RESPONSE

BEHAVIORAL REALISM ABOUT COLOR CONFUSION[†]

JERRY KANG[*]

CONTENTS

INTRODUCTION

In *Colorblind Capture*, Jonathan P. Feingold identifies an important error made by the Left. Even as it defends race conscious affirmative action, the Left has often acquiesced to the Right's framing of affirmative action as "racial preference," with predictable political and legal consequences. Feingold offers an alternative framing that better resists this "colorblind capture." By carefully explicating how and when race matters—before, during, and after admissions— he explains how we might rebrand affirmative action as a countermeasure, a corrective, indeed, an affirmative obligation, potentially mandated by antidiscrimination law itself.

Feingold's argument is sound, his analysis trenchant, and his recommendations thoughtful. He reminds us that even as the Left pragmatically engages advocacy in venues that have adopted a colorblind frame, the Left should do so strategically, with reservations, under a form of intellectual protest, so as not to reinscribe a "racial preference" narrative. I start my Response by wondering why the Left is so easily captured. Perhaps it's because in between the two competing theories Feingold lays out—*colorblindness* (on the Right) and *colorconsciousness* (on the Left)—there sits a muddled theory in the middle I call *color confusion*.

I.   COLOR CONFUSION: IN THE MIDDLE

A.   *Racial Relevance*

One of Feingold's theoretical contributions is his clearly identifying four elements that describe the colorblindness versus colorconsciousness approaches. The first element is "race = irrelevant." The colorblind approach sees race formally, as an "irrelevant biological fact or physical attribute."[1] By contrast, the colorconscious approach believes that "race ≠ irrelevant," that race has mattered structurally in the past and continues to matter today even though it is a social construct.[2]

For those admitting to *color confusion*, they agree with both approaches because they *distinguish between* "is" and "ought." On the one hand, descriptively, race *does* matter (for all the historical and present-day reasons commonly articulated). On the other hand, normatively, race *should matter less*, especially in the sense of decreasing discrimination based on race (i.e. negative treatment caused by negative attitudes or stereotypes against racial groups). For many in the *color confused* camp, race should also matter less in the identity politics sense, such that when people understand themselves, their racial category should be low on the ordinal list of attributes specified.

---

[1] Jonathan P. Feingold, *Colorblind Capture*, 102 B.U. L. REV. 1949, 1958 (2022).

[2] *Id*. at 1961.

B. *Racial Neutrality*

Feingold's second element is "facially race neutral = race neutral" for the colorblind, whereas the colorconscious suggest that "facially race neutral ≠ race neutral." For the *color confused*, again both sides seem right, and the answer depends on, among other things, the choice of scope in terms of (1) timeframe and (2) unit of analysis for the rights-holder.

On the one hand, if we focus narrowly on the timeframe *now* and see the *individual* as the relevant holder of rights, and thus, the proper unit-of-analysis, facial race-neutrality is typically assumed to be race-neutral. Think about an individual applying for a college loan from a private bank. Suppose that the application process is facially race neutral and for good measure also blinds the application of all race information (including potential inferences of racial identity). The ultimate bank decision, including specific financial terms, will reasonably be called race-neutral.

On the other hand, as soon as we zoom out to consider the *past*, things get murky. For instance, we could increase the aperture slightly to consider just the recent past in the individual's own life. What happens if we see examples of discrimination in employment or the issuance of speeding tickets that lowered a credit score. We could zoom out quite broadly and consider the distant past and notice how racial discrimination prevented the intergenerational accumulation of wealth and social capital. Finally, if we zoom out beyond the individual to consider the entire racial *group*, facial race neutrality will seem still less race-neutral. To put a fine point on it, if for centuries a group were held in bondage and kept uneducated and then suddenly freed, applying a facially neutral, race-blind loan application process could hardly be called race-neutral.

