Article

new media & society
1–25
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14614448251385923
journals.sagepub.com/home/nms
Sage

Moral disengagement and content moderation attitudes: Examining how apathy to online harms may disguise racially conservative beliefs

Alyvia Walters
Rutgers University, USA

Tawfiq Ammari
Rutgers University, USA

Shagun Jhaver
Rutgers University, USA

Abstract

Social media users' preferences for various content moderation interventions have been widely studied, but the implicit beliefs that connect to these preferences are less understood. Using a nationally representative survey data set, we investigate how end-users' attitudes toward moderating harmful speech online relate to their offline racial attitudes. We find that racially conservative beliefs are significantly positively related to participants indicating a distaste for concepts related to content moderation and cancel culture, suggesting that racial conservatism may be a crucial factor to consider in assessing these attitudes. We discuss our findings through the lens of moral disengagement theory, positing that supporting "freedom of expression" by way of disagreeing with content moderation and cancel culture may be a contemporary mechanism of morally disengaging with the harmful effects of racially insensitive speech.

Keywords

Cancel culture, online harassment, Pew Research, survey, free speech

Corresponding author:
Alyvia Walters, Library and Information Science Department, Rutgers University, 4 Huntington Street,
New Brunswick, NJ 08901, USA.
Email: alyvia.walters@villanova.edu

Introduction

Racism in the United States, while deeply rooted in global histories of oppression, has intensified in visibility and volatility over the past decade—both offline and online. The 2016 US presidential election marked a turning point, igniting a surge of racially charged discourse that escalated in both volume and vitriol (Giani and Mèon, 2021). By 2020, this simmering tension reached a boiling point: the COVID-19 pandemic fueled a wave of anti-Asian hate crimes (Bresnahan et al., 2023), while the police killing of George Floyd triggered nationwide protests and a reckoning with structural and interpersonal anti-Black racism (Nguyen et al., 2021). Far from subsiding, unrest has evolved into online backlash against critical race theory (Walters et al., 2024), coupled with a wave of anti-diversity, equity, and inclusion policies (Aratani, 2025). This underscores the urgent need to examine how race-based sentiment, online and offline, continues to shape public attitudes, policy, and everyday life.

Online discourses about race often mirror offline speech and actions, as the Internet—particularly social media platforms—provides shareable public spaces for these interactions to occur (Chou and Gaysynsky, 2021). However, there is an inadequate understanding of how offline attitudes about race relate to users' attitudes toward harmful online speech. Based on Baym (1995, 2010)'s assertion that offline worlds are continuously permeating online contexts, and vice versa, discerning the connection between offline racial attitudes and preferences for moderating harmful speech online can better help us interrogate racism's enduring power in our social worlds.

In September 2020, the Pew Research Center ran a large-scale survey study that allows us to explore this very nexus. Pew sought to collect data on American attitudes of the cultural moment: a moment that consisted of massive social unrest and widespread mis- and disinformation regarding both the COVID-19 pandemic and the national Black Lives Matter (BLM) protests that had erupted after the killing of George Floyd. This survey was specifically crafted to better understand three main issues: online harassment, race relations, and COVID-19. While Pew has published excellent public-facing work[1] on the many trends that these data identify, there is a unique pathway that the data allow us to consider, which has, to this point, been left unexamined: what relationship, if any, exists between racial attitudes and user perceptions of addressing online harms?

Through a quantitative analysis of this survey data, we investigate this relationship, ultimately arguing that the particular moment of this data collection—one which was rife with heightened racial tension—allows us to theorize that people with more racially conservative attitudes may "morally disengage" with these socially harmful views by reframing them in a more socially acceptable way: indicating distaste for content moderation.

Related work

Social media users' opinions on content moderation

Online content moderation is defined by Roberts (2017) as "the organized practice of screening user-generated content (UGC) posted to Internet sites, social media, and other online outlets, in order to determine the appropriateness of the content" (p. 12). For the purposes of this study, we are most concerned with platform-enacted content moderation mechanisms, as opposed to moderation actions taken by end-users or volunteer moderators (Jhaver et al., 2023) that sanction norm violations entirely from the site. The ways that these content moderation mechanisms operate are often opaque to end-users—that is, the people using social networking platforms—and in the absence of this knowledge, users often develop “folk theories” around how they believe content moderation works (Myers West, 2018).

Social media users’ opinions on the functions, responsibilities, and limits of content moderation tactics vary widely, and have been studied through a variety of lenses. Notably for the present study, these lenses have included partisan differences in support for moderation tactics, up to and including complete deplatformization (Alizadeh et al., 2022; Appel et al., 2023), and identity-based differences in impressions of moderation techniques and fairness (Haimson et al., 2021; Hawkins et al., 2023; Weber et al., 2024). In particular, prior research has found that Black social media users are more often targeted by online moderation practices than their peers, often when they are speaking on issues related to racial justice (Haimson et al., 2021). Peterson-Salahuddin (2024) shows how racially marginalized users often experience “overblocking” by automated moderation tools when they describe or speak out against racist experiences they have had, as these users are wont to use terms deemed “hate speech” triggers by automated systems. In so doing, cycles of racism and race-based sexism are continually reinforced.

In the US context, concerns around “freedom of expression” are often highlighted in studies on user-centered opinions on content moderation, as it is a highly salient ideal in American culture. To be clear, because platforms are privately owned, there is no legal impetus for them to uphold free speech rights—but research has documented relationships between users’ high valuation of their rights to freedom of speech and their distaste for certain types of content moderation online. Naab et al. (2021), for example, found that people with a high commitment to freedom of speech were less supportive of authoritative content restrictions on Facebook, and Weber et al. (2024) found that support for freedom of speech significantly lowered participants’ belief in the fairness of content moderation practices. Jhaver and Zhang (2023) found differences by speech category—users’ support for freedom of expression negatively influenced their desire for platform bans of hate speech, but they found no significant relationship between support for freedom of expression and desire for bans of violent or sexually explicit material online. However, other research has pointed to a less tidy relationship between freedom of expression and content moderation: Kozyreva et al. (2023) found that a majority of users would opt to quash harmful misinformation, even at the expense of freedom of speech. The current study will provide evidence that there is space and motivation for better understanding the mediating function that racially conservative views may serve in strong beliefs of freedom of expression, as ideologies concerning free speech underpin many arguments against content moderation and “cancel culture.”

Content moderation, cancel culture, and race

Content moderation and “cancel culture” are different—but interrelated—concepts, both of which are often discussed within the discourse of “freedom of speech.” While content moderation revolves around the official actions taken by platforms and users to flag and regulate online speech deemed inappropriate for the site, cancel culture is a social process that gained mainstream cultural prominence around 2019–2020—though its etymology is much more storied, and its use on “Black Twitter” and in connection to the #MeToo Movement began in 2015 and 2017, respectively (Picarella, 2024).

The definition of cancel culture is not settled, but it can be generally understood as:

> (a) the public shaming of unacceptable behavior, and (b) withdrawal of support, which are (c) motivated by wanting to see the target persons experience some form of consequence or penalty due to their actions . . . or to ensure these persons are socially banished. (Tandoc et al., 2024, p. 3)

This terminology has, over time, become politicized, often with the right accusing the left of weaponizing cancel culture against dissenters (even though people of all political persuasions engage in counterspeech and have done so for centuries; Bridges, 2021–2022; Picarella, 2024). In a 2020 speech at the American landmark Mount Rushmore, US President Donald Trump called cancel culture the newest “political weapon of the ‘far left,’” and claimed that it was the “very definition of totalitarianism” (Trump, 2020, qtd. in Bridges, 2021–2022).

