
RESPONSE

REDUCING HATE ONLINE: THE MYTH OF COLORBLIND CONTENT POLICY BY ÁNGEL DÍAZ[†]

RICHARD DELGADO* & JEAN STEFANCIC**

[†] An invited response to Ángel Díaz, *Online Racialization and the Myth of Colorblind Content Policy*, 103 B.U. L. REV. 1929 (2023).

* Distinguished Professor, Seattle University School of Law.

** Research Professor, Seattle University School of Law.

CONTENTS

INTRODUCTION	1987
I. VIRTUES OF THE DÍAZ ARTICLE	1989
II. NEXT STEPS	1991
III. THREE APPROACHES	1993
A. <i>The Emerging Norm Against "Punching Down"</i>	1993
B. <i>Fairness and Formality: A Confrontation Theory for Mitigating Internet Racism and Prejudice</i>	1994
1. Ambivalence	1995
2. The Social Contact Hypothesis	1996
C. <i>Struggle</i>	1998
CONCLUSION	1999

INTRODUCTION

Cyberspace bristles with racially disparaging material even more so, it appears, than most other channels of communication.¹ In a 2014 article, we sought to explain the abundance of vituperation in that forum and suggested a few measures to abate it.²

In *Online Racialization and the Myth of Colorblind Content Policy*,³ Ángel Díaz offers further explanation. Focusing on large social media platforms like Facebook/Meta, Twitter, and YouTube, he points out defects in the dominant approach to addressing racism on platforms like these, namely content moderation performed under colorblind rules,⁴ in which a moderator reviews texts for objectionable material, deletes it, and warns the sender to desist from posting similar passages in the future.⁵

Díaz's article is one of the first to apply critical race theory in this area.⁶ His conclusion is straightforward: content moderation will fail unless it takes into

¹ E.g., major newspapers, magazines, books, or TV news. See Richard Delgado & Jean Stefancic, *Hate Speech in Cyberspace*, 49 WAKE FOREST L. REV. 319, 322 (2014) (positing Internet's privacy and intimacy create free-for-all atmosphere in which users feel free to vent their likes and dislikes); Richard Ashby Wilson & Molly K. Land, *Hate Speech on Social Media: Content Moderation in Context*, 52 CONN. L. REV. 1029, 1042-45 (2021) (noting populist speech, on Internet or anywhere else, brims with such material and may have contributed to over 2000 hate crimes); see also Steven Lee Myers & Nico Grant, *Combating Disinformation Wanes at Social Media Giants*, N.Y. TIMES (Feb. 14, 2023), <https://www.nytimes.com/2023/02/14/technology/disinformation-moderation-social-media.html> (noting large Internet companies are cutting back online moderation because of high cost).

² Delgado & Stefancic, *supra* note 1, at 320-22, 334-42.

³ Ángel Díaz, *Online Racialization and the Myth of Colorblind Content Policy*, 103 B.U. L. REV. 1929 (2023).

⁴ See, e.g., *id.* at 1949-50 (noting moderation policies make little, if any, distinctions on the basis race).

⁵ Usually threats or disparaging material targeting "someone," "other people," "an individual," or "group of people." Díaz, *supra* note 3, at 1933. Other scholars point out that even a highly sophisticated, multi-tier (but colorblind) system of content review, such as Facebook's, is virtually "toothless," perhaps for this very reason. See Wilson & Land, *supra* note 1, at 1052 (describing Meta's Oversight Board as ineffective due to restrictions on scope of its review and slow speed at which review takes place).

⁶ E-mail from Ángel Díaz, Visiting Assistant Professor of L., USC Gould Sch. of L., to Richard Delgado, Professor of L., Seattle Univ. Sch. of L. (May 18, 2023, 7:37 PM) (on file with author) [hereinafter E-mail] (describing his article as first to apply critical race theory to online content moderation). A few others wrote around this time as well. See, e.g., Wilson & Land, *supra* note 1, at 1033 (applying critical race analysis to content moderation in 2021); Evelyn Douek, *The Siren Call of Content Moderation Formalism*, in SOCIAL MEDIA, FREEDOM OF SPEECH, AND THE FUTURE OF OUR DEMOCRACY 139, 139 (Lee C. Bollinger & Geoffrey R. Stone eds., 2022) (same). Before the advent of the Internet, scholars studied the role of hate speech in creating social division. See, e.g., MARI J. MATSUDA, CHARLES R. LAWRENCE III, RICHARD DELGADO & KIMBERLÉ WILLIAMS CRENSHAW, WORDS THAT WOUND: CRITICAL RACE THEORY, ASSAULTIVE SPEECH, AND THE FIRST AMENDMENT (1993).

account the history and logic of racism.⁷ Proceeding, as it currently does, in colorblind terms—banning, for example, *any* imminent threat—will merely magnify the advantages attendant to whiteness by making them appear natural and inevitable.⁸

But colorblind moderation does not merely shield white supremacists. Minorities who speak out against oppression may easily find themselves banned from a favorite site,⁹ especially if they do not speak the King’s English, use terms like “goddamn” or “racist,” or speak of wanting to bring down the current social order.¹⁰ Majority-group users, however, who disparage minorities via code

