The Relationship of Implicit Social Cognition and Discriminatory Behavior

Tessa E.S. Charlesworth & Mahzarin R. Banaji

Harvard University, Cambridge, Massachusetts


Author Note

This research was supported by the Mind Brain Behavior Interfaculty Initiative award, the Hao Family Inequality in America award, and the Foundations of Human Behavior Initiative award to M. R. Banaji and T. E. S. Charlesworth. The authors declare no competing financial interests. Correspondence concerning this article should be addressed to Tessa Charlesworth, Department of Psychology, Harvard University, Cambridge, MA 02138, at tet371@g.harvard.edu.
When viewing the state of the world today, social and behavioral scientists face a puzzling inconsistency: how can it be that evidence of discrimination persists in all significant aspects of life, from housing and jobs to healthcare and law enforcement, even though individuals and institutions adamantly stand for equality in treatment? Over the past two decades, research has demonstrated that at least part of the answer to this puzzle can be attributed to the implicit nature of biases – attitudes, beliefs, and identities that are less conscious and controllable but that nevertheless exist and shape behavior. Indeed, today, we take it as given that evidence is strong and substantial for the presence of implicit bias in the minds and behaviors of individuals. This chapter, however, reviews an emerging body of research that uses massive, aggregated data across millions of tests of implicit attitudes and beliefs to understand outcomes of socially significant systemic behaviors ranging from the police use of lethal force to infant healthcare to school suspensions and discipline. Methodologically, the studies quantify social and psychological processes in the real world and introduce data of unprecedented scope across geography and time. Theoretically, this melding of psychological evidence about implicit bias aggregated across millions of individuals alongside large-scale socially significant behaviors underscores a new and more robust meaning of the term systemic discrimination: the results show the various ways that implicit bias both shapes and is shaped by broad structural systems and outcomes.

*Keywords*: education, discrimination, healthcare, implicit attitudes, implicit bias, implicit stereotypes, policing, systemic discrimination
The Relationship of Implicit Social Cognition and Discriminatory Behaviors

For students of the social sciences, the term “bias” is commonly used to capture at least two distinct meanings. In one sense, a bias refers to a behavior or cognition that deviates from accuracy. If two individuals of differing ethnicity say Asian and White, are U.S. citizens, and both were born in and live in the United States, it is accurate to consider them both to be American. Yet if measures of implicit beliefs indicate that one group is perceived to be more American, i.e., that White Americans are more “American” than Asian Americans (Devos & Banaji, 2005; Devos & Mohamed, 2014), that belief can be said to be biased in the sense that it deviates from accuracy. A second meaning of bias refers to thoughts or behavior that deviates from one’s own consciously stated values. If one’s conscious values dictate that the best lab manager should be hired, but the evidence shows that male candidates are repeatedly hired even if they are equal to female candidates (Moss-Racusin et al., 2012), that behavior is taken to be biased because it is inconsistent with stated ideals.

Both forms of bias are known to be prevalent in shaping our cognitions and behaviors (Pager, 2007). For instance, experimental studies demonstrate that biases based on the social category of gender are abundant throughout the hiring and evaluation of women in science, technology, engineering, and mathematics (Charlesworth & Banaji, 2019). As an example, participants (drawn from a population of scientists) were randomly assigned to see either a male or female applicant profile for a science lab manager position (Moss-Racusin et al., 2012). Systematically, participants were more likely to recommend the male applicant be hired, given a higher salary, and given more mentorship compared to the identical and equally-qualified female applicant. Similarly, participants in another study were more likely to hire a male candidate to complete an arithmetic task over an identical female candidate, and, most surprisingly, continued
to select the male candidate even after receiving evidence that women outperform men on the task (Reuben et al., 2014).

A complementary method – audit studies, also called “experimental field studies,” in which real-world respondents do not know they are in a study (Pager, 2007) – have also been used to reveal biased behavior towards targets differing on race, sexual orientation, criminal history, and more. For instance, counselling professionals were more likely to initiate further counselling treatment for a potential client if the hypothetical client was White rather than Black (Shin et al., 2016), and hiring managers were more likely to call back an applicant if they were White rather than Black (Bertrand & Mullainathan, 2004). Similarly, applications from ostensibly heterosexual job candidates sent to real hiring managers across 1,769 jobs were more successful than applications from identically qualified but openly-gay candidates (Tilcsik, 2011). And hypothetical job candidate profiles that listed a previous criminal record were less likely to receive a call back than identical candidates without a criminal record, especially if the candidate with a criminal record was Black (Pager, 2003).

