Roles for Implicit Bias Science in Antidiscrimination Law

*Anthony G. Greenwald & Thomas Newkirk*

Declining scholarly interest in intentional discrimination may be due to rapid growth of interest in systemic biases and implicit biases. Systemic biases are produced by organizational personnel doing their assigned jobs, but nevertheless causing adverse impacts to members of protected classes as identified in civil rights laws. Implicit biases are culturally formed stereotypes and attitudes that cause selective harms to protected classes while operating mostly outside of conscious awareness. Both are far more pervasive and responsible for much greater adversity than caused by overt, explicit bias, such as hate speech. Scientific developments may eventually influence jurisprudence to reduce effects of systemic and implicit biases, but likely not rapidly. We conclude by describing possibilities for executive leadership in both public and private sectors to ameliorate discrimination faster and more effectively than is presently likely via courts and legislation.

Scholarly and scientific understanding of discrimination have developed greatly since implicit bias was introduced almost thirty years ago. Figure 1 illustrates the usage frequency of four discrimination-related terms that appeared in English-language books from 1959 to 2019. The plot reveals a long dominance of *intentional discrimination*, peaking in the first decade of the twenty-first century, followed by a more recent decline. Two terms rose to prominence in only the last twenty years, surpassing intentional discrimination by 2013: *implicit bias* and *systemic racism*. These trends signal a rapid societal assimilation of recent work by social scientists, psychological scientists, and legal scholars.

Implicit biases are a subset of one's social knowledge. They include mental associations that are the core of attitudes and stereotypes, acquired continuously, starting early in life. These associations are triggered automatically and without one's awareness during encounters with members of the demographic groups with which they are associated. When activated, the associated attitudes and stereotypes influence thoughts, judgment, and behavior that may thereby be biased toward or against members of those demographic groups. Implicit bias contrasts with explicit bias, a widely used label for consciously accessible beliefs that serve as a basis for (quite possibly) biased judgments and decisions.[1]

Figure 1
Four Discrimination-Related Terms in English-Language Books, 1959–2019
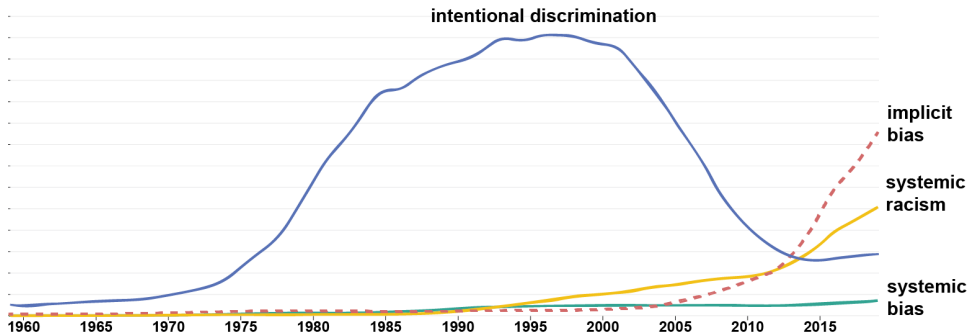
Usage, from 1959 through 2019, of four concepts prominent in English-language scholarly treatments of intergroup discrimination. Source: This plot was produced in Google Ngram by entering the four two-word terms, case insensitive, separated by commas, into the Ngram Viewer's search box, with smoothing set at three years. Google Ngram Viewer, https://books.google.com/ngrams/graph?content=intentional+discrimination%2Cimplicit+bias%2Csystemic+racism%2Csystemic+bias&year_start=1960&year_end=2015&corpus=en-2019&smoothing=3# (accessed December 15, 2023).

Systemic bias is a term we use in place of *systemic racism*, even though the latter term has had much more active use by legal scholars and social scientists since the 1980s (see Figure 1). We avoid using systemic racism both because systemic bias is not limited to race and because the *-ism* suffix connotes a negative mental attitude that is not a component of most of the phenomena now taken to exemplify systemic racism. Systemic biases are rooted in bureaucratic practices that are not in the human mind, but are codified in, among other places, corporate manuals and legislated regulations.

Both implicit biases and systemic biases can produce discrimination that occurs as intentional behavior, and both can occur, when not accompanied by explicit bias, without intent to harm. We are among a growing proportion of scholars and scientists who understand that, in combination (and likely also separately), implicit and systemic biases account for substantially more discriminatory harm than is due to explicit biases.

There are four empirically established properties of implicit biases, each with its own particular challenges: pervasiveness, predictive validity, lack of awareness, and resistance to change.

Pervasiveness. Multiple large studies using the Implicit Association Test (IAT)[2] have found that implicit biases are evident in many people. In this volume, Kirsten N. Morehouse and Mahzarin R. Banaji present in detail the evidence for the pervasiveness of implicit race bias, as measured by the IAT metric of racial preference for White relative to Black, a measure often identified as revealing "automatic White preference."[3] Combining data over fourteen years (2007–2020), Morehouse and Banaji observe that "2.1 of 3.3 million respondents automatically associated the attribute 'Good' (relative to 'Bad') more so with White than Black Americans."[4] By contrast, on a self-report measure of explicit White preference, only 29 percent preferred White relative to Black and "60 percent of respondents reported equal liking for both groups."[5] Data from many volunteers' performance on IAT measures have accumulated at the Project Implicit website, where visitors can choose to complete any of more than a dozen IAT measures of intergroup attitudes or stereotypes.[6] Visitors' performances on these IATs typically reveal that implicit biases are both stronger and more widely prevalent than explicit biases for measures concerning old compared with young, abled compared with disabled, gay compared with straight, male compared with female, Native American compared with White American, light skinned compared with dark skinned, thin compared with fat, and European American compared with Asian American. Numerous other attitude and trait dimensions have been tested and described in research publications, similarly often showing greater prevalence of implicit than explicit biases, but without numbers of respondents approaching the very large proportion of completed tests obtained and archived at the Project Implicit website.

