

# RETHINKING INTENT AND IMPACT: SOME BEHAVIORAL REALISM ABOUT EQUAL PROTECTION

*Jerry Kang*\*

I. IMPLICIT BIAS .....	627
II. BEHAVIORAL REALISM .....	635
III. THE EQUAL PROTECTION “MACHINE” .....	636
A. <i>Welcome to the Machine</i> .....	637
1. <i>Tracks of scrutiny</i> .....	637
2. <i>Track selector</i> .....	637
a. <i>Trigger 1: Intent to Harm Minorities (T<sub>1</sub>)</i> .....	637
b. <i>Trigger 2: Facial Racial Classification (T<sub>2</sub>)</i> .....	638
B. <i>Reverse Engineering</i> .....	639
1. <i>Function 1: A Sensor for Concealed Intent to Harm Minorities (F<sub>1</sub>)?</i> .....	639
2. <i>Puzzling Triggers</i> .....	640
a. <i>Reconciling T<sub>1</sub>: The Problem of Pointless Redundancy</i> .....	640
b. <i>Reconciling T<sub>2</sub>: Not about Protecting Minorities</i> .....	642
3. <i>Function 2: A Shield to Protect Minorities against Intentional Harm (F<sub>2</sub>)</i> .....	643
4. <i>Function 3: Avoiding Racial Balkanization (F<sub>3</sub>)</i> .....	644
IV. REDESIGN .....	646
A. <i>Redesign T<sub>1</sub>: From Intent to Causation</i> .....	646
B. <i>Abandon F<sub>1</sub>: Stop Using It to Justify T<sub>2</sub></i> .....	648
C. <i>Explore F<sub>3</sub> Balkanization</i> .....	649

## I. IMPLICIT BIAS

Our brains think through mental schemas, which are templates of knowledge that help us organize specific examples into broader categories.

---

\* © 2014 Jerry Kang. Professor of Law, Professor of Asian American Studies (by courtesy), Korea Times-*Hankook Ilbo* Endowed Chair in Korean American Studies and Law, Associate Provost. My thinking benefited from comments provided by Rachel Godsil, Anthony Greenwald, Linda Tropp, Eugene Volokh, participants of the Meador Lecture as well as the University of Texas School of Law Faculty Colloquium. Helpful research assistance was provided by Ashley Stein and the Hugh & Hazel Darling Law Library at UCLA.

When we see, for example, something with a flat seat, a back, and some legs, we categorize it as a “chair.” Regardless of whether it’s plush or wooden, with wheels or bolted down, we know what to do with a “chair.” Unless we’re antiquing or worry that it can’t bear our weight, we just sit down, no questions asked.

It could not be otherwise. Our brains are deluged with too much information every second to attend to each particular datum, to treat each facet of each object as unique. Instead, we must make sense of the data stream by chunking it into categories. The fact that we might do this to objects like chairs, or animals like dogs, does not much surprise us. However, things get more interesting when we learn that we have mental schemas for groups of human beings as well.

For example, when you see me come up to the stage, your brain immediately tags me into numerous social categories, such as those based on gender, age, role, and race. By visual inspection, you tag me #Asian, which means that every memory associated with that social category could potentially be activated.

A social psychologist would describe this cloud of information tagged #Asian in terms of attitudes and stereotypes. An *attitude is an association between a social group and an evaluative valence, either positive or negative*.<sup>1</sup> In other words, it’s an overall reaction toward the social category that ranges between positive and negative, good and bad, approach and avoid, thumbs up and thumbs down. Pause for a moment, and ask yourself what’s your attitude toward Asian men?

By contrast, a *stereotype is an association between a social group and a specific trait*.<sup>2</sup> For example, when you think of the category “bird,” you think of the trait “flies.” We know that not all birds fly (e.g., ostriches and penguins), but there is a strong association nevertheless. By the way, what are your stereotypes about Asians? For example, do you think it a little more likely (because I’m Asian) that I got a perfect score on my math SAT? Do you think it a little more likely that I’m a bad parallel parker? Do you think it a little more likely that I have a black belt in some exotic East Asian martial art? Do you think it a little more likely that I was raised by a tiger mother who forced me to practice violin six hours a day so that I became a virtuoso?

Decades of social psychological research suggest that your attitudes and stereotypes influence your interaction with me, at least on the margins.<sup>3</sup> So, if I want to know whether you discriminate—by which I mean, quite

---

1. Jerry Kang et al., *Implicit Bias in the Courtroom*, 59 UCLA L. REV. 1124, 1128 (2012).

2. *Id.*

3. MAHZARIN R. BANAJI & ANTHONY G. GREENWALD, BLINDSPOT: HIDDEN BIASES OF GOOD PEOPLE 133 (2013).

narrowly, that you treat me differently from someone else who provides an identical performance simply because of race—it would be helpful to know the attitudes and stereotypes in your head.

But when I inquire, I run into a “willing and able” problem. First, let’s suppose that you have complete introspective access to the attitudes and stereotypes in your head. Maybe you don’t feel especially positive toward Asian people—who knows why? Maybe you honestly think, based on personal experience, that Asians work really hard but don’t know how to have a good time. Would you share these views with me? Or would you be *unwilling*? After all, who needs to be harassed by the PC (politically correct) police? Second, and scientifically more interesting, what if there are attitudes and stereotypes you simply lack direct introspective access to? In other words, you can’t just close your eyes, ask yourself, “So, what do I really think about Asian Americans,” and get a complete answer. In other words you are *unable*.<sup>4</sup>

If such implicit attitudes and stereotypes (“implicit biases”) exist, then by definition we can’t simply ask people, even on anonymous surveys, for their honest self-reports. Instead of asking, we must somehow measure. Well, scientists have tried numerous ways to measure them. The most successful method uses differences in reaction-times, which leverage the fact that the speed of our response says something about the strength of mental connections.<sup>5</sup> The stronger any two concepts are connected in our minds, the faster we can sort them together. Thus, if we have a more positive attitude toward Whites as compared to Blacks, then we should sort White faces more quickly with good words than Black faces with good words. This basic logic drives the well-known Implicit Association Test (IAT).<sup>6</sup>

Over the past decade, much has been written about implicit bias generally and the IAT specifically in the legal literature.<sup>7</sup> Therefore, I will not repeat in detail what has already been carefully described. Suffice it to say that implicit biases have been found to be:

---

4. Self-knowledge is more limited than we generally suppose. See generally Timothy D. Wilson & Elizabeth W. Dunn, *Self-Knowledge: Its Limits, Value, and Potential for Improvement*, 55 ANN. REV. PSYCH. 493 (2004).

5. BANAJI & GREENWALD, *supra* note 3.

6. See Anthony G. Greenwald et al., *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 J. PERSONALITY & SOC. PSYCH. 1464, 1464–66 (1998) (introducing the Implicit Association Test). For more detailed exposition, see Brian A. Nosek, Anthony G. Greenwald & Mahzarin R. Banaji, *The Implicit Association Test at Age 7: A Methodological and Conceptual Review*, in AUTOMATIC PROCESSES IN SOCIAL THINKING AND BEHAVIOR 265 (John A. Bargh ed., 2007).

7. See, e.g., Kang et al., *Implicit Bias in the Courtroom*, *supra* note 1; Jerry Kang & Kristin A. Lane, *Seeing Through Colorblindness: Implicit Bias and the Law*, 58 UCLA L. REV. 465 (2010); Kristin A. Lane, Jerry Kang & Mahzarin R. Banaji, *Implicit Social Cognition and Law*, 3 ANN. REV. L. & SOC. SCI. 427 (2007); Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1491, 1506 (2005).

