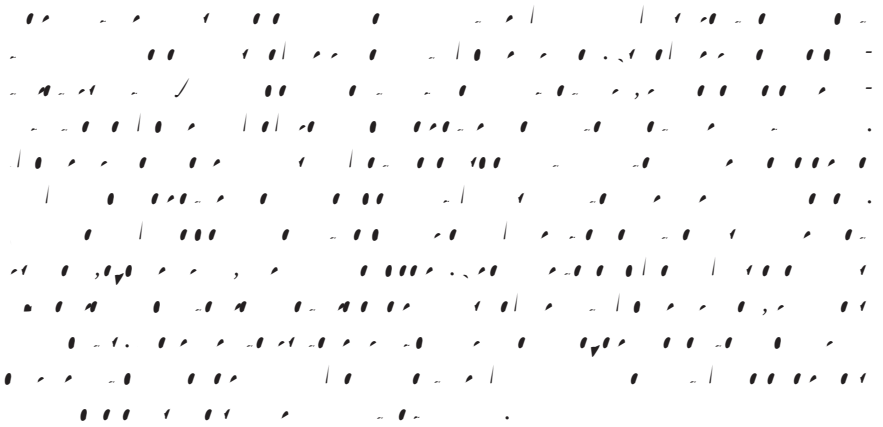
Roles for Implicit Bias Science in Antidiscrimination Law

Anthony G. Greenwald & Thomas Newkirk



S

is a term we use in place of , 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 - 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.

. 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.

. 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 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.

. 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-

, 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: . 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: . 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: . 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

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 socioeconomic 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

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 , 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 : 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
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 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 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,). 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

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, . , , , 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

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

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 prob

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.

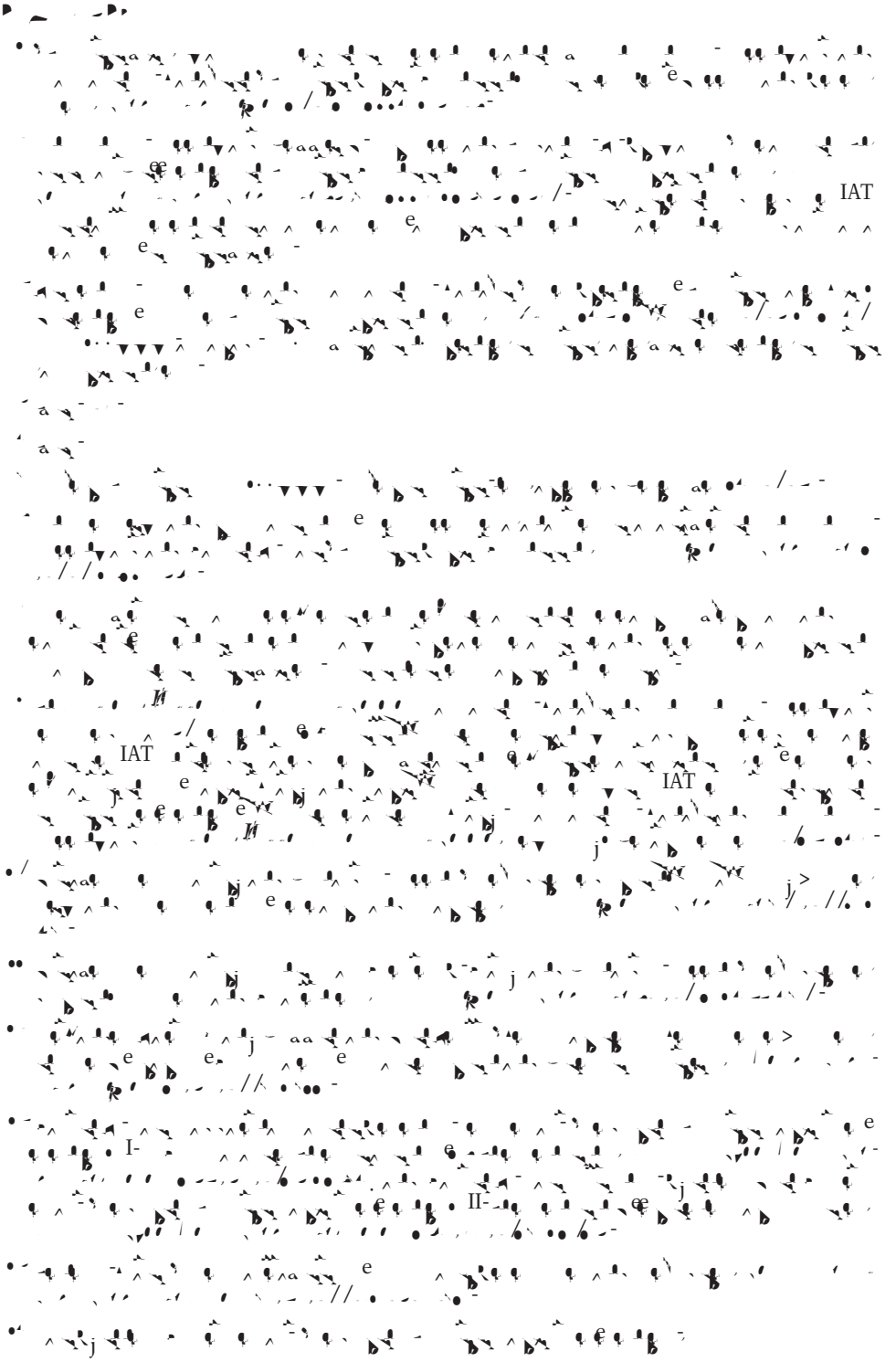
There is an essential second step after finding a disparity. A data analyst with the skills of an epidemiologist must also understand how interrelations among the personnel data variables can spuriously create or obscure appearances of a discriminatory disparity. Therefore, before suggesting that a discovered disparity must be repaired, a necessary step is to have the statistical expert assure that an

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?

Anthony G. (Tony) Greenwald

Thomas Newkirk


IAT
IAT
IAT