Predictive validity. Discriminatory behavior is reliably predicted by IAT measures of implicit biases. Three meta-analyses of predictive validity of IAT measures have supported this conclusion.[7] It is not presently possible (nor will it likely be in the foreseeable future) to conduct true experimental tests that could establish the interpretation that implicit bias is a *cause* of discriminatory behavior. On the other hand, as an explanation for the observed correlations of implicit bias measures with discriminatory behavior, this causal interpretation has only one competitor, which is that implicit biases and discriminatory behavior have shared causes. At present, and also for the foreseeable future, there is no practical method of using either laboratory or field experimentation to choose between the implicit-bias-as-cause theory and the shared-causes theory.[8] It is therefore reasonable to treat implicit bias either as itself a cause of discrimination or as an indicator of a not-yet-identified precursor of both IAT-measured bias and the discriminatory behavior measures with which IAT measures are found to be correlated.

Lack of awareness. Implicit biases produce discriminatory behavior in persons who do not know that they have discriminatory biases. The best anecdotal evidence for lack of awareness of discriminatory implicit biases is the large proportion of people who, on self-testing with one or more of the freely available on-

line IATs, are surprised – often distressed – to learn that their test scores indicate more-than-trivial strengths of associations indicative of implicit bias.[9]

Resistance to change. Research showing that long-established implicit biases resist change has recently been reviewed in several authoritative publications. We describe here those reviews' findings and their significance. In 2009, psychologists Betsy Paluck and Donald Green reviewed a large collection of studies of prejudice reduction efforts and concluded that "Entire genres of prejudice reduction interventions, including moral education, organizational diversity training, advertising, and cultural competence in the health and law enforcement professions, have never been [rigorously] tested."[10] In 2021, Paluck, Green, and colleagues reported a follow-up review of several hundred subsequent studies, leading to their conclusion that "much research effort is theoretically and empirically ill-suited to provide actionable, evidence-based recommendations for reducing prejudice."[11] The discouraging conclusions of these two large reviews were preceded by a similarly discouraging 2006 review by psychologists Alexandra Kalev, Frank Dobbin, and Erin Kelly, who concluded that "Practices that target managerial bias through feedback (diversity evaluations) and education (diversity training) show virtually no effect in the aggregate."[12] Two substantial multi-investigator collaborative studies by psychologist Calvin Lai and colleagues, of experimental interventions designed to weaken or eliminate long-established implicit biases, concluded that these biases "remain steadfast in the face of efforts to change them."[13] That conclusion by Lai and colleagues was in striking contrast to the more optimistic conclusion – that automatic stereotypes and attitudes were "malleable" – from a 2002 review of the earliest studies of experimental interventions.[14] All of the interventions examined in the 2002 review had been tested with posttests administered very near in time to the intervention. In Calvin Lai, Allison L. Skinner, Erin Cooley, and colleagues' 2016 report of studies with 6,321 participants, none of eight interventions that had previously been found to be effective when tested near immediately after intervention was found to be effective in tests after delays ranging from several hours to several days.[15] The review articles we've briefly summarized here, along with others that reviewed studies conducted in other settings, have themselves been summarized more thoroughly in a recent review, which did not alter the overall picture.[16] We conclude that evidence for the effectiveness of methods assumed to be capable of reproducibly moderating or eliminating implicit biases is lacking.

With these properties of implicit biases in mind, we outline four misunderstandings of scientific work on implicit bias, each immediately followed by its evidence-based correction ("proper understanding").[17]

Misunderstanding 1: *IAT measures assess prejudice and racism*. Proper understanding: IAT measures reveal associative knowledge about groups, not hostility toward them. The IAT and other indirect measures are better described as measuring *bi-*

ases, a term that does not imply prejudice, hostility, or intent to harm, all of which are part of the generally understood meanings of "prejudice" and "racism."

Misunderstanding 2: *Implicit measures are capable of predicting only automatic behavior that is done unthinkingly. They do not predict intentional behavior that is done deliberately*. This misunderstanding was sufficiently widespread that one can find it stated in multiple peer-reviewed psychological publications of the last twenty years. Proper understanding: As three independently conducted meta-analyses have demonstrated, IAT measures equally predict automatic (spontaneous) and intentional (deliberate) behavior.[18]

Misunderstanding 3: *Implicit biases are amenable to modification by experimental treatment interventions*. Proper understanding: As we described above, published experimental tests do not find that long-established implicit biases are reliably modifiable, let alone eradicable, by interventions. This misunderstanding resulted from early studies that examined only effects observable within minutes of administering a treatment intervention. The effects of interventions that produced those findings are now known not to be durable.[19]

Misunderstanding 4: *Group-administered antibias or diversity-training procedures can effectively manage problems that have been attributed to systemic or implicit bias*. Proper understanding: The most authoritative reviews of available studies have concluded that the evidence falls far short of justifying such claims.[20]

How much discriminatory adversity is caused by implicit and systemic biases? Looking at implicit biases first, consider that majorities of all samples that have been studied display the race attitude IAT's "automatic White preference" result. Likewise, majorities (often including majorities of women) associate men more than women with career and women more with family, men more with leader roles and women more with support roles, and men more with STEM disciplines and women more with arts or humanities disciplines. In educational and work settings, these implicit biases predispose teachers and managers to judge the work of White persons more favorably than that of Black persons, to judge men more capable of leadership than women, and to judge men superior to women in math and science disciplines. These observations are a small portion of the empirical support for a conclusion that discrimination-predisposing implicit biases are present in majorities of most populations and, therefore, when aggregated over all those affected, must account for much more damage than do openly expressed (explicit) biases, which are never evident in more than small-to-modest minorities of research samples.