1. Pervasive;
2. Strong in magnitude (as compared to explicit biases, measured by anonymous self-reports and put onto the same statistical units);<sup>8</sup>
3. Related to, but not the same as, explicit biases that are introspectively accessible but may be shielded from public view;<sup>9</sup> and
4. Predictive of behavior in real-world circumstances.<sup>10</sup>

If the science is new to you, I encourage you to learn more, and to do so skeptically. That said, I believe that after a fair reading of the research literature, one can no longer deny that “implicit biases” as meaningful psychological constructs exist. One can no longer deny that implicit biases can be measured by scientific instruments, such as reaction-time devices (of which the IAT is just one).<sup>11</sup> One can no longer deny that implicit biases have *some* predictive validity of behavior, such as body language, judgments, evaluations, and interactions.<sup>12</sup> The most important and unsettled question is how big are the effect sizes?

Just because something is statistically significant (which means that the particular experiment’s results are unlikely due to chance) does not mean that the findings are socially, ethically, or morally significant. For example, implicit biases may predict behavior better than chance alone. But if the effect sizes are too low, they may provide so little explanatory power so as to be practically insignificant.

Back in 2009, the first meta-analysis on the predictive validity of the IAT was published.<sup>13</sup> In socially sensitive domains, such as Black/White interactions, the meta-analysis found an effect size of Pearson’s (average)  $r=.236$ , which would explain 5.6% of the variance in behavior.<sup>14</sup> This is considered to be a small effect size; that said, it is always important to ask, “As compared to what?” Measures of explicit bias—presumably honest self-reports on anonymous surveys—had even *less* predictive power,

8. See, e.g., Brian A. Nosek et al., *Pervasiveness and Correlates of Implicit Attitudes and Stereotypes*, 18 EUR. REV. SOC. PSYCH. 1 (2007) (analyzing massive data set from Project Implicit).

9. See Anthony G. Greenwald & Brian A. Nosek, *Attitudinal Dissociation: What Does It Mean?*, in ATTITUDES: INSIGHTS FROM THE NEW IMPLICIT MEASURES 65 (Richard E. Petty, Russell E. Fazio & Pablo Briñol eds., 2008).

10. See Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, 97 J. PERSONALITY & SOC. PSYCH. 17, 19–20 (2009).

11. See Russell H. Fazio & Michael A. Olson, *Implicit Measures in Social Cognition Research: Their Meaning and Use*, 54 ANN. REV. PSYCH. 297 (2003) (reviewing different implicit measures).

12. See Greenwald, *supra* note 10; John T. Jost et al., *The Existence of Implicit Prejudice Is Beyond Reasonable Doubt: A Refutation of Ideological and Methodological Objections and Executive Summary of Ten Studies That No Manager Should Ignore*, 29 RES. ORG. BEHAV. 39, 41 (2009).

13. See Greenwald, *supra* note 10.

14. See *id.*, at 24 tbl.3.

(average)  $r=.118$ , which would explain only 1.4% of the variance in behavior.<sup>15</sup> It is worth noting that the lead author of this meta-analysis was Anthony Greenwald, who invented the IAT.<sup>16</sup> And, without casting aspersions, skeptical lawyers might discount these findings accordingly.

In 2013, another meta-analysis following a different methodology (e.g., which studies and results to count) computed a lower effect size between implicit bias measured by the IAT and behavior in Black/White interactions: (average)  $r=.15$ .<sup>17</sup> This study was published by a team hostile to the IAT specifically and implicit bias research generally.<sup>18</sup> Again, without casting aspersions, skeptical lawyers should always check for financial<sup>19</sup> and ideological motivations that might prompt one to discount these findings accordingly.

Most legal readers will be unfamiliar with Pearson's correlation "r," what the difference between .236 or .15 means mathematically,<sup>20</sup> and the current controversies over the proper methods of conducting meta-analyses.

---

15. *See id.*

16. *See* Greenwald, *supra* note 6.

17. *See* Frederick L Oswald et al., *Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies*, 105 J. PERSONALITY & SOC. PSYCH. 171, 178 (2013) (providing the meta-analytically estimated population correlation of 0.15 for the "race domain").

18. Two of the authors on the Oswald paper include Gregory Mitchell and Philip Tetlock, who have penned sharp criticisms. *See, e.g.*, Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023, 1108–10 (2006). *See id.* at 1029 ("Accordingly, implicit prejudice scholars work hard to claim the *mantle of science* as they advance their agenda.") (emphasis added); *id.* at 1031 (noting that "some . . . claim a presumption of correctness in their interpretations of this ambiguous evidence by *attaching the label 'science'* to their views.") (emphasis added); *id.* at 1029 (describing scientific rhetoric as "more honorific than descriptive"); *id.* at 1076 (complaining about "politicized research programs—in which hypothesis advocacy has supplanted hypothesis testing"). For responses, see Kang & Lane, *supra* note 7, at 504–09 (discussing "junk science" backlash). *See also id.* at 509 n.208 (identifying a pattern and practice of "strawman" techniques).

19. Oswald, Mitchell, and Tetlock flag that they are consultants for LASSC, LLC, "which provides services related to legal applications of social science research, including research on prejudice and stereotypes." *See* Oswald et al., *supra* note 17, at 171. The corporation's website notes Mitchell and Tetlock as founders. *See Professionals*, LASSC, <http://lassc.com/professionals/> (last visited Sept. 08, 2014). In describing their litigation services, they point out that their principals "have testified recently in a number of major employment discrimination matters regarding what social scientists know – and don't know – about the factors that promote and inhibit bias toward members of protected groups and about the flawed methods used by a number of experts who purport to apply social science research to specific cases." *Litigation*, LASSC, <http://lassc.com/litigation/> (last visited Sept. 08, 2014). In the spirit of full disclosure, I should note that I often receive speaking fees and honoraria when I talk about implicit bias to law firms and corporations. I also received a generous honorarium from the University of Alabama School of Law for giving this Meador Lecture.

20. Pearson's  $r$  is a measure of linear correlation between two variables, and can range from -1 to 1, with  $|r|=1$  representing perfect linear correlation, and 0 representing no correlation.

And, this is not the place to get mired in complex scientific and statistical debates.<sup>21</sup> Instead, let me simply make two observations.

First, notice that scientists cannot simply make their data say what they would like them to say. One might assume that those in favor of the “implicit bias” construct, the IAT, and its explanatory power might want to demonstrate the largest effect size possible ( $|r|=1$ ), but that team found only .236, which by standard convention is a small effect size. Conversely, one might assume that the most hostile critics of the IAT might want to demonstrate a nil effect size ( $r=0$ ), but that team found .15. These results underscore that in science, data misbehave. They certainly do not adhere blindly to the personal or political desires of the researchers that collected them.

Second, for purposes of this discussion, let’s just take the lower effect size estimate as accurate. What does it mean from a policy, legal, and moral perspective to assume that the IAT (which is hardly the only way to measure implicit biases) predicts racially discriminatory behavior at only  $r=.15$ ? Well, consider some baselines for comparison, starting with medical common sense. Do you think that taking aspirin reduces risk of death by heart attack? Do you think that receiving chemotherapy increases survival of breast cancer? Do you think that ever smoking increases your chance of lung cancer? Do you think that drinking alcohol during pregnancy leads to increased subsequent premature birth? What about lead exposure and child IQ? According to statistics compiled by Gregory Meyer and colleagues, here are the Pearson’s correlation values<sup>22</sup>:

---

21. See, e.g., Anthony G. Greenwald et al., *Statistically Small Effects of the Implicit Association Test Can Have Societally Large Effects*, J. PERSONALITY & SOC. PSYCHOL. (forthcoming 2015) (responding to Oswald meta-analysis).

