

American University Washington College of Law

Digital Commons @ American University Washington College of Law

Articles in Law Reviews & Other Academic Journals

Scholarship & Research

4-12-2021

The insights, uses, and ethics of social neuroscience in anti-discrimination law

Susan Carle

American University Washington College of Law, scarle@wcl.american.edu

Follow this and additional works at: https://digitalcommons.wcl.american.edu/facsch_lawrev




Part of the [Courts Commons](#), [Law and Psychology Commons](#), and the [Law and Race Commons](#)

Recommended Citation

Carle, Susan, "The insights, uses, and ethics of social neuroscience in anti-discrimination law" (2021).
Articles in Law Reviews & Other Academic Journals. 1968.
https://digitalcommons.wcl.american.edu/facsch_lawrev/1968

This Article is brought to you for free and open access by the Scholarship & Research at Digital Commons @ American University Washington College of Law. It has been accepted for inclusion in Articles in Law Reviews & Other Academic Journals by an authorized administrator of Digital Commons @ American University Washington College of Law. For more information, please contact kclay@wcl.american.edu.

The insights, uses, and ethics of social neuroscience in anti-discrimination law

Susan D. Carle* 

Washington College of Law, American University, Washington, DC, United States

*Corresponding author: e-mail address: scarle@wcl.american.edu

Contents

1. The useful insights of social neuroscience for anti-discrimination law	1
2. Using social neuroscience in the law to counter implicit bias	7
2.1 Minimizing the use of intent standards to prove unlawful discrimination	7
2.2 Fashioning remedies for discrimination	10
2.3 Broadening doctrines that cramp the application of anti-discrimination law	13
2.4 Expanding rights to act differently	16
3. Regulating the use of neuroscience in anti-discrimination law	20
3.1 Using neuroscience in particular cases	20
3.2 Using neuroscience to inform policy and reform anti-discrimination law	21
4. The ethics of using neuroscience in anti-discrimination law	24
5. Conclusion	30
Conflict of interest	31
References	31



1. The useful insights of social neuroscience for anti-discrimination law

The scientific discoveries of neuroscience, and more specifically the emerging field called social neuroscience (Cacioppo and Decety, 2011; Todorov et al., 2012), have provided important new insights to the field of anti-discrimination law and theory. By social neuroscience, I mean a broadly interdisciplinary perspective that uses natural science-based methods to inquire into how the “social brain” works (Matusall, Kaufmann, & Christen, 2011, p. 9). As an interdisciplinary field, social neuroscience borrows from many natural science traditions, including medicine, epidemiology, behavioral economics, animal behavior studies, experimental social

and developmental psychology, game theory and computer science, to learn more about how the brain responds to its social environment (Cacioppo & Decety, 2011, pp. 5–7).

The psychological disciplines have long played a role in law (Dailey, 2017). In an earlier era, the psychology that influenced legal theorists was often Freudian, as Dailey explores. In the 20th Century, Freud provided important contributions to legal theorists, including the so-called “American Legal Realists”—who are, somewhat confusingly, anti-realists in the philosophical sense, in that they viewed law as thoroughly socially constructed and primarily generated through politics rather than logical rules (Leiter, 2002). Theorists such as Jerome Frank drew from Freud to argue that the forces motivating judges and other legal actors were often lodged in the unconscious, and thus far more complex and opaque than law traditionally assumed (Frank, 2009).

Freudian psychology, based primarily in humanistic, hermeneutical traditions rather than natural science methods, differs greatly from neuroscience and related disciplines that focus on experimentally disprovable propositions. According to one observer’s characterization, the unconscious that neuroscientists describe is “relatively passive, like computer software” (Dailey, 2017, p. 231), as patterns of neurons fire to cause automatic associations. In contrast, Freud’s unconscious is wild, colorful, and free, engaged in a dynamic process of creation as the mind makes imaginative associations akin to artistic work. It may seem somewhat disheartening to move from literature and imagination to science and its tightly controlled disprovable hypotheses, as I will discuss later in the article. Nonetheless, in contemporary times, natural science-based approaches have become predominant in many of the psychological, philosophical, and social studies disciplines that influence law.

In one key respect, however, neuroscience and Freudian psychology share a fundamental insight highly relevant to law. That insight concerns the fact that much of human action originates in the unconscious. Freud posited this idea in his concept of the “id.” Neuroscientists have proved the existence of the unconscious, though the neuroscientists’ unconscious has, as already noted, a very different character than Freud’s. Neuroscience and related disciplines by this point have established beyond doubt that a great deal of what humans (and other creatures) are able to perceive, process in their brains, and then respond to comes into the brain and then motivates conduct without consciousness at any level. The way the eye perceives its environment is one good example, as popular neuroscience writer Leonard Mlodinow explains.

Unconscious processes organize the unmanageably complex data we receive through our senses, and especially through our eyes, in order to even begin to make our surroundings intelligible (Mlodinow, 2013, p. 35). The brain does this so quickly and automatically that we are not aware of all the processing that takes place before we can start to make sense of our environment in both its physical and social aspects.

Social neuroscientists have established that similar mechanisms account for much of human social behavior. The human brain perceives and responds to its social environment through mechanisms that often work prior to human consciousness or volition. Humans' judgments about what is safe and what is dangerous, and what is pleasant, worthwhile, and valuable and what—or who—is not, all occur with split-second timing, relying on so-called “fast” thinking processes (Kahneman, 2013). Without such “fast thinking,” we could not operate in the world; it would take far too long to deliberate about every perception and reaction.

At the same time, fast, non-deliberative thinking carries with it many irrationalities that introduce non-volitional biases of many kinds into human thinking and conduct (Kahneman, 2013). Humans exhibit these biases in the form of many irrational predilections. These include a preference for material goods they already have versus goods of the same value they could acquire, an “optimism” bias that anticipates that situations will work out better than they are actually likely to, and a bias toward thinking more highly of one's personal abilities than an objective assessment would indicate. Another is the brain's common practice of inventing false memories to fill in gaps in information that one has forgotten (Kahneman, 2013). Daniel Kahneman and Amos Tversky, as well as many other behavioral economists and experimental psychologists, have established the existence not only of these, but many other types of bias in what they call “fast,” or quick, intuitive thinking. Further research will undoubtedly lead to many more discoveries, but the basic proposition—that much of what the human brain does in the social realm is completely unconscious and often irrational—is beyond contestation among brain scientists today.

Such scientifically tested evidence, based on verifiable facts, may make especially useful contributions to law, given that law is also a discipline concerned with finding factual truth (albeit in quite a different way from science) (Haack, 2014). Neuroscience's emphasis on provable facts—or, better put, on disprovable hypotheses—would seem well suited to law's evidentiary focus on proving facts at specified standards of proof (such as more likely than not (50%+) or beyond a reasonable doubt (75%–95%)).

(Haack, 2014, p. 32–33)). Thus, I will argue below, neuroscientific evidence can prove quite useful in anti-discrimination cases. An even more high value use may be in guiding the reform of law-related policy, rules, structures and systems more generally. For social neuroscience to contribute to such system-wide reform, however, more than the mere weight of scientific authority will be required; social movement actors will have to embrace this project and infuse it with values and rhetoric in order to persuade in the political sphere. Thus, no matter how good the science becomes, its use in law as a general rather than case-specific matter will require advocates to move beyond the scientific realm and back into the humanistic territory of values, tropes, politics, and culture.

One of the ways in which social neuroscience can contribute to anti-discrimination law is in the field of implicit bias research. Along with biases in the way that human brains think about statistical prediction, material objects, the self, the past, and a host of other matters, researchers have proven the existence of inherent and inescapable biases in the way humans perceive and react to human difference (Cunningham, Haas, & Jahn, 2011; Greenwald, McGhee, & Schwartz, 1998). Put most bluntly, researchers have repeatedly found that humans can have uncontrollable aversions to those they perceive as different from their own groups and/or “in” groups with high social status (Haidt, 2013). Even fellow members of socially defined “out” groups, or groups that are socially constructed to have low status, tend to have these biases; members of low-status groups may internalize the negative views that groups with privilege and power have constructed of them (Morrison, Decety, & Molenberghs, 2012).

These biases have been shown to be extremely hard to extinguish, even with great effort at control. Some measures of bias, such as the often-used Stroop test, measure just this effort at control, which researchers theorize causes a slowdown in subjects’ abilities to engage in competing mental tasks (Cunningham et al., 2011, p. 220). Even with anti-bias training, persons continue to exhibit implicit biases, though they may be somewhat less likely to act on them, as discussed further below.

Biases can be trivial—for example, researchers have proved that human subjects prefer others who are defined as being on the same “team” based on what shirt they are assigned to wear (Molenberghs, 2013, p. 1532). Biases can be problematic but functional in some ways; for example, having strong in-group identification helps work teams work together especially well (Haidt, 2013, p. 275), but at the same time can create problems for managers in integrating those teams into a larger organization (Mlodinow, 2013, p. 172). Biases that are individually held can be pernicious but not combine

to produce structural subordination. At the other extreme, biases can be outright deadly, as documented in the many horrific genocides that have occurred in modern history all around the world (Sapolsky, 2017, pp. 570–79). Illegal discrimination results partly from biases that are pernicious and structurally embedded in social institutions, practices, and the distribution of key resources.