⁷ See Díaz, *supra* note 3, at 1936 (“[T]he refusal to address racism is often part of a conscious corporate strategy . . . to continue leveraging racist content for financial gain.”); see also Wilson & Land, *supra* note 1, at 1046 (observing some smaller platforms such as Gab, 4Chan, and 8Chan engage in little or no content moderation); *id.* at 1060-64 (observing that when moderating ignores power and context, it yields poor results). Critical race theorists have questioned colorblind approaches in many areas. See, e.g., RICHARD DELGADO & JEAN STEFANCIC, *CRITICAL RACE THEORY: AN INTRODUCTION* 27-28 (4th ed. 2023) (criticizing Supreme Court’s failure to incorporate race in legal analysis of civil rights questions); Kimberlé Williams Crenshaw, *This Is Not a Drill: The War Against Antiracist Teaching in America*, 68 UCLA L. REV. 1702, 1705 n.3 (2022) (noting critical race theory attempts to understand why colorblind rules seem to enforce systematic inequalities); Neil Gotanda, *A Critique of “Our Constitution Is Color-Blind,”* 44 STAN. L. REV. 1, 1 (1991) (highlighting how Supreme Court’s use of colorblind constitutionalism supports white supremacy); FOUNDATIONS OF CRITICAL RACE THEORY IN EDUCATION 5 (Edward Taylor, David Gillborn & Gloria Ladson-Billings eds., 3rd ed. 2023) (recounting how “separate but equal” policies in education resulted in systematic inequalities and limited opportunities for Black students); DELGADO & STEFANCIC, *supra*, at 134-38 (highlighting inadequacy of colorblind approaches in medical and law school admissions based on merit and standardized tests); DOROTHY ROBERTS, *FATAL INVENTION: HOW SCIENCE, POLITICS, AND BIG BUSINESS RE-CREATE RACE IN THE TWENTY-FIRST CENTURY* (2012) (critiquing how colorblind approaches to medicine recreate systemic racial inequities).

⁸ See Díaz, *supra* note 3, at 1929-43 (noting colorblind content moderation policies routinely allow coded language supporting white supremacy to remain on social media platforms, but effectively stop communities of color from using social media to renounce white supremacy explicitly); see also E-mail, *supra* note 6. Colorblind review will produce this result in several ways. See, e.g., Díaz, *supra* note 3, at 1935-36 (“By requiring explicit racial animus or undeniable calls to violence before company intervention, content policy largely shields the vast arsenal of attacks available to white voices who trade in the language of coded messages and dogwhistles.”). Moreover, “[t]reating all uses of racist speech equally erases a foundational purpose behind racialization as a tool for legitimating one group’s privilege over others.” *Id.* at 1941; see also *id.* at 1950 (“[C]olorblind hate speech and harassment rules [also] foster white supremacy [by] suppressing the voices of communities of color speaking out against it.”).

⁹ E.g., Díaz, *supra* note 3, at 1936 (“Conversely, communities of color are policed as violent, suspicious, and uncivilized.”); *id.* at 1969-74 (contrasting how content regulators treat those who speak out in favor of Palestine and Ukraine).

¹⁰ See *id.* at 1936-37 (noting companies often “use secret blacklists . . . to police racialized groups that are viewed as inherently dangerous” and that these groups often find themselves on the receiving end of “policies against terrorism and violent extremism.”); *id.* at 1945-46

words or circumlocution (as lazy, undeserving, having a poor work ethic, or un-American, for example) will pass muster.¹¹ Moderators will deem their speech mere humor or political commentary.¹² For Díaz, these flaws are systemic, not products of the occasional reviewer who is asleep at the switch, overworked, or secretly in league with white supremacy.¹³

I. VIRTUES OF THE DÍAZ ARTICLE

Díaz's article illustrates the need for color-conscious content moderation not merely by applying social theory¹⁴ but by a series of examples drawn from the world of online content review.¹⁵ He shows the failures of the current approach are not singular but systemic and products of the setting in which online communications take place, especially on large sites like Facebook, Twitter, TikTok, and Reddit.¹⁶ As such, better training of content moderators is unlikely to improve matters. Colorblind monitoring will continually overlook racism or even veiled threats, while suppressing indignant counterspeech by minorities under the guise of protecting civility and public safety.¹⁷ Some of the smaller, more raw sites do very little or no content monitoring,¹⁸ and one of the largest, Twitter, has cut back on content monitoring out of concern for its high cost.¹⁹

(noting popular conspiracy theorists like Alex Jones often passed muster because they were considered parts of a free market of ideas).

¹¹ *Id.* at 1935-36 (noting code words often survive scrutiny; only explicit terms are flagged).

¹² *See id.* at 1955 (noting marginalized communities are subject "to content removal or disfavored treatment by ranking algorithms," whereas racist humor often passes inspection).