22. Gregory J. Meyer et al., *Psychological Testing and Psychological Assessment: A Review of Evidence and Issues*, 56 AMER. PSYCH. 128, 130 tbl.1 (2001).

<b>Medical Relationship</b>	<b>Pearson's r correlation</b>
Aspirin → reduced risk of death by heart attack	.02
Chemotherapy → surviving breast cancer	.03
Ever smoking → subsequent lung cancer in the next 25 years	.08
Alcohol use → premature birth	.09
Low-level lead exposure → reduced childhood IQ	.12

All these correlations are smaller than the  $r=.15$  found in the hostile meta-analysis.

How about a sports analogy? In the ninth inning, with a player in scoring position, who would you send to bat—an All-Star with a batting average of .320 or a benchwarmer who is four standard deviations worse, with a batting average of .220? That's easy! Of course, the All-Star! But reduced to Pearson's  $r$ , Robert Abelson has demonstrated that the difference in these batters' skills would predict a hit or no hit in a single at-bat, with an effect size of only  $r=.113$ .<sup>23</sup> As Abelson cautioned,

[T]he attitude toward explained variance ought to be conditional on the degree to which the effects of the explanatory factor cumulate in practice . . . . In such cases, it is quite possible that small variance contributions of independent variables in single-shot studies grossly understate the variance contribution in the long run.<sup>24</sup>

Finally, let me share an example adapted from Mahzarin Banaji and Tony Greenwald.<sup>25</sup> Imagine a law firm "tournament" that requires an up-or-out decision *each month*. An evaluation of this sort, taken monthly, sounds ridiculously harsh; however, let's soften that review process by positing that the chances of being promoted up (versus being kicked out) are exceedingly high: 99%. Finally, suppose that it takes eight years to make income partner, after which you can enjoy a life of high remuneration

---

23. See Robert P. Abelson, *A Variance Explanation Paradox: When a Little Is a Lot*, 97 PSYCH. BULL. 129, 132 (1985).

24. *Id.* at 133.

25. BANAJI & GREENWALD, *supra* note 3, at 203.

and repose. (Any law partner will tell you that this is not exactly how the tournament works.)

How would Greg fare versus Jamaal in this tournament, if implicit biases alter success rates each month by a measly 1%? In other words, suppose that implicit biases—which could manifest in terms of ingroup favoritism that makes it easier to find mentors and sponsors, benefits of the doubt in the evaluation of ambiguous performance, access to better work that provides learning and critical face time opportunities—improve Greg's chance of monthly promotion from 99% to 99.5%. By the way, Greg is White. Conversely, they depress Jamaal's chances from 99% to 98.5%. By the way, Jamaal is Black. At a glance, this does not seem like much of a difference—a testament to how far society generally and legal culture specifically have come in terms of civil rights.

But what happens when this 1% gap in treatment, actuated by implicit bias, is compounded monthly for eight years? In this stylized hypothetical (assuming the chances of promotion are independent probabilities each month), Greg's chance of surviving an eight-year tournament is 0.995 raised to the 96<sup>th</sup> power (12 months per year  $\times$  8 years), which equals 61.8%. By contrast, Jamaal's chance of survival is 0.985 raised to the 96<sup>th</sup> power, which is only 23.4%. This is a breathtaking difference. How did something so small (a 1% implicit bias penalty) change into something so big (a 38.4% difference in making partner)? It's the power of compound interest, which is why you should start investing in your 401(k) plans when you're young. The exact same phenomenon is taking place here. Small differences in treatment compounded repeatedly over time pay big dividends for Greg and big losses for Jamaal.

If this is the first time you have been exposed to the science of implicit bias, much of this will sound foreign, incredible, and frankly unbelievable. I encourage you to approach all of this skeptically and learn more. It's also undoubtedly true that if I paid you \$800 per hour, you should be motivated enough and smart enough to pick apart some aspect of each and every scientific study that I've mentioned. No piece of research is perfect. And, to repeat, I am inviting your skepticism. After all, skepticism is a virtue in science. It was, for example, skepticism that all of our attitudes and stereotypes were entirely explicit and introspectively available that helped prompt the research into implicit social cognitions in the first place. We want to approach new material with eyes wide open.

That said, skepticism should be even-handed. It should not be naïvely and selectively trotted out against politically inconvenient findings<sup>26</sup> and

---

26. See Adam Benforado & Jon Hanson, *Naïve Cynicism: Maintaining False Perceptions and Policy Debates*, 57 EMORY L.J. 499 (2008).

not against, for example, the psychological “common sense” assumptions littered throughout the status quo, most of which have never been empirically tested. In other words, if we demand metaphysical certitude of one set of empirical claims, we should demand the same certitude of all comparable empirical claims.<sup>27</sup> Both eyes, regardless of the direction we are looking (left or right), should be wide open.

What gives me cause for some optimism is that, over time, the operation of normal science should provide a clearer empirical picture. Gradually, we will get better measures of implicit biases. We will have more accurate estimates of their effect sizes, and better understand their underlying mechanisms. More important, I am confident that we will discover better interventions that either decrease implicit biases or prevent their manifestation into discriminatory behavior.<sup>28</sup>

## II. BEHAVIORAL REALISM

Having briefly summarized the recent science of implicit bias, it is now time to examine their implications for the law. For the past decade, I have been calling for a commitment to “behavioral realism” within law and policy.<sup>29</sup> This involves a simple three-step process:

1. *Identify new science.* As legal analysts and policymakers, we should regularly seek to identify more accurate models of human behavior and decision-making, as uncovered by the experimental sciences.

2. *Excavate old law.* Next, we should critically excavate the “common sense” assumptions embedded within the law. Although we may not notice it, within the law, there is always some model of human behavior and decision-making presumed.<sup>30</sup> These models appear in the Model Penal Code, doctrines of tort liability (negligence versus strict liability), cap-and-trade environmental regulations, simply everywhere. Without some such model or theory of human behavior, we could not understand, punish, deter, coordinate, or incentivize people.

---

27. See, e.g., Jerry Kang, *The Missing Quadrants of Anti-Discrimination: Going Beyond the “Prejudice Polygraph,”* 68 J. SOC. ISSUES 314, 319 (2012). (making this claim by comparing the admissibility of implicit bias evidence as compared to forensic evidence, such as bite mark and hair specimen analyses, after the National Research Council concluded that “there is a notable dearth of peer-reviewed, published studies establishing the scientific bases, and validity of many forensic methods.”).

28. For examples of recommendations to combat implicit bias in the courtroom, see Kang, *supra* note 7, at 1169–86.

29. For more on behavioral realism, see Kang & Lane, *supra* note 7, at 490; Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CALIF. L. REV. 997, 997–1008 (2006).

30. Krieger & Fiske, *supra* note 29.

3. *Account for the gap.* Finally, compare the two. Whenever science provides an improved model of human behavior, compare that to the common sense model baked into the law. If we see a sufficiently large gap between the two, we should pressure the law to take account. An accounting can be provided in one of two ways. On the one hand, the law can be reformed to reflect the new scientific findings. Why not update the law, if it is predicated on a factual mistake? Regardless of whether you value efficiency or fairness or both, this kind of accounting seems eminently sensible.<sup>31</sup>

On the other hand, society may have a set of value commitments that preclude such factual updating. For example, suppose we lived in a religious state that believed in the truth of a canonical religious text, in which it is written that the Earth is only 5,000 years old. Suppose that scientists uncover evidence that the Earth is far older than 5,000 years. A religious state, even if it generally practices behavioral realism, may nevertheless decide not to change its laws or policies based on this particular finding. Even if the same scientific techniques that dated the planet (e.g., carbon dating) are regularly used and accepted in other contexts, a religious state could nevertheless reject this specific application because of its conflict with the holy text. The only requirement is that the state do so transparently. Society must “own up” to its decision to deny the relevance of certain facts in pursuit of particular values.