At the time this survey data was collected by Pew Research Center—and at the time of this speech by President Trump—cancel culture was a prominent term that was often connected to BLM and its associated movements for racial justice (Spicer, 2022). For example, many people decried the removal of Confederate monuments throughout the nation as a “cancel culture” effect of the BLM protests (Meesala, 2020). Although social “cancelation” in online environments does not always lead to top-down, platform-appointed moderation of an idea or entity, many leaders of major social media outlets, such as Meta’s Mark Zuckerberg, vowed to devote stricter attention to monitoring and removing racist language from their platforms[2] at that time (Bridges, 2021–2022). Because of the heightened awareness around racial issues, cancel culture, and platform content moderation during this period, the Pew survey data provide a unique opportunity to explore user attitudes at the intersection of race and online speech regulation.

Racism and moral disengagement online

Online racism has long been a problem—the 2020 BLM protests were simply a cultural flash that momentarily brought this issue to the surface. Prior research suggests that the increased anonymity of online spaces fosters racist discourse online (Keum and Miller, 2018), and that cyber-racism is able to thrive through the many communication channels that the digital age provides (Bliuc et al., 2018). Social media, in particular, has facilitated racist discourse, and Ng and Indran (2024) found nearly 100 million Tweets using racist hashtags from the past 15 years. Users’ racial identities also impact the experiences of racism online, as non-White users self-report experiencing racist content online at higher rates (Pew Research Center, 2021).

The ways that racism manifests have shifted over time, and many scholars argue that overt racism has long been “out of style,” socially. Concepts such as colorblind racism

(Bonilla-Silva, 2022) and post-race racism (Goldberg, 2015) have come to describe the ways that racism still thrives—interpersonally and structurally—in spite of this shift in its social "acceptability." A manifestation of this shift, many scholars have studied the ways that seemingly race-neutral policy positions often operate as racial proxies, providing socially palatable positions that actually forward racially harmful goals (Bonilla-Silva, 2022; Costley White, 2018; Gilens, 2009; Hall et al., 2013; Mendelberg, 2001; Shook & Lizarraga-Dueñas, 2024; Winter, 2008).

Earlier research suggests that the "styles" of offline and online manifestations of racism may be slightly divergent. Offline, racial bias is often systemic and implicit, operating through entrenched social institutions and interpersonal dynamics in many fields such as policing, healthcare, and employment discrimination. While social movements like BLM have helped shift explicit attitudes—especially among White Americans—implicit biases remain relatively stable and resistant to long-term change (Sawyer and Gampa, 2018). By contrast, online racism is often more overt and amplified by features such as anonymity, lack of accountability, and the algorithmic reinforcement of group norms. The Internet allows users to express racist ideologies more freely through toxic disinhibition and group polarization, often bypassing the social constraints that inhibit such expressions in offline settings (Keum and Miller, 2018a).

One way that social psychologists have theorized people's continuation of problematic patterns—even when their social or moral value is in question, as is the case with racism—is through "moral disengagement." Albert Bandura (1999) is credited with this theoretical development, in which he explains how people self-justify harmful acts through several different mechanisms (see Figure 1). According to Bandura (2011), there are eight such mechanisms: moral justification; exonerative comparison; euphemistic labeling; displacement of responsibility; diffusion of responsibility; minimizing, ignoring, or misconstruing the consequences; dehumanization; and attribution of blame.

For the purposes of our study, we are most interested in four key mechanisms through which, we argue, people with racially conservative beliefs could be "morally disengaging" with those views by reassigning these beliefs to a distaste for content moderation:

1. *Moral justification*, in which people justify harmful actions by claiming a higher moral purpose;
2. *Diffusion of responsibility*, in which people shirk responsibility for the harm they cause by perceiving themselves as one of innumerable faceless actors perpetuating the action;
3. *Minimizing, ignoring, or misconstruing the consequences*, in which people downplay the detriment of their actions;
4. *Attribution of blame*, in which people blame the victim for their own suffering.

As we will explain, the significant, positive relationship between offline racially conservative views and lack of support for online content moderation presents a possibility to extend moral disengagement theory to the realm of online content moderation, and it provides a vehicle for expressing a way that racial conservatism may covertly manifest through an expressed distaste for content moderation.

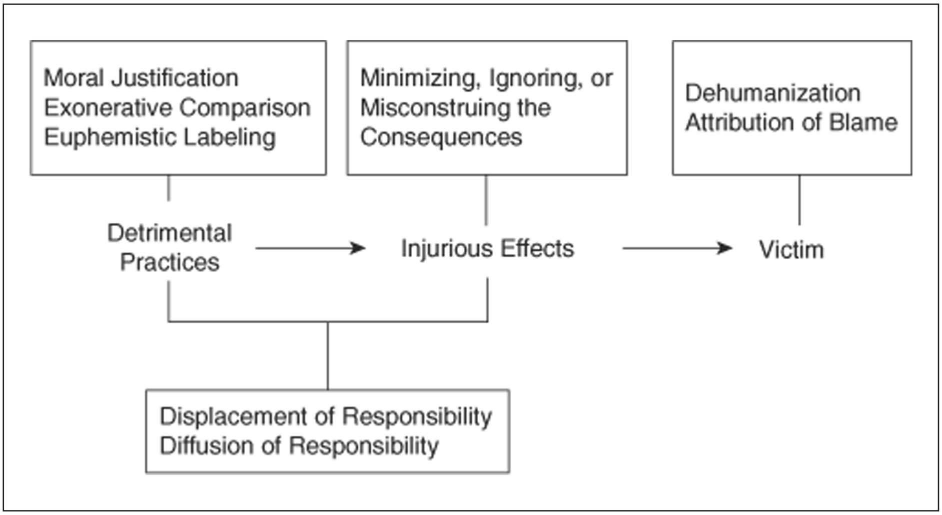

Figure 1. Albert Bandura's (2011) model of the "psychosocial mechanisms through which moral self-sanctions are selectively disengaged from detrimental conduct."

Methods

To investigate the relationship between racial attitudes and online content moderation preferences, we employ survey data collected by Pew Research Center's American Trends Panel (ATP) Wave 74 (Topics: Online harassment, Race relations, COVID-19), which was conducted from 8 to 13 September 2020. The ATP is a panel of approximately 10,000 respondents representative of the US population that Pew maintains in order to gather high-quality data about a variety of topics (Pew Research Center, 2025). Although secondary data analysis is not ideal, we argue that the analysis of this particular survey's data can provide valuable insights because of its sociopolitical context and topics of focus: online harassment and race relations. Pew Center's robust data collection and validation methods and large sample sizes add further value to this data set.

Nearly 4 months after the murder of George Floyd—and the subsequent months of large-scale protests, all tied to the Movement for Black Lives and eventually encompassing calls for justice for the murders of Ahmaud Arbury and Breonna Taylor—the nation was at a point of reckoning with the stakes of race, policing, and social inequality (Kishi and Jones, 2020). Evidence of this topic's discursive importance is plain in news data (see Appendix 1), but it is also obvious in the structuring of ATP Wave 74. Indicative of the historical moment, this wave of ATP included many questions about the BLM movement, respondents' orientations toward racial inequality in the United States, and policing.[3] As such, it stands to reason that this is a particularly interesting occasion in time to interrogate the connections between respondents' racial attitudes and their positionalities toward offensive content online, much of which, at the time, was likely race-based or reflective of racial ideologies. To effectively analyze this connection between offline values and online orientations toward content moderation and harassment, we only included survey respondents who indicated that they are social media users (n = 7453).