For systemic biases, consider that these are usually the result of a widely applied procedure that was (perhaps long past) created to serve organizational or governmental purposes, presumably without considering how it might affect demographic groups differentially.[21] Systemic biases occasionally receive attention

from public health organizations and news media. During 2021 and 2022, there was frequent reporting of racial disparities in health care outcomes for COVID-19, with substantial attention also given to disparities for groups differing in socio-economic status or age. These disparities are sometimes striking enough to be perceived as unfair and to generate protest, but even so, those who notice the disparities are often in no position to either modify them or influence others to take corrective action. Many discriminatory systemic biases that have not been noticed sufficiently to generate public protest will continue to occur – implemented routinely by myriad employees of governments, businesses, hospitals, schools, and other institutions who are only doing the work that they were hired, elected, or appointed to do. Systemic biases can appear in the form of policies, practices, regulations, and traditions that typically affect multiple (often many) people and frequently produce relatively small effects – but their small size does not mean that those effects are ignorable. The small effects to individuals accumulate, both because of the large number of people affected and because they can affect the same persons repeatedly in settings such as work, school, shopping, travel, paying rent, and paying interest on loans.

There is presently no way to estimate with precision either the percentage of the U.S. population affected by discriminatory implicit and systemic biases, or the magnitude of adversity produced by those discriminatory impacts. We expect it to be relatively modest at the level of individual episodes. Even so, the number affected must be vastly greater than the very small percentage of the U.S. population that now seeks or obtains legal or other governmental redress for discrimination. We know this partly from studies of the Equal Employment Opportunity Commission (EEOC) by the Center for Public Integrity, which investigated the dispositions of discrimination complaints submitted to the EEOC from fiscal years 2010 through 2017.[22] The Economics Policy Institute has a report that goes beyond examining just the EEOC's actions, considering also its problems in gaining congressional budgetary support.[23] One cannot avoid concluding that a great deal slips through large cracks in governmental programs for dealing with discrimination in the United States, even if one considers only discrimination occurring in employment. It is certainly much greater than what is described in reports by the EEOC and parallel state-based agencies. And this is in a system that presently does not yet attempt to deal with more than a small fraction of discriminatory impacts of implicit and systemic biases. We gave brief thought to generating hypothetical estimates of costs, both to those who suffer discrimination and to organizations that have responsibilities for remedying discrimination. However, we are so far from having access to data that could allow even approximate estimates that we must let that challenge await later efforts. When economists with appropriate expertise do undertake such an accounting, they will not find that task easy. Damages due to implicit and systemic biases typically leave no fingerprints, let alone dollar signs.

Discrimination occurs in multiple domains that receive little attention from legislators, regulators, and courts. We learn about these the same way others do : from news reporting via a variety of media. In health care, differential diagnosis and treatment of persons of color, elderly persons, and impoverished persons have been documented in data and reporting from the Centers for Disease Control and Prevention (CDC), the Department of Health and Human Services (HHS), academic researchers, and investigative reporters. In real estate, properties belonging to racial and ethnic minorities are typically undervalued by realtors, meaning that owners receive artificially low offers when selling their properties. Minority purchasers are also most likely to be shown available rentals and homes selectively in neighborhoods in which their ethnic groups have an established presence, if not a majority. In banking, loans to African Americans, women, and members of other protected classes are more likely to be denied and loan interest rates are likely to be elevated. In insurance, as in banking and real estate, members of protected classes receive inferior service and coverage, higher rates charged, and lower rates of success of claims made by them as policy holders. In policing, there are thousands of daily interactions between law enforcement and African Americans and other members of protected ethnic and racial classes that produce increased stops, arrests, arraignments, injuries, and deaths.

Most people (we include ourselves) remain unaware of the majority of discrimination occurring around them. When workers suspect that they are being discriminated against, they will often have difficulty convincing coworkers, or even friends and relatives, that it is indeed discrimination. Should they file a discrimination complaint with the EEOC or other agency, those agencies are very often poorly funded or subject to the enforcement (or nonenforcement) interests of the political party currently in power. In some cases, the EEOC or other agency will investigate the claim and, if the process of conciliation with the accused is unsuccessful, file a lawsuit directly. But even then, agencies litigate a small portion of those lawsuits. More often, agencies will leave it to individual claimants to pursue a lawsuit themselves. Once a complainant receives a notice of right to sue from these agencies, they may, with the help of an attorney, pursue the case in court. Finding a lawyer who understands the claim adequately or finding funds to pursue the claim poses another series of barriers. An expert on implicit bias may also be needed to convince a judge that the plaintiff's case is one for which a jury might award damages. Before a trial occurs, the plaintiff who has overcome all these obstacles must often also survive a defendant's request for summary judgment that can lead a judge to decide to end the proceedings immediately in favor of the defendant.

A suit strong enough to clear all these hurdles may lead the employer being sued to settle rather than face the probability of a jury finding for the plaintiff. Or uncertainty about a favorable outcome may prompt the plaintiff to accept a low

settlement offer. If a trial proceeds, something that happens for only a very small minority of cases, there still remains the barrier of rulings by the court on admissibility or sufficiency of evidence. In the end, a jury composed of persons who might not include a member of the plaintiff's protected class – the jury itself conceivably influenced by implicit biases – may reach a verdict against the plaintiff.

