Norton, M. P. J., & Asmundson, G. J. (2007). Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of anxiety disorders*, *21*(1), 105-117.

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

Chin, M. J., Quinn, D. M., Dhaliwal, T. K., & Lovison, V. S. (2020). Bias in the air: A nationwide exploration of teachers' implicit racial attitudes, aggregate bias, and student outcomes. *Educational Researcher*, 1-13.

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

Corneille, O., & Hütter, M. (2020). Implicit? What Do You Mean? A comprehensive review of the delusive implicitness construct in attitude research. In press at *Personality and Social Psychology Review*.

Cummins, J., Lindgren, K. P., & De Houwer, J. (2020). On the role of (implicit) drinking self identity in alcohol use and problematic drinking: A comparison of five measures. In press at *Psychology of Addictive Behaviors*.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological science*, *12*(2), 163-170.

Daumeyer, N. M., Onyeador, I. N., Brown, X., & Richeson, J. A. (2019). Consequences of attributing discrimination to implicit vs. explicit bias. *Journal of Experimental Social Psychology*, *84*, 103812.

De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers & A. W. Stacy (Eds.), The Handbook of Implicit Cognition and Addiction (pp. 11–28). Thousand Oaks, CA: Sage.

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*(7), 342-353.

De Houwer, J. (2019). Implicit bias is behavior: A functional-cognitive perspective on implicit bias. *Perspectives on Psychological Science*, *14*(5), 835-840.

Dentale, F., Vecchione, M., Ghezzi, V., Spagnolo, G., Szemenyei, E., & Barbaranelli, C. (2020). Beyond an associative conception of automatic self-evaluation: Applying the Relational Responding Task to measure self-esteem. *The Psychological Record*, 1-16.

Duncan, L. A., Schaller, M., & Park, J. H. (2009). Perceived vulnerability to disease: Development and validation of a 15-item self-report instrument. *Personality and Individual differences*, *47*(6), 541-546.

Fazio, R. H. (2007). Attitudes as object–evaluation associations of varying strength. *Social Cognition*, *25*(5), 603-637.

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*(6), 1013-1027.

Fazio, R. H., & Zanna, M. P. (1981). Direct experience and attitude-behavior consistency. In *Advances in Experimental Social Psychology* (Vol. 14, pp. 161-202). Academic Press.

Fishbein, M., & Ajzen, I. (1977). Belief, attitude, intention, and behavior: An introduction to theory and research.

Friese, M., Hofmann, W., & Schmitt, M. (2008). When and why do implicit measures predict behaviour? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology, 19*, 285-338.

Friese, M., Smith, C. T., Plischke, T., Bluemke, M., & Nosek, B. A. (2012). Do implicit attitudes predict actual voting behavior particularly for undecided voters?. *PloS One*, *7*(8), e44130.

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

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. In *Advances in Experimental Social Psychology* (Vol. 44, pp. 59-127). Academic Press.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are “implicit” attitudes unconscious?. *Consciousness and cognition*, *15*(3), 485-499.

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

Gladwell, M. (2005). Blink: The power of thinking without thinking.

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

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects.

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

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

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*(1), 17-41.

Greenwald, A. G., Smith, C. T., Sriram, N., Bar-Anan, Y., & Nosek, B. A. (2009). Implicit race attitudes predicted vote in the 2008 US presidential election. *Analyses of Social Issues and Public Policy*, *9*(1), 241-253.

Hahn, A., & Gawronski, B. (2018). Implicit social cognition. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, *4*, 1-33.

Hahn, A., & Gawronski, B. (2019). Facing one's implicit biases: From awareness to acknowledgment. *Journal of Personality and Social Psychology*, *116*(5), 769-794.

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

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, *143*(3), 1369-1392.

Haines, N., Kvam, P. D., Irving, L. H., Smith, C., Beauchaine, T. P., Pitt, M. A., ... & Turner, B. (2020). Learning from the reliability paradox: How theoretically informed generative models can advance the social, behavioral, and brain sciences. Unpublished manuscript.

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

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

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.

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit—explicit consistency?. *European Review of Social Psychology, 16*, 335-390.

Hoover, J., & Dehghani, M. (2020). The big, the bad, and the ugly: Geographic estimation with flawed psychological data. *Psychological Methods*, *25*, 412-429.

Howell, J. L., Gaither, S.E., & Ratliff, K.A. (2015). Caught in the middle: Defensive responses to IAT feedback among whites, blacks, and biracial black/whites. *Social Psychological and Personality Science*, *6*, pp.373-381. doi: 10.1177/1948550614561127

Howell, J. L., & Ratliff, K. A. (2017). Not your average bigot: The better-than-average effect and defensive responding to Implicit Association Test feedback. *British Journal of Social Psychology*, *56*, 125-145. doi: 10.1111/bjso.12168

Howell, J. L., Redford, L., Pogge, G., & Ratliff, K. A. (2017). Defensive responding to IAT feedback. *Social Cognition*, *35*, 520-562. doi: 10.1521/soco.2017.35.5.520

Inbar, Y., Westgate, E. C., Pizarro, D. A., & Nosek, B. A. (2016). Can a naturally occurring pathogen threat change social attitudes? Evaluations of gay men and lesbians during the 2014 Ebola epidemic. *Social Psychological and Personality Science, 7*, 420-427.

