

Diversity among Colleagues and Individual Decision Making in the Criminal Legal System

Allison P. Harris* Nicholas Ottone† Hannah Walker‡

Abstract

How does being surrounded by diverse colleagues impact individual decision making in the criminal legal system? Research suggests that group-level racial diversity can reduce racially disparate criminal legal outcomes by leading members of the dominant group to change their behavior towards minority-racial-group-member defendants. With an original interactive survey experiment, we explore the psychological mechanisms underlying this relationship by randomly assigning respondents to live online chats in which we manipulate the perceived racial composition of chat groups and discussion topics. Respondents then make sentencing decisions for five Black and five White defendant profiles. We find that assignment to diverse chats focused on criminal justice decreased respondents' selection of an incarceration sentence for Black defendants by 3.3 percentage points compared to homogenous criminal justice chats. Analysis of 28,000 chat messages suggests that relevant conversation with racial out-group members drives dominant racial group members' behavior changes as their colleague group becomes more racially diverse.

*Yale University

†Stanford University

‡University of Texas at Austin

The U.S. criminal legal system produces some of society’s starkest and most enduring racial disparities, particularly when comparing the experiences of Black and White individuals. Among the many suggestions reform-minded activists propose to address these disparities is greater representation of non-White, especially Black, officials within the institutions comprising the criminal legal system, including police departments, prosecutors offices, and courts. Calls to diversify criminal justice officials join broader exhortations to diversify workplaces of all kinds – even as diversity, equity and inclusion programming has been dismantled in both government and private industry. The contest over the value of diversity in public and private institutions raises questions scholars have not satisfactorily answered: does increased group-level racial diversity impact the behavior of group members? If so, how?

There are a few mechanisms by which increased organizational diversity might impact the behavior of individuals. It may be that individuals who are members of the majority group make decisions in ways that fundamentally differ from minority group members. Evidence that policy makers of different identities make systematically different decisions is mixed, especially in criminal legal settings governed by strict policy (Welch, Combs and Gruhl 1988; Scherer 2004; Cohen and Yang 2019; Spohn 1990; Schanzenbach 2005; Abrams, Bertrand and Mullainathan 2012; Steffensmeier and Britt 2001; Sharp 2014; Legewie and Fagan 2016; Ba et al. 2021). There is some evidence that Black representation in leadership in policing can improve outcomes for Black civilians (Wu 2021; Shoub and Christiani 2023), suggesting that organizational culture can be changed by diversity. Likewise, as they gain Black colleagues, White judges become less punitive toward Black defendants (Harris 2024). Thus, there is some evidence that diversity among criminal legal officials could reduce racial disparities in outcomes, but the process by which this would occur is more nuanced than we might expect.

Rather than race directly determining how officials make decisions, organizational diversity is likely to impact behavior via mechanisms associated with intergroup contact (Pettigrew 1998; Allport 1955). A substantial body of research demonstrates that interacting with out-group members can change how dominant group members perceive, think about, and behave toward the outgroup. In the best of circumstances, such inter-

actions can change behavior by reducing prejudice (Pettigrew 1998). But racial diversity may also change the types of *conversations* that occur among group members, shifting the considerations upon which individuals base their decisions (Sommers 2006). In sum, racial diversity among organizational members alters the dynamics within an institution thereby influencing the decisions individuals make. This would be consistent with the observation that White judges’ behavior is most affected by increasing diversity on the bench (Harris 2024) – but no research of which we are aware examines these more nuanced mechanisms, specifically.

To investigate how group-level diversity in a workplace might impact relevant organizational outcomes, we examine how individuals make decisions about culpability, and how they think about attendant sentencing. A person is arrested every three seconds in the United States¹ and state court systems processed over 16 million criminal cases in 2023 with criminal cases representing 24 percent of the caseload (*NCSC Annual Report 2023: Inspiring Future-Ready Courts* 2023). Experiences with the criminal legal system are increasingly common and racial disparities in outcomes enduring. Officials at all points in the criminal legal system make consequential decisions about a given case, which are subject to individual biases (implicit and explicit) and impact downstream outcomes.