Although survey data from 2020 is, at the time of publication, 5 years old, it remains relevant because the sociopolitical moves made in that cultural moment—a shift toward DEI-centered policies, critical discussions on the impacts of racism in the United States, and a reckoning with the country's racial history—is now being explicitly and systematically reversed (Aratani, 2025). Understanding how the 2020 survey insights may inform today's political and regulatory climate provides an empirical way of tracing race-based attitudinal shifts in the United States—and their lasting effects.

Measures

Racial attitudes are notoriously difficult to measure in survey studies, as it is unlikely that many respondents would openly admit to holding racist beliefs (Bonilla-Silva, 2022). This is often attributed to the "social desirability bias," or the tendency of respondents to self-report ideas, beliefs, and/or behaviors that they see as most socially desirable rather than truly reflective of their thoughts (Fisher and Katz, 1999).

However, based on the wide array of data that the ATP provides, we were able to construct a racial attitudes proxy scale[4] with high reliability ($\alpha = 0.87$; $n = 20$) to help counteract this bias and better gauge respondents' true racial attitudes. Based on a polarity scale, we averaged[5] each respondent's answers to 20 of the questions presented in the Pew survey.

Through iterative discussion among the researchers, these questions were chosen for their propensity to unveil respondents' racial attitudes—not simply their attention to the conversation on race and racism at the time. For example, we did not include the question, "In the past three months, how much attention have you been paying to [issues of race and racial inequality]?" A respondent's answer to this question would not reliably reveal their racial attitudes. Instead, we did include questions on a wide variety of issues such as race-based discrimination, approaches to combatting racism, and opinions on the BLM movement. According to prior research, individuals' attitudes in these topic areas do relate to their racist or anti-racist outlooks (Bonilla-Silva, 2022; Goldberg, 2015; Hawkins and Saleem, 2022; Holt and Sweitzer, 2018; West et al., 2021; White and Crandall, 2017). This fact, alongside the scale's reliability, gives us the support to posit that each respondent's resulting composite score (scaled from 0 to 1) provides a measurable representation of their racial attitudes and beliefs.

Our scale is based on 20 different metrics. While these include some binary-response questions, which are limited in their ability to evoke more nuanced views, they do indicate a preference for one attitude/position over another, and together they help us align our scale which is indicative of racial conservatism/progressivism.[6] We constructed this scale such that higher scores equate to more racially conservative views and lower scores equate to more racially progressive views ($M = .28$, $SD = 0.25$). See Appendix 2 for the precise list of items used to create this scale.

Other key independent variables in this analysis include age ($M = 2.61$, $SD = 0.95$, where category 1 reflects ages 18–29 and category 4 reflects 65+), race/ethnicity (68.5% White non-Hispanic), gender (41.1% men), education level ($M = 4.36$, $SD = 1.44$, where

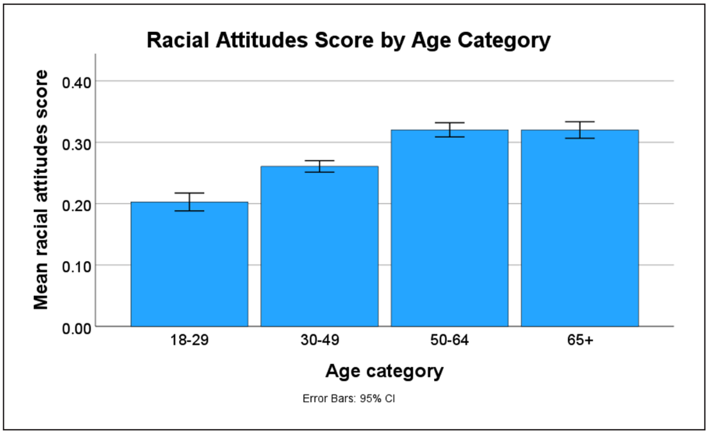

Figure 2. Racial attitudes scores by age.

category 1 reflects less than high school and category 6 reflects postgraduate), and political party affiliation (26.2% Republican, 40.7% Democrat, 25.0% Independent, 8.1% some other party).

Considering these independent variables—primarily the racial attitudes composite score, but also age, race/ethnicity, gender, education level, and political party affiliation—we test for associations with a variety of dependent variables which asked about respondents' orientations toward offensive online content and instances of harassment online. Taken together, these results help us better understand the relationships between offline racial attitudes and feelings about the social and personal costs of online harm.

Findings

Racial attitudes

Beginning with the assessment of offline racial attitudes, we found significant differences across each demographic category explored—age, gender, race/ethnicity, education level, and political ideology—regarding their scores on the racial attitudes scale (RAS, henceforth). In addition to analyzing the descriptive statistics on RAS scores across demographic groups (seen in Figures 2 to 6), we ran one-way ANOVA and Tukey tests. The results of these tests allowed us to investigate differences in RAS scoring within each demographic category.

Considering age, there are significant differences in RAS scores between all age groups except for one: the difference between the racial attitudes of 50- to 64-year-olds and those 65 and older holds no statistical significance (Figure 2). Consistently, we found that those in younger age groups hold more racially progressive views than their older counterparts. When looking at gender (Figure 3), men ($M=0.33$) hold significantly more racially conservative views than women ($M=0.26$; $p<.001$) and non-binary respondents ($M=0.18$, $p<.001$).

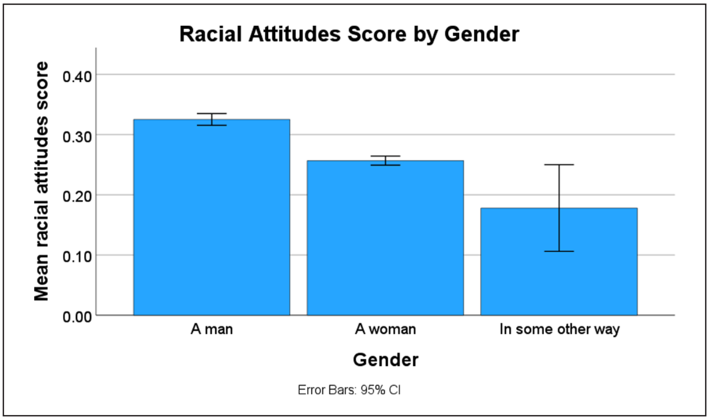

Figure 3. Racial attitudes score by gender.

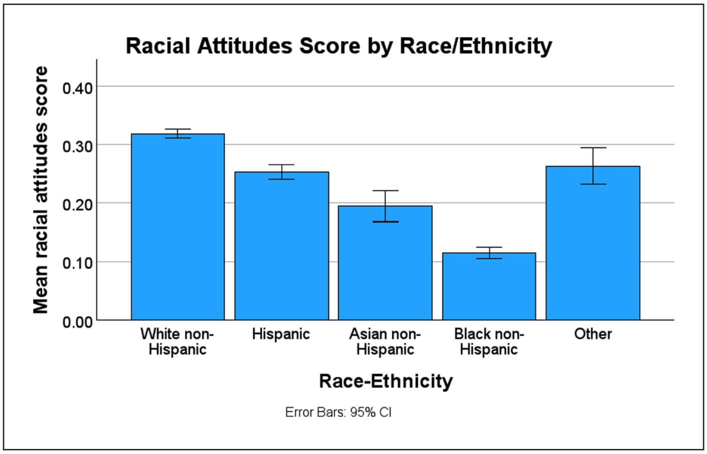

Figure 4. Racial attitudes score by race/ethnicity.

Shifting to race/ethnicity, we found that White non-Hispanic respondents ($M=0.32$) held significantly more racially insensitive views than each of the other racial groups (Figure 4). Conversely, Black non-Hispanic participants ($M=0.11$) held significantly more racially progressive views than each of their peers.