Social neuroscience literature shows a strong connection between implicit bias and the dynamics that cause people to conform to the norms of the groups in which they belong. At the same time that most people generally seek to conform to group norms, they also judge harshly both in-group nonconformists and outsiders. One well known experiment, dubbed the ultimatum game, illustrates these dual phenomena. This experiment involves giving one player, called the proposer, a sum of money, and then asking the proposer to offer to divide that sum between herself and another player, called the responder (Mendoza, Lane, & Amodio, 2014). The responder then chooses whether to accept or reject the proposal. If the responder accepts, the players split the money according to the proposer's offer. If the responder rejects the proposal, neither player receives any money. Experiments repeatedly find that proposers reward more money to responders who are arbitrarily designated as fellow in-group members than to responders who are identified as out-group members—even when an option is to award the money so that both groups end up with more (Mendoza et al., 2014, p. 663). In other words, humans' focus on in-group likeness is so strong that it outweighs the classic economic assumption that individuals act to maximize self-interest. And people punish outsiders even when there is no rational reason to do so.

Experiments involving political affiliations have similarly shown the effects of group identity on human judgment. These experiments have shown that people even perceive *facts* about the actions of political in-group members differently than facts about out-group members. In one experiment, researchers showed a video recording of a controversial, roughly played football game between Princeton and Dartmouth Universities to students from the two schools (Hastorf & Cantril, 1954). Students from Princeton counted Dartmouth players as having committed more fouls, whereas students from Dartmouth viewed the number of fouls committed by the two teams as equal. In effect, the researchers explained, the two groups of students had viewed “a totally different game” (pp. 129–30). Another experiment tested how subjects viewed video clips of anti-abortion protests at clinics (Kahan, Hoffman, Braman, Evans, & Rachlinski, 2012). This experiment found that subjects holding anti-abortion views saw protestors

commit fewer illegal acts than did subjects who favored access to abortion (Kahan et al., 2012, p. 884). These effects occur even when groups are newly created. Experimenters randomly divided participants into a “red” group and a “blue” group and then asked them to watch video clips to determine which team’s members pushed a button faster (Molenberghs, 2013, p. 1523). The experimenters had constructed the video clips so that the two groups were exactly equivalent on this measure. Each team, however, judged members of their own team as faster. These results are consistent with still another experiment that found that people are more willing to donate money to fellow in-group members as opposed to members of outsider groups who are in distress (Molenberghs, 2013, p. 1533).

Researchers in the fields of experimental psychology and neuroeconomics have also found, somewhat surprisingly, that human subjects make decisions more rapidly when they are making altruistic decisions rather than selfish ones. This suggests that altruism toward others in one’s group, rather than selfishness, is the more powerful automatic or intuitive response. As neuropsychologist Jonathan Haidt puts it, the evidence shows that we are “groupish,” rather than selfish, in our moral reasoning. Haidt argues that humans’ “hivish” moral nature makes it easier for us to work in groups and organizations (Haidt, 2013, pp. 258–84). By activating “pride, loyalty, and enthusiasm” among employees, for example, managers reduce the need for costly employee monitoring. Group-serving moral impulses and bonds of trust help cooperative endeavors succeed (Haidt, 2013, p. 275).

Researchers further theorize that disfavoring outsiders contributes to how human beings form positive self-identity and self-esteem (Molenberghs, 2013, p. 1532; Morrison et al., 2012, p. 2115; Abrams & Hogg, 1990, p. 196). Social identity theory proposes that positive self-development involves individuals’ positive identification with the groups to which they belong (Abrams & Hogg, 1990). In other words, persons identify with models of in-group behavior they view as admirable and, without conscious volition or even awareness, define outsiders as inferior.

Where groups are not interconnected in ways that reflect disparate power, privilege and access to resources, these dynamics may in essence cancel each other out as different groups view themselves as better than others. At least, in egalitarian conditions, in-group preferences do not become ossified in structures and institutions. But where insider groups possess disparate power, the mechanisms of implicit in-group preference and out-group bias produce unlawful discrimination. Writ large, implicit bias becomes discrimination.

Illegal discrimination against protected classes falls in this category of socially shared, and thus compounded, implicit biases. In modern, highly complex and diverse societies, implicit biases cause great harm. In a modern world, illegal discrimination generally involves irrational and counterproductive negative treatment of out-group members. Such discrimination introduces inefficiencies into economic relations and causes enormous psychological, economic, physical, and medical harm to those subject to it (Pager & Shepherd, 2008; Stone, 2012). Discrimination may lead to interpersonal and society-wide violence, loss of human creative and productive potential, physical and psychological harms perpetuated across generations through epigenetics and other mechanisms, as well as animosity, resentment, backlash, political and social conflict and much more. Demagogues exploit implicit biases and nations fall into disrepute because of them. Thus universal human rights declarations condemn discrimination against protected classes and all modern societies prohibit it (see [Universal Declaration of Human Rights, 2020](#), Art. 2). The question thus arises: how might neuroscience help law eradicate, or at least curtail, unlawful discrimination?



2. Using social neuroscience in the law to counter implicit bias

There are many potential uses of neuroscience research in anti-discrimination law. I here sketch just a few directions that I see as useful from the perspective of my own disciplinary background in anti-discrimination law, history and theory. In this short treatment, I will discuss four broad categories of uses that present special promise. These include: (1) abolishing or minimizing reliance on intent standards for proving unlawful discrimination; (2) fashioning more effective remedies for unlawful discrimination; (3) expanding doctrines about what counts as unlawful discrimination, based on new understandings of the harms discriminatory treatment causes; and (4) supporting a social movement to expand human rights law to recognize a general right to “act differently,” within the bounds of not harming others.

2.1 Minimizing the use of intent standards to prove unlawful discrimination

Many countries have moved beyond, or at least fine-tuned, their intent standards for proving discrimination. Canada, for example, weighs the harm caused by an allegedly discriminatory practice against the strength of the rationale for using it (Seiner, 2006, p. 119). Other countries, including the United States, still usually require a showing of intent to make out

disparate treatment-based discrimination in key areas of social resource allocation such as housing and employment. Many legal systems, including the United States, have non-intent based options for proving discrimination, such as “disparate impact” or “effects” standards that call for a showing that a challenged practice has unjustifiably harsh effects on persons in protected categories (Carle, 2017). These effects tests, however, typically require complainants to surmount high hurdles of proof in various respects. Thus, intent-based standards remain the predominant way of proving unlawful discrimination in many less progressive national systems of anti-discrimination law, including the United States.

The findings of social neuroscience push toward the elimination or modification of intent standards for proving discrimination. Social neuroscience shows that decisions based on implicit bias typically occur without conscious volition. It makes little logical sense to require a showing of malevolent intent to discriminate when science shows that decision makers most often do not act with malice. Instead, their acts take place at an unconscious level without any awareness of discriminatory intent, much less any volition or desire to act discriminatorily.

As already noted, some legal systems have blended intent and impact standards into hybrid approaches that assess the burden of a challenged practice in relation its benefits (Seiner, 2006). Under a science-based view, these approaches make far more sense because implicit bias research casts doubt on the validity of accusing most actors who have been alleged to have engaged in discrimination of having “intended” to discriminate.

In addition, intent requirements make it hard to prove discrimination because it is usually very difficult to prove what has gone on inside a respondent’s head (if an individual) or a collective decision making process (if a collective decision). Absent cases with so-called “smoking gun” evidence, where a respondent decision maker has openly admitted to prejudiced motives, it is very hard to establish what the decision maker’s thought process was. Since the burden of proof in such cases remains with the complainant, the complainant typically ends up losing because the facts simply do not establish what the respondent’s true motives were.

Under the lessons that can be drawn from implicit bias research, questions about the respondent’s conscious intent should be eliminated. Instead, statutory anti-discrimination schemes should require complainants to point to harm to a person or persons in a protected class that could be avoided by modifying a nonessential business practice or structure or by introducing practices that work to oppose unconscious bias, such as

training, priming at key decision points, and the like. These approaches would do far more to eradicate discrimination that current legal rules ignore, while at the same time eliminating the blame and finger-pointing that often is not justified given the non-volitional nature of much bias. Such reform would produce better social results along with less social friction and backlash against the important project of extinguishing discrimination's harm.

Eliminating or downplaying the intent standard in anti-discrimination law would also make it easier to settle or resolve cases. As noted, no one wants to admit to intentional discrimination because that is an admission of a moral wrong. Moreover, since, as neuroscience shows, many actors whose implicit bias may have led them to take discriminatory actions will be totally unaware of having committed wrongful acts, they will, quite understandably, resist demands that they admit to doing so. Accusing someone of malevolent "intent" by accusing them of intentional discrimination does not advance the ball in typical cases. Acting in good faith is, after all, how respondents typically experience their actions, so they are likely to insist on being vindicated rather agreeing to engage in problem-solving approaches. Even if in fact non-volitional implicit bias may also have been at play—and probably often is given that all persons harbor implicit biases—non-accusatory problem-solving modes of approaching discrimination problems often may be far more effective in the longer run.