Over the past decade, this commitment to behavioral realism has prompted a myriad of questions, analyses, answers, and recommendations in legal scholarship.<sup>32</sup> In a public lecture at NYU’s Straus Institute, I projected out the future of this research and catalogued what I called “The Grand Challenges for Implicit Social Cognition and the Law.”<sup>33</sup> In the remaining portion of this Meador Lecture, I will focus on just one of those grand challenges—Rethinking Intent and Impact—in the context of Equal Protection doctrine.

### III. THE EQUAL PROTECTION “MACHINE”

Let me start with a confession. I do not view myself as a constitutional law scholar. I neither revel in doctrinal minutiae nor enjoy total recall of

---

31. Of course, I do understand that constant legal revision would inflict administrative and other transaction costs. Everything in due moderation.

32. A Westlaw Next search conducted on September 08, 2014, in the “Law Reviews & Journals” database finds 210 hits for the exact term “behavioral realism.” The term “implicit bias” generated 1,175 hits.

33. See Jerry Kang, Address at The Straus Institute: The Grand Challenges of Implicit Social Cognition and the Law (Feb. 11, 2014) (video available at <http://www.nyustraus.org/events/Kangpublic.html>).

which justice made what claim, in what case, in which partial concurrence. For someone who is about to expound on constitutional law, this is a dangerous “bug” to concede.

But sometimes bugs can become features, and that is my gambit here. I want to approach Equal Protection doctrine like a stranger from a strange land, as if a martian landed in the United States and encountered a novel “machine” called Equal Protection doctrine. Curious, I want to open up the machine to try to figure out how it works, and what it does, by reverse engineering it. With minimal preconceptions about the machine’s operation, my challenge is to try to figure out its central purpose by examining its inputs, outputs, and moving parts. To keep the discussion manageable, I will keep the focus mostly on race.

### *A. Welcome to the Machine*

#### *1. Tracks of scrutiny*

If you have taken constitutional law in the past 20 years, your outline for Equal Protection doctrine probably looked something like this. First, there are tracks of scrutiny called “strict,” “intermediate,” and “rational.” When a state action is challenged for violating the Constitution’s guarantee of equal protection, you must select the appropriate track, then engage in the requisite means-ends analysis. For example, if you select the “strict scrutiny” track, you must demand a “compelling interest” that the state is pursuing through a technique deemed “narrowly tailored.” Although not impossible, in the context of race-based equal protection claims, this level of scrutiny is extremely hard to satisfy.

#### *2. Track selector*

How does one select the appropriate track? Any good outline will feature a track selector, a type of switch that shunts the analysis onto one of the three tracks. In race cases, there are two important triggers that constitute the switch.

##### *a. Trigger 1: Intent to Harm Minorities (T<sub>1</sub>)*

The first and classical trigger (T<sub>1</sub>) is the state’s *intent to harm minorities*.<sup>34</sup> Such intent is sufficient to track the case onto strict scrutiny.

---

34. I concede that this trigger might be unduly narrow. Arguably, it could be rephrased more generically and capaciously as intent to treat people differently (either better or worse) because of race

Arguably, this trigger was announced as early as the wartime internment cases, *Hirabayashi v. U.S.*<sup>35</sup> and *Korematsu v. U.S.*<sup>36</sup>—even though those cases failed tragically to apply in practice the strict scrutiny they encouraged in form.<sup>37</sup>

Conversely, if there is no such *intent to harm minorities*, then any collateral damage to minorities caused by the challenged state action is irrelevant.<sup>38</sup> Such harm will not trigger strict scrutiny. This is the conventional understanding of *Washington v. Davis*.<sup>39</sup> Disparate impact may be cognizable as a statutory matter,<sup>40</sup> but not under the Constitution's guarantee of equal protection.

*b. Trigger 2: Facial Racial Classification (T<sub>2</sub>)*

The second trigger (T<sub>2</sub>) is a *facial racial classification*, which is also sufficient to trigger strict scrutiny. Equally important, the fact that a facial racial classification is made with the intent to help, not harm, racial minorities does not matter. The Supreme Court told us so in cases such as

(whether they be minorities or White). I think most of my analysis remains sound even with such a substitution.

35. 320 U.S. 81 (1943). See *id.* at 100 (“Distinctions between citizens solely because of their ancestry are by their very nature odious to a free people whose institutions are founded upon the doctrine of equality. . . . [R]acial discriminations are in most circumstances irrelevant and therefore prohibited. . . .”). For historical context, see Jerry Kang, *Dodging Responsibility: The Story of Hirabayashi v. United States*, in *RACE STORIES* (Devon Carbado & Rachel Moran eds., 2008).

36. 323 U.S. 214 (1944). See *id.* at 216 (“[A]ll legal restrictions which curtail the civil rights of a single racial group are immediately suspect. . . [and] courts must subject them to the most rigid scrutiny.”).

37. See generally Jerry Kang, *Denying Prejudice: Internment, Redress, and Denial*, 51 UCLA L. REV. 933 (2004). Various scholars have tried to minimize or contextualize the harm that *Korematsu* did by pointing to *Endo*, a companion case, which held that indefinite detention of concededly loyal citizens was unlawful. See, e.g., Patrick O. Gudridge, *Remember Endo?*, 116 HARV. L. REV. 1933 (2003); Samuel Issacharoff & Richard H. Pildes, *Between Civil Libertarianism and Executive Unilateralism: An Institutional Process Approach to Rights During Wartime*, 5 THEORETICAL INQ. L. 1, 21–23 (2004) (adopting Gudridge’s interpretation). But as I’ve explained, *Endo* should not be seen as absolution for *Korematsu*. Instead, *Endo* was the Wartime Supreme Court’s *pièce de résistance* that allowed President Roosevelt and Congress to dodge all corporate responsibility for interning the Japanese Americans. See Jerry Kang, *Watching the Watchers: Enemy Combatants in the Internment’s Shadow*, 68 J. LAW & CONTEMP. PROBS. 255, 270–75 (2005).

38. See *Washington v. Davis*, 426 U.S. 231, 242 (1976).

39. *Id.* See also *Pers. Adm’r of Mass. v. Feeney*, 442 U.S. 256 (1979) (gender). For thoughtful and historically inflected analyses of these cases, see Ian Haney-Lopez, *Intentional Blindness*, 87 N.Y.U. L. REV. 1779 (2012); Reva Siegel, *Foreword: Equality Divided*, 127 HARV. L. REV. 1 (2013). Siegel sees *Feeney* as a decisive move “to restrict the ways that evidence of foreseeable impact could be used to prove unconstitutional purpose.” Siegel, *supra* at 17.

40. For example, Title VII permits causes of action for disparate *treatment*, in which the plaintiff must demonstrate an adverse employment action “because of” the plaintiff’s “race, color, religion, sex, or national origin.” 42 U.S.C. § 2000e-2(a)(1) (2012). It also permits causes of action for disparate *impact*, in which the plaintiff challenges facially neutral policies that produce a disparate impact on protected groups. See Civil Rights Act of 1991, 42 U.S.C. § 2000e-2(k); *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971).