How do implicit biases produce discriminatory behavior? As we summarized above, when mental associations about demographic groups are triggered automatically, the associated attitudes and stereotypes influence behavior that may be discriminatory against members of those groups. But how do *the courts* consider implicit bias as a basis for unlawful discrimination? Because courts give close attention to the role of intent in contemporary discrimination law, and because "intent" is used with a variety of meanings in jurisprudence, we apply a definition of "intent to discriminate" based on the legal definition of intent provided in *Black's Law Dictionary*: intent to discriminate is the mental resolution or determination to do an act that existing law classifies as discriminating against a member of a protected class.[24] Using this definition, we conclude that implicit biases influence decisions that may prove discriminatory, even when decision-makers cannot be fully aware of this influence and may not anticipate that their actions will produce discriminatory consequences. The scientific basis for this understanding comes from adaptation of a long-established *information-processing stages* analysis of choice decision-making, as shown in Figure 2.

The processes of perception and memory retrieval in the first two stages of Figure 2 are understood to occur automatically when encountering a person.[25] Upon perception of stimulus features adequate to distinguish a demographic category of an encountered person (for example, their race), long-established associations, including ones measurable via the IAT, are activated in memory. This activation can predispose (or *prime* in psychology terms) conscious thoughts in the third (thinking) stage, and this priming can further influence conscious judgments and decisions made about the person in the fourth stage. These intentional fourth-stage decisions produce behaviors that can have discriminatory beneficial or harmful consequences, without the decision-maker being aware that these influences are acting through conscious thoughts and judgments that have been influenced by automatically activated mental associations (implicit biases). Psychologists describe the influence of the first two (automatic) stages on thought content as a *bottom-up* influence, meaning that lower (more rapidly occurring) mental processes are influencing higher (later occurring) processes. The influences of thought on judgment (third stage) and decision (fourth stage) are the influences of higher mental processes on behavior (that is, *top-down*). Another useful description of the mechanics of the model in Figure 2 is that the operations of the first two stages occur outside of conscious awareness, but the products of those

Figure 2
Information-Processing Stages Theoretical Framework for Choice
Decision-Making
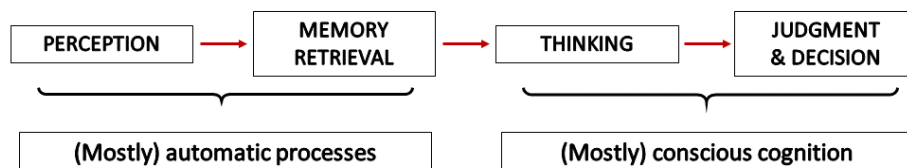
Source: Authors' image, as influenced by Edward E. Smith, "Choice Reaction Time: An Analysis
of the Major Theoretical Positions," *Psychological Bulletin* 69 (2) (1968): 77.

operations are known consciously as they shape judgment and behavior in the
third and fourth stages.

Any person with whom one interacts belongs to multiple demographic cate-
gories on dimensions of gender, race, ethnicity, weight, occupation, and socioeco-
nomic status, among others. All familiar demographic categories have multiple
associations that have been strengthened by overlearning since early childhood.
It is therefore not surprising that an IAT measure of an association of just one de-
mographic category with just one associated attitude or stereotype typically has
only small-to-moderate correlations with measures of discriminatory attitudes or
actions.

Two widely known Supreme Court decisions, *Brown v. Board of Education* in
1954 and *Price Waterhouse v. Hopkins* in 1989, featured critical evidence in-
volving what is now understood as implicit bias.[26] These cases well preced-
ed the scientific introduction of implicit bias in 1995.[27] The decision in *Brown* was
influenced by findings of an experiment showing that, when given the choice be-
tween playing with a White or a Black doll, young Black schoolchildren were much
more likely to choose the White doll. Their choices were implicit (indirect) expres-
sions of a racial bias, because the bias in favor of white skin color was an indirect
expression of the bias. In *Price Waterhouse*, the female plaintiff was turned down
for promotion to a high position for which she was well-qualified. The decision-
making executives at Price Waterhouse gave the explanation that her assertive
personality, something they regarded as appropriate for a male occupant of the
position for which she was being considered, was inappropriate as a trait of the

female plaintiff. This reaction to Hopkins's not conforming to the female gender stereotype of nurturance was an implicit (indirect) indicator that this gender stereotype had played a role in the firm's decision not to promote her.

Another widely known Supreme Court case, *Batson v. Kentucky* in 1986, concluded that peremptory dismissal of Black jurors solely on the basis of race constituted an equal protection violation. In a concurring opinion, Justice Thurgood Marshall wrote, "A prosecutor's own conscious or unconscious racism may lead him easily to the conclusion that a prospective black juror is 'sullen,' or 'distant,' a characterization that would not have come to his mind if a white juror had acted identically. A judge's own conscious or unconscious racism may lead him to accept such an explanation as well supported." Marshall's two examples of conscious or unconscious bias fit quite closely to our analysis, using Figure 2, of how implicit biases can influence judgments and decisions.

Recent scientific advances have led to a new understanding of how, when, and where discrimination occurs. Although scientific knowledge never achieves certainty, it does reach a point at which there is consensus. Scientific understanding of implicit bias is either at or close to that point in regard to the four established properties of implicit bias (pervasiveness, predictive validity of IAT, lack of awareness, and resistance to change). No one expects the American legal system to rapidly and efficiently accommodate new scientific understanding. Many proponents of change to the legal system would still say that change should not be rapid. It is (fortunately, many might say) not up to scientists to decide when new scientific understanding has developed to a point at which it should be put to use. In the world of federal jurisprudence in the United States, the Supreme Court primarily has that power, aided by the circuit Courts of Appeals.