C. *Racial Preference*

The third element is that being "race-conscious = racial preference" for the Right and that "race-conscious ≠ racial preference" for the Left. These views are natural entailments from each theory's second element. If you think that facial race neutrality is indeed neutral, then race consciousness breaks that baseline and seems like a preference. By contrast, if you think that facial race neutrality is not in fact neutral, then race consciousness will seem like a corrective that restores neutrality.

For the *color confused*, it's complicated and depends on the degree of departure that the race-conscious intervention represents. Leaning left, even if the proper unit of analysis is the individual (not the group), a well-calibrated departure from facial race neutrality that counters the ways in which race disadvantages now or has recently disadvantaged the individual will be seen as a countermeasure (not a preference). Moreover, if we shift the unit of analysis to the entire group, then even if in any individual case a race-conscious departure might be too large or too small, as long as the aggregate magnitude of corrective turns out to be roughly comparable to the aggregate magnitude of disadvantage suffered, race consciousness can be seen as a group-tailored countermeasure and not a preference.

But the color confused can also lean Right. If the departure seems too drastic, race consciousness will be viewed as overcorrection, which is not an outright "preference" but is still disfavored. Also, those who lean Right tend to anchor to the individual as the proper unit of analysis because they understand racial justice in terms of an individual's rights and not anything like a guarantee of demographically proportional outcomes across racial groups. They find little comfort in being told that individual-level errors wash out when summed over the larger aggregate.

D.   *Racial Hierarchy*

Feingold's final element is that for the colorblind, "racial hierarchy = natural & legitimate" and for the colorconscious, "racial hierarchy ≠ natural & legitimate." The *color confused* find themselves again somewhere in the middle. They view racial disparities in status and power as having multiple causes, not all of them tightly or recently connected to racial discrimination. They will often speculate that hierarchies arise from groups having different preferences, aptitudes, immigration trajectories, and random path dependencies that are inflected with race but not entirely determined by race. (Some will, to be brutally honest, also assume average biological and/or nearly fixed cultural differences, e.g. "Asian students do badly in law school because they weren't raised to be argumentative, but Jewish students are just the opposite"—a sentiment I've heard a colleague express.)

They will also concede that racist history and structures matter but don't know what to do about it. In particular, they worry about who should shoulder the specific burdens that might be triggered in relatively zero-sum games, such as college admissions for their children. In sum, the *color confused* respond to racial hierarchy with concern but also doubt that such hierarchy always and totally reflects race discrimination. Furthermore, the *color confused* don't know how to distribute the burdens associated with undoing racial hierarchy.

*****

Following Feingold's schematic, here then is my descriptive account of the *color confused*:

Race *is* relevant although it generally *should not be*.

Facially race neutral is more race neutral when we focus on *the individual, right now*. It's less race neutral when we think about the *entire group and the past*.

Race consciousness is *less* a race preference if the intervention is well-calibrated to counter race-caused disadvantage that *a specific individual has actually suffered*. Race consciousness is *more* a race preference if the intervention *overshoots*. Unfortunately, individual disadvantage is hard to

quantify, and there's reasonable disagreement over the right unit of analysis (individual versus group).

Racial hierarchy *should generally concern us* but in any specific domain, lack of representation *may not be caused* by race discrimination.

To avoid misreading, I underscore that this is a descriptive account, not my personal normative commitments.

II.   EXPLAINING CAPTURE

What is the analytical payoff of identifying this middle category of *color confusion*, which blends elements of both colorblindness and colorconsciousness? First, I think it accurately describes a large swatch of our fellow Americans who do not feel entirely comfortable pigeon-holed as Right or Left. Second, it helps explain why the Left can be easily "captured" toward colorblindness.

Consider, for instance, the examples of colorblind capture that Feingold highlights from the past, drawing on the work of Devon Carbado.[3] In *Regents of the University of California v. Bakke*,[4] Justice William Brennan, joined by Justices Byron White, Thurgood Marshall, and Harry Blackmun, embraced a structuralist account and protested against "let[ting] color blindness become myopia which masks the reality"[5] of racial disadvantage.[6] This sounds Left. Nevertheless, Justice Brennan characterized race-conscious admissions as a racial preference (the "preferential use of race").[7] This sounds Right.