*City of Richmond v. J.A. Croson Co.*<sup>41</sup> and *Adarand Constructors, Inc. v. Peña*.<sup>42</sup> Good intentions, which may motivate affirmative action programs that employ facial racial classifications, do not prevent strict scrutiny.

In sum, there are two separate and independent triggers.  $T_1$  is sensitive to *intent to harm minorities*;  $T_2$  is sensitive to *facial racial classification*. Either trigger alone is sufficient to slot the case onto strict scrutiny. Interestingly, both triggers ignore certain facts as constitutionally irrelevant. To repeat,  $T_1$  ignores harm to minorities in the form of collateral damage or disparate impact (and detects only the intent to harm minorities);  $T_2$  ignores the intent to help minorities (and detects only the existence of facial racial classifications).

### B. Reverse Engineering

Now, imagine that you have just landed from Mars, and stumble across this doctrinal “machine.” You are told that it’s a very special machine, and are provided a minimal user manual—essentially, the above description. How would you learn more about the machine—its purpose and design? Suppose you are also a tinkerer. No one is looking. Why not crack the outer casing open, take a peek inside? Identify its inputs and outputs, examine its sensors and algorithms, and reverse engineer the machine?

#### 1. *Function 1: A Sensor for Concealed Intent to Harm Minorities (F<sub>1</sub>)?*

Let’s start with the crucial algorithm that is called means-ends scrutiny. What is it, and what is it used for? Why do we inspect the state’s interests against some normative standard: compelling, important, or lawful? Why do we inspect the state’s chosen methods for narrow, intermediate, or rational tailoring? And why strict scrutiny in particular, which requires compelling interests and narrow tailoring?

One classical explanation<sup>43</sup> for strict means-ends scrutiny is that it acts as a sensor to help detect something illicit and concealed. What exactly? Its function is to smoke out concealed (but in psychological terms, quite explicit) intent to hurt minorities.<sup>44</sup> Let’s call this the  $F_1$  hypothesis.

41. *Id.*

42. 515 U.S. 200 (1995).

43. This theory was popularized by John Hart Ely. See JOHN HART ELY, *DEMOCRACY AND DISTRUST* 146–50 (1980).

44. “Explicit” means that the attitude and stereotype is available through direct introspection. In other words, it is privately accessible. “Concealed” means that these explicit social cognitions are, however, not shared with others, including the court.

How might the “narrow tailoring” requirement help sense what has been purposely concealed? If some government technique is poorly tailored to the proffered governmental interest, then a judge has reason to be skeptical. She might ask:

Wait, your goal was to get to this point X, but you got there through this weird, inefficient detour through Y and Z? Well, given that there are so many more direct and efficient routes to X, I’m not so sure that’s where you truly intended to go. Maybe you actually wanted to visit Y or Z. Maybe you’re lying. In any event, try again.

This reasoning helps explain why the “narrow tailoring” requirement can help act as a sensor for concealed intent to harm minorities.

What about the “compelling interest” requirement? This requirement acts as a backstop for the above analysis. If there is no independent constraint on what state interests are permitted, then the state could manufacture a state interest that enjoys a tautologically perfect fit. In other words, the state could simply announce in court that their interest was precisely to take a path through Y and Z. Then lo and behold, their adopted mechanism—a path that travels through points Y and Z—would be perfectly (not just narrowly) tailored to that specific state interest. To prevent such gaming, a court could inquire: “What interest do you have in following some route through Y and Z?” If there is no independently compelling reason, a court could become skeptical about the state’s true motivation.

In sum, by demanding both an important destination (the compelling interest) and an exceedingly direct path towards that destination (narrow tailoring), the Equal Protection Machine serves a particular function,  $F_1$ : *It acts as a sensor that detects concealed intent to harm racial minorities.*

## 2. *Puzzling Triggers*

As martian tinkerers, we might now have a smile on our face. We have a working hypothesis on what this equal protection “machine” does. It’s like a Geiger counter, which detects radiation that is invisible to unaided human senses. But now let’s examine the track selector, with its two triggers.

### a. *Reconciling $T_1$ : The Problem of Pointless Redundancy*

Upon casual inspection, this  $F_1$  “sensor” function of strict scrutiny seems to plug and play nicely with  $T_1$  of the track selector. Recall this first

trigger: if your *intent is to harm minorities*, you will receive strict scrutiny. Building in such a trigger makes sense because as the saying goes, “Where there’s fire, there’s fire.” In other words, let’s use “intent to harm minorities” as a trigger to turn on a sensor that smokes out concealed “intent to harm minorities.” Makes sense.

Actually, this doesn’t make any sense at all. The correct saying is, “Where there’s *smoke*, there’s fire.” If you already know there’s smoke (i.e.,  $T_1$  has triggered because there is an intent to harm minorities), why would you use an elaborate means-ends scrutiny to look for more *smoke* (i.e. intent to harm minorities) when you have already found it? This is circular. Perhaps pointlessly redundant. It’s like having a Geiger counter that automatically turns on when ionizing radiation hits a frequency of 100 counts per minute. And what does it do after turning on? It measures the environment to sense whether there is ionizing radiation at the frequency of 100 counts per minute.

The  $F_1$  hypothesis (understanding the function of strict scrutiny to be a sensor that detects concealed intent to harm minorities) makes  $T_1$  (the trigger based on intent to harm minorities) redundant. Martian engineers do not design their machines to have useless, redundant parts. True, sometimes redundancy is built-in, as a fail-safe. But the redundancy between  $T_1$  and  $F_1$  seems more fundamental and problematic than the redundancy of airbags and seatbelts, or even belts and suspenders.<sup>45</sup>

Hmmm. Maybe  $F_1$  isn’t really the function of the Equal Protection Machine after all.

---

45. One could object that there really isn’t a redundancy because if  $T_1$  is satisfied, there isn’t any subsequent means-ends scrutiny. In other words, if there’s “intent to harm minorities” that has caused discrimination, then that’s the end of the story: there’s no actual evaluation of compelling interests and narrow tailoring. After all, if the intent is to harm minorities, how could that amount to a lawful much less compelling interest? But listen to how the court describes racial Equal Protection doctrine in *Feeney*: “A racial classification, regardless of purported motivation, is presumptively invalid and can be upheld only upon an extraordinary justification. . . . *This rule applies as well* to a classification that is ostensibly neutral but is an obvious pretext for racial discrimination.” *Pers. Adm’r of Mass. v. Feeney*, 442 U.S. 256, 272 (1979) (citing *Yick Wo v. Hopkins*, 118 U.S. 356 (1886)) (emphasis added).

We have seen means/ends-like strict scrutiny analysis being conducted on the basis of something like  $T_1$  being fired in related doctrinal areas ranging across districting, gender discrimination, and religion. *See, e.g., Bush v. Vera*, 517 U.S. 952, 993 (O’Connor, J., concurring) (explaining that strict scrutiny is appropriate after a finding that districting was caused “predominantly due to the misuse of race”); *Feeney*, 442 U.S. at 273 (“[T]hese precedents dictate that any state law overtly or covertly designed to prefer males over females in public employment would require an exceedingly persuasive justification to withstand a constitutional challenge under the Equal Protection Clause of the Fourteenth Amendment.”); *Church of the Lukumi Babalu Aye, Inc., v. City of Hialeah*, 508 U.S. 520, 533 (1993) (suggesting that facially neutral statutes adopted with an intent to harm religious minorities could be constitutional if they survive strict scrutiny). For a perceptive analysis of strict scrutiny, see Richard H. Fallon, Jr., *Strict Judicial Scrutiny*, 54 *UCLA L. REV.* 1267, 1310–11 (2007).