Perhaps the best that scientists can do to advance a goal of change is to point out areas in which established science is at odds with current legal precedents in discrimination law.[28] In discrimination law, we see four areas of such discrepancy that might eventually prompt changes in law or jurisprudence.

First, present scientific understanding does not fit with the present difference in court precedents for what constitutes discrimination in employment law versus equal protection law. Both bodies of law have a requirement of intent; plaintiffs must demonstrate that the defendant intentionally discriminated. In employment law (Title VII), the intent requirement translates to the proposition that the defendant committed an action that caused adversity to the plaintiff under circumstances indicating that the plaintiff's protected class status was a causal factor. When the cause is implicit bias, the defendant may not understand the adversity-producing action as discriminatory – perhaps considering it appropriate, given the defendant's implicitly biased judgment of the plaintiff's job performance. In equal protection law (based primarily on the Fourteenth Amendment's declara-

tion that "No state shall . . . deny equal protection of the laws to any person within its jurisdiction"), the intent requirement translates to the proposition that the defendant did the action purposefully to cause harm to a member (or members) of the plaintiff's protected class. This purposeful intent requirement in equal protection law creates a high bar that reduces the likelihood of plaintiffs succeeding in equal protection cases, such as when newly legislated voting procedures impair the voting opportunities of members of a protected class. There is no apparent basis in scientific understanding of discrimination's mental underpinnings for this difference between evidence requirements of employment law and equal protection law.

A second science-based concern is the often-insurmountable requirement to demonstrate purposeful intent in equal protection cases. The legislators and other officials who create laws and regulations that may have been shaped by implicit or explicit biases may not have purposefully intended to create the resulting adversities, or they may have been careful not to leave evidence of purposeful intent. Violations of equal protection resulting from many governmental actions may therefore not have a path to redress in courts. Similarly, adversities resulting from systemic biases may only rarely exceed the purposeful intent requirement in the equal protection domain. It is difficult to understand why, for example, a state's discriminatory redistricting legislation that denies Black Americans proportional representation in legislative bodies should be treated as an equal protection violation only if plaintiffs can show that the enacting legislators were purposefully trying to reduce Black Americans' opportunities to vote.

A third science-based concern is the recent shift away from the use of *disparate impact* (and toward *disparate treatment*) as the legal criterion for identifying discrimination in employment law. Disparate impact is "The adverse effect of a facially neutral practice (esp. an employment practice) that nonetheless discriminates against persons because of their race, sex, national origin, age, or disability and that is not justified by business necessity. Discriminatory intent is irrelevant in a disparate-impact claim." Disparate treatment is "The practice, esp. in employment, of intentionally dealing with persons differently because of their race, sex, national origin, age, or disability. To succeed on a disparate-treatment claim, the plaintiff must prove that the defendant acted with discriminatory intent or motive."[29] Disparate impact (for which intent is not required) has long been regarded as the appropriate criterion to use when plaintiffs in employment suits claim discrimination due to a "pattern or practice" of the defendant (this translates to systemic bias, as used in this essay). Because of the need to demonstrate the defendant's intent when the court requires the disparate treatment criterion, those suits are necessarily more difficult for plaintiffs than are suits heard under the disparate impact requirements. As shown in a 2011 article by psychologist Lauren B. Edelman and colleagues, "disparate treatment has become far more

prevalent in civil rights cases over time," increasing from about 15 percent of cases in federal District Courts in 1970 to about 95 percent in 1997.[30] This was happening coincidentally with scholarly literature being on the verge of showing a decline in focus on intentional discrimination (see Figure 1). Courts' increasing focus on disparate treatment (for which evidence of intent is required) is at odds with recent social scientific and epidemiological work revealing the widespread operation of implicit and systemic biases, which can produce discrimination without accompanying evidence of intent to discriminate against members of protected classes. This would not happen if discrimination were, in the law, identified as behavior that causes adversity to protected classes rather than being identified with a state of mind that might (or might not) cause such adversity.

The fourth science-based concern is that, in employment discrimination class actions, implicit bias is not now recognized as a basis for establishing the existence of commonality, which is a requirement for certification of a class of plaintiffs in a discrimination suit. In the Federal Rules of Civil Procedure, the commonality requirement serves to assure that plaintiffs grouped into a class share the same basis for complaint against the employer.[31] To scientists, the pervasiveness of implicit biases seems a plausible and appropriate basis for commonality, but no plaintiff has yet tested this reasoning in a U.S. court.

How to deal with the great amount of discrimination that continues to occur in employment? The specifications of Titles VI and VII of the Civil Rights Act of 1964, including modifications added in subsequent congressional amendments and in Supreme Court and circuit Courts of Appeals precedents, fall well short of covering what scholarly and scientific work now identify as sources of employment discrimination. It is not simply the noncoverage of discriminatory impacts resulting from implicit and systemic biases. It is also that the Equal Employment Opportunity Commission's capacity does not come close to the EEOC's goals as stated in the Equal Employment Opportunity Act of 1972. The text of that Act starts with "The Commission is empowered … to prevent any person from engaging in any unlawful employment practice as set forth in section 703 or 704." Sections 703 and 704 contain the main statements of unlawful employment practices in the 1964 law's centerpiece, Title VII.

Writing this essay gave us some optimism that the science of implicit bias may be leverageable to improve prospects for plaintiffs to base effective discrimination suits at least partly on implicit bias evidence. Despite making good progress on that goal, much of what we learned in the process prompted us to consider prospects for effective efforts to address problems of discrimination outside the justice system, including both private-sector executives and officials in public-sector executive roles. We start with a short list of problems that can be addressed by actors in these nonjudicial, nonlegislative roles.

