

When bias is implicit, how might we think about repairing harm?

Mahzarin R Banaji¹, R Bhaskar¹ and Michael Brownstein²

A recent Supreme Court decision — *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.* — creates an opening to consider models for repairing the effects of unintended harm. We mention some results from the science of unconscious bias, consider the nature of n-to-n harm, cite recent philosophical arguments about responsibility for carrying implicit bias, and note the legal status of *intent* versus *impact* in civil rights law. Based on the opportunity presented by *Inclusive Communities*, we present three options for repairing unintended harm, placing emphasis on litigation-minimizing solutions, especially insurance.

Addresses

¹ Department of Psychology, Harvard University, United States

² Department of Philosophy, CUNY/John Jay College of Criminal Justice, United States

Corresponding author: Banaji, Mahzarin R
(mahzarin_banaji@harvard.edu)

Current Opinion in Psychology 2015, 6:183–188

This review comes from a themed issue on **Morality and ethics**

Edited by **Francesca Gino** and **Shaul Shalvi**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 29th August 2015

<http://dx.doi.org/10.1016/j.copsyc.2015.08.017>

2352-250X/© 2015 Elsevier Ltd. All rights reserved.

On June 25, 2015, the United States Supreme Court held that the Fair Housing Act of 1968 [1] requires remedies to organizational practices that have disparate impact on social groups covered by the statute. In *Texas Department of Housing and Community Affairs v. The Inclusive Communities Project, Inc.*, the Court held that the disparate impact a policy or practice produces is sufficient to merit remedy. In other words, the Court held that even if no intent to discriminate can be discerned, potential harm as revealed by disparate impact can be challenged [2].

Inclusive Communities is important in at least two ways. First, for most of the Court's recent history, in the area of civil rights, the idea that assessment of harm hinges on clear and demonstrable intent has been held with dogged persistence. *Inclusive Communities* is important because it reverses that direction. Second, in *Inclusive Communities*, Justice Kennedy, writing for the majority, made explicit reference to a particular psychological state of mind,

stating that disparate impact is sufficient basis for a challenge because 'unconscious prejudice and disguised animus' can mask discrimination.¹

The science

Unconscious forms of bias pose a problem for legal theory and practice because law's guiding model of human behavior assumes that sane, ordinary, adult behavior is the result of conscious and intentional decision-making. Yet a rich and robust body of research in experimental social psychology — the field from which the concept of 'implicit bias' has emerged — challenges that assumption. We know that human minds evolved unique and specialized ways of processing information, some of which are capable of producing self-reflective, deliberate, conscious thought in accordance with moral codes and intentions, while other mental computations are achieved in a more automatic, unconscious, and implicit manner [3*,4].

In 1995, Greenwald and Banaji [5] proposed that the study of *implicit social cognition* deserved new attention to understand core aspects of the mind: attitudes, stereotypes, and self-based cognition. Alongside, they issued a demand for the development of new methods that could robustly access implicit social cognition. The demand for new methods was partly fulfilled with the invention of the Implicit Association Test (IAT; [6]) and it has come to be, among other methods, a viable way to reveal the presence of implicit bias.² A signature result from research using the IAT is that people who have no intention to discriminate may still do so in their behavior toward others who vary in age, gender, class, race/ethnicity, sexuality, religion, and nationality among other social groupings (see [3*] for review).

The amount of published and replicated evidence showing the presence of differential treatment in the domains of employment, housing, financial lending and healthcare as a function of group membership is staggering. Even a focus on just one protected category, race, reveals an overwhelming amount of evidence from every social science [8–12]. The unique contribution of modern psychological research has been to show that such differences may emanate less from animus and more from implicit, less conscious, mental processes. For example, anti-Hispanic IAT bias predicted attitudes toward illegal

¹ p. 13 Slip opinion, decided June 25, 2015 [2].

² We leave aside discussion of whether the biases revealed by the IAT are completely unconscious or relatively unconscious. For discussion, see [7].

and legal immigration [13]; ER and resident physicians with stronger anti-Black IAT bias were less likely to prescribe a particular medical procedure to Black patients [14]; stronger anti-Black IAT bias among physicians led to more negative experiences with Black patients [15]; stronger anti-Arab IAT bias predicted hiring decisions [16]; anti-obese IAT bias was related to less likelihood of interviewing obese candidates [17]; associations of ‘mentally ill’ with ‘dangerous,’ were correlated with stronger endorsements of societal control mechanisms [18]; at the level of countries, an IAT gender-science measure predicted gender differences in performance on math achievement tests [19]; anti-Black IAT biased influenced (correctable) bias in trial judges [20]; voters with stronger anti-Obama IAT scores were more likely to oppose his policies — but not when the same policies were attributed to President Clinton [21].