We use evidence from an interactive survey experiment that simulates the adjudication and sentencing process criminal legal officials routinely confront. Respondents were presented with fictional defendant profiles featuring either Black or White individuals. The profiles included some details about the fictional individuals’ cases, and respondents were asked to reflect on the appropriateness of different sentencing options. But prior to completing the sentencing exercise, individuals were first invited to participate in a conversation with a group for which the perceived racial composition was experimentally manipulated. We are primarily interested in how White individuals’ behavior changes as a consequence of group diversity. Accordingly, we test how experimentally varying group composition (homogeneous or diverse) impacts how a respondent thinks about appropri-

¹<https://www.vera.org/publications/arrest-trends-every-three-seconds-landing/arrest-trends-every-three-seconds/overview>

ate sentencing. Consistent with intergroup contact theory, we find that when respondents engaged in relevant conversations about criminal justice with out-group members, they made more equitable sentencing decisions than did those individuals who engaged in conversations with a homogenous group.

This study has important implications for how we understand diversity and the way it operates in many types of organizations. We focused our study on how people view sentencing decisions, placing ordinary citizens in the context of the criminal legal system. On one hand, the sentencing process is overseen by judges, and respondents are far from being legal actors themselves. But the findings pair with research elsewhere demonstrating that, among judges, dominant group members change their behavior in response to increased racial diversity on the bench and provide evidence for mechanisms associated with inter-group contact.² Scholars have found evidence that intergroup contact can improve outgroup perceptions across a range of contexts – evidence from our interactive experiment complements this body of work. The fact that we tasked everyday citizens, who are not themselves criminal legal actors, with evaluating appropriate sentencing and found adjustments to individual level behavior that mirror those observed among judges from analyses of administrative data suggests the broad applicability of our claims. Group-level diversity should be expected to change the behavior of dominant group members across a variety of organizational contexts.

Proponents of increasing diversity (on many dimensions) argue that the benefits of diverse groups include increased innovation over homogenous groups and a greater sense of inclusion among members who are in the minority (e.g. *10 Reasons Why DEI Is Important for Businesses and How it's Evolving* 2024). Opponents of programs that seek to increase diversity argue that such programs promote demographics at the expense of expertise, or that such programs do not succeed in making workplaces and people more tolerant (e.g. Butcher 2023). Our findings suggest that the impact of racial diversity in a group extends beyond innovation and inclusion, and that it does not primarily derive from diversity

²Even as everyday citizens do not themselves make sentencing decisions, they do interact with information and entertainment about the criminal legal system on a regular basis, all of which informs how they view the fairness of the system itself.

trainings, from which scholars have not found clear links to prejudice reduction (Paluck et al. 2021). Our findings also further call into question critiques that pit racial diversity against qualifications: we highlight how members of majority racial groups respond to membership in a diverse group regardless of skills or other personal characteristics. Our innovative, interactive online experiment improves upon recent observational analyses of this topic, which do not allow for the identification of the mechanisms that might underlie the relationship between racial diversity and decision making. In this paper, we identify the social psychological processes that might lead White officials to alter their behavior as they gain Black colleagues in the context of sentencing decisions on a criminal trial court. In the sections that follow we outline our argument and theory, describe our research design and present results.

Argument and Theory

This paper identifies the means by which racial diversity in a group can influence the punishment decisions of individuals who identify with the majority racial group, focusing on two possibilities: intergroup contact and discussion. We argue that even relatively short instances of intergroup contact between majority and minority racial group members can lead majority group members to change their behavior, particularly when the content of the interactions is relevant to subsequent tasks individuals must complete. In US criminal legal institutions, which are known to produce wildly disparate outcomes for Black and White civilians, we might expect that interactions with Black colleagues could lead White officials' decisions to become less racially disparate.

Many factors may influence individuals' behavior that correlate both with increasing diversity in a group and individual behavioral changes, and these factors are not easily captured with observational analyses. The survey experiment we present in the next section assesses the psychological processes that underly the relationship between racial diversity among judges and majority-racial-group member judges' sentencing behavior. We focus specifically on differences in White individuals' decisions with respect to Black and White defendants when they are part of a racially homogenous group compared to

when they are in a group that includes Black people. We limit our definition of majority- and minority-racial groups to White and Black in this study for two reasons: first, because of the theoretical importance of Black-White racial disparities throughout U.S. history, especially in the criminal legal system; and second, because this type of racial diversity has been the focus of most existing literature on racial identity, descriptive representation, and racial diversity in government institutions.