*b. Reconciling T<sub>2</sub>: Not about Protecting Minorities*

What about that other trigger? Unfortunately, F<sub>1</sub> makes no greater sense when it is examined in light of T<sub>2</sub>, the trigger based on facial racial classifications. Most obviously, given that the racial classification is facial, nothing's really concealed. Race is front and center. Moreover, in the real world, T<sub>2</sub> is always triggered in the context of White plaintiff challenges to affirmative action programs that are clearly designed to *help*, not hurt, minorities. Even if you think these programs are poorly designed, subject to gaming, create more costs than benefits, and even end up harming those who are their immediate beneficiaries, it still remains a stretch to argue that they were adopted with the intent that minorities suffer.<sup>46</sup>

But maybe I'm being naïve by presuming that just because something is called an "affirmative action" or "diversity" program, it could not have been designed to hurt minorities. Isn't this precisely what Justice O'Connor cautioned against in *Adarand*?

We think that requiring strict scrutiny is the best way to ensure that courts will consistently give racial classifications that kind of detailed examination, both as to ends and as to means. *Korematsu* demonstrates vividly that even "the most rigid scrutiny" can sometimes fail to detect an illegitimate racial classification . . . . Any retreat from the most searching judicial inquiry can only increase the risk of another such error occurring in the future.<sup>47</sup>

From O'Connor's perspective, *Korematsu* teaches us that we are so incompetent at detecting concealed intent to harm minorities that we must use strict scrutiny (as a sensor) whenever facial racial classifications are made.<sup>48</sup> Never trust your gut. Instead, use the F<sub>1</sub> function of the strict scrutiny machine!

This lesson is hard to take seriously for many reasons. Here's just one: The same swing Justices who designed and implemented T<sub>2</sub> (demanding strict scrutiny lest we authorize more concentration camps<sup>49</sup>) are entirely

---

46. The only clear counterexample is when affirmative action programs for underrepresented minorities are intermixed with negative action programs against Asian Americans. See generally Jerry Kang, *Negative Action Against Asian Americans: The Internal Instability of Dworkin's Defense of Affirmative Action*, 31 HARV. C.R.—C.L. L. REV. 1 (1996).

47. *Id.* at 236.

48. *Id.*

49. I realize that this is an emotionally loaded term. But President Franklin D. Roosevelt, Dillon Myer, who directed the War Relocation Authority, and the U.S. Supreme Court Justices in the *Korematsu* case referred to the Japanese American "relocation centers" as "concentration camps." See ERIC YAMAMOTO ET AL., RACE, RIGHTS AND REPARATION: LAW AND THE JAPANESE AMERICAN INTERNMENT 168 (2d ed. 2013). If most of us have never learned, how can we never forget?

copacetic with programs that Kim Forde Mazrui calls “alternative action.”<sup>50</sup> Alternative action programs function like affirmative action programs—with the same goals and purposes—but have been designed by lawyers sophisticated enough to avoid facial racial classifications. The Supreme Court has repeatedly blessed alternative action programs, such as the Top Ten Percent plan adopted in Texas,<sup>51</sup> which increases minority student enrollment by leveraging the brute fact of residential, and thus racial segregation in public schools.<sup>52</sup>

But think on this. The fundamental justification for installing  $T_2$  was that we are unable to distinguish between the underlying intent behind concentration camps and campus diversity. But if we are so psychologically clueless, why should anything change simply because clever lawyers have coached bureaucrats how to reach the same results, without facial racial classifications? Evil state actors can pursue their sinister ends without such crude devices. Remember, San Francisco city officials targeting Chinese laundrymen needed no facial racial classification in *Yick Wo v. Hopkins*.<sup>53</sup> How do we know that the Top Ten Percent plan is not *Yick Wo*?

In sum, if we take  $T_2$  (facial racial classification) seriously—in particular its fetishization of the facial classification—we have more reason to think that  $F_1$  (sensing concealed intent to harm minorities) cannot be the true function of Equal Protection doctrine.

### 3. *Function 2: A Shield to Protect Minorities against Intentional Harm ( $F_2$ )*

Let’s take a fresh look at means-ends scrutiny to discover some other function. Yes, on the one hand, it does seem to work as a sensor that helps detect concealed intent to harm minorities. On the other hand, maybe means-ends scrutiny is less about information forcing (improving sensing powers) and more about substantive justification. After all, by demanding a compelling interest, strict scrutiny has the effect of putting up a higher substantive bar on certain forms of state action. Whereas normally any old rational state interest will do, when strict scrutiny is applied, a far more normatively important and empirically urgent need must be demonstrated. The benefits must be compelling.

---

50. Kim Forde Mazrui, *Alternative Action*, 13 MICH. J. RACE & L. 461 (2008).

51. *Fisher v. Univ. of Tex.*, 133 S. Ct. 2411, 2415 (2013).

52. In 1997, while George W. Bush was governor, Texas enacted a law guaranteeing admission to any state university to all students graduating in the top ten percent of their high school class. TEX. EDUC. CODE § 51.803(a) (2014).

53. 118 U.S. 356 (1886).

In addition, whereas typically any half-assed legislative fix will be good enough for government work, in some cases, the state will have to show far superior craftsmanship, avoiding the harms and costs of over-inclusiveness and under-inclusiveness, as well as demonstrating that no other readily available engineering option could have solved the problem better. Again, this higher quality control standard puts up a substantive bar against sloppy or unmindful state action.

Seen in this light, the function of strict means-ends scrutiny is less to be an information sensor and more to function as a substantive shield against certain types of actions. As analogies, think about a thermal shield that protects critical and fragile components from overheating. Think about a firewall that protects sensitive processors and storage from malware and viruses.

What precisely is being shielded from what? Well, even a martian can learn some US history, and understand where the Reconstruction amendments came from. Maybe the function of strict scrutiny is simply to shield racial (and other numerical) minorities from harm inflicted by a majoritarian state intending to inflict just such harm. Let's call this the  $F_2$  hypothesis.

On this view, strict scrutiny throws up a skeptical cost-benefit analysis requirement that demands especially high benefits (compelling interest) and especially good design (narrowly tailored), minimizing the costs experienced by minorities. On this view, the principal function of strict scrutiny is not to act as a sensor that detects concealed intent to harm racial minorities—although, this might be a side benefit. Instead, it functions substantively, to shield minorities from intentional harm.

Notice how  $F_2$  interfaces seamlessly with  $T_1$ . If the function of strict scrutiny is to shield minorities from intentional harm ( $F_2$ ), it makes all the sense in the world to throw up the shield when you have identified an intent to harm minorities ( $T_1$ ).

#### *4. Function 3: Avoiding Racial Balkanization ( $F_3$ )*

But does  $F_2$  also jive with the other trigger,  $T_2$ ? Unfortunately, no. Recall that the second trigger is the use of facial racial classifications, which arises almost always in the context of race-conscious affirmative action programs. But, as explained above, there is little reason to think that we need a substantive shield to protect racial minorities from race-conscious affirmative action plans.