Efforts intended to remediate suspected or claimed discrimination in large organizations presently use training methods that are not established as effective. If they serve the organization at all, these training efforts do so by projecting the appearance that the organization's leaders are trying to eliminate or control discrimination. This almost always misleading (as it turns out) appearance can be counterproductive when it deflects leaders from seeking more effective methods.

Those who discover evidence of discrimination rarely occupy positions that enable them to work cooperatively with leaders of the organization in which they have uncovered discrimination. They are more likely to be seen as whistle-blowing enemies of the organization, possibly also becoming targets for retaliation. CEOs of large organizations may have little internal motivation and little external pressure to scrutinize the organization's personnel databases to identify discriminatory disparities that would be both easy to identify and straightforward to repair, once identified.

Many organizations assign responsibility for dealing with discrimination not to top-level leaders, but to organizationally subordinate human relations and legal departments, the personnel of which may have greater motivation to please their supervisors than to rock the organizational boat by investigating, discovering, and calling for remediation of discrimination within the organization.

We did not initially intend for this essay to propose private-sector remedies for discriminatory disparities due to implicit and systemic biases. That plan developed when we became aware of an underused remedial strategy, disparity-finding, that has three attractive properties: 1) it is easy to describe, 2) it is straightforward to administer, and 3) it can be deployed outside the American justice system.

Even though not previously named, the disparity-finding method is well known to epidemiologists, who use it frequently to find and identify public health problems.[32] These discoveries not only reveal health care disparities, but can also make apparent who is in the best position to fix the disparities. Consider this example: An epidemiologist working at Institute I discovers a health care disparity at Hospital H, where members of Group A are noticeably more likely to suffer from affliction X than are members of Group B. Alas, the researchers at Institute I may have no power to direct administrators or staff at Hospital H to undertake feasible remedies. For example, epidemiologists working for the CDC and for other research agencies uncovered numerous health care disparities during the COVID-19 pandemic. However, the CDC could not direct public health agencies or governmental officials in various localities to invest funds or otherwise take steps needed to implement fixes, even though it was often obvious what fixes would be required. This example makes clear why it is optimal for the work of disparity-finding to be the responsibility of executive personnel within the organization in which the disparity exists.

One often reads about investigative journalists uncovering discrimination, especially police profiling that clearly amounts to racial or ethnic bias. The journalists who reveal these problems are not in positions that enable them to implement fixes. That is the general problem in many contexts: the people with data capable of revealing the problem lack authority to intervene to fix the problem. Remarkably, this problem need not exist in many situations in which implicit biases or systemic biases are causing discriminatory disparities. In a business organization, the personnel data are owned by the company that employs the affected workers. In a police department, the data on racial characteristics of drivers and pedestrians stopped and searched by police officers, as well as the footage from body cameras operated by those officers, are in the possession of those police departments. In a university, records of qualifications and performances of students or staff who may be disadvantaged by implicit or systemic biases are in possession of the university itself. If the business organization, the police department, or the university employs a data scientist with appropriate quantitative skills, there should be no difficulty in using available data to uncover discriminatory disparities and report findings to administrative executives who can take responsibility for fixing them. How often does this sequence of disparity-finding followed by repair occur in organizations in which unrecognized discriminatory disparities exist? To the best of our knowledge – mainly because we almost never hear about it – the answer is "rarely."

All medium-to-large workplaces in the United States maintain personnel data as required by the EEOC and also as needed to keep their businesses operating. The available information usually includes employee demographics and data on employees' educational qualifications, productivity, job title, years employed, salary, raises, promotions, absences, performance evaluations, awards received, and discipline administered. If demographic disparities exist, the available personnel data likely contain evidence of them, and that evidence should not be difficult to find.

There is an essential second step after finding a disparity. A data analyst with the skills of an epidemiologist must also understand how interrelations among the personnel data variables can spuriously create or obscure appearances of a discriminatory disparity. Therefore, before suggesting that a discovered disparity must be repaired, a necessary step is to have the statistical expert assure that an identified disparity does not have a straightforward nondiscriminatory explanation. As just one easy to understand example, Group A might differ from Group B by having both 1) higher salaries and 2) stronger performance evaluations, even though the two groups are indistinguishable in qualifications, years employed, and other possibly relevant variables. This might be a basis for judging that Group A's greater average salary is explained by their superior performance and is therefore not discriminatory. However, that conclusion should await examining other possibilities, especially whether the performance evaluations were made objectively by a validated method or, instead, subjectively by the same manager who

determined each employee's pay. If the latter, the more plausible interpretation may be that the manager is discriminating in favor of members of Group A. For this reason, evaluations of performance are ideally done using objective criteria (that is, no subjective evaluation involved) by persons who play no role in deciding on pay or promotion.

There are reasons to believe that discriminatory disparities will be found almost whenever disparity-finding is undertaken. The two settings that produce most of the publicly known examples of disparity-finding occur in policing and health care, and that disparity-finding has been done mostly by outside agencies. In policing, watchdog/citizens' organizations and investigative reporters use FOIA (Freedom of Information Act) requests to obtain data access. In health care, the data may be voluntarily provided by hospitals or other medical institutions, or available in public archives such as those maintained by HHS or the CDC. Unfortunately, those efforts may not have enough data access to establish whether revealed disparities are discriminatory or nondiscriminatory. Also, an outside agency that obtains the information generally has no authority to grant, force, or enforce an effective fix of a discovered disparity. On the other hand, when disparity-finding is done within an organization that maintains its own personnel database, the finding is in the hands of those best positioned both to identify a plausible nondiscriminatory cause and to devise a fix if they cannot identify a nondiscriminatory explanation.