¹³ Consequently, Díaz argues colorblind content policies operationalize "a racialized system that doles out a measured hand for the powerful and an iron fist for the marginalized." *Id.* at 1935. In other cases, substandard moderating is a product of cost cutting when an Internet company deems the expense of moderating teams too great. *See* Myers & Grant, *supra* note 1; *see also* sources cited *infra* notes 19, 40.

¹⁴ Díaz, *supra* note 3, at 1936 ("I conclude by advancing an alternative model of race-conscious content policy.").

¹⁵ *See supra* notes 8-12 and accompanying text (considering importance of color-conscious content moderation for safeguarding marginalized communities' ability to use social media without undue interference); *see also infra* notes 18-20.

¹⁶ Díaz, *supra* note 3, at 1936 ("[P]latforms are consistently caught wrong-footed in attempts to moderate discourse that is rife with racial bigotry."); *see also id.* at 1943 (intimating failure may be "part of company strategy to appease conservative politicians").

¹⁷ *Id.* at 1935; *see also* E-mail, *supra* note 6 (considering how colorblind hate speech rules encourage coded language supporting white supremacy, but censor posts denouncing racism due to public safety concerns).

¹⁸ *See* Wilson & Land, *supra* note 1, at 1046 (noting smaller platforms including Gab, 4Chan, and 8Chan choose not to heavily moderate content shared by users).

¹⁹ *Id.* at 1048-52 (noting Facebook is most aggressive in this respect, but its content-review program still ends up "toothless"); *see also* Shannon Bond & Ari Shapiro, *Twitter's Former Safety Chief Warns Musk Is Moving Fast and "Breaking Things,"* NPR (Dec. 2, 2022, 3:40 PM), <https://www.npr.org/2022/12/02/1140355862/twitters-former-safety-chief-warns-musk-is-moving-fast-and-breaking-things> [<https://perma.cc/M6MY-298A>] ("The [budget]

Díaz suggests a systemic approach can reduce the kinds of harm that colorblind approaches perpetrate in other fields such as public health, where they can easily end up suppressing holistic or natural medicine favored by some minority providers²⁰ or overlooking dermatology textbooks that display how afflictions manifest on white but not Black skin.²¹

He also hopes future investigation will show how growing up in all-white neighborhoods and attending segregated schools may shape the worldview of future Internet managers.²² He gives the example of one high-level executive who saw no problems in consulting with conservative figures like Glenn Beck in formulating moderation policy²³ and another who saw none in a YouTuber making explicitly anti-Semitic statements.²⁴

He points out that even when fair-minded executives lay down the rules, the daily work of content review is often carried out by low-paid members of marginalized communities sitting in front of terminals in far-off lands.²⁵ How effective is content moderation likely to be if the one performing it lacks a command of idioms like “no way,” “s___hole country,” or “low on the totem pole”?²⁶ Ableism, sexism, and ageism, for example, may not come to light if one’s main tool to combat this type of speech is a list of words like “cripple,” “decrepit,” or “spaz,” but this tool lacks the necessary nuance to capture exhortations to vote for leaders who are strong and virile, for example, which may easily amount to the same thing.²⁷

cuts include contract content moderators, the company’s human rights team and investigators working to curb political manipulation and child sexual abuse material.”).

²⁰ E-mail, *supra* note 6 (highlighting “role of race and racism in misinformation policy (such as health and political misinformation), and how it can be used to suppress dissent and entrench dominant actors,” as well as “racial segregation in forming the worldview of executives who make the biggest moderation decisions regarding high profile actors (i.e., politicians and other public figures.)”).

²¹ See Nada Hassanein, *Images of Dark Skin Are Absent from Medical Texts. Dermatologists Are Changing That*, USA TODAY (Dec. 7, 2022, 9:46 AM), <https://www.usatoday.com/story/news/health/2022/12/05/dermatologists-skin-of-color-underrepresented-medical-training/10748390002/> [https://perma.cc/BM2H-DPKD] (explaining how medical textbooks fail to prepare doctors to treat patients by primarily providing images of skin conditions on white skin).

²² Díaz, *supra* note 3, at 1948-49 (noting white executives that attended segregated schools bring this bias into their job duties and fail to moderate white supremacist content because they do not view it as dangerous).

²³ *Id.* at 1948-49.

²⁴ *Id.* at 1949; see also *id.* at 1948 (discussing high-level conservative influence on moderation policy).

²⁵ *Id.* at 1946 (noting much content review is carried out in places like Hyderabad); see also Wilson & Land, *supra* note 1, at 1056-57 (describing content moderators working in warehouses in Philippines, Ukraine, and Ireland in distressing conditions for low pay).

²⁶ See Wilson & Land, *supra* note 1, at 1067 (noting many idiomatic or slang expressions whose meaning may easily escape foreign-born speaker).

²⁷ See *id.* (discussing this risk).