One way to explain this clashing mix-and-match is to admit that it's easy to get confused. In other words, the Justices' framing reflects a *color confused* approach. Recall how the color confused think about the first element (racial relevance). In their view, race *does* matter (descriptively, discrimination continues) although it *shouldn't* matter so much in most domains of both public and private life (normatively, discrimination is bad and too much racial identity-politics is also bad).

Per the third element (racial preference), the University of California, Davis admissions system could be critiqued as not well-enough tailored to an individual applicant's specific accumulation of racial disadvantage, in which case it could be criticized as undue racial preference. Justice Brennan lays out

---

[3] *See generally* Devon W. Carbado, *Footnote 43: Recovering Justice Powell's Anti-Preference Framing of Affirmative Action*, 53 U.C. DAVIS L. REV. 1117 (2019).

[4] 438 U.S. 265 (1978).

[5] *Id*. at 327 (Brennan, White, Marshall & Blackmun, JJ., concurring in the judgment in part and dissenting).

[6] *Id*. at 324-379.

[7] *Id*. at 375.

clearly that "a determination of the *degree of preference* to be given is unavoidable."[8]

Notice also how easy it is to slip between different units of analysis. When we think about structural obstacles, relevant to the first element of racial relevance, we naturally think about obstacles that apply to the entire racial category. However, when we think about remedies, relevant to the third element of racial preference, we naturally think about specific individuals being admitted to or crowded out from a university. As explained above, the group-level unit of analysis leans us Left; the individual-level unit of analysis leans us Right. We often flit between the two, without realizing it much less being able to explain the change in levels. We're all just a bit confused.

In the end, when considering difficult matters of racial justice for which we have conflicting intuitions, we should not be surprised by confusion. We regularly vacillate between descriptive and normative accounts, individual and group units of analysis, and present and past timeframes without appreciating the internal tensions or incoherencies. We muddle through. Welcome to the color confused.

III. DECREASING CONFUSION THROUGH BEHAVIORAL REALISM

In Part III of his article, Feingold provides guidance to the Left on how to defend race-conscious admissions while evading colorblind capture. Notice carefully how Feingold makes the case that race matters at three different times—before, during, and after admissions. *Before* admissions, Feingold highlights the intervenors' brief in the case against the University of North Carolina at Chapel Hill ("UNC"), which detailed UNC's specific racist history and argued that "historical context has present-day manifestations that make students of color feel unwelcome on UNC's campus."[9] *During* admissions, Feingold highlights the challenge of stereotype/identity threat, which leads to mismeasure of latent ability, and implicit bias, which means that ambiguous performance will be interpreted consistent with our implicit social cognitions.[10] Finally, *after* admissions, Feingold emphasizes how having a critical mass for a racial group can create a more equal learning environment, by countering racial headwinds.[11]

If we unpack these moves, we see that the before-admissions analysis trades on history and a process that Devon Carbado and I have recently called "racial sedimentation."[12] Interestingly, for both during and after admissions, Feingold

---

[8] *Id.* at 378 (emphasis added).

[9] Feingold, *supra* note 1, at 1993 (quoting Defendant-Intervenors' Proposed Findings of Fact and Conclusions of Law at 43-44, Students for Fair Admissions v. Univ. of N.C., 319 F.R.D. 490 (M.D.N.C. 2017) (No. 1:14-cv-00954)).

[10] *Id.* at 1994.

[11] *Id.* at 2002-03.

[12] Devon Carbado & Jerry Kang, Why Implicit Bias? (Sept. 18, 2022) (unpublished manuscript) (on file with author).

draws heavily on new facts from the mind sciences about implicit bias and identity threat. In other words, he makes a fundamentally empirical case that race continues to matter right now in both fair individual admissions (during admissions) and in giving entire groups a fair shot at success in college (after admissions).