Education follows a steady pattern in which those with more education are significantly more likely to hold more racially progressive views (Figure 5), and political party is similarly straightforward (Figure 6), with Republicans ($M=0.54$) holding significantly more conservative views than each of the other political affiliations. Democrats ($M=0.11$), however, hold significantly more progressive views than all groups.

Given that these findings align with expected opinions as per prior research (Carian, 2022; Hagendoorn and Nekuee, 2018; Tietjen and Tirkkonen, 2023), they offer a validity check and raise confidence in our use of RAS scores as a proxy for attitudes about race.

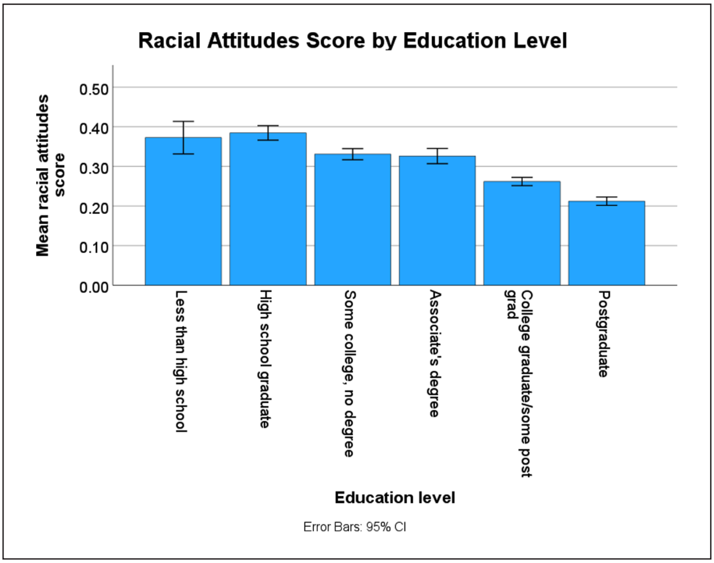

Figure 5. Racial attitudes score by education level.

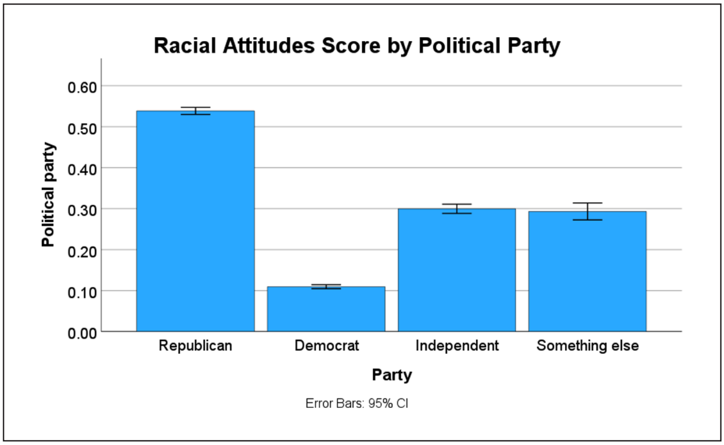

Figure 6. Racial attitudes score by political party affiliation.

Racial attitudes and opinions on offensive online content

We tested whether racial attitudes, captured through the RAS, have a significant association with respondents' opinions toward offensive content online. Specifically, focusing on these three Pew survey questions, we investigated how responses to them correlated to racial attitudes:

Table 1. Odds ratio of predicting OFFENSE and CANCEL CULTURE responses based upon demographics and racial attitudes.

Variable	Offense	Cancel culture
Step 1		
Age	.762***	.850***
Education	.883***	1.015
Man	–	–
Woman	.828*	.750***
Other gender	.724	.746
White NH	–	–
Black NH	1.159	.807
Hispanic	1.232	.776*
Asian NH	1.274	.562*
Other race	1.404	1.092
Republican	–	–
Democrat	.741*	.678**
Independent	1.058	1.001
Other party	1.146	1.207
R^2	.229***	.209***
Step 2		
RAS score	.006***	69.481 ***
R^2 change	.146	.126
N	6742	6636
Intercept	127.002***	.297***
Total R^2	.375***	.335***
Omnibus tests of multiple coefficients	$p < .001$	$p < .001$

All variables were binary dummy variables except for two ordinal variables: age and education. Odds ratios reported are from the full model, that is, controlling for all variables in the model. R^2 reported is Nagelkerke R^2. RAS score for OFFENSE item is using the RAS score + 1, and then its inverse square in order to meet the logistic regression assumption of linearity.
*$p < .003125$; **$p < .000625$; ***$p < .0000625$, to account for the Bonferroni correction.

Question 1: Which comes closer to your view, even if neither is exactly right? (1 = Offensive content online is too often excused as not a big deal; 2 = Many people take offensive content they see online too seriously). This question is coded as OFFENSE in Table 1.

Question 2: In general, when people publicly call out others on social media for posting content that might be considered offensive, are they more likely to . . . (1 = Hold people accountable for their actions; 2 = Punish people who didn't deserve it). This question is coded as CANCEL CULTURE[7] in Table 1.

Question 3: Thinking of some experiences that might happen to people when they use the Internet, how much of a problem, if at all, is people being harassed or bullied? (1 = Major problem; 2 = Minor problem; 3 = Not a problem). This question is coded as HARRASS in Table 2.

Table 2. Odds ratio of predicting HARRASS responses based upon demographics and racial attitudes, as compared with those who responded that online harassment was a "major problem."

Variable	HARRASS minor problem	HARRASS not a problem
Step 1		
Age	1.094*	1.128
Education	.944*	.765***
Man	–	–
Woman	.673***	.531***
Other gender	.556	.399
White NH	–	–
Black NH	1.268	3.413***
Hispanic	.987	1.237
Asian NH	.810	.638
Other race	.837	1.368
Republican	–	–
Democrat	1.081	1.099
Independent	1.028	.958
other party	.994	1.387
intercept	–.690***	–3.106***
R^2	.054***	.054***
Step 2		
RAS score	2.775***	29.552***
R^2 change	.012	.012
N	6769	
Total R^2	.076***	

All variables were binary dummy variables except for two ordinal variables: age and education. Odds ratios reported are from the full model, that is, controlling for all variables in the model. R^2 reported is Nagelkerke R^2.
*$p < .003125$; **$p < .000625$; ***$p < .0000625$, to account for the Bonferroni correction.

In order to test these relationships, we ran hierarchical binomial (for Questions 1 and 2) and multinomial (for Question 3) logistic regressions. To meet the linearity assumptions test for the OFFENSE item, we transformed the RAS continuous variable by adding 1 to each RAS score, and then calculating its inverse square.

The results, displayed in Tables 1 and 2, indicate that more racially conservative respondents are significantly more likely to believe that (a) people take offensive content online too seriously; (b) calling others out on social media is more likely to punish people who don't deserve it (rather than to hold them accountable for their actions); and (c) people being harassed or bullied online is not a major problem.

Furthermore, younger people are significantly more likely to believe that users take the things they see online too seriously, as are those with lower levels of education. Women and Democrat respondents, however, are more likely than men and Republicans, respectively, to believe that offensive content online is "too often excused as not a big deal." Demographic factors also have significant correlations to how ATP respondents understood the function of "calling others out" online: older respondents were significantly more likely to believe that calling people out holds people accountable. The same can be said for women, Hispanic, Asian non-Hispanic, and Democrat respondents (as compared with men, White non-Hispanic, and Republican counterparts).

When considering how big of a problem harassment and bullying is online, gender and education presented notable correlations. Higher levels of education relate to stronger beliefs in the seriousness of online harassment, as does being a woman (in comparison to being a man).