Additional difficulties arise with the increasing use of automated moderation under algorithms trained to carry out ostensibly colorblind content review.²⁸ Machine learning systems appeal to managers of Internet sites forced to review large amounts of material because they cut costs and eliminate the middleman.²⁹ But suppose the systems rely, as many do, on banks of material that are themselves rife with hate speech?³⁰ Purging the data that train such a model would presumably require a moderator of the moderator, but would the machine charged with this task (the instructor of the instructor) not itself require a watchful supervisor of some sort?³¹ Díaz performs an invaluable service by pointing out these and similar difficulties with the emerging system of colorblind content moderation.

II. NEXT STEPS

If Díaz succeeds in showing that online content moderation needs to proceed conscious of the history and current role of race,³² he is less successful in putting forward strategies for assuring that Internet executives make the requisite changes.

²⁸ E-mail, *supra* note 6 (warning of “development of content moderation algorithms trained on colorblind policy enforcement, and how they further automate white supremacist logics under the facade of unbiased neutrality”); *see also* Wilson & Land, *supra* note 1, at 1066 (warning of same).

²⁹ Díaz, *supra* note 3, at 1946 (“The ongoing pursuit of speed and scale reduces complex human tendencies into a set of formulas.”).

³⁰ With powerful new tools like these, “[b]ad guys are always early adopters.” Thomas L. Friedman, Opinion, *We Are Opening the Lids on Two Giant Pandora’s Boxes*, N.Y. TIMES (May 2, 2023), <https://www.nytimes.com/2023/05/02/opinion/ai-tech-climate-change.html> (noting this makes case for government regulation, something that big social media platforms strenuously oppose: “If you are worried that A.I. systems will compound discrimination, privacy violations, and other divisive societal harms, the way social networks do, you want regulations now”). Indeed, one online executive, asked what her number one fear was, replied: “regulation.” Díaz, *supra* note 3, at 1947.

³¹ We are reminded of a tale we heard from language theorist Stanley Fish. One evening he and his wife were having dinner with a guest they knew was unaccustomed to having dogs around. Before he arrived, they sternly warned their child not to feed the dog at the table. After the group started eating, the child dropped a portion of meat under the table, which the dog noisily gulped down, to the chagrin of the parents. They told the child to stop, and he agreed not to do so again. Moments later, he placed a morsel on his lap. The dog rose up on his hind feet and eagerly devoured it. When the parents glared, the child insisted he was not feeding the dog; the dog was feeding itself. They told him to stop doing that, too, but the child left the table, ostensibly to go to the bathroom, leaving his plate on a nearby coffee table. When the dog gobbled its entire contents, Fish smacked the child and sent him to his room. The point, of course, is that behavioral enforcement requires that parties on both sides be members of the same community of meaning and share common objectives and values—in the story, a peaceful meal. *See also* Wilson & Land, *supra* note 1, at 1033, 1058–62 (noting many words acquire meaning only in context).

³² As he does. *See, e.g.*, Díaz, *supra* note 3, at 1932–34 (discussing double standards in content moderation in Twitter’s moderation of reactions to Queen Elizabeth’s death).

He does note media executives live in fear that the government will enter the field with heavy-handed regulations that take much of the fun—and, maybe, profit—out of what they do.³³ But beyond noting that government regulation would be unpopular with the industry and might strike conservatives as smacking of censorship,³⁴ he might have cited a further reason why top-down regulation is apt to fail: except for a few highly protected areas such as child pornography, the Supreme Court has “weaponized” the First Amendment to serve conservative, pro-business interests.³⁵ As Catharine MacKinnon and others have pointed out, the Amendment is unlikely to find much use for protecting the interests of minorities, the disabled, ordinary consumers, or the poor.³⁶

This leaves private methods, where the First Amendment does not operate.³⁷ Accordingly, much of Díaz’s article addresses voluntary approaches to content moderation imposed by the industry itself.³⁸

As a legal realist,³⁹ he realizes that social media companies are unlikely to welcome the opportunity to self-regulate with open arms. After all, why should

³³ See *id.* at 1947-48; sources cited *supra* note 30 (noting industry’s fear of governmental regulation).

³⁴ Díaz, *supra* note 3, at 1947-49 (describing lack of enthusiasm for regulation by online platform industry executives, often in conjunction with fear of appearing politically biased against conservatives).

³⁵ See Catharine MacKinnon, *Weaponizing the First Amendment*, 126 U. VA. L. REV. 1223, 1223 (2020) (“This Article traces how and why the First Amendment has gone from a shield of the powerless to a sword of the powerful in the past hundred years.”); see also Díaz, *supra* note 3, at 1945 (noting removals on YouTube have been “primarily limited to pornography . . . and incitement to violence”). Internet speech that merely speaks sympathetically of Palestinians or other disfavored groups can easily find itself disapproved for glorifying a disfavored cause. See *id.* at 1969-73. By contrast, white supremacist or pro-Israel posts often are treated gently. See *id.* at 1973.