In short, a standard based on the effects of decisions on members of protected categories burdened with group histories of structural and other forms of discrimination could make it easier to recruit those with power—the ones who are making discriminatory decisions—to join the team working toward solutions rather than remaining defensive and resistant members of the opposition. Institutional leaders may find it far more palatable to be recruited to help dismantle practices or systems that produce discrimination rather than to be called upon to defend actions they experience as having been made in good faith. It is, after all, structures and history that produce implicit bias. An important step in eradicating implicit bias requires examining and reforming the institutions, ideas, and structures that encode certain types of prejudice against others into human thinking.

Of course, sometimes social justice activists very much want to prove the moral wrong of intentional, conscious and blatant animus against person in protected categories. I have documented these impulses in the face of blatant, open racism in studying the history of the development of U.S. federal anti-discrimination law. In early stages of that particular history, more radical racial justice advocates wanted to pursue intent-based discrimination

standards, while more moderate, sociologically inclined reformers wanted to engage in voluntarist, effects-focused, problem-solving approaches (Carle, 2011). These camps only later switched sides, so that today in the United States it is usually the most left-leaning anti-discrimination advocates who push for effects-based or disparate impact theories while more moderate and conservative jurists defend intent-based standards as all that should exist under US law (see, e.g., Ricci v. DeStefano, 2009). But to be sure, situations in which conscious and purposeful bad intent does exist certainly are not gone from the face of the discrimination landscape, and these are important to address through law as well. They can best be addressed, however, by including a special category of particularly bad discrimination based on a showing of open and conscious hostility and animus. This category of truly bad actors, who have admitted or displayed particularly malevolent and egregious prejudice, could be subject to higher damages or other penalties, such as extra fines and/or punitive damages. The problem with present standards is that the on/off switch of liability and duty to rectify harm gets triggered only if “intent” is shown, leaving many decisions and actions based on implicit bias without remedy and thus essentially unaddressed by law. In conclusion, consistent with neuro-scientific findings that bias is often unconscious and non-volitional, anti-discrimination law could:

- Lower the moral opprobrium attached to garden-variety discriminatory actions;
- Continue to provide and even enhance legal remedies for discrimination;
- Focus anti-discrimination law on dismantling discriminatory institutions and systems, including institutions of unconscious prejudice encoded in actors’ brains;
- Save punitive sanctions for egregious cases of blatant prejudice.

These approaches should lead to more successful individual cases, more progress in dismantling discrimination, and more social benefit for all the reasons discrimination constitutes a social harm, as already discussed above.

2.2 Fashioning remedies for discrimination

Social neuroscientists have not only studied how implicit bias occurs but also what can prevent actors with power over others from acting on their implicit biases. Unfortunately, that research shows that it is extremely difficult to exterminate implicit bias (Lai et al., 2014, pp. 1780–82). In comparing 17 studies of interventions to lessen implicit bias, Lia and his colleagues found that the majority of those interventions were

completely ineffective, and the ones that were somewhat effective were only weakly so. However, other studies have shown that sometimes professionals can be trained not to act on their implicit biases even though those biases remain. For example, research shows that police officers can be trained to avoid pulling the trigger due to implicit racial bias even though the same training of lay people cannot produce this result (Correll et al., 2007).

Another important finding is that implicit bias can most effectively be reduced by offering trainings and reminders close to the point of an actor's decision making or other potentially discriminatory action. For example, researchers have shown that doctors are most likely to refrain from manifesting implicit bias in their decisions about dispensing pain medications if they receive a reminder about the possible effects of implicit bias right before or at the time they are making prescription decisions (Green et al., 2007)

These observed effects involve the phenomenon of priming. Priming refers to "the effect in which recent experience of a stimulus facilitates or inhibits later processing of the same or a similar stimulus." (American Psychology Association Dictionary, "priming"). Such priming, or presenting a reminder that taps into a person's "fast" social thinking, can work both ways. Priming can increase bias, of course; for example, young adults asked to create sentences from the words "Florida, forgetful, bald, grey or wrinkle," later walk down a hallway more slowly than members of a control group (Kahneman, 2013, pp. 53–54). But priming can also decrease bias if constructed in artful ways. Thus, a reminder to the decision maker of that actor's desire to avoid bias can help produce less biased results if effectively timed so that it happens right before or at the point of decision, as in the example just given above about doctors prescribing pain medication.

These insights can also be helpful by indicating new ways of formulating effective remedies in anti-discrimination cases. For example, an employer found liable for discrimination in hiring or promotion decisions could be required to hold implicit bias training sessions for decision makers and to set up subsequent reminder systems that would activate soon before or at important junctures in hiring or promotion processes. Such an employer could also be asked to audit its systems and processes to determine where it was using practices that were not necessary and contributed to disparate impact harms. Employers could be required to produce self-audits and reports that measure results in eliminating barriers and bias. Employers might often find that such steps introduced greater rationality and productivity in their systems in general, leading to benefits for all (Guinier & Strum, 2019).

In summary, the research on the limited but partial effectiveness of trainings and similar techniques to reduce implicit bias suggests that courts should think more creatively about using these remedial measures in anti-discrimination cases. Along with remedies to make those subjected to discrimination whole to the extent possible in law (such as through restitution, back and front pay, reinstatement, and similar measures in employment contexts), lawyers should consider asking for, and judges should consider ordering, measures such as implicit bias training. Such training should be followed by reminder modules or other systems that prime decision makers to be conscious of possible tendencies toward implicit bias, timed to occur right before or at the moment of decision making. Other important processes the literature on implicit bias points to as necessary to produce less biased institutions and systems involve the institution of procedures to review, double check and audit aspects of institutional functioning that are most likely to be sites where either implicit bias and/or effects-based discrimination may occur. Many creative techniques can be envisioned to detect and counter fast thinking biases of many types as well as effects-based discrimination. However, introducing those ideas into the judicial system will take much work, and may even call for a “consciousness revolution” in law, as I will discuss further below.

Another way in which social neuroscience can help in fashioning anti-discrimination remedies is in illuminating more about the harm discrimination causes. Brain imaging studies have shown that individuals who are ostracized from social groups experience intense pain (Borsook & MacDonald, 2013; Pond, Richman, Chester, & DeWall, 2014). As Pond and his fellow researchers have shown, even minor incidents, such as being left out of a social game being played on a computer, produce these results. Long-term exclusion has far more profound consequences, including intense physical and psychological harm that can produce many anti-social effects as well as destroying individuals’ self-esteem (Leary, Kowalski, Smith, & Phillips, 2003; Williams, 2007).

Another large medical and epidemiological literature documents the connections between racism and adverse health consequences, including depression, anxiety, psychological stress, cardiac disease, and hypertension (Paradies et al., 2015). New discoveries in epigenetics show intergenerational transfer of adverse consequences from such trauma (Kuzawa & Sweet, 2009). This research shows that the social harm of discrimination is far greater than legal systems currently recognize. Courts should admit this research into evidence to allow complainants to prove higher damages in

anti-discrimination cases. One danger here, however, is that courts may regress to requiring such showings of acute pain in order to grant damages for pain and suffering at all. They should not do so: the research shows that all persons subject to sustained discrimination experience social pain, so that pain and suffering awards should be standard in all cases. Again, as I will discuss below, this research should not only be used in individual legal cases but also to produce systemic change in legal rules and systems.

Similarly, medical and epigenetic studies reveal that imposing restitution in individual cases does not go nearly far enough in forcing respondents to “internalize” the costs (as an economist would say) of discriminatory systems. In order to use legal systems as one helpful tool (not the only one, of course) in the project of eradicating discrimination, modifications in legal rules must be made to more fully capture the nature of discrimination’s harm. Law has had an important role to play in improving social consciousness of this harm, but at this historical juncture it may be time for social neuroscience to move into the lead in guiding law further along this path. Discrimination not only causes loss of individual resources such as housing, jobs, medical care, education and the like, but also causes additional social harms that are not sufficiently accounted for in current remedial doctrines.

2.3 Broadening doctrines that cramp the application of anti-discrimination law

Another set of ways that social neuroscience can help in the field of anti-discrimination law involves adjusting other doctrines that also tend to constrict the ability of complainants to prove their cases. Courts often impose these limiting rules as a way of clearing anti-discrimination cases off their dockets, but the findings of social neuroscience counsel again application of such restrictive doctrines. Doing so ignores what brain science shows about how discrimination occurs.