Maybe, then, there is actually a third substantive function served by the Equal Protection Machine. Maybe the modern function of equal protection is to prevent racial balkanization. In other words, the doctrine requires a higher substantive bar for state actions that will have the effect of creating

interracial strife, division, or resentment. I'm drawing heavily from recent work by Reva Siegel, who has made a powerful case that the swing Justices are concerned most precisely with this.<sup>54</sup>

To be clear, this is not  $F_1$ , a sensor function looking to uncover concealed intent to harm minorities. Also, it is not  $F_2$ , a shield that protects minorities from intentional harm by the state. Rather, it is a third function,  $F_3$ , that acts as a shield that protects the American people, including Whites, from state actions that threaten racial balkanization.

This interpretation makes sense of  $T_2$  under the plausible empirical assumption that facial racial classifications pose a greater risk of balkanization than alternative action programs that produce similar consequences, but are less "in-your-face." Certainly, Justice Kennedy's commentary in *Parents Involved in Community Schools v. Seattle School District No. 1*<sup>55</sup> reflects this view. Arguably, this functional interpretation also makes sense of  $T_1$  because state action that is intended to harm racial minorities risks racial balkanization as well.

\* \* \*

Thinking like a martian tinkerer has brought us to this point. We found something called an Equal Protection "Machine." We studied its internal parts, and tried to rationalize how they might work together. We started with a reasonable hypothesis, but ultimately were forced to reject  $F_1$ —equal protection strict scrutiny is not about sensing concealed intent to harm minorities.

Instead, the function of equal protection is to act as a substantive shield that is especially sensitive to particular harms. If we highlight  $T_1$ , which was the first trigger historically to be implemented, those harms are intentional harms suffered by racial minorities. But, in modern times, these equal protection cases have gone largely extinct. Frankly, it's hard to find state action that is being given strict scrutiny because of  $T_1$ . Instead, most of the action is in  $T_2$ , which was more recently designed and implemented, in the context of striking down affirmative action programs that utilized facial racial classifications. If we highlight this trigger, equal protection functions as a substantive shield to protect against what a majority of the Supreme Court believes to threaten racial balkanization.

---

54. Reva B. Siegel, *From Colorblindness to Anti-Balkanization: An Emerging Ground of Decision in Race Equality Cases*, 120 YALE L.J. 1278, 1282 (2011) (pointing out three blocks of Justices, including one that is anxious to "preserve social cohesion"). See *id.* at 1297 (adopting the term "balkanization" as a third theory, distinguished from anti-classification and antisubordination).

55. 551 U.S. 701 (2007). See *infra* note 67.

## IV. REDESIGN

Here's the recap so far. In Part I, I briefly summarized findings from implicit social cognition that reveal that implicit biases exist, are pervasive, strong in magnitude, and influence behavior in real-world settings. In Part III, I explained why the modern Equal Protection Machine functions not as an information sensor, but as a substantive shield. In the past, that shield was more concerned with protecting racial minorities from state action intending to harm them. In modern times, the shield ostensibly protects all Americans, including especially the racial majority, from state action that risks racial and ethnic balkanization. In this Part IV, my task is to see what the science of implicit social cognition has to say about this technologically deconstructed machine, in light of my commitment to behavioral realism outlined in Part III.

*A. Redesign T<sub>1</sub>: From Intent to Causation*

Nearly every lefty worth her weight in salt protests *Washington v. Davis*.<sup>56</sup> If courts insist on looking for such evil intent before selecting strict scrutiny, they will never find it. This is not only because such animus, when it exists, is kept carefully under wraps. It is also because there is almost always some plausible motivation for government action that is not sadistic racism. Accordingly, T<sub>1</sub> has fallen into desuetude.

What might a behavioral realist redesign of T<sub>1</sub> look like? The first question is why are we focusing only on *explicit* intent to harm minorities? We have long known that an explicit intent to harm is not necessary to produce harm. Ever since the first year of law school, we were sensitized to the existence of multiple mental states, such as those identified in the Model Penal Code: purposeful, knowing, reckless, and negligent.<sup>57</sup> Each one of these mental states can produce devastating consequences. Negligent driving regularly kills.

More to the point, the science of implicit social cognition teaches us *implicit* cognitions can be the “but for” cause of benefit to some groups (e.g., ingroups) and harm to other groups (e.g., outgroups). Even though many of us embrace colorblindness as a moral ideal, we nevertheless behave in ways that are unwittingly race conscious. We are not colorblind, at least implicitly. Instead, we are afflicted with blindsight. In this sense, race remains a “but for” cause for discriminatory behavior. Why should constitutional law not care about these harms? If racial minorities suffer in

---

56. 426 U.S. 229 (1976) (requiring proof of discriminatory purpose in order to trigger strict scrutiny of facially neutral state action).

57. MODEL PENAL CODE § 2.02(2)(a)–(d) (2001).

part because of attitudes and stereotypes that most of us have, why should we care that they are not directly accessible through introspection but are measurable by scientific instruments?

To be crystal clear, I'm not talking about racial discrimination in the distant past crying out for corrective justice—stories of slavery, conquest, exclusion, dispossession, and internment. There is reason to think that people don't care much about the distant past. That was then, this is now. Also, I'm not talking about some amorphous "societal discrimination" that amounts to a brooding cloud of White guilt that forever justifies wealth transfers to undeserving people of color without limitation.<sup>58</sup> Instead, I'm talking about implicit bias-actuated discrimination that is measurable, here and now.<sup>59</sup>

One reason to be wary might be because the causal relationship between implicit bias and discrimination is hard to detect in any specific case and feels probabilistic. I agree. But, being probabilistic is not the same thing as being amorphous or without constraint. Remember, quantum mechanics is probabilistic as well, but that does not make that branch of physics undisciplined, unrigorous nonsense. Cancer risk from radiation exposure is also probabilistic, but again, that does not make cancer risk invisible to tort law.

Another reason to be wary might be the fear that we will find constitutional violations everywhere.<sup>60</sup> I understand this anxiety. But, jurists should not forget how creative they can be, to forge pragmatic doctrines.<sup>61</sup> Most important, let's distinguish between disparate impact *alone* and disparate impact *tied causally* (albeit probabilistically) to implicit biases that have been shown to influence behavior in similar domains. Suppose that scientists eventually develop a set of best practices that materially mitigate such implicit bias-actuated harms within some specific institutional context. Assume further that this knowledge is widespread and the practices become widely adopted. When a specific state actor nevertheless declines to adopt these best practices, is it so implausible that upon constitutional challenge something beyond "rational basis" review should be activated? Would it be that difficult to redesign the appropriate

---

58. But this was the "societal discrimination" that the Supreme Court rejected in *Wygant v. Jackson Board of Education*, saying that "[s]ocietal discrimination, *without more*, is too amorphous a basis for imposing a racially classified remedy" because a "court could uphold remedies that are ageless in their reach into the past, and timeless in their ability to affect the future." *Wygant v. Jackson Bd. of Educ.*, 476 U.S. 267, 276 (1986) (plurality opinion) (emphasis added).

59. Jerry Kang & Mahzarin Banaji, *Fair Measures: A Behavioral Realist Revision of 'Affirmative Action'*, 94 CALIF. L. REV. 1063 (2006) (emphasizing the value of a presentist framing).

60. See *Washington v. Davis*, 426 U.S. at 248.

61. For general discussion about multiple quadrants of legal intervention, see Kang, *supra* note 27, at 317–25.

trigger, the appropriate track of scrutiny, with modest forward-looking injunctive remedies?

*B. Abandon  $F_1$ : Stop Using It to Justify  $T_2$*

The  $F_1$  hypothesis is that the function of the equal protection machine is to sense concealed intent to harm minorities. I've already made the case why that hypothesis should be rejected.<sup>62</sup> Nevertheless, courts still regularly invoke this explanation to justify specific components of the Equal Protection Machine. Specifically, it is used to justify  $T_2$ , which states that any facial racial classification must trigger strict scrutiny. We must never forget how we interned 120,000 Japanese Americans.