I deeply value this approach, which is an example of *behavioral realism*. This approach insists that legal analysis and decision-making should incorporate a scientifically up-to-date model of human behavior. In prior work, I've elaborated a three-step process of what this entails.[13] Put briefly, first, identify a more accurate, upgraded model of human decision-making and behavior. In my work, I have focused on discoveries from implicit social cognition, a branch of experimental social psychology. Second, compare that upgraded model to the "commonsense," legacy understanding embedded within status quo law and doctrine. Third, when the gap between the upgraded and legacy models grows too large (however defined), revise our legal thinking to incorporate the upgraded model or at least explain transparently why we cannot.[14]

Behavioral realism believes that *facts matter*.[15] If we have substantially misunderstood some problem—for example, not realizing that the problem of race discrimination is caused partly by implicit bias, which is not just a polite label for strategically hidden, covert explicit bias—we should learn from that mistake and iteratively update our legal and policy response. How might a behavioral realist approach illuminate the four elements of Feingold's schematic, especially for the *color confused*? Let's consider each element.

A.   *Racial Relevance*

On the first element of racial relevance, the color confused believe that race *is* relevant although it generally *should not be*. Do new scientific facts about how we process race alter the way we understand this element? Absolutely. Implicit bias research over the past twenty-five years provides compelling evidence to support the descriptive claim that race indeed *is* relevant. Crucially, it's relevant not just in the structural sense that an explicitly racist past has left behind material conditions, accumulating over time, that are unjust fruits of a poisonous tree that continue to structure opportunity in the present. Even more important, the widespread prevalence of implicit bias makes it difficult for individuals to assert their exceptionalism, for example, that even though race matters "out

---

[13] *See generally, e.g.*, Jerry Kang, *Implicit Bias, Behavioral Realism, and the Purposeful Intent Doctrine*, *in* THE OXFORD HANDBOOK OF RACE AND LAW IN THE UNITED STATES (Devon Carbado, Emily Houh & Khiara M. Bridges eds., forthcoming 2023); Jerry Kang, *Rethinking Intent and Impact: Some Behavioral Realism about Equal Protection*, 66 ALA. L. REV. 627 (2015); Jerry Kang, *The Missing Quadrants of Antidiscrimination: Going Beyond the "Prejudice Polygraph,"* 68 J. SOC. ISSUES 314 (2012).

[14] For more details, see Jerry Kang & Kristin Lane, *Seeing Through Colorblindness: Implicit Bias and the Law*, 58 UCLA L. REV. 465, 490-492 (2010).

[15] *See* Jerry Kang, *The Realities of Race*, 358 SCIENCE 1137, 1138 (2017).

there," it does not matter "in here," at our precious, self-congratulatory firm, school, house of worship, or organization. In short, behavioral realism confirms that *race is relevant* because implicit bias resides in each of us, right now and can be repeatedly confirmed via measurements through objective instruments.

B.   *Racial Neutrality*

On the second element of facial race neutrality, the color confused waffle on whether facially race neutral is indeed race neutral as a function of timeframe (now versus past) and unit of analysis (individual versus group). What does a behavioral realist approach contribute here? I think it allows us to add a new line of argument that can persuade even those who insist on focusing exclusively on the individual (not the group) and right now (not the past).

The evidence of implicit bias demonstrates that notwithstanding facial neutrality, biased application of facially neutral standards is the rule not the exception.[16] The mind sciences have demonstrated, in bodies of literature ranging from confirmation bias to shifting standards, that we tend to see what we expect to see and justify our decision-making process with fluid real-time adjustments in criteria and weighting.[17] We do this without self-awareness. And the way we tend to do this often correlates with implicit biases that we cannot easily detect in ourselves, especially through direct introspection. In other words, even if we pragmatically accept a "Right-ish" frame and focus on the *individual* and the *now*, we must reckon with the finding that implicit bias threatens our individual judgments right now. Facially race neutral is presumptively *not* race neutral.[18]