³⁶ See MacKinnon, *supra* note 35, at 1223-25; see, e.g., Richard Delgado, *Toward a Legal Realist View of the First Amendment*, 113 HARV. L. REV. 778, 786-87 (2000) (discussing majoritarian quality of rulings in this area). Will the mighty First Amendment even protect women? *Contra* Dobbs v. Jackson Women’s Health Org., 142 S. Ct. 2228, 2258 (2022) (implying answer for women seeking information about abortion services is no); but see Dessie Otachliska, *Free-Speech Post-Dobbs: The Constitutionality of State and Federal Restrictions on the Dissemination of Abortion-Related Information*, N.Y.U. J. LEGIS. & PUB. POL’Y QUORUM (2023), <https://nyujlpp.org/quorum/otachliska-free-speech-post-dobbs/> [https://perma.cc/NL93-RRJF].

³⁷ See Wilson & Land, *supra* note 1, at 1039 (“We are now in a brave new world in which the First Amendment applies to only a fraction of public discourse about the issues of the day.”).

³⁸ E.g., Díaz, *supra* note 3, at 1979-83 (considering voluntary approaches).

³⁹ See E-mail, *supra* note 6, endorsing legal realist themes, including that “content policy is written and interpreted to protect the dominant social, political, and economic advantages attendant to whiteness” as well as a theory of racial capitalism that “connects the profitability of white racism, the regulatory benefit of protecting politicians who trade in bigotry, and the racial biases that inform how platforms conceptualize the harms of online speech.” *Id.*

a highly profitable social media site like Facebook or YouTube impose restrictions on itself when rich backers like Elon Musk oppose them?⁴⁰ His discussion of remedies, unsurprisingly, is brief, urging little more than that reformers proceed with a sense of urgency,⁴¹ abandon a false sense of neutrality,⁴² be on the lookout for race-laced humor and code words,⁴³ train themselves to “see race,”⁴⁴ and learn from history.⁴⁵ While useful, these maxims are short on specifics and do not explain why social media companies should feel compelled to put them into effect.

But here his analysis ends—and ours begins. We offer three reasons why such companies could, and should, carry out thoroughgoing race-conscious reform in the hope that future scholarship—maybe Díaz’s—will examine such measures in greater detail.

III. THREE APPROACHES

A. *The Emerging Norm Against “Punching Down”*

Public discourse has begun to take seriously the intuition that humor and satire targeting weak populations like gay people, the handicapped, or minorities, stand on a different footing from those that target the high and the mighty.⁴⁶ Satire up, for example, stands on a different plane from satire down. As we once noted, a root meaning of “humor” is humus—bringing low, down to earth—a caveat that all leading satirists, such as Jonathan Swift, Molière, and Russell Baker, scrupulously observed.⁴⁷ These and other classic writers, going back to Aristophanes, reserved their slings and arrows for puffed-up generals, kings, and aristocrats, not the poor and lowly.⁴⁸

⁴⁰ See *supra* note 19 and accompanying text; Jeff Zymeri, *Meta Will Offer Facebook Users Greater Control over Content Moderation*, NAT’L REV. (Apr. 5, 2023, 8:09 PM), <https://www.nationalreview.com/news/meta-will-offer-facebook-users-greater-control-over-content-moderation/> [https://perma.cc/84LX-PGU5] (reporting Facebook to allow users to decide to what extent its content is moderated). As Díaz puts it, “[l]earning from the lived experience of impacted communities is an essential starting point for race-conscious content policy.” Díaz, *supra* note 3, at 1979 (footnote omitted). But why should profitable companies carry out this inversion? See discussion *infra* Sections III.A-B.

⁴¹ See Díaz, *supra* note 3, at 1980.

⁴² See *id.* (“To properly address the harms of racism, platform policies must abandon colorblind policies that treat all hate speech and harassment equally.”).

⁴³ See *id.*

⁴⁴ See *id.* at 1977.

⁴⁵ *Id.* at 1961-62.

⁴⁶ See, e.g., Richard Delgado & Jean Stefancic, *Scorn*, 35 WM. & MARY L. REV. 1061, 1063 (1994) (“[I]t is never permissible to use destructive humor at the expense of someone weaker or of a lower station than oneself.”).

⁴⁷ *Id.* at 1063-68 (discussing root meaning of “satire”).

⁴⁸ *Id.* at 1091 (noting Roman generals riding at head of victory parade would often assign servant to accompany them, whispering from time to time, “thou art but a man.”).

Because online platforms probably do not wish to appear to be bullies, it would seem feasible for their leaders to insist that users avoid “punching down.” The concept is easy to explain; even a moderator in a distant country can easily grasp this idea by analogy to familiar norms in their own society, such as the prohibition against disparaging Dalits in India⁴⁹ or indigenous-looking people in Mexico.⁵⁰ A clearly stated warning could give pause to some of the worst offenders, particularly after a few highly publicized cases in which the norm was enforced successfully.

B. *Fairness and Formality: A Confrontation Theory for Mitigating Internet Racism and Prejudice*

A second approach would complement the one against punching down. It would promote voluntary compliance by reminding users of commonly accepted concomitants of good citizenship. Because this approach does not seem to have been tried in the area of Internet regulation, we describe it in some detail.

A “fairness and formality” critique, which first entered the world of legal scholarship in connection with the choice between ordinary litigation and alternative dispute resolution,⁵¹ may prove helpful here as well. The critique builds on an insight first put forward by social scientists Gunnar Myrdal and Gordon Allport.