Because such experiments and audit studies control for all features of the hypothetical candidates except the candidate’s salient identity (e.g., their race, gender, or sexual orientation), such evidence conclusively points to the pervasiveness of biased behaviors and discrimination. That is, respondents in all studies were more likely to treat a candidate favorably (e.g., hire them, offer them medical treatment) when the candidate came from a typically-preferred, high-status or dominant social group (e.g., male, White, straight), relative to when they came from a typically not-preferred, low-status, or minority social group (e.g., female, Black, gay). Crucially, this evidence for discrimination is not an isolated phenomenon among a few “bad apples” - a common explanation for discrimination that suggests it would be solved if the few “bad apples”
were rooted out. Although individual differences no doubt exist in the frequency or severity of discriminatory behavior, the bulk of the evidence shows that such behavior is far more pervasive than explained by the “few bad apples” account. For instance, both men and women favored the male over female candidate for a STEM job (Moss-Racusin et al., 2012), hiring managers across most U.S. States favored a straight over gay job candidate (Tilcsik, 2011), and hiring managers across most occupations favored a White over Black candidate, generally regardless of the job requirements (Bertrand & Mullainathan, 2004).

The pervasiveness of discriminatory behaviors is difficult to square with the finding (often from these very same studies) that few people explicitly endorse biased beliefs. For instance, despite widespread evidence of gender discrimination that treat women as less hirable or capable than men, many people explicitly endorse the opposite beliefs and state that women are more intelligent than men (Storage et al., 2020). Additionally, the very same individuals whose behavior reveals unequal medical decisions to Black and White patients most often express equitable explicit beliefs about Black and White patients (Green et al., 2007). Such discordant patterns of results inevitably prompts the question: if discriminatory behaviors cannot be entirely attributed to explicit attitudes and beliefs, what is the source of such discrimination? What explains this inconsistency between expressed and revealed biases?

Over the past two decades, research and theory on implicit attitudes and beliefs – attitudes and beliefs that are less accessible to conscious introspection and deliberate control (Greenwald & Banaji, 1995) – can provide a compelling answer to the inconsistency between pervasive discrimination and low explicit biases. These discriminatory behaviors are being shaped and maintained, at least in part, by underlying implicit biases that exist and persist in the minds of individuals and in the culture. Indeed, as we review below, the evidence that implicit
attitudes and beliefs predict individual behaviors (e.g., hiring decisions, seating distance) is substantial, stemming from hundreds of studies reviewed in depth across major meta-analyses (Greenwald et al., 2009; Kurdi et al., 2019; Oswald et al., 2013).

But the role of implicit bias in behaviors does not stop with the individual decision-maker. In line with aforementioned evidence showing the spread of discriminatory behaviors across most people, the social sciences have long recognized that bias and discrimination are widely embedded in the broader systems of society – systems of healthcare, policing, education, and more. Recently, using massive data of implicit attitudes and beliefs aggregated across millions of respondents, an emerging body of studies has begun to quantify the relationship between implicit bias and such socially significant systemic outcomes, whether racial gaps in infant health outcomes (Orchard & Price, 2017) or the gender gap in science and math achievement (Nosek et al., 2009). These studies have not yet been summarized together. Thus, the goal of the current chapter is to provide an introduction to this new and growing empirical evidence that is revealing the tight relationships between implicit cognition and systemic discriminatory outcomes.

A brief definition of implicit and explicit attitudes and beliefs

In 1995, building from a wealth of literature showing that our conscious, introspective minds are not the whole story of our cognitions (Nisbett & Wilson, 1977), Greenwald and Banaji (1995) proposed a distinction between two forms of social attitudes and beliefs. On the one hand, there are explicit attitudes and beliefs – thoughts and feelings about social groups that are relatively more controlled, deliberate, and reflective of conscious, personal values. On the other hand, there are implicit attitudes and beliefs – thoughts and feelings about social groups that are
relatively automatic, uncontrolled, and inaccessible to introspective access. We cannot easily look into our own minds to understand or control these kinds of implicit cognitions.