What processes might explain a relationship between group-level racial diversity and reduced disparities in White individuals' decisions? Both of the potential psychological mechanisms we explore here are grounded in the extensive literature on intergroup contact (Selvanathan and et al. 2018; Enos 2014; Paluck, Green and Green 2019; Kalla and Broockman 2020; Nathan and Sands 2023; Allport 1955; Majumdar 2026). Intergroup contact theory posits that relationships that develop between in- and out-group members can change attitudes individuals hold about the group of which they are not a member. Under the best of circumstances, intergroup contact can reduce prejudice towards out-group members (Allport 1955). Researchers have demonstrated that contact must be sustained and collegial—simple exposure to out-group members not accompanied by a meaningful relationship can instead invoke feelings of threat (Enos 2014). Yet, scholars have found evidence that intergroup contact can reduce prejudice against a wide variety of stigmatized groups, including people with disabilities, people who identify as LGBTQIA, immigrants, members of differing religious groups, women, as well as along lines of race (Pettigrew and Tropp 2006; Nathan and Sands 2023; Majumdar 2026).

We might therefore expect that in a government agency (such as a courthouse), where officials regularly have collegial interactions with each other, intergroup contact could effectively reduce prejudice or heighten the salience of the value of racial equality sufficient to change White officials' behavior. A recent meta-analysis of studies examining intergroup contact in a controlled setting, where it is randomly assigned, casts some doubt on whether intergroup contact can reduce *racial* prejudice, however (Paluck, Green and Green 2019). Paluck, Green and Green (2019) note that very few studies have examined this specific relationship, and those that do have only evaluated populations under the age of 25 and have yielded very modest findings. It is important to keep in mind, however,

that behavior change does not require an update to deeply held beliefs (Paluck et al. 2021). This is particularly true if we are considering officials' behavior toward civilians. A White official may alter their behavior when interacting with Black civilians without updating their beliefs about Black people more generally.

Contact alone may not be enough to change individuals' behavior toward racial out group members. Another possibility is that membership in a racially diverse group might alter the information exchanges that occur during conversations among group members. Scholars have identified discussion as a vehicle through which peer influence can lead to prejudice reduction in a variety of settings (Alimo 2012; Dessel 2010; Gurin, Nagda and Zúñiga 2013; Paluck and Green 2009). If Black people are more likely to discuss race or inequality, exposure to such information could increase the salience of race and racial inequality for White group members who otherwise may not so frequently think about these subjects (Dottolo and Stewart 2013). In the context of public officials working in an institution, White officials, influenced by these conversations, might adjust their behavior such that it is less disparate and less punitive toward Black civilians (Harris 2024).

Allport's original theoretical insight was not that intergroup contact is a panacea, but instead that it can be particularly potent under certain conditions (2019; 1955). When group members hold equal status in a given situation, share common goals, engage in some intergroup cooperation, and have the support of laws and customs, intergroup contact should be especially operative to impact attitudes and behaviors. And, some evidence suggests that conversations need not be deep and sustained to have a positive impact on attitudes towards an out-group – even light-touch text exchanges were demonstrated to have positive outcomes (Majumdar 2026). These conditions characterize coworker relationships in several settings, and, importantly, those among judges in a courthouse, teachers in a school, and assistant prosecutors assigned to the same division. It is therefore theoretically plausible that interactions among diverse colleagues themselves shape behaviors among individual group members vis-a-vis the mechanisms described by intergroup contact theory, and we generate the following two expectations. First, we expect that White individuals who are part of a racially diverse group will be less punitive toward Black defendants than those who are members of an all-White

group. Second, we expect this relationship to be especially pronounced when the content of group conversations is related to crime and punishment.