Both scholars point out that “many Americans behave as though they were ambivalent—of two minds—about race.”⁵² The “American Creed” emphasizes the equal worth of all citizens, a value inherent in fundamental tenets of

⁴⁹ See, e.g., Donald Johnson & Jean Johnson, *Jati: The Caste System in India*, ASIA SOC’Y, <https://asiasociety.org/education/jati-caste-system-india> [https://perma.cc/ED33-JL2D] (last visited Nov. 10, 2023).

The Indian Constitution has outlawed the practice of Untouchability and the Indian Government has established special quotas in schools and Parliament to aid the lowest jatis. Caste discrimination is not permitted in gaining employment and access to educational and other opportunities. But this does not mean that caste is illegal or has faded away. Caste groups as political pressure groups work very well in a democratic system. Caste may provide psychological support that people seem to need. Economists and political scientists are finding that caste is no real barrier to economic development or political democracy.

Id.

⁵⁰ See, e.g., Virginia Mercado, *Mexico’s Racial Divide*, D+C (Mar. 8, 2021), <https://www.dandc.eu/en/article/mexicos-ideology-mestizaje-or-racial-mixing-observes-ingrained-racism> [https://perma.cc/PP23-7D4B] (discussing Mexico’s official position and informal exceptions in everyday treatment of *indios* and Blacks).

⁵¹ See Richard Delgado, Chris Dunn, Pamela Brown, Helena Lee & David Hubbert, *Fairness and Formality: Minimizing the Risk of Prejudice in Alternative Dispute Resolution*, 1985 WIS. L. REV. 1359, 1359 (1985) [hereinafter *Fairness and Formality*] (arguing informal setting of alternative dispute resolution increases possibility of fostering racial and ethnic prejudice, in comparison to formal and traditional adjudication methods).

⁵² Richard Delgado, *Alternative Dispute Resolution: A Critical Reconsideration*, 70 SMU L. REV. 595, 596-97 (2017) [hereinafter *Alternative Dispute Resolution*]; *Fairness and Formality*, *supra* note 51, at 1375, 1382-83.

democratic and Judeo-Christian teachings.⁵³ Representing a collective national conscience commanding high respect, according to Myrdal, “no other norm could compete in authority over people’s minds.”⁵⁴

1. Ambivalence

The Creed’s authority stems from the Constitution and social institutions such as churches, schools, and courts.⁵⁵ Although these institutions also reflect local habits and biases, they impel citizens to exhibit more fairness than some might otherwise display.⁵⁶ During a Fourth of July picnic, for example, many Americans might find themselves standing next to a person of color or someone whom they know to be gay. They may smile, put an arm around the other’s shoulder and say, “Isn’t the band great?”⁵⁷ After the parade ends, they may invite the other person to a barbecue in their yard—or, if they are religious, to a function next Sunday at their church.⁵⁸

On a different occasion, in a moment of intimacy or with friends in a bar or other social setting, that same person may blithely recount a racist joke or sidetrack a job application from a well-qualified person of color.⁵⁹ The contradiction between the American Creed and the reality of class- and race-based prejudice, on the Internet or anywhere else, probably registers most forcefully within institutions like the U.S. Army, which enforces norms of racial fairness out of longstanding commitment and institutional self-interest—the need for Black or Latino soldiers.⁶⁰ But these same feelings of self-doubt also lodge themselves in the consciences of particular individuals, so that “the average American . . . experiences moral uneasiness and . . . guilt” when tempted to behave badly toward minorities, immigrants, or the poor.⁶¹ Allport demonstrated this conflict in a well-known study, and many others have shown the same in different settings.⁶²

These studies show persons afflicted by prejudice resolve their resulting internal qualms in different ways, including repression (denial); defense

⁵³ *Fairness and Formality*, *supra* note 51, at 1383 (citing GUNNAR MYRDAL, AN AMERICAN DILEMMA 80 (20th anniversary ed. 1962)).

⁵⁴ MYRDAHL, *supra* note 53, at 23.

⁵⁵ *Fairness and Formality*, *supra* note 51, at 1383.

⁵⁶ *Id.*

⁵⁷ *Alternative Dispute Resolution*, *supra* note 52, at 597.

⁵⁸ *Id.*

⁵⁹ *Id.*

⁶⁰ *Id.* at 598 (noting specifically military’s formality, system of hierarchical control, and written rules for promotion).

⁶¹ GORDON ALLPORT, THE NATURE OF PREJUDICE 330 (25th anniversary ed. 1979) (internal quotations omitted); see also *Fairness and Formality*, *supra* note 51, at 1383 (discussing this widely shared conviction).

⁶² *Fairness and Formality*, *supra* note 51, at 1384 (discussing Allport’s study, where college students were asked about experiences with members of minority groups and only 10% of students who admitted feeling prejudice did so without feelings of guilt or conflict).