None of these results would be as surprising if the measures of attitudes and stereotypes were obtained via self-report. We would conclude that those who bear animus toward a group or idea are acting rationally on that preference or belief, hence the correlation. Data on implicit or unconscious bias are surprising and even troubling because individuals and even professionals, whose conscious values reveal no intent to harm, nevertheless show systematic and selective patterns of decision-making that result in differential treatment.

For research on implicit bias to have direct relevance for considerations of intent *vs.* impact arguments in the policy decisions, it is perhaps of use to view the range of conclusions that can be drawn from the empirical evidence that has gathered over the past thirty years (see [22]). The most significant of them for the discussion at hand is the idea that implicit bias (a) is pervasive, (b) is dissociated from conscious intent and values, (c) reflects preferences for one’s own group or dominant groups in society, and (d) influences behavior. It is these reasons more than any others that provide the foundation for an understanding of human behavior that can directly motivate looking deeper as Justice Kennedy argues, when disparate impact is observed. See [Table 1](#) for a more complete list of established results about implicit bias.

Harm: individual-to-individual vs. n-to-n

The question of responsibility is a thorny one when unintended harm occurs. Fortunately, moral philosophers have recently engaged this question in the context of implicit bias. Two volumes titled *Implicit Bias and Philosophy* [23,24*] include several viewpoints on exactly this question. It is surprising that every chapter dealing with the topic of responsibility takes the position that even though harm due to implicit bias may be unintended, responsibility for remedying the harm lies firmly with the agent [25–29]. The arguments to support this position

Table 1

Significant conclusions from research on implicit bias (IB) showing attitudes and stereotypes by group membership (age, gender, sexuality, race, ethnicity, religion, nationality, physical characteristics) as well as the involvement of self-identity.

- IB is universal; it is a byproduct of fundamental features of thinking
- IB varies by individual and by group
- IB reveals ingroup preference
- IB reveals dominant group preference
- IB is dissociated from conscious intent/values
- IB is also associated to conscious intent/values
- IB is present in children; at times, to the same extent as adults
- IB also varies developmentally
- IB in behavior has been linked to neural activation
- IB predicts behavior, including behavior in natural settings
- IB is temporarily malleable; it is responsive to particular interventions

range from comparisons to other situations of negligence, the distinction between guilt (not necessary) and reparations (necessary), to the conjecture that as evidence of implicit bias has become both scientifically clear and easily available in the public domain, it is one’s responsibility to be aware of it and act on it.

Important as these arguments are, they are restricted to cases of individual-to-individual actions. Indeed, in support of the philosophers’ positions, there is evidence to suggest that implicit social cognition is knowable and malleable. In fact, some methods of intervention, such as positive forms of contact, can change even unconscious bias [30]. This is a worthy path to develop as it can lead to changes in an individual’s behavior.

In this paper, however, we focus on a different level of harm-doing. We recognize that organizationally mediated disparate impact, that is, policies and practices in areas such as housing, education, medical care, and financial lending, have unique characteristics that deserve exploration of alternative methods for determining responsibility. In such cases, the actions may be termed n-to-n. That is, the sources of unintended harm are many, and the effects are experienced by many. In n-to-n actions, the central challenge is not in determining who is ‘guilty’ (as philosophers have keenly noted) but rather crafting the best mechanics of repair.

Intent vs. impact

Ordinary humans believe that it is important to separate acts of intentional harm from acts of unintentional harm. For example, if A intentionally plans to kill a girl and B accidentally kills a girl, the intentional harm-doer is obviously more morally compromised. But consequences also matter, so that if two individuals get equally drunk and drive, with one subsequently hitting a tree and the other hitting a girl, the extent of the harm done (rather than intent) matters for determining punishment [31,32*].

Download English Version:

<https://daneshyari.com/en/article/879432>

Download Persian Version:

<https://daneshyari.com/article/879432>

[Daneshyari.com](https://daneshyari.com)