Racial attitudes and beliefs in the effectiveness of platform-enacted moderation

The final relationship we analyzed was that between racial attitudes and faith in platform-enacted sanctions for harmful online speech. In order to measure this, we averaged participants' scores on three items in the Pew survey which sought people's opinions on the most common platform moderation mechanisms ($M=1.86$; $SD=0.76$; $\alpha=.769$):

> Question: How effective, if at all, do you think the following steps would be in helping to reduce harassment or bullying on social media?
>
> Item 1: Users getting temporarily suspended if they bully or harass others (1 = Very effective; 2 = Somewhat effective; 3 = Not too effective; 4 = Not at all effective).
>
> Item 2: Users getting permanently suspended if they bully or harass others (1 = Very effective; 2 = Somewhat effective; 3 = Not too effective; 4 = Not at all effective).
>
> Item 3: Social media companies proactively deleting bullying or harassing posts (1 = Very effective; 2 = Somewhat effective; 3 = Not too effective; 4 = Not at all effective).

The composite score variable is coded as OPINION in Table 3.

Through a two-step linear regression, we found a significant, positive relationship between respondents' RAS scores and their average attitudes toward these three platform-enacted content moderation mechanisms ($\beta=.364$; $p<.000067$). This indicates that racially conservative attitudes correlate to a lack of trust in the effectiveness of common content moderation strategies; see Table 3.

Discussion

By analyzing this Pew data set through the lens of the racial attitudes scale (RAS), we are able to come to a better understanding of how racial attitudes correlate to content moderation preferences and ideologies surrounding "cancel culture." Content moderation and cancel culture are two related mechanisms—one focusing on the use of platform-based methods for reporting, removing, or otherwise sanctioning online speech, the other focusing on social networking site (SNS) users' participation in counterspeech—with

Table 3. Relationship between racial attitudes and faith in content moderation strategies.

Variable	Opinion
Step 1	
Age	–.162***
Education	.039*
Man	–
Woman	–.072***
Other gender	.000
White NH	–
Black NH	–.028
Hispanic	–.110***
Asian NH	–.027
Other race	.003
Republican	–
Democrat	–.009
Independent	.016
other party	.055***
R^2	.105***
Step 2	
RAS score	.364***
R^2 change	.067
N	6767
Total R^2	.172***

All variables were binary dummy variables except for two ordinal variables: age and education.
Coefficients reported from the full model (final beta controlling for all variables in the model).
R^2 reported is Nagelkerke R^2.
*$p < .0033$; **$p < .00067$; ***$p < .000067$, to account for the Bonferroni correction.

undergirding philosophies that favor curtailing harmful speech online. Both processes are highly contested and politicized, with detractors often arguing that moderation and "cancellation" impede on citizens' freedom of speech and supporters often pointing to the need to create safer online environments for all (Norris, 2023). This study's findings contribute to overall understandings of how racial attitudes and content moderation preferences at large might be correlated, both empirically and theoretically.

Racial attitudes and content moderation preferences

Our findings provide empirical evidence of correlations between social media users' offline racial beliefs and their attitudes toward moderating harmful speech online. Respondents with RAS scores closer to one—that is to say, respondents who held more racially conservative beliefs—were more likely to feel that people take harmful content online too seriously and that harassment and/or bullying online is not a big problem. Essentially, individuals who hold racially conservative beliefs are more likely to minimize online harm, indicating a connection between offline racial conservatism and the trivialization of online harassment.

This trend also holds true when considering the impacts of cancel culture. When survey participants scored higher on the RAS, they were more likely to believe that cancel culture is detrimental, citing that it is more likely to “harm innocent people” than to “hold people accountable.” This correlation may suggest that users with racially conservative beliefs are more likely to highly value freedom of speech, as curtailing freedom of speech is the oft-cited issue that people take with cancel culture (Norris, 2023). White and Crandall (2017) found a similar phenomenon, finding that explicit racism is a predictor of what they call the “free speech defense” of racist expression. Some would also argue that the weaponization of “cancel culture” and “freedom of speech” by the political right has become a “dog whistle” for signaling anti-woke politics (Romano, 2021), and the correlation found in this study may suggest just such a phenomenon.

Furthermore, these analyses indicate that racially insensitive attitudes mapped onto a lack of faith in three common platform-enacted sanctions of harmful speech: temporary account suspension, permanent account suspension, and proactive deletion of harmful posts. These mechanisms have been empirically recognized as effective solutions and have been widely deployed to address online harm since the early days of online regulation (Kiesler et al., 2012). Our analysis shows that higher RAS scores correlated to less trust in the effectiveness of these measures. This could indicate that end-users with racially conservative attitudes do not believe in the utility of these tools due to their lack of support for constraining speech, including harmful speech. However, it could also mean that end-users with less progressive attitudes simply do not believe content moderation is effective from a pragmatic standpoint (i.e. a disbelief in its ability to actually curtail harmful speech; general platform distrust) rather than an ideological one. Future studies could look to delineate the standpoints from which these users are forming their beliefs about the effectiveness of moderation strategies.

Overall, this set of findings indicates that holding more racially conservative beliefs has a significant correlation to general distaste for content moderation and cancel culture, at least during a moment of heightened race-based discourse in Fall 2020. Users with higher RAS scores showed a hesitancy for social regulation of speech (e.g. distaste for cancel culture), a minimization of the severity of harmful online speech (e.g. belief that people take content online too seriously; online bullying/harassment is not a big problem), and a lack of faith in platform-enacted sanctions (e.g. distrust in the effectiveness of suspensions and bans). While these findings have face validity—it makes sense that users who hold racially conservative beliefs may find it problematic to have posts reifying their own beliefs “silenced” via platform content moderation or social cancelation—they also provide empirical evidence that when we are performing studies on content moderation preferences, we must consider racial attitudes as potentially impactful biases. This also suggests that recommendations derived from population-wide analyses that do not account for such biases may not meaningfully serve the moderation needs of the social media users who are most vulnerable to identity-based attacks.

Racial attitudes, content moderation preferences, and moral disengagement theory

Through these findings, we can argue that preferences for less content moderation correlate to racially conservative beliefs. Given the historical moment in which these survey responses were collected—September 2020, immediately following sustained nationwide BLM protests, anti-Asian hate crimes in the face of the Covid pandemic, and crackdowns on racist language online—we argue that these findings indicate that users' self-proclaimed distaste for content moderation may actually be a way of "morally disengaging" from their socially unacceptable beliefs. By supporting a more socially palatable ideology—disdain for sanctioning (harmful) speech online—respondents may be cloaking their support of racially insensitive speech through several different mechanisms of Albert Bandura's (1999) moral disengagement theory (see Figure 1).

Respondents with higher RAS scores could be engaging in *moral justification* (Bandura, 2011) through indicating that cancel culture is more likely to punish people who do not deserve it rather than hold people accountable. The "higher moral purpose," in this case, is protecting the "innocent"—these respondents were more likely than their racially progressive counterparts to believe that cancelation does not actually work toward holding people accountable but rather punishes the undeserving. Second, more racially conservative respondents could arguably be engaging in *diffusion of responsibility* tactics through their disbelief in the efficacy of popular content moderation techniques such as banning users or removing posts. By indicating a lack of faith in the main platform-enacted solutions that we currently have available to stem the proliferation of harmful content online, it becomes, in a sense, everyone and no one's responsibility to manage this speech. If these tools are not working, then it us on "us" to hold guilty parties responsible, but with what tools and in what way is entirely amorphous. This line of argumentation effectively diffuses the responsibility to sanction insensitive speech online—and to not post it in the first place—into an untenable and inconsequential task. Those with higher RAS scores could also be understood as *blaming the victims* for their own suffering (Bandura, 2011) by being more likely to indicate that people take offensive content online too seriously. If people were less sensitive to harmful content online, this logic goes, then they would not feel so victimized. Finally, higher RAS scores also correlate to a *distortion of consequences* (Bandura, 2011). Those with higher RAS scores were more likely to indicate that bullying and harassment online are not that big of a problem. No matter how we define how "big of a problem" bullying and harassment online are, this is patently untrue. Not only did thousands of respondents in this survey indicate having experienced online harassment in some form, but scholars have also indicated the serious health consequences of online harassment (Stevens et al., 2021). Thus, understanding online harassment as "not that big of a deal" certainly qualifies as a distortion of consequences—in this scenario, an undervaluation of how serious and far-reaching the effects of online harm can be.