(rationalization); and compromise or partial resolution.⁶³ They may insist, for example, that they refused to hire a well-qualified Black man because they didn't like his bright tie or loud laugh (denial).⁶⁴

They may also react by disparaging the victim,⁶⁵ a form of rationalization that enables a discriminator to reduce discomfort by asserting, for example, that Black schoolchildren should not receive placement in advanced classes because this would only frustrate them.⁶⁶ A third group adopts a situation-specific approach, like Southern racists of a former era who behaved kindly toward their household help but would never invite a Black associate to dinner in their home.⁶⁷ In modern times, such a person might use racial language with close friends but avoid it in a public setting.⁶⁸

2. The Social Contact Hypothesis

Each of these mechanisms (denial, rationalization, and situation-specificity) enables a discriminator to maintain distance—either in the world or in his thoughts—between himself and the people against whom he harbors prejudice. Thus, the basic premise of a much-heralded approach to mitigating prejudice holds that frequent contact with people of different types reduces its sway,⁶⁹ particularly when the contact takes place between equals in pursuit of common objectives and is carried out over an extended period from an early age.⁷⁰ The idea is that people learn from experience that members of other groups are very much like their own: some are smart, some stupid. Some trustworthy, others not. Some are good at sports; others are uncoordinated doofuses one wouldn't want on their team. Team sports, military service, and the Boy Scouts are often held up as settings offering the optimal conditions.⁷¹

Because the Internet is notoriously lacking in opportunities for users to meet people with different views and backgrounds from their own,⁷² and because neighborhood segregation compounds the problem, a backup approach has found favor, particularly in legal circles.

⁶³ *Id.* at 1384-85.

⁶⁴ *Id.* at 1384.

⁶⁵ *Id.* at 1384-85 (citing social science studies of this mechanism).

⁶⁶ *Id.* at 1385 (citing this and similar examples).

⁶⁷ *Id.* (citing social science literature).

⁶⁸ *Id.*

⁶⁹ ALLPORT, *supra* note 61, at 281.

⁷⁰ *See id.* (discussing his well-known "social contact" hypothesis); *see also Fairness and Formality*, *supra* note 51, at 1385-89 (discussing examples such as team sports or military service in an integrated unit).

⁷¹ *Id.* (noting maximized effects when "sanctioned by institutional supports"). Members of a successful, long-lasting first-year study group of mixed-racial membership might experience much the same.

⁷² *See* CASS SUNSTEIN, *REPUBLIC.COM* 71 (2001) (noting Internet tends to increase social division).

3. The Confrontation Theory

This second approach, associated with the early work of one of us, posits that providing conspicuous reminders that a situation calls for the higher standard of behavior that reserve for occasions of state will induce most observers to act accordingly, even without conscious thought.⁷³

Studies show that prejudiced persons are least likely to act on their beliefs in settings that confront them with the discrepancy between their ideals and their personal antagonism toward out-groups.⁷⁴ With litigation, for example, the Anglo-American system incorporates a host of formalities:

[T]he flag, the black robes, the ritual—[that] remind those present that the occasion calls for the higher, “public” values, rather than the lesser values embraced during moments of informality and intimacy. In a courtroom trial the American Creed, with its emphasis on fairness, equality, and respect for personhood, governs. Equality of status, or something approaching it, is preserved—each party is represented by an attorney and has a prescribed time and manner for speaking, putting on evidence, and questioning the opposing side. . . . The rules of procedure maintain distance between the parties. Counsel for the parties do not address one another, but present the issue to the trier of fact.⁷⁵

The rules operate to preserve the formality of the event by prescribing in detail the manner and sequence in which it is to unfold.⁷⁶ The intuition that the formality of adversarial adjudication reduces prejudice is corroborated by empirical studies comparing outcomes for minority people and women, finding a decided advantage for those who litigate, rather than mediate, their disputes.⁷⁷

Adversarial procedure thus counteracts bias on the part of the parties who come before the court. But it also seems to counteract *decisionmaker* bias because it combats the natural human tendency to make snap judgments.⁷⁸ It seems likely the mechanisms in question increase fairness for both groups in much the same way: they trigger unconsciously that the situation is one calling for the higher standards of decency we exhibit during the Fourth of July picnic.⁷⁹

With online content moderation, this leads to a simple solution: place Americanization reminders prominently on the website—perhaps on the sign-in

⁷³ See *Alternative Dispute Resolution*, *supra* note 52, at 596-99 (using example of in-court adjudication); see also *Fairness and Formality*, *supra* note 51, at 1386-1400 (setting out basis of this approach in social theory).

⁷⁴ *Fairness and Formality*, *supra* note 51, at 1386-87 (“[A] prejudiced person is more likely to act in prejudiced fashion when on familiar ground or with friends than when participating in a public function.”).

⁷⁵ *Id.* at 1388.

⁷⁶ *Id.*

⁷⁷ See *Alternative Dispute Resolution*, *supra* note 52, at 598-99 and accompanying notes. Arbitration seems to have an even worse record in this respect. *Id.* at 599-600.

⁷⁸ *Fairness and Formality*, *supra* note 51, at 1389-90.