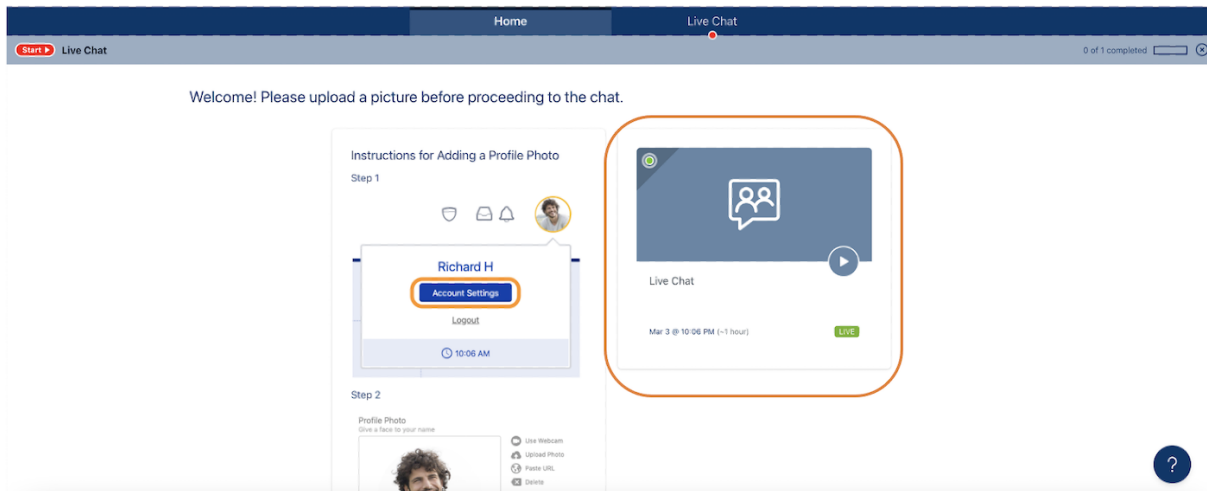
Research Design

From January 2023 through December 2023, we worked with Amazon Mechanical Turk’s Cloud Research to recruit 1,842 White respondents to participate in a study about judicial decision making. Respondents were asked to pretend they were judges. They then participated in a two-part study that consisted of an online activity and a survey. After respondents completed the online consent process, they were randomly assigned to one of five conditions:

1. **Control:** Respondents complete a brief non-chat-based online task and then complete a seven-minute criminal sentencing survey.
2. **Racially homogeneous group, crime and punishment chat:** After uploading a profile photo, respondents engage in a ten-minute live chat about crime and punishment in the United States with a racially homogeneous group. Afterwards, they complete a seven-minute criminal sentencing survey.
3. **Racially diverse group, crime and punishment chat:** After uploading a profile photo, respondents engage in a ten-minute live chat about crime and punishment in the United States with a racially diverse group. Afterwards, they complete a seven-minute criminal sentencing survey.
4. **Racially homogeneous group, environment chat:** After uploading a profile photo, respondents engage in a ten-minute live chat about climate change with a racially homogeneous group. Afterwards, they complete a seven-minute criminal sentencing survey.
5. **Racially diverse group, environment chat:** After uploading a profile photo, respondents engage in a ten-minute live chat about climate change with a racially diverse group. Afterwards, they complete a seven-minute criminal sentencing survey.

After assignment, respondents received a link to a study website on the Recollective platform. Recollective is a web-based qualitative research platform. In the control condition, respondents were asked to answer a few questions about their experience taking the survey, after which they were given a code that allowed them to complete the survey in Qualtrics.³ In the treatment conditions, respondents navigated to a landing page on the Recollective site (see Figure 1) and followed instructions to upload a profile photo. The landing pages were identical across conditions, and we customized them to minimize attrition in the transition from Qualtrics to Recollective.⁴ Respondents then entered a live chat. As seen in Figure 2, study administrators post the chat’s prompt, which then “floats” at the top of the screen for the duration of the ten-minute chat so that respondents can easily refer to it.