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.

Jiang, C., Vitiello, C., Axt, J., Campbell, J., & Ratliff, K. A. (2020). An examination of ingroup preferences among people with multiply socially stigmatized identities. *Self and Identity.*

Jost, J. T. (2019). The IAT is dead, long live the IAT: Context-sensitive measures of implicit attitudes are indispensable to social and political psychology. *Current Directions in Psychological Science, 28*(1), 10-19.

Karpinski, A., Steinman, R. B., & Hilton, J. L. (2005). Attitude importance as a moderator of the relationship between implicit and explicit attitude measures. *Personality and Social Psychology Bulletin, 31*, 949-962.

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

Kurdi, B., & Banaji, M. R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes?. *Journal of Experimental Psychology: General*, *146*(2), 194-213

Kurdi, B., Ratliff, K. A., & Cunningham, W. A. (2020). Can the Implicit Association Test serve as a valid measure of automatic cognition? A response to Schimmack (2020). In press at *Perspectives on Psychological Science*.

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*(5), 569-586.

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

Lowe, C., Beach, M. C., & Roter, D. L. (2020). Individuation and implicit racial bias in genetic counseling communication. *Patient Education and Counseling*, *103*(4), 804-810.

MacInnis, C. C., Page-Gould, E., & Hodson, G. (2017). Multilevel intergroup contact and antigay prejudice (explicit and implicit) evidence of contextual contact benefits in a less visible group domain. Social Psychological and Personality Science, 8, 243-251.

Maximiano-Barreto, M. A., Fabrício, D. D. M., Luchesi, B. M., & Chagas, M. H. (2020). Factors associated with levels of empathy among students and professionals in the health field: a systematic review. Trends in Psychiatry and Psychotherapy, *42*(2), 207-215.

Moors, A., & Houwer, J. D. (2007). *What is Automaticity? An analysis of its component features and their interrelations.* In J. A. Bargh (Ed.), *Frontiers of Social Psychology. Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (p. 11 50). Psychology Press.

Neville, H. A., Gallardo, M. E., & Sue, D. W. E. (2016). *The myth of racial color blindness: Manifestations, dynamics, and impact*. American Psychological Association.

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., & Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*(1), 101–115.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues*, *58*(1), 161-176.

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*(1), 36-88.

Ofosu, E. K., Chambers, M. K., Chen, J. M., & Hehman, E. (2019). Same-sex marriage legalization associated with reduced implicit and explicit antigay bias. *Proceedings of the National Academy of Sciences, 116*(18), 8846-8851

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*(2), 171-192.

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. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: measurement, theory, and applications* (pp. 1 15). New York, NY: Guilford Press.

Payne, B. K., Vuletich, H. A., & Brown-Iannuzzi, J. L. (2019). Historical roots of implicit bias in slavery. *Proceedings of the National Academy of Sciences, 116*(24), 11693-11698.

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.

Prairie, T. M., Wrye, B., Bowman, A. S., Weatherby, N., & Thareja, G. (2019). Does Location of Practice or Religiosity Predict Negative Physician Attitudes or Beliefs Toward LGB+ Individuals?. *Journal of religion and health*, *58*(6), 2208-2218.

Rae, J. R., & Greenwald, A. G. (2017). Persons or situations? Individual differences explain variance in aggregated implicit race attitudes. *Psychological Inquiry*, *28*(4), 297-300.

Rae, J. R., Newheiser, A. K., & Olson, K. R. (2015). Exposure to racial out-groups and implicit race bias in the United States. *Social Psychological and Personality Science*, *6*, 535-543

Ratliff, K. A., Lofaro, N., Howell, J. L., Conway, M. A., Lai, C. K., O'Shea, B., Smith, C. T., Jiang, C., Redford, L., Pogge, G., Umansky, E., Vitiello, C., Zitely, H. (2020). Documenting Bias from 2007-2015: Pervasiveness and Correlates of Implicit Attitudes and Stereotypes II. Unpublished Manuscript.

Ratliff [Ranganath], K. A., & Nosek, B. A. (2008). Implicit attitude generalization occurs immediately; explicit attitude generalization takes time. *Psychological Science*, *19*(3), 249-254.

Ratliff [Ranganath], K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology, 44*, 386-396.