The distinction in format between implicit and explicit attitudes and beliefs also requires a distinction in measurement. Explicit attitudes and beliefs, being controlled and accessible to cognitive awareness, can be self-reported through the typical tools of social surveys or Likert items. We can ask respondents “do you prefer elderly people or younger people?” or “to what extent do you think younger people are smarter than elderly people,” and receive answers that reflect explicit cognitions. Implicit attitudes and beliefs, however, being less accessible to conscious awareness and controllable, inherently cannot be measured in such direct ways. Instead, implicit attitudes and beliefs are measured through indirect measures, including response time tasks such as the widely-used Implicit Association Test (Greenwald et al., 1998); for a recent review, see (Greenwald et al., 2020). Today, in the more than two decades that have followed Greenwald and Banaji’s (1995) initial conceptualization, the evidence continues to be compelling for both explicit and implicit bias (Greenwald & Banaji, 2017), revealing that they are related but distinct mental constructs (Cunningham et al., 2001; Nosek & Smyth, 2007).

The relationship of implicit bias and individual discriminatory behavior

Implicit and explicit attitudes and beliefs don’t sit idle in the mind; they also reveal themselves in our behaviors. Indeed, when studies of the relationship between implicit bias and behaviors are well-powered and performed with precision, the correlation between an individual’s IAT score and their discriminatory behavior is of moderate to large magnitude, \( r \sim 0.40 \) (Kurdi et al., 2019), indicating that the relationship is above the majority of effect sizes in the field of psychology (Funder & Ozer, 2019). Moreover, though both implicit and explicit
cognitions often relate to individual’s behaviors, the two cognitions show incremental predictive validity, meaning that each explains variance above and beyond the other (Kurdi et al., 2019), lending confidence to the unique role played by implicit bias.

Notably, the majority of evidence for such relationships between implicit bias and behavior come from studies of individuals. As an example of such a study, Green and colleagues (2007) assessed individuals’ implicit pro-White/anti-Black attitudes and found that stronger implicit pro-White/anti-Black attitudes correlated with less treatment for hypothetical Black patients with cardiovascular disease ($B = -0.19$), but more treatment for White patients ($B = 0.17$). Over hundreds of such studies, reviewed across three meta-analyses (Greenwald et al., 2009; Kurdi et al., 2019; Oswald et al., 2013) implicit attitudes and beliefs are shown to play a key explanatory role in understanding what cognitions dispose some people to act in more or less discriminatory ways. Recently, however, investigations have turned to a new type of discriminatory behavior that is revealed through socially significant behaviors aggregated across millions of people.

**The relationship of implicit bias and systemic discriminatory behaviors: an overview**

Prompted, in part, by the new availability of big data documenting implicit bias across millions of people around the globe (through the Project Implicit demonstration website at [https://implicit.harvard.edu](https://implicit.harvard.edu)), the past few years have seen more than a dozen studies testing the role of aggregated implicit bias in both shaping and being shaped by systemic outcomes. In addition to the theoretical implications elaborated throughout this chapter, the new focus on aggregated systems or behaviors offers methodological advantages to identify the role of implicit cognition in discrimination. Aggregation allows for greater precision in the estimates of both
explicit cognition and systemic behaviors and can thereby reveal the underlying strong relationships that may previously have been obscured by imprecise and noisy estimates taken from single individuals (B. K. Payne et al., 2020). Furthermore, big data allows for investigations of implicit cognition in real-world behaviors measured at-scale, greatly improving ecological validity and enabling more generalizable conclusions across broader samples and geographic locations.

Notably, nearly all of the studies examining aggregate relationships between implicit cognition and behavior have adopted correlational designs: implicit attitudes or stereotypes are aggregated across geographic locations (counties, states, countries, etc.) and correlated with aggregated systemic outcomes (e.g., rates of lethal force by police, gender gaps in math tests) across those same geographic regions, while controlling for a variety of structural factors (e.g., demographic representation, GDP, etc.). Because such correlations are generally noncommittal regarding the direction of the relationship, the studies can be grouped into two complementary approaches in thinking about the meaning of the relationship. First, there is a set of studies (see Table 1) that take discriminatory systemic behavior as the dependent variable to be explained by measures of implicit bias. For instance, this sort of study may look at how country-level differences in implicit gender stereotypes help explain country-level differences in the outcome of gender gaps in standardized tests (Nosek et al., 2009). Because our goal in this chapter is to show how implicit cognition can illuminate systemic discrimination, this first set of studies constitutes our primary focus. Below, we review and group these studies according to the outcome across socially significant domains of education, healthcare, and policing.