Harvard University had the opportunity to concede this point and influentially endorse an empirical case for race-conscious countermeasures that offset implicit-bias-actuated discrimination. But it rejected this path[19] because any such concession would risk its litigation position and brand. In this way, Harvard performed as a "Left-ish" brand (defending race conscious admissions) by denying facts (about how implicit bias in high-discretion systems threatens fairness) that are crucial to the long-term success of the Left's substantive project. Asian Americans caught in the middle—who recognize that both discrimination persists against them (with preference given to Whites over

---

[16] *See, e.g.*, Kang & Lane, *supra* note 14, at 481-90.

[17] *See* Jerry Kang, Mark Bennett, Devon Carbado, Pam Casey, Nilanjana Dasgupta, David Faigman, Rachel Godsil, Anthony G. Greenwald, Justin Levinson & Jennifer Mnookin, *Implicit Bias in the Courtroom*, 59 UCLA L. Rev. 1124, 1135-68 (2012).

[18] Some might infer, then, that "blinding"—removing social category information from admissions processes entirely—is thus the panacea. As I've written, blinding may help in certain circumstances but it has its own set of dangers, including passing-through (and indeed morally laundering) discrimination of others. *See generally*, Jerry Kang, *What Judges Can Do About Implicit Bias*, 57 Ct. Rev. 78 (2021) (discussing alternative strategies of dimming and temporary cloaking).

[19] *See* Feingold, *supra* note 1, at 2001.

identically situated Asian Americans)[20] and that race conscious remedies are necessary to create an equal learning environment on campus—rightly feel ambivalent about Harvard's stance.

C.   *Racial Preference*

On the third element of racial preference, the color confused believe that whether race consciousness is a preference or not depends on the degree of calibration. Moreover, as explained previously, the color confused tend to focus on the harms suffered by the specific recipient of the race conscious intervention. Suppose we accept this framing and meet the color confused where they are. What does a behavioral realist approach add?

One contribution is to provide a more precise gauge of how much even well-intentioned, facially neutral admissions processes can disadvantage racial minorities. Various meta-analyses have tried to estimate the effect size of implicit bias on discretionary judgments[21] and of identity threat on testing performance.[22] As rough and contested as these statistical measures are, they do give us a ballpark figure of how much discrimination might be taking place. A well-tailored race-conscious intervention designed to prevent a specific vector of implicit-bias-actuated discrimination or to accurately offset a statistically probable mismeasure will seem more like a legitimate countermeasure and less like a blunderbuss preference.

These are arguments that I and others, including Feingold, have previously made.[23] To be clear, all that any such tightly calibrated intervention can do is prevent or offset discrimination taking place right now to a specific individual. It does nothing to ameliorate structural disparities created in the past that do not need present-day acts of race discrimination to persist or even gain momentum

---

[20] I called this "negative action" against Asian Americans. *See* 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, 3 (1996).

[21] *See generally, e.g*., Anthony G. Greenwald, T. Andrew Poehlman, Eric Luis Uhlmann & Mahzarin R. Banaji, *Understanding and Using the Implicit Association Test* (pt. 3), 97 J. PERSONALITY & SOC. PSYCH. 17 (2009); Frederick L. Oswald, Gregory Mitchell, Hart Blanton, James Jaccard & Philip E. Tetlock, *Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Research*, 105 J. PERSONALITY & SOC. PSYCH. 171 (2013).

[22] *See generally, e.g*., Joel T. Nadler & M.H. Clark, *Stereotype Threat: A Meta-Analysis Comparing African Americans to Hispanic Americans*, 41 J. APPLIED SOC. PSYCH. 872 (2011); Gregory M. Walton & Steven J. Spencer, *Latent Ability: Grades and Test Scores Systematically Underestimate the Intellectual Ability of Negatively Stereotyped Students*, 20 PSYCH. SCI. 1132 (2009); Hannah-Hanh D. Nguyen & Ann Marie Ryan, *Does Stereotype Threat Affect Test Performance of Minorities and Women? A Meta-Analysis of Experiment Evidence*, 93 J. APPLIED PSYCH. 1314 (2008).