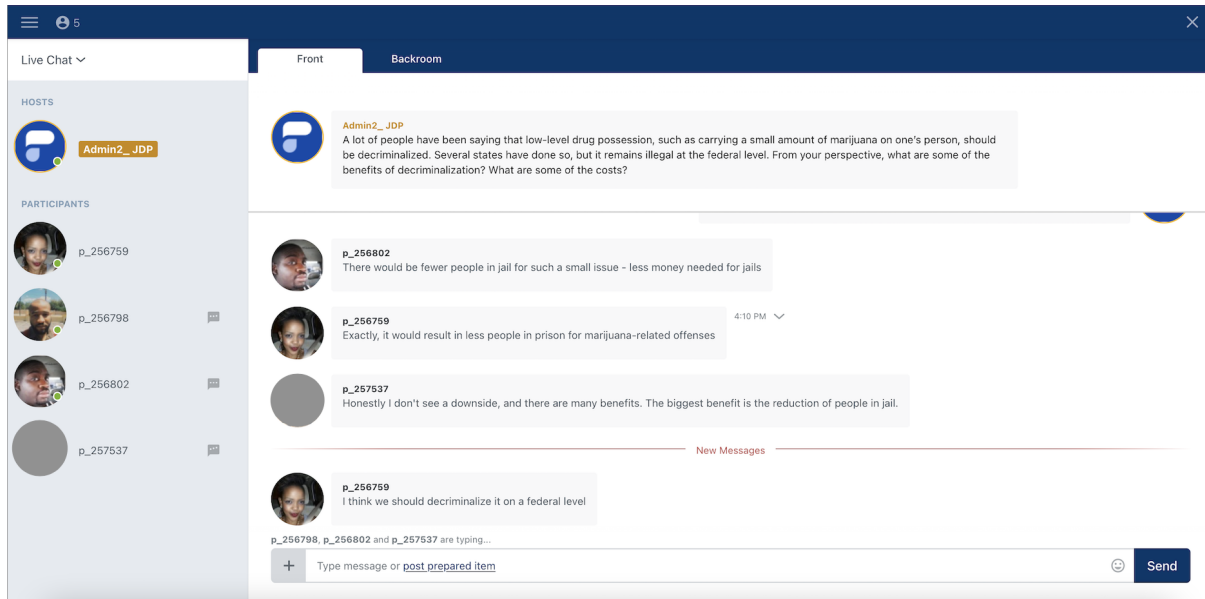
Figure 1: Treatment Condition Landing Pages



³This exercise was designed to replicate the technological hurdle of shifting to another online platform, which treated respondents are also asked to do. Therefore, we expect this to minimize differential attrition between treatment conditions

⁴We piloted this study over Zoom and experienced higher levels of attrition compared to the Recollective platform.

Figure 2: Live Chat Example



The criminal justice prompt reads as follow:

A lot of people have been saying that low level drug possession, such as carrying a small amount of marijuana on one's person, should be decriminalized. Several states have done so, but it remains illegal at the federal level. From your perspective, what are some of the benefits of decriminalization? What are some of the costs?

The environment prompt reads as follows:

Greta Thunberg is a climate activist who became mobilized after learning about the growing problem of global warming. She has said that, "Some people say that the climate crisis is something we have all created. But that is just another convenient lie. Because if everyone is guilty then no one is to blame," and "But I don't want your hope. . . I want you to panic. I want you to feel the fear I feel every day. And then I want you to act." Do you agree with Thunberg's perspective? Why or why not?

In order to manipulate group composition we rely on confederates who participate in our live chats alongside respondents. The confederates are paid undergraduate

research assistants who participated in training sessions to learn about the project and their role in it.⁵ In diverse chat conditions, confederates upload selfies of Black individuals so that White respondents perceive a racially diverse group. In homogeneous chat conditions, confederates upload selfies of White individuals so that White respondents perceive a racially homogeneous group.⁶ Confederates were trained to engage with survey respondents as if they, too, were participating in the study and to take positions on the prompt. Chats consisted of two to four confederates and one to five respondents. A research assistant also served as the administrator of each chat room, ensuring the prompt was pasted to the top of the chat, keeping time, monitoring the chat, and answering respondent questions should they arise.

At the end of a chat, the administrator shared a code which would allow respondents to advance in the Qualtrics survey. All respondents completed the same seven-minute survey in which they were asked to judge ten defendant profiles. Each profile contained a photograph, a brief description of the felony for which the defendant had been convicted, the felony’s corresponding class, and a box explaining the different felony classes. Defendant profiles varied on two dimensions: race (Black or White) and felony class (Class 2 or Class 3). For each profile, respondents were asked whether a short term in prison would be appropriate or whether a sentence more or less severe would be appropriate. Respondents then answered a set of questions that collected information on demographics and racial attitudes.