Ratliff, K. A., Redford, L., Conway, J., & Smith, C. T. (2019). Engendering support: Hostile sexism predicts voting for Donald Trump over Hillary Clinton in the 2016 US presidential election. *Group Processes & Intergroup Relations*, *22*(4), 578-593.

Ravary, A., Baldwin, M. W., & Bartz, J. A. (2019). Shaping the body politic: Mass media fat shaming affects implicit anti-fat attitudes. *Personality and Social Psychology Bulletin, 45*(11), 1580–1589.

Redford, L., & Ratliff, K. A. (2016). Perceived moral responsibility for attitude-based discrimination. *British Journal of Social Psychology*, *55*(2), 279-296.

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*(17), 8255-8260. https://doi.org/10.1073/pnas.1808307116

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

Rydell, R. J., McConnell, A. R., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science*, *17*(11), 954-958.

Sabin, J. A., Riskind. R. G., & Nosek, B. A. (2015). Health care providers' implicit and explicit attitudes toward lesbian women and gay men. *American Journal of Public Health, 105*, 1 831-1841.

Schimmack, U. (2020). The Implicit Association Test: A method in search of a construct. In Press at *Perspectives on Psychological Science*.

Schmidt, K. & Axt, J.R. (2016). Implicit and explicit attitudes toward African Americans and Barack Obama did not substantively change during Obama's presidency. *Social Cognition*, *34*, 559-588.

Schmukle, S. C., & Egloff, B. (2005). A latent state-trait analysis of implicit and explicit personality measures. *European Journal of Psychological Assessment*, *21*(2), 100-107.

Smith, C. T., Calanchini, J., Hughes, S., Van Dessel, P., & De Houwer, J. (2019). The impact of instruction-and experience-based evaluative learning on iat performance: A quad model perspective. *Cognition and Emotion*.

Smith, C. T., & De Houwer, J. (2014). The impact of persuasive messages on IAT performance is moderated by source attractiveness and likeability. *Social Psychology*.

Smith, C. T., De Houwer, J., & Nosek, B. A. (2013). Consider the source: Persuasion of implicit evaluations is moderated by source credibility. *Personality and Social Psychology Bulletin*, *39*(2), 193-205.

Smith, C. T., & Nosek, B. A. (2011). Affective focus increases the concordance between implicit and explicit attitudes. *Social Psychology*, *42*, 300-313.

Smith, C. T., Ratliff, K. A., & Nosek, B. A. (2012). Rapid assimilation: Automatically integrating new information with existing beliefs. *Social Cognition*, *30*(2), 199-219.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*(2), 108-131.

Starck, J. G., Riddle, T., Sinclair, S., & Warikoo, N. (2020). Teachers are people too: Examining the racial bias of teachers compared to other American adults. *Educational Researcher*. *49*(4), 273-284.

Vianello, M., & Bar-Anan, Y. (2020). Can the Implicit Association Test measure automatic judgment? The validation continues. In press at *Perspectives on Psychological Science*.

Vuletich, H.A. & Payne, B.K. (2019). Stability and change in implicit bias. *Psychological Science, 30*(6), 854-862.

Westgate, E. C., Riskind, R. G., & Nosek, B. A. (2015). Implicit preferences for straight people over Lesbian Women and Gay Men weakened from 2006 to 2013. *Collabra*, *1*, (Art.1).

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101-126.

Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning? *Quarterly Journal of Experimental Psychology*, *67*(11), 2105-212

Table 1. Task Names, IAT Category Labels, and Stimuli

	Target Labels (stimuli type)	**Attribute Labels (stimuli type)**
Age Attitudes	Young People-Old People (images)	Good-Bad (words)
Arab-Muslim Attitudes	Other People-Arab Muslims (names)	Good-Bad (words)
Disability Attitudes	Abled People-Disabled People (images)	Good-Bad (words)
President Attitudes	Current President-[random assignment from a set of previous Presidents]	Good-Bad (words)
Race Attitudes	White People-Black People *or* European Americans-African Americans	Good-Bad (words)
Religion Attitudes	Christianity-Judaism; Christianity-Islam; Islam-Judaism [at random]	Good-Bad (words)
Sexuality Attitudes	Straight People-Gay People (images, words)	Good-Bad (words)
Skin Tone Attitudes	Image of a person with lighter or darker skin tone (images)	Good-Bad (words)
Transgender Attitudes	Transgender People-Cisgender People (images)	Good-Bad (words)
Weight Attitudes	Thin People-Fat People (images)	Good-Bad (words)
Gender-Career Stereotypes	Male-Female (names)	Career-Family (words)
Gender-Science Stereotypes	Male-Female (names)	Science-Liberal Arts (words)
Native-Foreign Stereotypes	White American-Native American (images/words)	American-Foreign (images)
Asian-Foreign Stereotypes	European American-Asian American (images)	American-Foreign (images)
Race-Weapons Stereotypes	White Americans-Black Americans (images)	Harmless Objects-Weapons (images)