Anti-discrimination cases have been subject to a great many, often irrational, limiting doctrines. These problems extend to far more matters than I can discuss here. A few examples, based on my own legal system in the United States, can highlight some of these problems. A more thorough discussion would require anti-discrimination lawyers in various legal systems to compare examples, a project I hope might happen in a future, as yet unplanned, conference. In the United States, anti-discrimination doctrine requires an employee to experience a “material” harm, such as a loss of pay, promotion, or job altogether. This material harm is what triggers an actionable employment discrimination claim (Modjeska, 2017, § 1:2,

pp. 1–4). This doctrine arises through courts' interpretation of statutory law. It is the existence of a material harm, also known as a “tangible” action, that establishes that the employee has experienced discrimination in the “terms and conditions of employment,” as provided in the statutory language of federal anti-discrimination laws such as Title VII of the Civil Rights Act of 1964 ([Civil Rights Act, 1964](#)).

Although this doctrine correctly states the statutory requirement, courts have sometimes applied the material harm test too restrictively, dismissing claims of discrimination based on exclusion from informal work groups, trainings, social opportunities, or the assignment of workers to less attractive work within a job classification ([Carle, 2019](#), p. 711). Where excluding persons from informal opportunities and/or assigning them less desirable work gets in the way of their job success, however, material harm *has* occurred. Better understanding the relationships among in-group bias, social exclusion, and discrimination, as discussed above, could make courts more attuned to how unlawful workplace discrimination takes place. Mere trivial complaints should not make out an actionable claim, but courts should find actionable claims based on facts showing long-term, repeated and persistent exclusion and social shunning, where complainants establish that these practices are linked to their membership in traditional outsider identity groups and have adverse implications for job success. Moreover, harmful impacts on complainants' psychological and physical well-being, as already discussed above, establish the damages arising from such practices.

Another example involves anti-retaliation law. Most anti-discrimination laws prohibit retaliation against persons who raise anti-discrimination claims. In the United States, courts are somewhat more permissive about the rules for showing retaliation, but still impose onerous restrictions as to the acceptable *manner* of employees' conduct in opposing discrimination ([Carle, 2019](#), pp. 215–17; [Carle, 2016](#), p. 190). Again, social neuroscience findings can be helpful here.

An interesting experimental social psychology literature documents a phenomenon researchers have labeled the “black sheep” effect. A series of experiments conducted by Jose M. Marques and Vincent Y. Yzerbyt found that, in comparison to a control group, in-group members evaluated in-group members who deviate from group norms or standards far less favorably than they evaluated out-group members who deviated in the same way ([Marques & Yzerbyt, 1988](#)). Thus, they found, group members' judgments about fellow in-group members were more extreme, in both directions, than their judgments about out-group members. If group members

perceived fellow in-group members as conforming to the group's norms and standards, they evaluated them as better than out-group members, but if they deviated from the group's norms and standards, fellow in-group members judged them more harshly. Marques and Yzerbyt suggest that this so-called black sheep effect may reflect a process by which group members define and esteem "good" exemplars of their group and at the same time strongly reject "bad" ones, because their deviance from the group's values damages the self-esteem that comes with positive self-identification with one's identity group (pp. 289–91).

The black sheep effect, through which groups tend to be particularly harsh toward internal dissenters (or "deviants" from social norms) (Marques & Yzerbyt, 1988), has special relevance in the anti-retaliation context. So-called "whistleblowers"—i.e., those who point out institutional wrongdoing, whether because an institution has engaged in discrimination or in other types of misconduct—typically are insider members of a team. Members of such groups, bound together by codes of secrecy and loyalty, find the presence of whistleblowers among them particularly repugnant and thus are extremely prone to retaliate (Vaughn, 2012, pp. 63–65). For this reason, as courts recognize to some extent but still not sufficiently well, claims that employers or other institutional actors have retaliated against someone who has "blown the whistle" on their misconduct must be treated with special care.

In other words, as social neuroscience findings regarding the black sheep effect attest, retaliation against those who accuse an employer or other institutional actor of committing illegal acts or moral wrongs is particularly likely, even by otherwise lawful actors. To encourage and protect employees who speak out against perceived wrongdoing, courts should err on the side of providing more generous protections against retaliation.

Indeed, a consilience, meaning an "agreement between the approaches to a topic of different academic subjects, especially science and the humanities" (Oxford English Dictionary, 2020), has emerged between the empirical findings of social psychologists and other science-based researchers, on the one hand, and anti-discrimination scholars and other civil rights policy advocates, on the other. This consilience between neuroscience and progressive social policy discourse pushes toward greater protection for workplace dissenters as well as those who act differently in other scenarios. As biologist, writer, and Nobel Laureate Edmund O. Wilson has pointed out, a general consilience has emerged between the social and biological sciences on the general matter of protecting diversity of all

types in both the human and the environmental realms (Wilson, 1998). Similarly, both social neuroscience and recent legal theory increasingly converge on propositions counseling in favor of greater protection of a right to act differently as a more general principle as well.

2.4 Expanding rights to act differently

Perhaps most importantly, social neuroscience points to the need for an expansion of human rights to include the right of human beings to act differently, so long as their actions do not impinge on the rights of others. This is a longstanding principle of liberal philosophers such as John Stuart Mill, who famously argued in *On Liberty* that individuals should be permitted to do whatever they want as long as they do not harm others in doing so (Mill, Bromwich, Kateb, & Elshtain, 2003, p. 139). This idea coincides with the far more recent literature on neurodiversity in the disability rights context, which points to the benefits to society derived from the vastly different ways in which human brains are “wired” to exhibit different strengths, talents, and ways of thinking (Armstrong, 2011). Similarly, social neuroscience and related social psychology literatures are replete with studies demonstrating the benefits of group processes that value diversity of all kinds. Such appreciation for diversity in all manner of human differences—including background, gender, skills, and other aspects of social identity—can lead to better group dynamics, outcomes, and even more accurate factual findings.

Experiments show that diversity of thinking in groups improves the accuracy of collective outcomes, although only if the participants are each required to think independently. If they remain subject to group influences, or “group think,” those influences may cause individuals to follow the crowd and agree to suboptimal outcomes, even outcomes that are objectively, or factually, wrong (Berns, 2010; Phillips, 2014; Surowiecki, 2005; Treynor, 1987).

A host of often-replicated classic experiments, conducted by researchers such as Stanley Milgram, Muzafer Sheriff and Solomon Ash, have proved the powerful social psychological effects of humans’ desire to conform. Thus, in one classic example from the 1950s experimental psychologist Solomon Asch found that when experimental subjects were placed in a group in which they first heard other participants (who were in fact confederates of the experimenter) give wrong answers to a factual question as to whether the length of two lines was the same, these subjects were far more likely to

give wrong answers than those in a control group who did not hear others giving wrong answers (Berns, 2010, pp. 89–91). On the other hand, when each member of a group was asked to give privately their own estimates of the number of candies in a jar, and the researcher then took the average of all of these individual responses, this aggregated estimate proved better than even the best estimates of the individual estimators (Berns, 2010, p. 91).

This experimental psychology literature on the benefits of diversity and the downsides of group conformity coincides with the appreciation of the need for diversity in the biological sciences (Wilson, 2012, pp. 80–81). As Wilson, who is by no means a leftist, writes

[p]erhaps the time has come ... to adopt a new ethic of racial and hereditary variation It would give proper measure to our species' genetic variation as an asset, prized for the adaptability it provides all of us during an increasingly uncertain future. Humanity is strengthened by a broad portfolio of genes that can generate new talents, additional resistance to diseases, and perhaps even new ways of seeing reality. For scientific as well as for moral reasons, we should learn to promote human biological diversity for its own sake instead of using it to justify prejudice and conflict.

Wilson (2012, p. 80–81).

In short, a consilience has arisen among the political theory literature of classic liberal philosophers, the biological sciences, and social neuroscience. This consilience provides another boost to arguments supporting the promotion of diversity writ large. Indeed, it is somewhat ironic that, while a subject of great debate among social commentators today, the benefits of diversity are largely noncontroversial to a host of researchers based in a wide array of natural science-based disciplines.

Legal anti-discrimination theorists are similarly pursuing lines of inquiry that support expanding and generalizing a broader right to act differently. Here I will provide a brief summary, drawing primary from the US legal theory I know best. In recent decades, legal scholars who have focused in this direction include Kenji Yoshino, who introduced the term “covering” to describe how discrimination can work by requiring persons to hide aspects of their identity in order to be socially accepted (Yoshino, 2007). In other words, he argues, in contemporary cosmopolitan social conditions, more privileged outsiders may be “included, but only if we behave like insiders—that is, only if we cover” (p. 22). Yoshino inquires why, half a century after US law imposed broad anti-discrimination mandates, so many people still feel the need to disguise their “authentic selves” in this way (Yoshino, 2007, p. 184). In personally reflective sections of his

book, Yoshino focuses on two aspects of his own identity, as a gay man and as a person of Japanese descent who was raised in the United States with an ambivalent relationship to his ethnic heritage. Interweaving personal narrative with legal and theoretical analysis, including comparative analysis of German law, Yoshino argues for an anti-discrimination paradigm that would grant greater protection to what he calls, drawing on his translation of the German phrase, a “right to personality” (p. 189). Yoshino’s vision for the future has won many dedicated followers. A recent contribution that builds from Yoshino is Zachery Kramer’s book, appropriately titled *Outsiders*. Using engaging examples, Kramer similarly argues for a discrimination-based right to expression of one’s personality (Kramer, 2019, pp. 4–5).