Following a pilot period to refine the data collection process, we pre-registered this experiment prior to any randomization or data collection.⁷ We evaluate the following pre-

⁵Our standard training introduced research assistants to the research question and design, including the purpose of each experimental condition. We also told them what their role would be and initiated a ten-minute practice chat so they could familiarize themselves with the platform. We finished by discussing potentially difficult scenarios, such as inflammatory and abusive language. Our training materials are available upon request.

⁶We conducted a survey prior to the launch of this project to assess how respondents perceived multiple selfies. We then selected for inclusion those selfies that respondents rate as most similar across a variety of characteristics.

⁷Please see our anonymized pre-registration at this link:

registered hypotheses:

1. Respondents assigned to diverse groups will sentence more leniently than those in the control.
2. Respondents assigned to diverse groups that discuss the criminal legal system will sentence more leniently than those in the control.
3. Respondents assigned to diverse groups will sentence more leniently than those in other conditions.
4. Respondents assigned to diverse groups that discuss the criminal legal system will sentence more leniently than those in other conditions.

In the sections that follow we present the results of these evaluations as well as other results that provide context for our findings.

Survey Results

Our sample of 1842 pretend judges was relatively young and mostly female. To assess the extent to which our sample compares to would-be judges, Table 1 presents descriptive statistics of our sample compared to lawyers in the U.S. and the general population.⁸ The individuals in our sample are younger than lawyers and the general population. Women are also better represented in our sample.

First, we present plots of the mean responses to the appropriateness of a prison sentence question—our main outcome variable—across the control, grouped diverse, and grouped homogenous conditions in Figure 3. We separate these results by defendant race and felony class: Black defendants with a Class 2 felony, Black defendants with a Class 3 felony, White defendants with a Class 2 felony, and White defendants with a Class 3 felony. https://osf.io/tbwvx/overview?view_only=697453444c354c6a8c5edd1a4cdd4925.

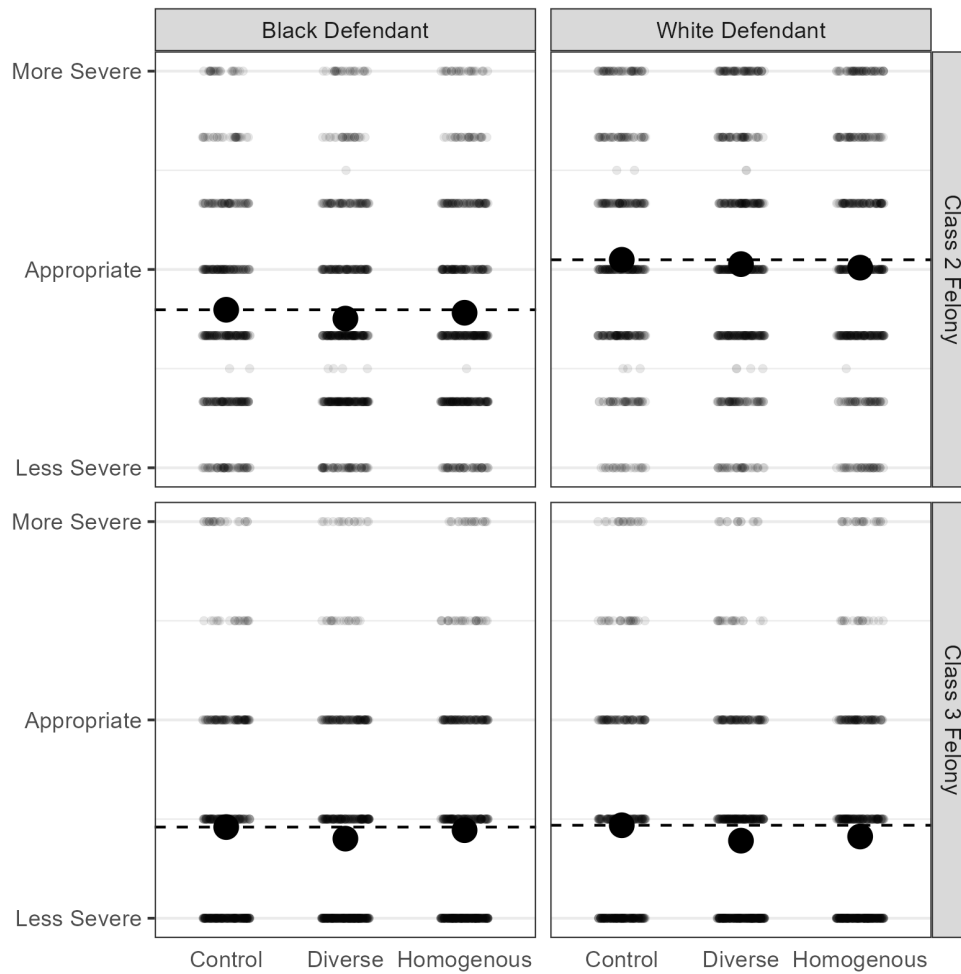
⁸Lawyer demographic information are collected from the American Bar Association Profile of the Legal Profession 2023 and the National Association for Law Placement 2022 Report on Diversity in U.S. Law Firms. Demographic information for the general population are collected from the 2022 U.S. Census estimates for the age and sex composition of the United States.