A second type of study in this area considers the relationship between implicit bias and discriminatory systemic behaviors by flipping the equation and identifying the sources of
implicit bias itself as the outcome (i.e., dependent variable). For instance, studies of this type may look at how country-level differences in the representation of fat people help explain the country-level differences in the implicit anti-fat/pro-thin attitudes (Marini et al., 2013) or how demographic and physical features of the environment (e.g., the number of mental health care providers or the number of lakes) predict the magnitude of implicit biases (Hehman et al., 2020). Although these studies primarily seek to understand the sources of implicit bias (rather than the contribution of implicit bias to behaviors) the correlational nature of the studies means that they can still provide complementary insight into the coupling of implicit cognition and specific discriminatory behaviors. As such, we review this group of studies together in the final section.

**Disparities in educational achievement, school discipline and upward economic mobility**

In the first paper to use aggregated Project Implicit data obtained from individual minds, Nosek and colleagues (2009) investigated whether country-level differences in implicit male-science/female-arts stereotypes correlated with gender gaps on 8th grade standardized mathematics achievement tests. The authors found that, across hundreds of thousands of respondents overall, countries with higher implicit associations between male-science (and female-arts) showed larger gender gaps (where boys outperformed girls) on standardized science and math tests ($r = .6$). Additionally, this relationship persisted after controlling for country-level explicit associations, underscoring the incremental predictive validity of implicit cognitions in explaining a socially significant outcome of gender differences in performance.

A few years later, a similar analysis examined the correlation between country-level implicit male-science/female-arts stereotypes and the representation of women in the STEM workforce and STEM tertiary education (Miller et al., 2015). Although this study technically
used the implicit stereotypes as the outcome rather than the predictor, we interpret the results alongside the study by Nosek and colleagues because of the correlational nature of the study, and the fact that gender representation in STEM could theoretically be both a predictor and an outcome of bias (Charlesworth & Banaji, 2019). Here, the researchers found that countries with higher representation of women in STEM tertiary education had weaker implicit male-science/female-arts stereotypes ($R^2 = 0.26$), with similar effects observed for explicit stereotypes ($R^2 = 0.21$). Additionally, this relationship between women’s representation in STEM and implicit stereotypes overrode the relationship observed between achievement and implicit stereotypes. That is, when women’s representation in STEM was controlled for, the relationship between achievement and implicit stereotypes was eliminated. Thus, although there are individual bivariate correlations between country-level implicit bias and the outcomes of gender gaps in STEM achievement and representation, the role of representation may be primary.

It is worth dwelling briefly on why we interpret these sorts of results as “discrimination”. Of course, the aggregated testing gap between boys and girls in an entire country does not arise from any single actor producing discriminatory behavior (e.g., a specific biased examiner or teacher). Rather, the outcomes arise from the systemic presence of beliefs that are so pervasive that we refer to them as being “in the air.” We use this analogy in the same way that Claude Steele referred to “a threat in the air” when explaining women’s underperformance on mathematics tests after stereotypes were evoked (Steele, 1997). This air reveals itself, for example, through ambient cues that signal who belongs in science and math (Cheryan et al., 2009) and in the stereotypes revealed in the word associations across billions of words constituting our collective language (Caliskan et al., 2016; Charlesworth et al., 2021). In turn, such systemic cues may evoke stereotype threat among girls (Spencer et al., 1999) or stereotype
lift among boys (Walton & Cohen, 2003). In short, these studies build an understanding that stereotypes are widely embedded in our society and therefore have the potential to shape the discriminatory behaviors of all those interacting in that society. This spread may be especially true for implicit stereotypes (B. K. Payne et al., 2017) because such hidden, indirect beliefs can pervade and persist more easily in the face of conscious values and ideals against them. In this way, the results summarized in this chapter give new meaning to the idea of systemic discrimination because they quantify how discriminatory outcomes arise from the “air” of bias aggregated across thousands of people in a region.