Still other leading anti-discrimination scholars turn to history for instances in which law expanded (not necessarily but contingently; there is nothing inevitable about this) to make more room for supporting human difference. In her award-winning book, *Vagrant Nation*, Dean Risa Goluboff traces a decades’ long campaign to strike down so-called “vagrancy laws,” which criminalized those socially constructed as having the status of so-called deviants. Goluboff traces the origins of anti-vagrancy laws to 16th Century English concepts of everyone having a proper place; those lacking social power who threatened to move out of their proper place faced prosecution for no other reason than this, whether they were “out of place socially, culturally, politically, racially, sexually, economically or spatially” (pp. 1–2). The people against whom local law enforcement applied these laws in the 20th Century United States included poor and homeless persons—and especially Black persons, even when employed—as well as non-geographically attached persons encompassed under the traditional image of the “hobo.” Law enforcement targeted a wide range of others under vagrancy laws because they did not conform to social norms. As Goluboff puts it, “the ‘queer,’ the ‘Commie,’ the ‘uppity’ black man, the ‘scruffy’ young white one,” all embodied difference; the police and others who enforced law were “trained to see difference as dangerous, to see the unusual as criminal” (p. 3).

Goluboff thus focuses on the commonality, in the form of shared persecution through vagrancy laws, underlying various forms of deviance. Vagrancy law bound together political dissenters, including war protestors, communists, and irascible political contrarians, all of whom were prosecuted under their authority, as well as those who dissented from the racial caste system: “[If you are f]or integration[,] [y]ou’re a Vagrant” (p. 123). Police applied vagrancy laws to dignified African American ministers taking part in civil rights protests, mixed-race groups in the South, and persons in

the "wrong" racial neighborhoods all over the country. Sexual minorities also came under these laws' reach, in prosecutions of so-called "vag" lewd charges against gay men and arrests of persons of both sexes for dressing in the clothes traditional for the opposite sex. Vagrancy laws were applied to women having sex outside marriage and persons committing the then-taboo act of having sex across racial lines. Police defined hippies, beatniks, and other members of the counterculture of the 1950s and 1960s as dangerous and "vagged" them because they violated conventional norms about styles of dress, hair, and lifestyle (pp. 53–55, 123, 295).

Goluboff does not focus on the application of vagrancy and similar laws to persons with disabilities, but another book fills in that important gap. In *The Ugly Laws*, Susan Schweik documents how civic leaders used vagrancy and other laws to banish from public spaces persons with disabilities others viewed as unsightly. As in Goluboff's narrative, [Schweik \(2009\)](#) shows how persons with power used laws to exclude and penalize persons regarded as repugnant due to their perceived differences. Both books trace various strains of the complex, decades-long coalition-building, and, as it happened, eventually successful activism to abolish vagrancy statutes and ugly laws. These initiatives contribute to the history underlying broader recognition of a general right to act differently.

Recognizing such a more general right, however, will require somewhat different strategies than the ones Goluboff and Schweik document, for a number of reasons. First, there is an inherent paradox in mandating tolerance (in order words, protecting persons' rights to act differently), because any rule that requires tolerance tends itself to be intolerant as enforced. As Yoshino notes, not all anti-discrimination and fairness goals can be accomplished directly through so-called "hard" law, meaning, basically, law that can be enforced in court ([Yoshino, 2007](#), p. 192). Thus, best strategies for expanding general rights to act differently may involve using neuroscience findings to persuade through "soft" law. By soft law, I mean such quasi-legal techniques, now a subject of great interest by legal scholars, as voluntary standards, best practices guidance, social boycott (and reward) campaigns, and the like (see, e.g., [Bradlow & Hunter, 2019](#); [Levin & Mather, 2018](#)). These strategies use social pressure—which, as previously discussed, social neuroscience shows to be very effective in producing conformity in groups (see, e.g., [Berns, 2010](#), p. 246; [Blass, 2000](#)). Behavioral economics and experimental social psychology may be helpful not only in expanding protections in the field of anti-discrimination law, but also in informing the strategies that might best help us get there.



3. Regulating the use of neuroscience in anti-discrimination law

There are two ways neuroscience might be used in civil anti-discrimination cases. (I do not here consider the use of neuroscience in criminal cases, though a fascinating, and burgeoning, literature exists on that topic (see, e.g., [Hughes, 2011](#); [Larson, 2016](#); [Winslade, 2002](#)). One is in particular anti-discrimination cases and the other is in law and policy reform on a broader, possibly even system-wide, scale.

3.1 Using neuroscience in particular cases

Using expert evidence and peer reviewed neuroscience studies would not be at all unusual in anti-discrimination litigation. Litigators in anti-discrimination cases routinely use many types of science-based evidence, including introducing into evidence statistical findings by industrial psychologists, as well as actuarial data, mathematical economists' modeling results, demographics of labor and applicant pools, and much more. Science-based studies and testimony by experts about such studies frequently come into cases to help prove both disparate impact and disparate treatment discrimination, as well as to prove causation, general and specific damages, physical and psychological harm, pain and suffering, and the like.

There thus should be nothing special about using neuroscience in particular lawsuits in the same way that one would use any of these other expert evidence based on findings produced through accepted scientific methods. The findings of neuroscience are well respected as scientific evidence, and expert scientific evidence stemming from neuroscience should readily achieve admission into evidence in individual cases, with some ethics caveats that I will discuss in a later section of this article.

Legal systems have evolving standards (though often comparatively different ones ([Taylor, 1996](#))), for evaluating expert testimony. In the United States, the standards for the admissibility of expert testimony are articulated in a case called [Daubert v. Merrell Dow Pharmaceuticals, Inc. \(1993\)](#). *Daubert* requires courts to assess five factors in deciding on the admissibility of scientific evidence. Those factors include:

1. Whether the claim or theory has been tested;
2. Whether the research has been subject to sufficient peer review and publication;

3. Its known or potential error rate and the existence and maintenance of standards controlling its operation; and
4. Whether the findings are widely accepted in a respected scientific community.

The use of social neuroscience can be regulated under these same standards. Implicit bias research has clearly been accepted by a respected scientific community, making it admissible as expert testimony in individual cases so long as the particular studies being relied on meet the requirements listed above or comparable ones as framed in the legal jurisdiction in question.

3.2 Using neuroscience to inform policy and reform anti-discrimination law

A second way of using social neuroscience in anti-discrimination law would be to draw from it to inform policy and law making on a system-wide basis. This is the spirit in which I have suggested using social neuroscience to reform anti-discrimination law through the specific proposals I have sketched above. To achieve system-wide reforms, it is not enough simply to introduce neuroscientists' findings in particular lawsuits. If the structure or set of rules through which those cases are processed does not change in response to the policy implications of what neuroscience shows, then there will be no broad, lasting change in the law in the directions I suggest. For example, if the legal rule is that intent must be shown in order to prove a case of disparate treatment, there will be no finding of a legal violation even if implicit—i.e., unconscious, non-volitional—bias can be established through the use of social neuroscience testing methods such as the Stroop test. The overall set of rules, doctrines and policies in the law need to be reformed in order for the findings of neuroscience to have maximum impact. The question thus becomes, can we expect neuroscience to have an impact on law in this system-wide way?

Those who are not law trained may have an optimistic hope that the powerful facts about human social behavior that neuroscience establishes, including but not limited to those covered in my brief overview above, will, by the sheer force of their logic and the reliability (albeit partial and temporary (Kuhn & Hacking, 2012)) of scientific evidence, lead to change in law. As a scholar of legal change, however, I very much doubt this can be the case. Facts and logic typically do not change law; social movements do (see, e.g., Brown-Nagin, 2012). Historical experience and politics also factor into change in law, as the Legal Realists taught so many years ago.

Thus, I will propose: social neuroscience needs to operate as something more than science if it is to be incorporated into the underpinnings of policy and law. This proposition will undoubtedly make some brain scientists uneasy. It is likely to make ethicists uneasy too, for reasons I will discuss later in this article. Historically, neuroscience has been used for a variety of political goals, some positive or benign and some very far from that, including in some neuroscientists' partnership with the genocidal programs of the Third Reich (Loewenau & Weindling, 2016; Illes & Bird, 2006, p. 1). Thus, the intersection of neuroscience and law writ large necessarily involves questions of ethics, values and cultural ideas.

As I have suggested, encouraging a partnership between neuroscience and law reform requires a social movement. There are promising signs of policy makers' increasing interest in social neuroscience broadly conceived, especially in the findings about implicit bias, which, in the racial justice movement especially, have sprung to the fore in public consciousness (see, e.g., Kendi, 2019). The implications of implicit bias research play a significant role in calls for institutional reform, especially in policing and other aspects of the criminal justice system.

The problem with advocating for the use of social neuroscience to reform policy is that policy advocacy is not something most scientists routinely do. Scientists do research and then document and publish results. To be credible and have integrity, scientists do not, and usually should not, have a vested interest in how their results will be used. To be sure, they choose topics and research directions based on their values and personal intellectual interests, but they must be willing to acknowledge the limits of their findings and their possible flaws. Advocacy, on the other hand, works in a very different way, using the techniques of rhetoric to persuade and simplify rather than to qualify, interrogate, and make more complex.