Business considerations, especially the standard goal of maximizing profit, may suppress willingness of organizational leaders to undertake routine (such as annual) disparity-finding scrutiny of their personnel data. Ideally, the CEO assigns responsibility for disparity-finding work to an executive whose annual bonus will increase directly as a function of success in identifying previously unrecognized disparities and determining whether they are discriminatory.

Advocates of internal disparity-finding should be aware that the organization's leaders will be concerned (appropriately) that employees who learn of uncovered disparities may use that knowledge to launch discrimination suits. For that concern not to discourage businesses from undertaking disparity-finding, courts can recognize a "self-critical analysis" privilege that protects the company from having its self-discovered evidence used against it. In practice, however, courts rarely grant this privilege, in effect motivating employers to neglect routine disparity-seeking scrutiny of their personnel data. Legal scholar Deana Pollard Sacks and others have pointed out that, if courts allow this self-critical analysis privilege, this could be very helpful in reducing unwanted discrimination, such as can result from not yet recognized implicit and systemic biases.[33]

We imagine how future historians may view progress of American treatment of discrimination since the Civil Rights Act of 1964. They might see that Act itself as a central piece in two centuries of legislation that

increased legal protections of civil rights beyond the Fifth Amendment's (1791) declaration that "No person shall be . . . deprived of life, liberty, or property, without due process of law" and the Fourteenth Amendment's (1868) assertion that "No state shall . . . deny equal protection of the laws to any person within its jurisdiction." Some important later pieces of legislative progress include the Fifteenth (1870), Nineteenth (1920), and Twenty-Fourth (1964) constitutional amendments, the Equal Pay Act (1963), the Age Discrimination in Employment Act (1967), and the Americans with Disabilities Act (1990). Concurrent with legislative developments since the middle of the twentieth century, decisions of the U.S. Supreme Court gradually limited the scope of antidiscrimination laws. Concurrently, but outside the legal system, scientists and scholars were establishing that much more discrimination than was previously apparent to the legal system was occurring in forms that were often not intended to harm and that were not readily apparent either to their perpetrators or to their victims. Remedy for those forms of discrimination – implicit and systemic biases – was not then easily available within the U.S. justice system.

Can we predict the next few sentences of this future history? An even more interesting question: what might be done now to shape the content of those sentences?

ABOUT THE AUTHORS

Anthony G. (Tony) Greenwald, a Fellow of the American Academy since 2007, is Professor Emeritus of Psychology at the University of Washington. He provoked modern attention to the psychological self with his 1980 article, "The Totalitarian Ego." His 1990s methods contributions made unconscious cognition and subliminal perception orderly research topics. His 1995 creation, the Implicit Association Test, enabled observation of unconscious associative knowledge, revamping understanding of stereotypes and prejudice.

Thomas Newkirk is a Partner at Newkirk Zwagerman. For thirty-five years, he has represented individuals who have been denied equal employment opportunities. He specializes in incorporating the science of implicit bias into the practice of law to advance the goal of achieving equality for all protected groups.

ENDNOTES

1 "Implicit bias" was first used in its present meaning by Anthony G. Greenwald and Mahzarin R. Banaji in "Implicit Social Cognition : Attitudes, Self-Esteem, and Stereotypes," *Psychological Review* 102 (1) (1995): 4–27.

2 Anthony G. Greenwald, Debbie E. McGhee, and Jordan L. K. Schwartz, "Measuring Individual Differences in Implicit Cognition : The Implicit Association Test," *Journal of Personality and Social Psychology* 74 (6) (1998): 1464–1480. This article introduced the IAT, originally presenting it as a measure of association strengths, later understood also as a measure of implicit biases.

3 Kirsten N. Morehouse and Mahzarin R. Banaji, "The Science of Implicit Race Bias : Evidence from the Implicit Association Test," *Dædalus* 153 (1) (Winter 2024): 21–50, https ://www.amacad.org/publication/science-implicit-race-bias-evidence-implicit -association-test.

4 Ibid., 28.

5 Ibid.

6 Project Implicit, https ://www.projectimplicit.net (accessed December 15, 2023).

7 An overview and comparison of the three meta-analyses is available in Anthony G. Greenwald and Calvin K. Lai, "Implicit Social Cognition," *Annual Review of Psychology* 71 (2020): 419–445.

8 The problem is that true experiments require assigning research subjects at random, early in life, to environments that would create the attitude and stereotype associations that comprise implicit biases. This is neither practical nor ethical.

9 In *Blindspot : Hidden Biases of Good People*, Mahzarin R. Banaji and Anthony G. Greenwald reported that 40 percent of 1.5 million White Americans who had completed the Race attitude IAT online displayed the combination of explicit egalitarianism (self-reported equal liking of racial Black and racial White) together with IAT results indicating implicit preference for White relative to Black. Mahzarin R. Banaji and Anthony G. Greenwald, *Blindspot : Hidden Biases of Good People* (New York : Delacorte Press, 2013), 158.

10 Elizabeth Levy Paluck and Donald P. Green, "Prejudice Reduction : What Works ? A Review and Assessment of Research and Practice," *Annual Review of Psychology* 60 (2009): 356.

11 Elizabeth Levy Paluck, Roni Porat, Chelsey S. Clark, and Donald P. Green, "Prejudice Reduction : Progress and Challenges," *Annual Review of Psychology* 72 (2021): 533–560.

12 Alexander Kalev, Frank Dobbin, and Erin Kelly, "Best Practices or Best Guesses ? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies," *American Sociological Review* 71 (4) (2006): 611.