As shown in this study, a minimization of the severity of harmful speech online, a hesitancy for social regulation of this speech, and a lack of faith in platform-enacted sanctions can be correlated to racially conservative beliefs. This arguably indicates that there may be an ideological slide between racial conservatism and distaste for content moderation online—moderation through both platform-enacted means and social means (such as "cancel culture"). Although we recognize that the moment of Pew's data collection may have made this correlation stronger, as racial tensions were higher than normal at that time, it is worth considering how Bandura's (1999) moral disengagement theory continues to explain explicit and implicit ties between racism, content moderation, and distaste for cancel culture.

Conclusion

At the time of this study's publication, we are facing a moment in which platform-based content moderation is undergoing major change. Elon Musk, now a major federal government influence, named fewer restrictions on content moderation among his goals for the platform when he bought Twitter and rebranded it to X in 2022. These changes in moderation strategies have increased hate speech dramatically on the platform (Hickey et al., 2023). Similarly, Meta has recently announced their rollback of third-party fact-checking and a rewriting of policies on the definition and removal of hateful content, allowing for "more speech," according to Meta's Chief Global Affairs Officer in a January 2025 press release (Kaplan, 2025). Meta is doing away with restrictions on a variety of topics "that are the subject of frequent political discourse and debate," including identity-related topics (Kaplan, 2025). The press release goes on to say, "It's not right that things can be said on TV or the floor of Congress, but not on our platforms" (Kaplan, 2025). With the recognition that "anti-DEI" measures are a front-stage, high-priority policy implementation in President Trump's second term (Aratani, 2025), what "can be said on . . . the floor of Congress" does not bode well as a litmus test for content moderation and online safety for historically marginalized users. Understanding that racially conservative attitudes are correlated to preferences for light content moderation practices is a necessary lens through which we should be analyzing these shifts—and ultimately narrativizing their impacts.

Future work should meaningfully interrogate the connections between racially conservative attitudes and unequivocal support for freedom of speech—both online and offline—as the present study would indicate there is likely a positive correlation between the two. Although Pew's survey data did not give us the tools to answer that particular question, as we were limited by the items that they chose to include, ties between racial conservatism and distaste for content moderation and cancel culture provide compelling evidence that the freedom of speech connection should be investigated. Future studies that collect data directly for the purpose of investigating such relationships should be able to better deploy survey items that explicitly measure the variables of interest. Another limitation of this study is that the data can only capture a particular moment in time—a limitation all survey data is plagued by—and the September 2020 moment is both a strength and a weakness. Its strength is in its ability to capture the heightened racial tension of the moment (Chong and Druckman, 2007), which we argue framed much of the discourse and mental models on content moderation around race-based discourse, but understanding if these relationships hold beyond that time is necessary to understand and should be attended to in future work.

We already know there can be deleterious effects from experiencing harmful online content—particularly that which is race-based (Bresnahan et al., 2023; Saha et al., 2019). Without concerted platform checks on that speech, we are likely to see those effects amplify. Understanding racial conservatism as a bias that relates to preferences for "more speech," as Meta would say, predicts a difficult cultural moment for confronting the balance of online safety and freedom of expression—a balance that has, for now, tipped decidedly toward the latter.

Acknowledgements

The authors thank Serita Sargent for their support in the creation of the racial attitudes scoring mechanism utilized in this project and Kiran Garimella for assistance in early stages of the project.

Data availability statement

Please see Pew Research Center's full data set here: https://www.pewresearch.org/science/dataset/american-trends-panel-wave-74/

Ethical considerations

Per Rutgers University's IRB, approval was not needed for this study, as our use of a fully anonymized secondary data set is not considered human subject research.

Consent to participate

Not applicable.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by an award from the National Science Foundation's Division of Information and Intelligent Systems (Grant No.: 2329394).

ORCID iDs

Alyvia Walters https://orcid.org/0000-0002-8862-4832

Tawfiq Ammari https://orcid.org/0000-0002-1920-1625

Shagun Jhaver https://orcid.org/0000-0002-6728-7101

Notes

1. Please see https://www.pewresearch.org/dataset/american-trends-panel-wave-74/ for Pew's published reports borne from this dataset.
2. Of note, these policies have since been rolled back.
3. See Pew's publicly accessible dataset for a complete list of survey questions: https://www.pewresearch.org/science/dataset/american-trends-panel-wave-74/
4. Bonilla-Silva's (2022) cornerstone work *Racism Without Racists* employs similar methodologies when considering survey data. Bonilla-Silva states that, in fact, surveys may serve as a limitation in the direction of conservatism because "respondents work hard to choose the "right" answers (i.e. those that fit public norms). For instance, though a variety of data

suggest racial considerations are central to whites' residential choices, more than 90 percent of whites state in surveys that they have no problem with the idea of blacks moving into their neighborhoods." (p. 11) Thus, we employ this proxy methodology while also acknowledging that ideologies in the direction of steeper racist beliefs may actually be at play.

5. Given that there is no strong literature to suggest how or if any of the included metrics should be weighted more heavily than others, we chose to simply average participants' responses to the scale measures, as is common practice identified by Greco et al. (2018).
6. For the purposes of this study, we define "racial conservatism" as a perspective that includes colorblind or race-neutral views alongside a racially resentful perspective. It builds from conservative elites' decades-long articulations of positions which distanced themselves from considering race in the policymaking process (Engelhardt, 2019; King and Smith, 2014; Lowndes, 2009; Mendelberg, 2001). It incorporates liberal ideals like equal opportunity, choice, and individualism, and minimizes discrimination as explanation for potentially race-related matters (Bonilla-Silva, 2022; Engelhardt, 2021). Racial conservatism also includes a view that we are now "post-race" (Goldberg, 2015) and society is truly a meritocracy—the best, most qualified people rise to the top. Thus, if processes violate meritocratic norms, as in instances of affirmative action, racial conservatism sees this as constituting reverse discrimination (Bonilla-Silva, 2022). By deeming skin color irrelevant, racial conservatives are unlikely to ascribe importance to race when addressing social problems, and may overtly support regressive policies or practices.
7. Pew also coded this question as cancel culture in the dataset.

References

Alizadeh M, Gilardi F, Hoes E, et al. (2022) Content moderation as a political issue: the Twitter discourse around Trump's ban. *Journal of Quantitative Description: Digital Media* 2: 23.

Appel R, Pan J and Roberts ME (2023) Partisan conflict over content moderation is more than disagreement about facts. *Science Advances* 9: 6799.

Aratani L (2025) What we know so far about Trump's orders on diversity, equity and inclusion. *The Guardian*, 26 January. Available at: https://www.theguardian.com/us-news/2025/jan/26/trump-executive-orders-dei

Bandura A (1999) Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review* 3(3): 193–209.

Bandura A (2011) Moral disengagement. In: *The Encyclopedia of Peace Psychology*. DOI: 10.1002/9780470672532.wbepp165.