In considering how neuroscience can affect social policy in the area of anti-discrimination law, the following reality must be acknowledged: law will not change simply because science shows it should. This is a plain fact that even a neuroscientist as brilliant as neurologist, biologist, primatologist, popular science translator and writer Robert Sapolsky fails to realize. In a recent popular book, *Behave: The Biology of Humans at Our Best and Worst*, Sapolsky rails against the way that criminal law requires the supposition that human beings have "free will" and thus should be punished for bad acts because they had the ability to resist committing them (pp. 580). As Sapolsky points out, brain scientists know that human beings often cannot avoid carrying out the bad acts they commit. Bad acts may

bepredetermined by the individual's brain wiring, including experiences that have irrevocably shaped that individual's brain. Why then, he asks, does law rest on the very obvious "fiction" of free will?

One could take issue with Sapolsky's assumption that free will is utterly a fiction and, indeed, neuroethicists have proposed other approaches to understanding the free will versus determinism conundrum (e.g., [Schaffner, 2002](#), p. 30). Schaffner, for example, proposes not a "sweeping determinism" but instead what he terms "creeping" or partial reductionism, in which free will with respect to any choice is the default position that can be modified by excusing conditions on a case-by-case basis ([Illes & Bird, 2006](#); [Schaffner, 2002](#)). But regardless of what neuroscience continues to discover about the balance between free will and predetermined, brain-based causes for an individual's behavior, law can never fully succumb to a deterministic view of human conduct because law represents the articulation, reflection, enforcement and propagation of a set of values and norms held by a particular social group. Law presumes the fiction of free will because the societies that rely on that concept believe in punishment for morally bad acts.

Indeed, this is why law typically requires a showing of intentional bad actions to impose liability for discrimination, as discussed above. It may be that law can be persuaded not to impose standards of moral wrong on all discriminatory acts, however. In the past, vagrancy laws punished outsiders simply for the status of being "deviant" in some socially salient way ([Goluboff, 2016](#)), as discussed above, but today many social systems worldwide no longer regard deviance that is based on non-volitional human difference as a punishable wrong. So too can law decide not to rely on proof of "bad" intent as the prerequisite for using the courts' remedial powers to rectify practices that cause discriminatory effects, as I have argued above.

Social change is due to many factors, including historical, social, political and economic developments, improved scientific understandings of the brain and other aspects of human functioning, and, most importantly, social movements that have fought for change in the way law treats vulnerable and marginalized communities. Extending that fight even more generally, to embrace more forms of acting differently that do not cause harm to others will require further social activism. Such change may be harder to achieve because the proposition that all persons should have the right to act differently, being general in its scope, may not motivate particular communities to action in the same way that the identity-based movements of past moments in history generally have. But evolving human rights norms, and coalitions

among many identity groups, somewhat along the lines of what Goluboff excavates in her study of anti-vagrancy law in the United States, may help produce such a movement.

Of course, the opposite historical trajectory is equally if not more likely given current worldwide political conditions. Those conditions suggest a worldwide movement toward more intolerance, “othering” or “cossification” of marginalized others, and a generalized stirring up of hatred of difference as some political leaders tap into the irrational thought processes that lead to these results. What the future holds in this respect is anyone’s guess as I write these words. My point is that the findings of social neuroscience at least provide the empirical support for a movement to protect, value, and indeed encourage human difference, and that those who support such a movement can and should draw on social neuroscience to help support this work.

Put most simply, law is a complex set of cultural artifacts embodying a legal system’s values, judgments, conceptual constructions and deep assumptions—and, yes, also its prejudices, biases, fast thinking and irrationality. Law is not logic at its core, and thus logic alone will not move law. Particular legal systems may move toward expanded rights protections or in the opposite direction or, most likely, some complex combination of both. There is nothing inevitable about any direction, so human agency will be required to make the policy arguments and seek to influence systems external to but related to law, such a public opinion, private practices, and changing norms and values. All of these factors can in turn shape law in the directions neuroscience findings support.



4. The ethics of using neuroscience in anti-discrimination law

So far, I have suggested some of the benefits of using social neuroscience in anti-discrimination law. Yet, as in all things, benefits also come with drawbacks, and these must be considered as well. Here I discuss a few that I find particularly salient. One involves the problem of reducing complex human behavior—i.e., the wildly irrational human mind and dynamic, unpredictable unconscious of Freud’s quasi-literary psychology—to the deterministic assumptions of natural science as applied to humans, as the introduction discussed. What happens to imagination, transgression, human freedom? What happens to the “fiction” of free will that Sapolsky so opposes as a neuroscientist, but upon which law’s basic assumptions of moral

responsibility rest? I think the answers to these ethical questions are fairly straightforward, as I have already discussed above. Science is not, and never will or can be, the only system of ideas operating in a social system. There will always be room for human brains to respond to and be influenced by the humanistic disciplines, and law seems likely to continue to impose consequences based on concepts of moral responsibility even if neuroscience helps law see that the scope of “free will” is narrower than law current assumes (Winslade, 2002).

A second, more serious ethics concern involves the use of neuroscience testing on individuals as evidence of motive or intent. If anti-discrimination law does not change to eliminate intent requirements in the way I suggest above, then implicit bias testing could be used on individuals to establish and impose penalties for the existence of bias, albeit of an unconscious and non-volitional nature. Although by definition implicit bias is not “intentional” prejudice, because there is no conscious, volitional intent in this regard, there is considerable danger that such evidence could be used in particular cases. This would constitute a disturbing misuse of the lessons of neuroscience, but one can certainly imagine this occurring if political and social circumstances allowed, including (or even especially) in highly politically progressive social environments. Thus, one important ethical limitation of neuroscience evidence in anti-discrimination law is that it should never be used against individuals to establish “bad” action or thoughts.

A third concerning possibility is that the insights of neuroscience could be used for goals antithetical to those I outline here. Just as social neuroscience can help illuminate paths toward eliminating illegal discrimination, as I have argued above, the discipline’s discoveries could also be used for contrary purposes. Just as social neuroscience can aid progressive reformers, social neuroscience can provide insights to those who wish to inflame hatred by exploiting the human brain’s tendency to be suspicious, fearful, and even disgusted when it unconsciously detects social difference among people. As discussed, perceptions of trivial differences between in-groups and out-groups can lead to negative social reactions to out-groups as the brain non-volitionally reacts to the group as inferior, or even detestable and worthy of persecution at the extreme.

To be sure, political demagogues appear to have little trouble inflaming hatred of “the other” in order to exploit social division for their political gain, even though these political figures typically are not fans of the neuroscience literature—or, indeed, of science at all. They do not need an expert

guide to sow division and hate because their own fast thinking gives them an intuitive feel for how to play this type of politics. But just as the findings of social neuroscience may come to help institutional reformers train decision makers to reduce their chances of acting on implicit biases, so too can one imagine those who wish to inflame hatred and dissention using science-based techniques to improve their success in this regard. The many recent examples of disinformation campaigns used to influence the politics of foreign countries provide early warnings of this enormous negative potential.

Similarly, institutional leaders may use the insights of social neuroscience to select institutional participants who exhibit the personality, or brain-based characteristics, they prefer. These preferred characteristics could just as easily include such features as fierce in-group loyalty, tendencies toward extreme conformity, unwillingness to challenge authority, strong implicit biases and similar traits. In other words, neuroscience could be used as easily to promote reactionary tendencies as the progressive ones I champion above, such as open-mindedness, tolerance, and a positive interest in persons not like oneself. These are the characteristics I have been arguing institutions should encourage. My proposals call for encouraging human tendencies toward appreciating—and in fact reveling in—human difference, but those arguments rest on my own moral preferences or “tastes.”

To further explore this ethics issue, it will be helpful to turn to another experiment-based subfield of neuroscience called “moral cognitive neuroscience.” This field explores the correlations between personality tendencies, political orientations, and styles and priorities in human moral thinking (Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005). Like social neuroscience, the field draws on the findings of experimental psychology, cognitive neuroscience, cultural anthropology, primate studies, developmental psychology, and more, to understand more scientifically how human moral thinking takes place. In order to apply the insights of that field here, it will be helpful to take a step back and outline some of the key debates taking place within it.

Two of its leaders, both of whom have helped popularize its ideas through accessible books, are Jonathan Haidt, a business school professor at New York University, and Joshua Greene, a member of the psychology department and Center for Brain Science at Harvard University. As these two scholars explain in a coauthored article, their goals are to show that “moral reasoning matters, but it matters in social context,” and that “much, though not necessarily all, moral judgment makes use of processes

specifically dedicated to social cognition” (Greene & Haidt, 2002). Haidt and Greene’s theories differ from each other in some respects, but both offer many useful insights on using social neuroscience in law reform.