13 Calvin K. Lai, Maddalena Marini, Steven A. Lehr, et al., "Reducing Implicit Racial Preferences : I. A Comparative Investigation of 17 Interventions," *Journal of Experimental Psychology : General* 143 (4) (2014): 1765 ; and Calvin K. Lai, Allison L. Skinner, Erin Cooley, et al., "Reducing Implicit Racial Preferences : II. Intervention Effectiveness across Time," *Journal of Experimental Psychology : General* 145 (8) (2016): 1014.

14 Irene V. Blair, "The Malleability of Automatic Stereotypes and Prejudice," *Personality and Social Psychology Bulletin* 6 (3) (2002): 242–261.

15 Lai, Skinner, Cooley, et al., "Reducing Implicit Racial Preferences."

16 Anthony G. Greenwald, Nilanjana Dasgupta, John F. Dovidio, et al., "Implicit Bias Remedies : Treating Discriminatory Bias as a Public Health Problem," *Psychological Science in the Public Interest* 23 (1) (2022): 7–40.

17 A more detailed presentation of these misunderstandings is in Greenwald, Dasgupta, Dovidio, et al., "Implicit Bias Remedies."

18 To illustrate, in the first of the three meta-analyses reported by Anthony G. Greenwald, T. Andrew Poehlman, Eric L. Uhllmann, and Mahzarin R. Banaji, each discriminatory bias measure that was tested for prediction by an IAT measure was rated for the extent to which the behavior was difficult to control consciously (0 = maximally spontaneous ; 10 = maximally controllable). Choice of whom to vote for in a presidential election was a measure that was judged easy to control, whereas nonverbal behaviors such as eye blinks, speech hesitations, or body orientation variations were judged as highly spontaneous. Anthony G. Greenwald, T. Andrew Poehlman, Eric L. Uhllmann, and Mahzarin R. Banaji, "Understanding and Using the Implicit Association Test : III. Meta-Analysis of Predictive Validity," *Journal of Personality and Social Psychology* 97 (1) (2009): 17.

19 A detailed review of studies of interventions can be found in Greenwald, Dasgupta, Dovidio, et al., "Implicit Bias Remedies." That review identified only a few isolated studies that found a measurable intervention effect after the day on which the intervention was administered. The review describes those exceptional results but was unable to identify any component of those studies that appeared responsible for findings of apparent durability. This leaves a presently unsolved mystery.

20 For an overview that includes summaries of the authoritative reviews mentioned in this paragraph, see Greenwald, Dasgupta, Dovidio, et al., "Implicit Bias Remedies."

21 It is easily conceivable that many systemic racial biases were created deliberately with the expectation that they would produce race or gender disparities.

22 Maryam Jameel and Joe Yerardi, "Despite Legal Protections, Most Workers Who Face Discrimination Are on Their Own," The Center for Public Integrity, February 28, 2019, https ://publicintegrity.org/inequality-poverty-opportunity/workers-rights/workplace -inequities/injustice-at-work/workplace-discrimination-cases. This report states, "Though the law places the burden on employees to prove discriminatory intent or impact, when hard evidence of unequal treatment exists, it is often buried in personnel records only the employer can access." See also Maryam Jameel, "More and More Workplace Discrimination Cases Are Closed Before They're Even Investigated," The Center for Public Integrity, June 14, 2019, https ://publicintegrity.org/inequality-poverty -opportunity/workers-rights/workplace-inequities/injustice-at-work/more-and-more -workplace-discrimination-cases-being-closed-before-theyre-even-investigated.

23 Jenny R. Yang and Jane Liu, "Strengthening Accountability for Discrimination," Economy Policy Institute, January 19, 2021, https ://www.epi.org/unequalpower/publications /strengthening-accountability-for-discrimination-confronting-fundamental-power -imbalances-in-the-employment-relationship.

24 *Black's Law Dictionary*, 11th edition (St. Paul, Minn. : Thomson Reuters, 2019). This elaboration of *Black's* definition of intent applies it to the context of discrimination law. This definition neither asserts nor implies that motive to inflict harm is part of intent.

25 This point was established in a widely referenced theoretical article published in 1977 : Richard E. Nisbett and Timothy D. Wilson, "Telling More Than We Can Know : Verbal Reports on Mental Processes," *Psychological Review* 84 (3) (1977): 231–259.

26 *Oliver Brown, et al. v. Board of Education of Topeka, et al.*, 347 U.S. 483 (1954); *Price Waterhouse v. Ann B. Hopkins*, 109 U.S. 1775 (1989).

27 See Greenwald and Banaji, "Implicit Social Cognition."

28 Established science is never unanimously accepted by scientists. An essential and valuable aspect of the scientific method is to question existing understanding. Science becomes problematic when scientists stop questioning their favorite theories, which may be their own theories.

29 These definitions are from *Black's Law Dictionary*, 11th edition.

30 Lauren B. Edelman, Linda H. Krieger, Scott R. Eliason, et al., "When Organizations Rule: Judicial Deference to Institutionalized Employment Structures," *American Journal of Sociology* 117 (3) (2011): 888–954.

31 A prerequisite for class certification stated in Rule 23(a) of the Federal Rules of Civil Procedure is that "there are questions of law or fact common to the class."

32 This epidemiological technique was only recently given the label "disparity finding." See Greenwald, Dasgupta, Dovidio, et al., "Implicit Bias Remedies," 25.

33 Deana A. Pollard, "Unconscious Bias and Self-Critical Analysis: The Case for a Qualified Evidentiary Equal Employment Opportunity Privilege," *Washington Law Review* 74 (1999): 913–1031.