Baym NK (1995) The emergence of community in computer-mediated communication. In:Jones SG (ed.) *Cybersociety: Computer-Mediated Communication and Community*. London: SAGE, pp. 138–163.

Baym NK (2010) *Personal Connections in the Digital Age*. London: Polity Press.

Bliuc A, Falukner N, Jakubowicz A, et al. (2018) Online networks of racial hate: a systematic review of 10 years of research on cyber-racism. *Computers in Human Behavior* 87: 75–86.

Bonilla-Silva E (2022) *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America*. Lanham, MD: Rowman & Littlefield.

Bresnahan M, Zhu Y, Hooper A, et al. (2023) The negative health effects of anti-Asian stigma in the U.S. during COVID-19. *Stigma and Health* 8(1): 115–123.

Bridges KM (2021–2022) Language on the move: "Cancel Culture," "Critical Race Theory," and the digital public sphere. *Yale Law Journal Forum* 131: 767–798. Available at: https://heinonline.org/HOL/P?h=hein.journals/yljfor131&i=776

Carian E (2022) "No seat at the party": mobilizing White masculinity in the men's rights movement. *Sociological Focus* 55(1): 27–47.

Chong D and Druckman JN (2007) Framing theory. *Annual Review of Political Science* 10: 103–126.

Chou WS and Gaysynsky A (2021) Racism and xenophobia in a pandemic: interactions of online and offline worlds. *American Journal of Public Health* 111: 773–775. Available at: https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2021.306230

Costley White K (2018) *The Branding of Right-wing Activism: The News Media & the Tea Party*. Oxford: Oxford University Press.

Engelhardt AM (2021) The content of their coverage: contrasting racially conservative and liberal elite rhetoric. *Politics, Groups, and Identities* 9(5): 935–954. https://doi.org/10.1080/21565503.2019.1674672

Fisher RJ and Katz JE (1999) Social-desirability bias and the validity of self-reported values. *Psychology & Marketing* 17(2): 105–120.

Giani M and Mèon P (2021) Global racist contagion following Donald Trump's election. *British Journal of Political Science* 51(3): 1332–1339.

Gilens M (2009) *Why Americans Hate Welfare*. Chicago, IL: University of Chicago Press.

Goldberg DT (2015) *Are We All Postracial Yet?* London: Polity Press.

Greco S, Ishizaka A, Tasiou M, et al. (2018) On the methodological framework of composite indices: a review of the issues of weighting, aggregation, and robustness. *Social Indicators Research* 141: 61–94.

Hagendoorn L and Nekuee S (2018) *Education and Racism: A Cross National Inventory of Positive Effects of Education on Ethnic Tolerance*. London: Routledge.

Haimson OL, Delmonaco D, Nie P, et al. (2021) Disproportionate removals and differing content moderation experiences for conservative, transgender, and Black social media users: marginalization and moderation gray areas. *Proceedings of the ACM on Human-computer Interaction* 2021: 21–35.

Hall S, Critcher C, Jefferson T, et al. (2013) *Policing the Crisis: Mugging, the State and Law and Order*. 2nd ed. London: Palgrave Macmillan.

Hawkins I and Saleem M (2022) How social media use, political identity, and racial resentment affect perceptions of reverse racism in the United States. *Computers in Human Behavior* 134: 107337.

Hawkins I, Roden J, Attal M, et al. (2023) Race and gender intertwined: why intersecting identities matter for perceptions of incivility and content moderation on social media. *Journal of Communication*, 73: 6539–6551.

Hickey D, Schmitz M, Fessler D, et al. (2023) Auditing Elon Musk's impact on hate speech and bots. *Proceedings of the International AAAI Conference on Web and Social Media* 17(1): 1133–1137.

Holt LF and Sweitzer MD (2018) More than a black and white issue: ethnic identity, social dominance orientation, and support for the Black Lives Matter movement. *Self and Identity* 19(1): 16–31.

Jhaver S and Zhang AX (2023) Do users want platform moderation or individual control? Examining the role of third-person effects and free speech support in shaping moderation preferences. *New Media & Society* 27(5): 2930–2950.

Jhaver S, Zhang A, Chen Q, et al. (2023) Personalizing content moderation on social media: user perspectives on moderation choices, interface design, and labor. *Proceedings of the ACM on Human-Computer Interaction* 7: 1–33.

Kaplan J (2025) More speech and fewer mistakes. *Meta*, 7 January. Available at: https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

Keum BT and Miller MJ (2018) Racism on the Internet: conceptualization and recommendations for research. *Psychology of Violence* 8(6): 782–791.

Kiesler S, Kraut RE, Resnick P, et al. (2012) Regulating behavior in online communities. In: Kraut RE, Resnick P and Kiesler S (eds) *Building Successful Online Communities: Evidence-Based Social Design* (pp. 112–159). Cambridge, MA: MIT Press.

King DS and Smith RM (2014) "Without regard to race": critical ideational development in modern America politics. *The Journal of Politics* 76(4): 958–971.

Kishi R and Jones S (2020) Demonstrations and political violence in America: new data for summer 2020. Armed Conflict Location & Event Data Project, 3 September. Available at: https://acleddata.com/2020/09/03/demonstrations-political-violence-in-america-new-data-for-summer-2020/

Kozyreva A, Herzog SM, Lewandowsky S, et al. (2023) Resolving content moderation dilemmas between free speech and harmful misinformation. *Proceedings of the National Academy of Sciences of the United States of America* 120(7): e2210666120. https://doi.org/10.1073/pnas.2210666120

Lowndes JE (2009) *From the New Deal to the New Right: Race and the Southern Origins of Modern Conservatism*. New Haven, CT: Yale University Press.

Meesala S (2020) Cancel culture: a societal obligation or infringement on free speech? *UAB Institute for Human Rights Blog*, 4 December. Available at: https://sites.uab.edu/humanrights/2020/12/04/cancel-culture-a-societal-obligation-or-infringement-on-free-speech/

Mendelberg T (2001) *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton, NJ: Princeton University Press.

Myers West S (2018) Censored, suspended, shadowbanned: user interpretations of content moderation on social media platforms. *New Media & Society* 20(11): 4366–4383.

Naab TK, Naab T and Brandmeier J (2021) Uncivil user comments increase users' intention to engage in corrective actions and their support for authoritative restrictive actions. *Journalism & Mass Communication Quarterly* 98(2): 566–588.

Ng R and Indran N (2024) Social media discourse on ageism, sexism, and racism: analysis of 150 million tweets over 15 years. *Journal of the American Geriatrics Society* 72(10): 3149–3155.

Nguyen TT, Criss S, Michaels EK, et al. (2021) Progress and push-back: how the killings of Ahmaud Arbery, Breonna Taylor, and George Floyd impacted public discourse on race and racism on Twitter. *SSM: Population Health* 15: 100922.

Norris P (2023) Cancel culture: Myth or reality? *Political Studies* 71(1): 145–174.

Peterson-Salahuddin C (2024) Repairing the harm: toward an algorithmic reparations approach to hate speech content moderation. *Big Data & Society* 11(2): 245333.

Pew Research Center (2021) The state of online harassment. Available at: https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/

Pew Research Center (2025) The American Trends Panel. Available at: https://www.pewresearch.org/the-american-trends-panel/

Picarella L (2024) Intersections in the digital society: cancel culture, fake news, and contemporary public discourse. *Frontiers in Sociology* 9: 1376049.

Roberts ST (2017) Content moderation. *Encyclopedia of big data*. Available at: https://escholarship.org/uc/item/7371c1hf

Romano A (2021) The second wave of "cancel culture": how the concept has evolved to mean different things to different people. *Vox*, 5 May. Available at: https://www.vox.com/22384308/cancel-culture-free-speech-accountability-debate

Saha K, Chandrasekharan E and De Choudhury M (2019) Prevalence and psychological effects of hateful speech in online college communities. In: *Proceedings of the 10th ACM conference on Web Science*, Boston, MA, 30 June–3 July.