The central insight of moral cognitive neuroscience borrows from and extends the implications of Tversky and Kahneman’s findings about fast and slow thinking. Not surprisingly, moral judgment, viewed through the lens of brain science, has an extensive automatic, intuitive, and unconscious dimension. Haidt and Greene differ in their views as to the balance between intuitive, emotional or automatic thinking about moral issues—what Greene labels “automatic” thinking, versus what he terms deliberative thinking, which Greene labels as thinking in a “manual” mode. Haidt believes that in moral judgment there is very little to the manual mode, or deliberate, rational thinking. He reaches this conclusion based on his experimental work showing that human beings generally make moral judgments very quickly, as characteristic of automatic or intuitive processes (Haidt, 2013, p. 28). In his view, the brain “does morality” just as it does other automatic judgment processes such as social exclusion. In support of this view, Haidt points out that when experimenters ask people to explain how they made moral judgments, these subjects insist that they make them on the basis of rational deliberation; however, when researchers ask them to explain the deliberation leading to their judgments, the subjects cannot present convincing rationales (Haidt, 2013, p. xiv). Instead, their explanations appear strategic—in other words, ad hoc explanations devised after the fact of judgment (Haidt, 2013, p. 44).

Haidt’s findings echo the arguments of the American Legal Realists (introduced above), who argued that judges and other legal decision makers do not reach legal judgments based on rational deliberation but by intuition, and then backfill with legal logic to support their intuitive conclusions, using whichever of many often inconsistent legal argumentation devices they need to support their conclusion. Thus, Jerome Frank, who had a deep interest in psychology, as already noted, famously declared that judges’ legal decisions are based on what they had for breakfast, which he meant as a facetious way of proposing that legal decision making is based on unconscious, intuitive factors (i.e., fast or automatic thinking) rather than on rational deliberation (i.e., deliberative or manual thinking). Again, a consilience emerges in theories about how human thinking works: Haidt’s experiments support the insights of the Legal Realists from nearly century ago.

Haidt proposes six moral dimensions with corresponding emotions that form the foundations of moral thought. These include polarities such as

loyalty versus betrayal—the very tradeoff involved in whistle-blowing, as I have discussed above—as well as five others that are not necessary to explain here. Extensive testing has shown that participants place different relative weights on these different dimensions or polarities, even within the same culture, due to factors such as personality differences (Haidt, 2013, pp. 178–79). Pointing to the bitter divisions between political and social conservatives versus liberals in the United States today, Haidt argues that these polarities explain the intensive disagreements in many political cultures. Different persons have different moral palates or “tastes,” and, because moral judgment is intuitive and emotional, members of warring political groups cannot change each other’s minds on moral questions through rational argument.

Greene, in turn, argues for what he calls a “dual process” theory of moral judgment (Greene, 2013). According to Greene, both intuitive and deliberative processes are at work in moral decision making. When the two processes arrive at the same moral judgment, the dual process system works wonderfully. However, in other situations, the two processes conflict. Greene believes he has captured evidence of such conflict between automatic and manual mode moral thinking through his ingenious experimental adaptation of a classic ethics problem known as the “Trolley Problem.” That problem, designed by philosopher Judith Jarvis Thomson, posits that a trolley without a driver has escaped from a station and is hurtling toward five workers who are standing on the track (Thomson, 1985). The moral question is whether it would be acceptable for a person knowing the situation to move a lever in a switching station to divert the trolley onto another track on which only one person is standing, so that only one rather than five will be killed. Greene found through repeated experiments that most people state that this would either be acceptable or perhaps even morally required, apparently on the utilitarian ground that it is better to kill only one person rather than five. Yet, as Thomson points out, this same moral intuition does not always hold. Most people say that it would not be morally acceptable for a doctor to kill one healthy patient in order to use his lungs, heart, and kidneys to save five others.

Greene and his colleagues take Thomson’s basic scenario and give it another twist. Their first scenario is the same as Thomson’s; their second asks if it would be equally morally acceptable to physically push a person onto the tracks at the station to stop the trolley from hitting the five people standing on the tracks further on. As one might expect, many fewer experimental subjects agree that it would be morally acceptable to push someone onto

the tracks, even though using one's hand to move the lever to switch the train onto the track on which only one person is standing involves a person exercising moral agency, just in a less physical and intensely immediate way.

Greene uses these experimental results to make a number of points. First, like Haidt, Greene argues that his experiments capture the fact that moral thinking is intuitive, emotional and based on automatic processes. He suggests that the reason so few people will accept pushing someone on the tracks in the face of an oncoming train to prevent the death of five, which makes rational sense in utilitarian terms, is because humans' "automatic" settings abhor violence toward fellow members of what he calls their circle of regard (Greene, 2013, p. 217–18). Moreover, Greene argues, humans use faulty heuristics in moral judgment, including one that distinguishes in an automatic and rapid way between intent and side effects. The act of moving the lever to switch the train onto a different track appears to have the intent of saving four human lives, with the unfortunate side effect of killing someone on that track, while the act of pushing someone into a train one can see coming appears to have the intent of killing someone, which intuitively feels immoral, even though in fact both acts are the same. In short, Greene concludes, rapid, intuitive thinking in the context of moral judgment uses heuristics just as other types of rapid, intuitive thinking do. These heuristics may generally be useful—as, for example, when one must make a split-second decision about the morally right thing to do in a tragic emergency—but are also inflexible and sometimes irrational, as shown in the experiment described above. Heuristics ignore the harmful side effects of decisions that feel moral or right, what Greene terms "modular myopia" (Greene, 2013, p. 224).

Greene gleans even more from his trolley experiment when he conducts the experiment with subjects in brain scanners in order to explore where in their brains moral decision making takes place. He asks his subjects first to think about their intuitive response, and then to think about why pushing someone into the oncoming train to save five lives may rationally be the better moral outcome. Greene finds that subjects thinking intuitively activate the parts of their brains that process social thinking. Moreover, people asked to think about the reasons for pushing someone in front of an oncoming train to save five lives have a hard time doing so. Their response times slows, showing that they are operating under a cognitive load. Here, Greene explains, automatic and manual mode thinking processes are in conflict (Greene, 2013, pp. 115–17).

I want to end by suggesting that all social reformers must be cognizant of the tendency toward moral myopia that Greene identifies. Eradicating discrimination is likely to have negative side effects, which in the viewpoints of libertarians and like thinkers involve moral harms. For example, barring discrimination helps prevent its profound social harm, but also requires encroaching on individual actors' free agency. Moral cognitive neuroscience teaches the importance of self-reflection, detachment and skepticism—even, or especially, about the moral principles one holds most dear.

With these guiding insights in mind, further dialogue with those who disagree should take place in bringing social neuroscience to bear on the task of eradicating discrimination. There should be limits to the use of neuroscience in law, even for such worthy purposes as enhancing law's ability to identify and deter discrimination. Those using social neuroscience for such ends may be subject to fallibilities in moral thinking just as those opposed to these political objectives are subject to moral errors of their own. Further dialogue, appealing to both deliberative logic and moral emotion, will be necessary.



5. Conclusion

In modern social conditions, human beings often deal with persons who are different from themselves. Modern societies, which are politically based on pluralism and economically based on labor specialization, could not exist without a rich variety of differences among people (Rosen, 1983). The non-volitional brain processes that can react negatively to perceived differences are maladaptive in present social conditions. Neuroscience—and especially the field of social neuroscience, which focuses directly on how the brain engages in social behavior—can help law diminish discrimination based on identity characteristics that are specially protected because of past histories of subordination. Although this task will not be easy, it is an imperative one, and advances in social neuroscience can help accomplish it.

The findings of neuroscience must gain traction among judges, policy makers, law makers and opinion makers. Law can be influenced by science but it is not a science, contrary to the views of our 18th century forbearers. Law is full of irrationality just as the human brain is, and necessarily so since it is a construct or collective artifact of human life. Thus neuroscientists must continue to produce science, and to challenge, improve, critique and revise

it, while social activists, lawyers, opinion makers, law makers and judges continue to champion its benefits so that it becomes part of the cultural context through which law can improve equality and fairness in life opportunities for all.

Conflict of interest

I have no known conflicts of interest to disclose.