Following these early studies on gender gaps in STEM achievement and representation, recent studies have looked to also explain race gaps in educational achievement and school discipline. Specifically, Chin and colleagues (2020) found that higher county-level pro-White/anti-Black implicit bias among teachers (a sufficient subsample of the Project Implicit data) correlated with larger Black/White gaps in standardized Math and English testing for 3rd – 8th grade students. Furthermore, county-level relationships also emerged between teacher’s implicit bias and Black/White gaps in student discipline, such that high-bias counties had greater disproportionate discipline of Black students across K-12 (Chin et al., 2020). Similar results were obtained with a larger Project Implicit sample and more indicators of student discipline ranging from in-school suspensions to arrests (Riddle & Sinclair, 2019), although the relationships here were generally stronger for explicit bias than for implicit bias. These results are particularly poignant because early experiences of school discipline interrupt future opportunity and increase interactions with policing and prisons, creating the so-called “school-to-prison” pipeline (Smith, 2009).
Relatedly, when implicit Black/White race attitudes are aggregated within U.S. Census tracts, greater aggregated county-level implicit biases are found to correlate with less future opportunity and upward social mobility (i.e., making more money than their parents) among historically disadvantaged groups (Chetty et al., 2020). That is, Black American boys or girls living in a high implicit bias neighborhood, have a lower chance of upward mobility than if you had, by luck or coincidence, grown up in a relatively low implicit bias neighborhood. These data reinforce the analogy that such aggregate measures of implicit cognition reveal the biased “air” that permeates a community, and the potential for affecting systemic discrimination.

**Disparities in spending on healthcare, rates of disease, and infant health outcomes**

A second group of outcomes centers on the domain of health, including healthcare spending, rates of disease, and infant health. For instance, Leitner and colleagues (2018) used health data on U.S. state-level Medicaid spending from across the United States to examine whether states that are lower in implicit anti-Black attitudes also had greater Medicaid spending. Because Medicaid spending is more likely to preferentially affect Black than White Americans due to persistent and significant racial gaps in income inequality and other health insurance coverage (Smedley et al., 2003), states investment in programs like Medicaid is, in part, reflect their intention to reduce racial inequality. In line with this perspective, states with lower implicit anti-Black attitudes were also found to be higher Medicaid spending states (Leitner et al., 2018), likely facilitating greater equity in healthcare treatment and coverage.

Looking even earlier in the lifespan, researchers have also investigated the relationship between county-level pro-White/anti-Black implicit attitudes and the persistent racial gap in infant health outcomes in the U.S. (Orchard & Price, 2017). Because of a complex set of
compounded stressors including lower prenatal health care and nutrition, and repeated experiences of discrimination, Black American mothers in the U.S. are 1.6 times more likely to have preterm births and twice as likely to give birth to infants with low birth weight. *Why?* Using aggregated county-level implicit race attitudes, Orchard and Price quantified the “biased air” that helps shape these racial gaps in health outcomes: disparities in preterm births were 14% larger in high implicit bias counties (those that were +1 standard deviation above the mean), and disparities in low birth weight were 29% larger in high bias counties. Again, the evidence points to the power of a region with high implicit biases in shaping and maintaining greater disparate outcomes between groups.

**Disparities in police use of lethal force**

Finally, perhaps one of the starkest displays of life-or-death consequences emerges in the relationship between pro-White/anti-Black implicit attitudes and the disproportionate use of police lethal force towards Black Americans (Hehman et al., 2018). First, using fact-checked data from the *Guardian*’s reports of U.S. individuals killed by police, Hehman and colleagues observed that Black Americans constituted approximately 23% of all deaths from police lethal force, but only 13% of the population, indicating disproportionate overrepresentation in deaths at the hands of police. The new result is that such disproportionate overrepresentation can be explained by the strength of implicit bias in a region: the greater the implicit anti-Black/pro-White attitudes and implicit Black-weapon stereotypes in a region, the greater the likelihood of lethal use of force by police toward Black Americans. For instance, a Black American who happened to be living in a region with relatively high implicit Black-weapon/White-object
stereotypes was more likely to be killed by police than if they resided in a region with relatively lower implicit biases.