[23] *See, e.g*., Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of "Affirmative Action,"* 94 CALIF. L. REV. 1063, 1067-68 (2006); Jonathan Feingold, Note, *Racing Towards Color-Blindness: Stereotype Threat and the Myth of Meritocracy*, 3 GEO. J.L. & MOD. CRITICAL RACE PERSPS. 231, 260-66 (2011).

over time. It also does nothing to respond to even explicit discrimination visited upon the applicant in earlier stages of their life.

Finally, for those in the color confused category who are willing to entertain group-based considerations, a behavioralist approach might offer some evidence-based metrics that improve upon what might otherwise be pure speculation. For example, in *Grutter v. Bollinger*,[24] Justice Sandra Day O'Connor speculated that in twenty-five years (in the year 2028) the need for race conscious affirmative action may simply disappear.[25] Mahzarin Banaji and I argued, in the alternative, that if we must have some metric, why not use a more evidence-based measure, such as a time when there are no differences in implicit racial attitudes toward White and Black racial groups.[26] According to recent modeling by Tessa E.S. Charlesworth and Mahzarin Banaji, that neutrality might be seen in as early as fifty years (2073), or frighteningly, implicit bias may double in size by 2123.[27]

D. *Racial Hierarchy*

On the final element of racial hierarchy, the color confused believe that racial hierarchy generally *should concern us* but in any specific domain, lack of representation *may not be caused* by race discrimination. Consistent with the analysis of elements one and two, the new facts from the mind sciences provide greater reason for us to be skeptical of the legitimacy of racial disparities in status hierarchies.

Allow me to tease out one interesting information asymmetry that can trick us into thinking some hierarchy is natural and legitimate. Many in the muddled middle have approved of expanding opportunity through race conscious affirmative action and equal opportunity programs. Adopting these programs is usually contentious and a part of a very explicit decision-making process, which is highly salient.

By contrast, we are unaware of the myriad, subtle ways in which implicit bias can accumulate to undermine a person's educational trajectory before, during, and after admissions. Because these biases are implicit, and because we tend to view ourselves as well-meaning and objective, we never appreciate how we implicitly contribute to the worse treatment of the very same people whom we have explicitly tried to benefit. This is the information asymmetry: we are hyperaware of explicit race conscious remedies, but we are largely clueless of implicit race discrimination. So when we encounter persistent hierarchies, we notice only the multiple efforts to fix the problem and become deflated and think

---

[24] 539 U.S. 306 (2003).

[25] *Id.* at 343 ("We expect that 25 years from now, the use of racial preferences will no longer be necessary to further the interest approved today.").

[26] *See* Kang & Banaji, *supra* note 23, at 1115-17.

[27] Tessa E.S. Charlesworth & Mahzarin R. Banaji, *Patterns of Implicit and Explicit Attitudes* (pt. 1), 30 Psych. Sci. 174, 182 (2019).

that perhaps the hierarchies are "natural" or inevitable. Becoming equally aware of implicit forces can check against that interpretive tendency.

CONCLUSION

Feingold provides useful conceptual architecture, identifies an important problem of colorblind capture, and provides guidance on how the Left might avoid the racial preference frame. My contribution identifies a middle group between the colorblind and the colorconscious, a group which I think many Americans fall into. For these *color confused*, mixing and matching elements, arguably incoherently, should not be surprising. My other contribution is to demonstrate the value of behavioral realism, which takes seriously the idea that facts matter. The new facts from implicit social cognition provide additional tools with which to persuade the muddled middle. It allows the Left to meet them on their own terms and still tilt them toward accepting tailored race-conscious remedies. Taking the science of implicit bias and identity threat seriously makes clear that well-designed race conscious interventions are countermeasures to race discrimination and not racial preferences.