⁷⁹ See *supra* text and notes 57-58.

page—so that users encounter them when they open it and sign in. Whenever the site managers send a message, even a routine one, to users—for example, in connection with billing or a change in address—they should accompany them with emoji or logos suggestive of Americanization.⁸⁰

C. *Struggle*

Predictably, some users will be hardcore white nationalists and refuse to act on even a series of such cues. They may even think that punching down is perfectly acceptable and more fun than engaging in combat with one's equals. In that event, we urge a simple strategy—struggle, including street politics. Díaz looks for formulas to solve online harassment and settles on a few, none particularly promising in our eyes.⁸¹ But life is full of clashing interests, battles among competing factions, and interest groups.⁸² The approaches that he proposes—like ours—may do some good, particularly in the short term. In the long run, though, the only solution may be to fight harder. Regulation is unlikely to help for the reasons mentioned earlier, including the current Court's retrograde interpretation of the First Amendment.⁸³ Big Media are unlikely to help, either—Internet companies are mainly motivated to advance their own profits and power. And, of course, the Americanization approach we suggest will not work with Internet users who care little about the underlying values of a democratic society.

Therefore, we urge website managers interested in a fair fight build in a multitude of opportunities for contestation. Because one never knows in advance who is likely to find a text offensive, indignant groups need to be ready to come

⁸⁰ Examples include American flags, pictures of the Marines at Iwo Jima, likenesses of George Washington, photographs of the White House on a spring day, and the Liberty Bell.

⁸¹ See sources cited *supra* note 40. Díaz proposes regulators abandon neutrality, see Díaz, *supra* note 3, at 1980, take a dim view of humor and tricky code words, see *id.*, “see race,” *id.* at 1977, and learn from history, *id.* at 1981.

⁸² See, e.g., KARL MARX, CAPITAL: A CRITIQUE OF POLITICAL ECONOMY (Frederick Engels ed., Samuel Moore & Edward Aveling trans. 1909) (1867) (discussing status of workers under capitalism); see also Matthew Desmond, *Why Poverty Persists in America*, N.Y. TIMES, <https://www.nytimes.com/2023/03/09/magazine/poverty-by-america-matthew-desmond.html> (Apr. 3, 2023).

[S]ocial scientists have a fairly coolheaded way to measure exploitation: When we are underpaid relative to the value of what we produce, we experience labor exploitation; when we are overcharged relative to the value of something we purchase, we experience consumer exploitation. For example, if a family paid \$1,000 a month to rent an apartment with a market value of \$20,000, that family would experience a higher level of renter exploitation than a family who paid the same amount for an apartment with a market valuation of \$100,000. When we don't own property or can't access credit, we become dependent on people who do and can, which in turn invites exploitation, because a bad deal for you is a good deal for me.

Id.

⁸³ Namely, opposition by the industry itself, see *supra* notes 30, 33 and accompanying text, as well as fear of the role of the state as censor. See Díaz, *supra* note 3, at 1947-49.

forward and defend their own interests quickly, before their tormentors build up a head of steam. The managing elite can make their task easier by incorporating American flags and other such symbols everywhere so that they can sense the light breeze building in their sails.⁸⁴

CONCLUSION

Readers reluctant to discard the mannerly approach of law and legal regulation should also remind themselves that many early radical lawyers refused, on principle, to bring cases to court; they put their faith in organizing and social movements.⁸⁵ Judges were The Man, and they knew it. The problem really is structural, just as Díaz says.⁸⁶ But he, perhaps, does not appreciate the full implications of his own intuition: the system will change only if it is in its own interest to do so. As Derrick Bell pointed out, milestone events like *Brown v. Board of Education*⁸⁷ tend to arrive only when momentary coincidences of Black and majority interests call them forth.⁸⁸ Managers of online sites are likely to resist even the small Americanization measures we suggest. But when the first unruly crowd bursts into the board room at their annual meeting, they may begin to see things differently. Putting the American flag prominently on a few places in the platform is, after all, just a small thing.

⁸⁴ See generally Lani Guinier & Gerald Torres, *Changing the Wind: Notes Toward a Demosprudence of Law and Social Movements*, 123 YALE L.J. 2740 (2014) (providing metaphor for social change).

⁸⁵ GERALD ROSENBERG, *THE HOLLOW HOPE* 420-29 (2d ed. 2008) (noting lawyers can act as “flypaper”); Richard Delgado & Jean Stefancic, *The Social Construction of Brown v. Board of Education: Law Reform and the Reconstructive Paradox*, 36 WM. & MARY L. REV. 547, 549, 567 (1994) (same).

⁸⁶ See Díaz, *supra* note 3, at 1937-43 (Part I: Social Media and the Structural Maintenance of White Supremacy).

⁸⁷ 347 U.S. 483 (1954).

⁸⁸ Derrick Bell, *Brown v. Board of Education and the Interest Convergence Dilemma*, 93 HARV. L. REV. 518, 523 (1980) (referencing “deeper truth about the subordination of law to interest-group politics with a racial configuration”).