References

- Abrams, D., & Hogg, M. A. (1990). Social identification, self-categorization and social influence. *European Review of Social Psychology*, 1(1), 195–228.
- Armstrong, T. (2011). *The power of neurodiversity*. Perseus Books Group.
- Berns, G. (2010). *Iconoclast: A neuroscientist reveals how to think differently*. Harvard Business Press.
- Blass, T. (Ed.). (2000). *Obedience to authority: Current perspectives on the Milgram paradigm*. Lawrence Erlbaum Associates Publishers.
- Borsook, T. K., & MacDonald, G. (2013). *Social Pain*. Oxford University Press.
- Bradlow, D., & Hunter, D. (2019). *Advocating social change through international law: Exploring the choice between hard and soft international law*. Brill.
- Brown-Nagin, T. (2012). *Courage to dissent: Atlanta and the long history of the civil rights movement*. Oxford University Press.
- Cacioppo, J. T., & Decety, J. (Eds.). (2011). *The Oxford handbook of social neuroscience* (1st ed.). Oxford University Press.
- Carle, S. (2011). A social movement history of title VII disparate impact doctrine. *Florida Law Review*, 63(1), 251–300.
- Carle, S. (2016). Angry employees: Revisiting insubordination in title VII cases. *Harvard Law & Policy Review*, 10(1), 185–227.
- Carle, S. (2017). Analyzing social impairments under title I of the Americans with disabilities act. *UC Davis Law Review*, 50(3), 1109–1180.
- Carle, S. (2019). Acting differently: How science on the social brain can inform anti-discrimination law. *University of Miami Law Review*, 73(3), 655–730.
- Civil Rights Act, (1964) 42 U.S.C. § 2000e.
- Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial Bias in the decision to shoot. *Journal of Personality and Social Psychology*, 92(6), 1006–1023.
- Cunningham, W., Haas, I., & Jahn, A. (2011). Attitudes. In *The Oxford handbook of social neuroscience*. New York, NY: Oxford University Press.
- Dailey, A. C. (2017). *Law and the unconscious: A psychoanalytic perspective*. Yale University Press.
- Daubert v. Merrell Dow Pharmaceuticals, Inc. (1993) 509 U.S. 579.
- Frank, J. (2009). *Law and the modern mind*. Transaction Publishers.
- Goluboff, R. L. (2016). *Vagrant nation: Police power, constitutional change, and the making of the 1960s*. Oxford University Press.
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., et al. (2007). Implicit Bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, 22(9), 1231–1238.
- Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Science*, 6(12), 517–523. [https://doi.org/10.1016/S1364-6613\(02\)02011-9](https://doi.org/10.1016/S1364-6613(02)02011-9).
- Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. The Penguin Press.

- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.
- Guinier, L., & Strum, S. (2019). *Trial by firefighters*. The New York Times. Retrieved September 28, 2020, from <https://www.nytimes.com/2009/07/11/opinion/11guinier.html>.
- Haack, S. (2014). *Evidence matters: Science, proof, and truth in the law*. Cambridge University Press.
- Haidt, J. (2013). *The righteous mind: Why good people are divided by politics and religion*. Vintage Books.
- Hastorf, A. H., & Cantril, H. (1954). They saw a game: A case study. *Journal of Abnormal and Social Psychology*, 49, 129–134.
- Hughes, E. (2011). The empathic divide in capital trials: Possibilities for social neuroscientific research. *Michigan State Law Review*, 2011, 541–565.
- Illes, J., & Bird, S. (2006). Neuroethics: A modern context for ethics in neuroscience. *Trends in Neurosciences*, 29(9), 511–517.
- Kahan, D. M., Hoffman, D. A., Braman, D., Evans, D., & Rachlinski, J. J. (2012). “They saw a protest”: Cognitive illiberalism and the speech-conduct distinction. *Stanford Law Review*, 64(4), 851–906.
- Kahneman, D. (2013). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kendi, I. X. (2019). *How to be an antiracist* (1st ed.). One World.
- Kramer, Z. (2019). *Outsiders: Why difference is the future of civil rights*. Oxford University Press.
- Kuhn, T. S., & Hacking, I. (2012). *The structure of scientific revolutions* (4th ed.). The University of Chicago Press.
- Kuzawa, C. W., & Sweet, E. (2009). Epigenetics and the embodiment of race: Developmental origins of US racial disparities in cardiovascular health. *American Journal of Human Biology*, 21(1), 2–15.
- Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., et al. (2014). Reducing implicit racial preferences: A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, 143(4), 1765–1785.
- Larson, D. (2016). Fair and implicitly impartial jury: An argument for administering the implicit association test during Voir dire. *DePaul Journal for Social Justice*, 3(2), 139–171.
- Leary, M. R., Kowalski, R. M., Smith, L., & Phillips, S. (2003). Teasing, rejection, and violence: Case studies of the school shootings. *Aggressive Behavior*, 29(3), 202–214.
- Leiter, B. R. (2002). *American legal realism*. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.339562>.
- Levin, L. C., & Mather, L. (2018). The impact of international lawyer organizations on lawyer regulation. *Fordham International Law Journal*, 42(2), 407–480.
- Loewenau, A., & Weindling, P. J. (2016). Nazi medical research in neuroscience: Medical procedures, victims, and perpetrators. *Canadian Bulletin of Medical History*, 33(2), 418–446.
- Marques, J., & Yzerbyt, V. (1988). The black sheep effect: Judgmental extremity towards Ingroup members in inter- and intra-group situations. *European Journal of Social Psychology*, 18(3), 287–292.
- Matusall, S., Kaufmann, I. M., & Christen, M. (2011). *The emergence of social neuroscience as an academic discipline*. Oxford University Press.
- Mendoza, S. A., Lane, S. P., & Amodio, D. M. (2014). For members only: Ingroup punishment of fairness norm violations in the ultimatum game. *Social Psychological and Personality Science*, 5(6), 662–670.
- Mill, J. S., Bromwich, D., Kateb, G., & Elshaint, J. B. (2003). *On Liberty*. Yale University Press.
- Mlodinow, L. (2013). *Subliminal: How your unconscious mind rules your behavior*. Vintage Books.

- Modjeska, A. (2017). *Employment discrimination law* (3rd ed.).
- Molenberghs, P. (2013). The neuroscience of in-group Bias. *Neuroscience & Biobehavioral Reviews*, 37(8), 1530–1536.
- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, 6(10), 799–809.
- Morrison, S., Decety, J., & Molenberghs, P. (2012). The neuroscience of group membership. *Neuropsychologia*, 50(8), 2114–2120.
- Oxford English Dictionary. (2020). Consilience. In *Oxford online dictionary*. <https://www.lexico.com/en/definition/consilience>.
- Pager, D., & Shepherd, H. (2008). The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Review of Sociology*, 34(1), 181–209. <https://doi.org/10.1146/annurev.soc.33.040406.131740>.
- Paradies, Y., Ben, J., Denson, N., Elias, A., Priest, N., Pieterse, A., et al. (2015). Racism as a determinant of health: A systematic review and meta-analysis. *PLoS One*, 10(9). <https://doi.org/10.1371/journal.pone.0138511>.
- Phillips, K. (2014). *How diversity makes us smarter*. Scientific American. <https://doi.org/10.1038/scientificamerican1014-42>.
- Pond, R., Richman, S., Chester, D., & DeWall, N. (2014). Social pain and the brain: How insights from neuroimaging advance the study of social rejection. In T. D. Papageorgiou, G. I. Christopoulos, & S. M. Smirnakis (Eds.), *Advanced brain neuroimaging topics in health and disease—Methods and applications*. InTech.
- Ricci v. DeStefano, (2009) 557 U.S. 557.
- Rosen, S. (1983). Specialization and human capital. *Journal of Labor Economics*, 1(1), 43–49.
- Sapolsky, R. M. (2017). *Behave: The biology of humans at our best and worst*. Penguin Press.
- Schaffner, K. F. (2002). Neuroethics: Reductionism, emergence, and decision-making capacities. In S. Marcus, & A. Charles (Eds.), *Neuroethics: Mapping the field: conference proceedings, May 13–14*, Dana Press.
- Schweik, S. M. (2009). *The ugly laws: Disability in public*. NYU Press.
- Seiner, J. A. (2006). Disentangling disparate impact and disparate treatment: Adapting the Canadian approach. *Policy Review*, 25(1), 96–142.
- Stone, K. (2012). Taking in strays: A critique of the stray comment doctrine in employment discrimination law. *Missouri Law Review*, 77, 149–197.
- Surowiecki, J. (2005). *The wisdom of crowds*. Anchor Books.
- Taylor, R. (1996). A comparative study of expert testimony in France and the United States; philosophical underpinnings, history, practice, and procedure. *Texas International Law Journal*, 31, 181–213.
- Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, 94, 1395–1415.
- Todorov, A., Stewart, L. H., Ajina, S., Getov, S., Bahrami, B., & Rees, G. (2012). Unconscious evaluation of faces on social dimensions. *Journal of Experimental Psychology: General*, 141(4), 715–727.
- Treynor, J. L. (1987). Market efficiency and the bean jar experiment. *Financial Analysts Journal*, 43(3), 50–53.
- Universal Declaration of Human Rights. 2020 https://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf Retrieved September 15, 2020, from.
- Vaughn, R. (2012). *The successes and failures of whistleblower laws*. Edward Elgar Publishing.
- Williams, K. D. (2007). Ostracism. *Annual Review of Psychology*, 58(1), 425–452.
- Wilson, E. O. (1998). *Consilience: The unity of knowledge* (1st ed.). Knopf: Distributed by Random House.
- Wilson, E. O. (2012). *The social conquest of earth*. Liveright.
- Winslade, W. (2002). Traumatic brain injury and legal responsibility. In S. Marcus, & A. Charles (Eds.), *Neuroethics: Mapping the field: Conference proceedings, May 13–14*, Dana Press.
- Yoshino, K. (2007). *Covering: The hidden assault on American civil rights*. Random House.