This study and others reviewed in this section belong to a new generation of evidence. Their impact emerges from the fact that they reveal a relationship between implicit bias and discriminatory behavior even though the biases represent the region, not the individual police officers (or health care providers, teachers, and so on) whose actions constitute the behavioral data. Because these analyses quantify the influence of widespread attitudes and stereotypes in socially significant behaviors these data allow us to advance the proposition that implicit bias is not located in a few individuals or a few segments of society but rather that it is systemic in nature.

Table 1
Select Studies Using Implicit Cognition to Explain Systemic Discriminatory Behaviors

<table>
<thead>
<tr>
<th>Authors</th>
<th>Geographic aggregation level</th>
<th>IAT topic</th>
<th>Societal outcome</th>
<th>Key result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nosek et al, 2009</td>
<td>Countries</td>
<td>Gender-science IAT (male-science/female-arts)</td>
<td>8th grade standardized math/science testing</td>
<td>Stronger male-science/female-arts associations correlated with larger gender gaps on 8th grade testing, $r = 0.60$</td>
</tr>
<tr>
<td>Leitner et al, 2016</td>
<td>U.S. Counties</td>
<td>Black-White race IAT (Black American-bad/White American-good)</td>
<td>Black American and White American death rates from circulatory diseases</td>
<td>Among Black Americans, stronger pro-Black/anti-White implicit attitudes correlated with higher death rates from circulatory diseases, $B = 0.11$</td>
</tr>
<tr>
<td>Leitner et al, 2018</td>
<td>U.S. States</td>
<td>Black-White race IAT (Black American-bad/White American-good)</td>
<td>State spending on Medicaid disability programs (more likely to assist Black than White Americans)</td>
<td>Among White Americans, stronger pro-White/anti-Black implicit attitudes correlated with less spending on Medicaid, $B = -0.33$</td>
</tr>
<tr>
<td>Orchard &amp; Price, 2017</td>
<td>U.S. Counties</td>
<td>Black-White race IAT (Black</td>
<td>Black-White gap in infant low birth</td>
<td>Stronger pro-White/anti-Black implicit attitudes</td>
</tr>
<tr>
<td>Study</td>
<td>Location</td>
<td>Methodology</td>
<td>Findings</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------</td>
<td>-------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Riddle &amp; Sinclair, 2019</td>
<td>U.S. Counties</td>
<td>Black-White race IAT (Black American-bad/White American-good)</td>
<td>Stronger pro-White/anti-Black implicit attitudes correlated with larger Black-white gaps in infant birth weight and preterm births. Counties with +1 SD attitudes had 14% larger gap in low birth weight, and 29% larger gap in preterm births.</td>
<td></td>
</tr>
<tr>
<td>Chin et al, 2020</td>
<td>U.S. Counties</td>
<td>Black-White race IAT (Black American-bad/White American-good) and school discipline</td>
<td>Stronger pro-White/anti-Black implicit attitudes among teachers correlated with larger Black-White gaps in school discipline.</td>
<td></td>
</tr>
<tr>
<td>Chetty et al, 2020</td>
<td>U.S. Counties</td>
<td>Black-White race IAT (Black American-bad/White American-good)</td>
<td>Upward mobility among Black and White boys and girls (roughly defined as making more money than their parents) correlated with less upward mobility among Black boys and girls. Black boys and girls living in counties with +1 SD in implicit bias earn ~0.8 percentiles less income when they grow up.</td>
<td></td>
</tr>
<tr>
<td>Hehman et al, 2018</td>
<td>U.S. Community-based statistical area (CBSA)</td>
<td>Black-White race IAT (Black American-bad/White American-good)</td>
<td>Police lethal force towards Black Americans vs. White Americans correlated with greater disproportionate use of police lethal force, $B = .35$.</td>
<td></td>
</tr>
</tbody>
</table>