I would like the Court to stop using this rhetorical move. Personally, I find it an offensive exploitation of one of the most grotesque violations of civil rights in modern American history. But my idiosyncratic preferences are neither here nor there. Moreover, you should be asking what does this have to do with implicit social cognition?

Well, to the extent that courts continue to repeat the  $F_1$  trope, behavioral realism allows me to call their bluff. If there is such a serious desire to adopt better algorithms and instruments to “sense” different forms of concealed explicit intent, why not embrace implicit social cognition, which deploys advanced scientific instruments that probe even deeper beneath the surface? Why only search for concealed explicit bias using a machine kludged together by lawyers in the 20<sup>th</sup> century when we could also look for subterranean implicit bias and their “but for” behavioral manifestations via instruments designed and tested by scientists in the 21<sup>st</sup>?

Of course, the answer may be that we don't really care about all forms of race-contingent discrimination. We only care about race-contingent discrimination that is caused by explicit attitudes and stereotypes, not implicit ones. But this stance, which makes much of the sharp distinctions between explicit and implicit social cognitions, seems inconsistent with our unwillingness to draw distinctions elsewhere in the doctrine, for example, distinguishing between benign and malign facial race classifications. From a psychological perspective, why would we ignore the difference between benign and malign, but make all the difference between explicit and implicit? When we proclaim to be desperate to engineer the best detection machine possible, why would we then immediately sabotage our own efforts and say that we don't want the sensor to be that good, to pick up subtle, probabilistic phenomena in the form of implicit bias-actuated discrimination?

---

62. See *supra* Part III.B.2.

*C. Explore F<sub>3</sub> Balkanization*

Upon reverse engineering the Equal Protection “Machine,” we came to the conclusion that its dominant modern function is to act as a substantive shield that essentially throws up a weighted cost-benefit analysis requirement. Back in the day, that shield emphasized (or overweighed) intentional harm to minorities. But today, the shield emphasizes balkanization threats to the American polity. In practice, this means that courts show tremendous solicitude to those who, from their subjective point of view, resent preferential treatment given to undeserving people, especially members of outgroups.

If this is indeed how the machine operates, legal analysis would benefit from greater behavioral realism about what causes balkanization and what mitigates it. Here are a few questions:

- What is balkanization precisely?
- How can we measure it?
- How does it differ from resentment or negative attitudes toward other groups?
- What’s the relationship between balkanization and explicit bias, or implicit bias?
- Can the same program increase balkanization from the perspective of some groups, but decrease it from the perspective of others?
- How strong is the relationship between facial racial classification and balkanization? Are there mediating variables?
- Do alternative action programs balkanize less? Or might honesty be the better policy?<sup>63</sup>
- What matters more: form or substance?<sup>64</sup>
- How strong is the relationship between lack of critical mass and balkanization?
- Do speech codes, diversity policies, or diversity trainings decrease or increase balkanization?

---

63. *Cf. Gratz v. Bollinger*, 539 U.S. 244, 305 (2003) (Ginsburg, J., dissenting) (“If honesty is the best policy, surely Michigan’s accurately described, fully disclosed College affirmative action program is preferable to achieving similar numbers through winks, nods, and disguises.”).

64. *Cf. Shaw v. Reno*, 509 U.S. 630, 647 (1993) (“[W]e believe that reapportionment is one area in which appearances do matter.”). Siegel, *supra* note 54, at 1334 (“From the vantage point of the antibalkanization principle, which is concerned with the lived experience of social relations, appearances matter centrally.”).

- What is the relationship, if any, between the amount that traditional merit standards are sacrificed and the amount of resentment created?
- Does the narrative of justification alter the amount of balkanization? What happens when we talk about historical injustice, White privilege, implicit bias, or neuroscience?
- Does exposure to “debiasing agents”<sup>65</sup> alter balkanization?
- What are the short-term versus long-term trade-offs? In other words, might some programs that increase balkanization in the short-term actually decrease balkanization in the long-term, or vice versa?
- What are the local versus global trade-offs? In other words, might some programs that increase balkanization in the local environment actually decrease balkanization in the global environment, or vice versa?

We know so very little.<sup>66</sup> But the Court has already created constitutional doctrine on the basis of its common sense assumptions about the nature and causes of balkanization. For example, in *Parents Involved*, Justice Kennedy wrote:

If it is legitimate for school authorities to work to avoid racial isolation in their schools, must they do so only by indirection and general policies? Does the Constitution mandate this inefficient result? Why may the authorities not recognize the problem in candid fashion and solve it altogether through resort to direct assignments based on student racial classifications? So, the argument proceeds, if race is the problem, then perhaps race is the solution.

The argument ignores the *dangers* presented by individual classifications, dangers that are *not as pressing* when the same ends are achieved by *more indirect* means. . . . Governmental classifications that command people to march in different

---

65. See Kang & Banaji, *supra* note 59, at 1109 (defining debiasing agent as “an individual with characteristics that run counter to the attitudes and/or the stereotypes associated with the category to which the agent belongs.”).

66. To deepen our scientific understanding, see Linda R. Tropp, et al., *The Use of Research in the Seattle and Jefferson County Desegregation Cases: Connecting Social Science and the Law*, 7 *Analyses of Soc. Issues & Public Policy* 93 (2007). For an early, thoughtful intervention in the law reviews, see Linda Hamilton Krieger, *Civil Rights Perestroika: Intergroup Relations after Affirmative Action*, 86 *CAL. L. REV.* 1251 (1998).

directions based on racial typologies can cause a new divisiveness. The practice can lead to *corrosive discourse*, where race serves not as an element of our diverse heritage but instead as a bargaining chip in the political process. On the other hand race-conscious measures that do not rely on differential treatment based on individual classifications present these problems to a *lesser degree*.<sup>67</sup>

Whether something is a “pressing danger” is at least partly an empirical question. Whether some action is viewed by the relevant audience as “indirect” and how much that matters are empirical questions. Whether certain discourse is “corrosive” and to what “degree” are also empirical questions. A behavioral realist would not indefinitely trust gut feelings to answer such questions. To steal a line from a Republican President, “trust, but verify.”<sup>68</sup>

Sometime in the future, an amicus brief will be filed with the scientific consensus on what causes balkanization, and what does not.<sup>69</sup> Maybe there will be no consensus. Maybe it will be just a muddle. But one could imagine findings that run contrary to the common sense assumptions so far expressed. What will the Court do? Will it reform constitutional doctrine accordingly? Or, will it simply insist on its own fundamental faiths?

As behavioral realists, our job is to hold lawmakers, including judges, to account. And we will know them by their deeds.

---

67. *Parents Involved in Cmty. Sch. v. Seattle Sch. Dist. No. 1*, 551 U.S. 701, 796–97 (2007) (Kennedy, J., concurring in part and concurring in judgment) (emphasis added).

68. Remarks on Signing the Intermediate-Range Nuclear Forces Treaty, 2 PUB. PAPERS 1455 (Dec. 8, 1987).

69. This was already done in a brief I co-authored with Stuart Banner and Rachel Godsil, with respect to stereotype threat literature in the *Fisher* case. See Brief of Experimental Psychologists as Amici Curiae in Support of Respondents, *Fisher v. Univ. of Tex.*, 133 S. Ct. 2411 (2013) (No. 11-345). Although signed onto by leading stereotype threat scholars, such as Claude Steele, no justice cited the brief.