Sawyer J and Gampa A (2018) Implicit and explicit racial attitudes changed during Black Lives Matter. *Personality and Social Psychology Bulletin* 44(7): 1039–1059.

Shook LM and Lizarraga-Dueñas LI (2024) Of DEI and denials: a critical discourse analysis of Texas' 88th legislative session. *Education Policy Analysis Archives* 32: 8590.

Spicer R (2022) The marketplace of ideas, cancel culture, and misunderstanding the First Amendment. *Communication and Democracy* 56(2): 192–197.

Stevens F, Nurse JRC and Arief B (2021) Cyber stalking, cyber harassment, and adult mental health: a systematic review. *Cyberpsychology, Behavior, and Social Networking* 24(6): e0253.

Tandoc EC, Tan Hui Ru B, Lee Huei G, et al. (2024) #CancelCulture: examining definitions and motivations. *New Media & Society* 26(4): 1944–1962.

Tietjen RR and Tirkkonen SK (2023) The rage of lonely men: loneliness and misogyny in the online movement of "involuntary celibates" (incels). *Topoi* 42: 1229–1241.

Trump DJ (2020) Remarks at an Independence Day celebration at the Mount Rushmore National Memorial in Keystone, South Dakota (speech). Available at: https://www.govinfo.gov/content/pkg/DCPD-202000494/pdf/DCPD-202000494.pdf

Walters A, Ammari T, Garimella K, et al. (2024) Online knowledge production in polarized political memes: the case of critical race theory. *New Media & Society* 27(9): 4997–5021.

Weber I, Goncalves J, Masullo GM, et al. (2024) Who can say what? Testing the impact of interpersonal mechanisms and gender on fairness evaluations of content moderation. *Social Media + Society*. Epub ahead of print 26 November. DOI: 10.1177/20563051241286702.

West K, Greenland K and van Laar C (2021) Implicit racism, colour blindness, and narrow definitions of discrimination: why some White people prefer "All Lives Matter" to "Black Lives Matter.." *British Journal of Social Psychology* 60: 1136–1153.

White MHII and Crandall CS (2017) Freedom of racist speech: ego and expressive threats. *Journal of Personality and Social Psychology* 113(3): 413–429.

Winter NJG (2008) *Dangerous Frames: How Ideas about Race & Gender Shape Public Opinion*. Chicago, IL: University of Chicago Press.

Author biographies

Alyvia Walters is a Visiting Assistant Teaching Professor at Villanova University. Her research interests center around the intersections of social media, content moderation, and politicized communication. Walters holds a PhD from Rutgers University's School of Communication, Information, and Media.

Tawfiq Ammari is an Assistant Professor of Library and Information Science at the Rutgers University School of Communication and Information. He is a mixed methods researcher who connects critical theory from Science, Technology, and Society studies (STS) with computational social science techniques to advocate for equity and progressive social change in online context with a focus on trauam-informed design. For more information, visit: https://sites.comminfo.rutgers.edu/tammari/

Shagun Jhaver is an assistant professor in the School of Communication & Information at Rutgers University. He takes a user-centered approach to studying platforms' design, moderation policies, and algorithmic procedures, with a research goal of making the internet safer and fairer for everyone. Jhaver holds a PhD from Georgia Tech's School of Interactive Computing and an MS from the University of Texas at Dallas. For more information, visit https://shagunjhaver.com/

Appendix 1

Race-based news discourse in 2020

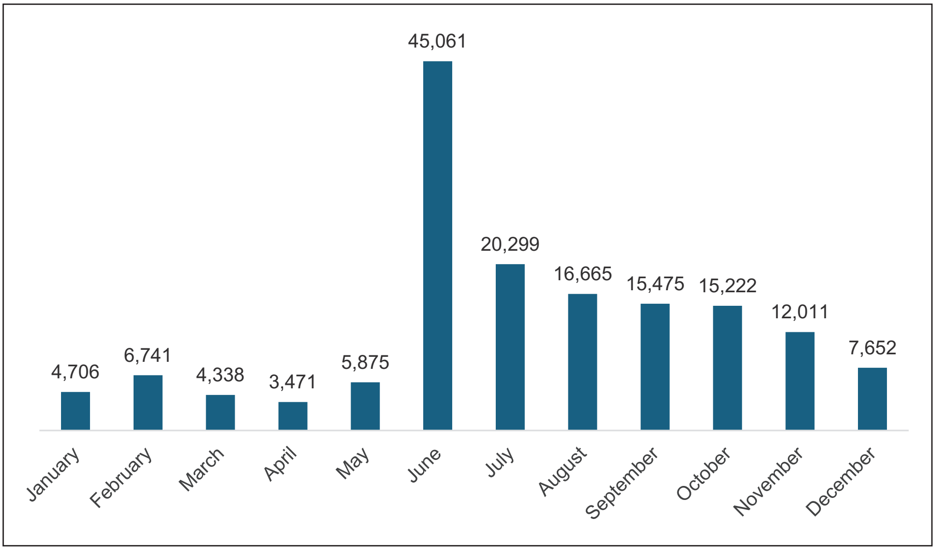

Figure 7. News-based discourse on racism in the United States in 2020.
Data gathered from Dow Jones database Factiva, which has access to over 30,000 US-based news publications. Keywords "race relations," "racial inequality," "racism," "white supremacy," "race-based harm," and "racist" used to conduct search.
Note that George Floyd was murdered on 25 May and the ATP Wave 74 Survey ran from 8 to 13 September.

Appendix 2

Survey items included in the racial attitudes scaling mechanism

(Cronbach's alpha = .87, N = 20)
Variables relating to racial discrimination:

Overall, how does each of the following affect people's ability to get ahead in our country these days?

- Being White
- Being Black
- Being Hispanic
- Being Asian
  - Helps a lot
  - Helps a little
  - Hurts a little

    - Hurts a lot
    - Neither helps nor hurts

When it comes to racial discrimination, which do you think is the bigger problem for the country today?

- People seeing racial discrimination where it really does NOT exist
- People NOT seeing racial discrimination where it really DOES exist

When it comes to giving Black people equal rights with White people, do you think our country has . . .

- Gone too far
- Not gone far enough
- Been about right

In general in our country these days, would you say that Black people are treated less fairly than White people, White people are treated less fairly than Black people, or both are treated about equally in each of the following situations?

- In hiring, pay and promotions
- In stores or restaurants
- When applying for a loan or mortgage
- In dealing with the police
- When voting in elections
- When seeking medical treatment
    - Black people are treated less fairly than White people
    - White people are treated less fairly than Black people
    - Both are treated about equally

Variables relating to combatting racism:
In general, do you think there is too much, too little, or about the right amount of attention paid to race and racial issues in our country these days?

- Too much attention
- Too little attention
- About the right amount of attention

How important, if at all, do you think it is for people in our country to do each of the following?

- Educate themselves about the history of racial inequality in our country
- Confront other people when they say or do something racist
- Support businesses that are owned by racial or ethnic minorities
- Attend protests or rallies focused on issues related to racial equality

- Choose to live in communities that are racially and ethnically diverse
- Have conversations about race with people who are not the same race as them
  - Very important
  - Somewhat important
  - Not too important
  - Not at all important

Variables relating to Black Lives Matter:

From what you've read and heard, how do you feel about the Black Lives Matter Movement?

- Strongly support
- Somewhat support
- Somewhat oppose
- Strongly oppose