Examining implicit bias as the outcome explained by systemic predictors
Implicit bias not only shapes systemic outcomes across aforementioned domains of education, healthcare, and policing but is also shaped by the demographics and structural features of the environment. Indeed, knowing that the level of implicit biases varies meaningfully in magnitude across geography, the question naturally arises: where does such variation come from in the first place? A handful of recent studies have begun to shed light on the answer by using aggregated implicit attitudes and beliefs as the dependent variables predicted from a range of systemic predictors. The majority of these studies have focused on the role of regional demographic representation – the frequency or diversity of specific groups such as Black Americans or fat people – to explain geographic variation in implicit bias. In some cases, the relationship is intuitive and aligns with expectations of intergroup contact theories, in which greater intergroup contact will correlate with less bias (Allport, 1954; Pettigrew & Tropp, 2006). For instance, higher frequency of Asian Americans in U.S. metropolitan areas was found to correlate with lower implicit stereotypes association Asian American with “foreign” and European American with “American” (Devos & Sadler, 2019). Similarly, as aforementioned, higher country-level representations of women in STEM correlated with lower implicit male-science/female-arts stereotypes (Miller et al., 2015).

However, the relationship between demographic representation and implicit bias is not always straightforward. In fact, a higher representation of fat people in a country (presumably allowing for more outgroup exposure) was found to correlate with higher country-level implicit anti-fat/pro-thin attitudes (Marini et al., 2013), and greater frequency of Black Americans in a U.S. state was found to correlate with higher state-level implicit pro-White/anti-Black attitudes among White Americans (Rae et al., 2015). Perhaps these counterintuitive findings could be explained by considering not just the quantity of intergroup contact but also the quality of that
contact (Paluck et al., 2018). Indeed, areas with higher representations of a marginalized group may counterintuitively result in more negative intergroup interactions, thereby perpetuating negative implicit bias. More nuanced insights into the relationship of representation and aggregate implicit bias may therefore come from considering not only raw frequencies of minority representation but also the integration and variety of minority groups. For instance, when looking at multiple indicators of context racial diversity (e.g., integration, prevalence, and variety of groups), implicit stereotypes associating Black Americans with weapons were indeed found to be weaker in U.S. metropolitan areas with greater integration and a larger variety of minority groups (Sadler & Devos, 2020).

It is also notable that the relationships between demographic representation and implicit bias have deep and complex roots in the historical patterns of representation across the country. Recent data show that the greater the proportion of enslaved to free people in the Southern states in the 1860s, the greater the implicit anti-Black/pro-White bias among White Americans in those areas today, 160 years later (Payne et al., 2019). In fact, the correlation between historical rates of slavery from 1860 and state-level implicit race bias today \((r = 0.87)\) was nearly three times larger than the relationship between contemporary Black American representation and implicit race bias \((r = 0.32)\). Studies like this suggest that, when we think about group-based discrimination, we must also think about their effects as extending forward in time as well, translated through the continuous presence of social structures and reminders of inequality (e.g., the presence of confederate monuments; Payne et al., 2019). The result also suggests that today’s Americans who live in regions of greater historical legacies of slavery must be acquiring the particles embedded in their social atmosphere. Systemic discrimination is a useful term in this
case as it helps capture the pervasiveness of discriminatory treatment as it extends across both space and time.

Indeed, time has featured as an important variable in several recent studies examining the temporal relationships between implicit bias and structural variables. For instance, although the implicit Asian American=foreign stereotype had been slowly decreasing over the past decade, racial slurs tweeted during the beginning of the Covid-19 pandemic were found to have reversed the trend and coincided with sharp spikes in implicit bias (Darling-Hammond et al., 2020). Such findings, taken alongside others, show how specific actions in the form of words and slurs can shape implicit attitudes and beliefs. Other data similarly suggest that levels of implicit bias fluctuate in response to the actions and events in the world – whether they are legislation (Ofosu et al., 2019), social movements (Sawyer & Gampa, 2018), pathogens such as Ebola (Inbar et al., 2016), or fat-shaming tweets (Ravary et al., 2019). Each of the studies mentioned accumulates evidence of the relationship between world events and change in implicit bias. These macro-level events have the potency to shape the thoughts and feelings inside human minds which, in turn, re-shape the state of society.

**Implications for understanding implicit bias and systemic behaviors**

The studies we reviewed here belong to a new thrust in research that holds the potential to develop a more robust understanding of the sources and consequences of systemic discrimination. Across over a dozen studies, the evidence shows consistent relationships between aggregated implicit cognition across millions of individuals and socially significant outcomes from academic performance to upward mobility to health and mortality. Such findings reveal the tight coupling between the contents of our minds and our structural worlds.
The data also encourage a new perspective on implicit cognition: perhaps, implicit attitudes and beliefs are to be understood as reflections of the culture and the representations that paradoxically are hidden from conscious awareness, yet pervasively embedded in the structures that surround us (B. K. Payne et al., 2017). This perspective stands in contrast to the early notion that an IAT score reveals an individual’s trait-like personality or permanent individual inclinations (Greenwald et al., 1998) nor that implicit bias offers a *bona fide* pipeline of individual differences (Fazio et al., 1995). Instead, the data reviewed in this chapter support the notion that implicit cognition is an ideal measure for understanding the thumbprint of the culture on the mind. Recently, Payne and colleagues (2017) have summarized this new perspective in the “Bias of Crowds” model to suggest that individual measures of implicit bias are a noisy indicator of the true signal of bias embedded in the broader culture. Although some of that individual “noise” may end up being informative in revealing *which* individuals reflect more versus less accurate signals of their environment, the more useful point is that by aggregating across people, we are better able to understand the true signals of bias in our world.

Although this perspective – that implicit bias reflects the representations embedded in our surrounding cultures – is supported by this newest evidence from big data studies, we note that the argument also receives support from complementary research on children’s implicit bias and the spread of implicit bias in language. For instance, evidence of children’s implicit biases show just how quickly they incorporate the dominant cultural beliefs about their groups. While all children start out with a preference for groups that are similar and familiar to them (Charlesworth & Banaji, 2021) they gradually grow into the associations embedded in their environment. With age, children from minority social groups even come to prefer the dominant or high-status group in their society (Dunham et al., 2007, 2008, 2013, 2014; Newheiser et al., 2014). Thus, children’s
implicit preferences and stereotypes come to reveal the “thumbprints” of their cultures. Additionally, recent applications of natural language processing have shown how implicit, hidden representations of social groups are strongly and widely embedded throughout the language of our culture – the books, TV shows, movies, and even everyday conversations that adults and children encounter (Caliskan et al., 2016; Charlesworth et al., 2021; Charlesworth & Banaji, 2021). Thus, while the theorizing about implicit attitudes and beliefs as reflecting culture is not altogether unprecedented, what is new is the availability of large-scale data and analytic methods to *quantify* evidence for the operation of implicit bias in socially significant outcomes.

**Concluding remarks**

Although the concept and measurement of implicit bias began in psychology, it has since permeated many disciplines from medicine (Green et al., 2007) to computer science (Caliskan et al., 2016), to business (Banaji et al., 2003) to law (Kang & Banaji, 2006). One particularly fruitful interdisciplinary approach has come from the intersection of psychology which is focused on measurements of individuals and those social sciences that have typically focused on larger units of society, particularly economics (Carlana, 2019; Chetty et al., 2020). At this intersection, new combinations of variables are being studied that interweave measures of the thoughts and feelings inside individual minds with measures of broader outcomes such as the opportunity of upward mobility, the likelihood of police shootings, or the health of infants. Evidence has also accumulated for the reverse relationships, in which structural variables, such as the frequency and diversity of demographic representations (e.g., of women, Black Americans), help shed light on the magnitude of implicit bias as an outcome itself.
Together, this emerging body of work shows the tight coupling between implicit bias at the level of the individual and socially significant outcomes at the level of society. The strength of relationships being what it is – with at least small-to-moderate effect sizes compounded over millions of people that experience hundreds of such interactions (Greenwald et al., 2015) – this evidence cannot be set aside. We must take seriously the clear link between implicit bias and systemic discrimination. Moreover, these new studies remind us that discrimination is not always a simple person-to-person act but arises throughout the systems that shape the lives that inhabit these systems such as education, healthcare, or policing. Tackling discrimination therefore requires that we not only address individual decisions but also societal practices, systems, policies, and laws. With such goals, it is clearly an exciting and pressing time for social and behavioral scientists to collaborate in the study of implicit bias and discrimination – we are, in this moment, consolidating our joint expertise in theory and methods to better understand the nature of the mind, society, and the way that each reflects and reinforces the other.
Cross References
References


https://doi.org/10.31234/osf.io/bf97c