Table 1: Demographics

	Survey	Lawyers	Gen. Pop.
Age: 18-24	4.9%	0.1%	10.1%
Age: 25-34	31.4%	18.9%	13.6%
Age: 35-44	34.6%	29.6%	13.1%
Age: 45-54	17.4%	20%	12.2%
Age: 55-64	8.5%	17.6%	12.8%
Age: 65+	3.3%	13.7%	17.1%
Female	55.8%	38.7%	50.7%

felony. Large black dots indicate the means, and dashed lines represent the mean in the non-chat control group.

Figure 3: Appropriateness of Prison Sentence by Group Composition



Note: Large black dots indicate the mean of “appropriateness of prison sentence” outcome among all respondents assigned to that treatment group and for profiles corresponding to that facet (e.g., in the top left, the mean for Black defendants with class 2 felonies). Faded dots in the background are individual respondents’ averages. Dashed lines indicate the mean of the control group in that facet.

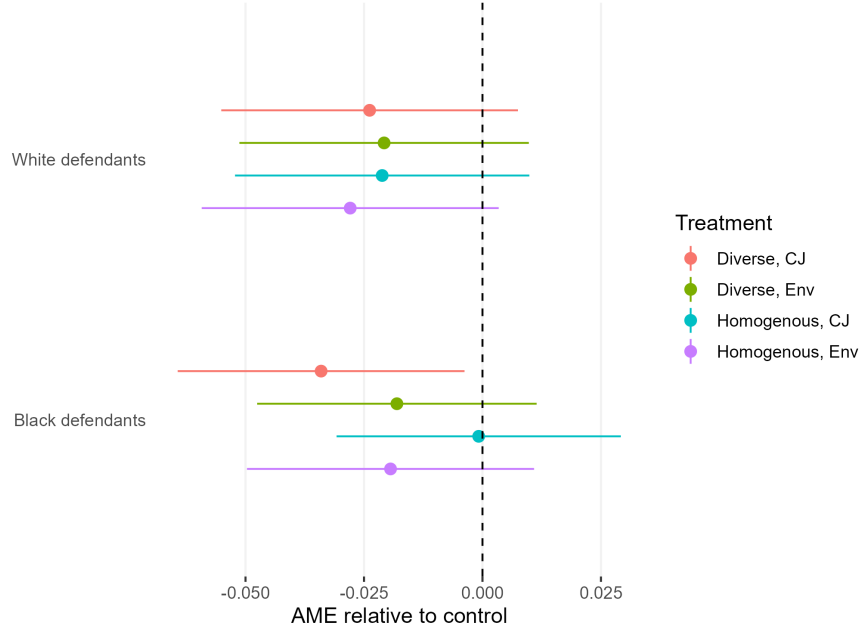
Assignment to a diverse group leads to less comfort with a prison sentence for Black defendant profiles of both felony classes and for White defendant profiles where the conviction is for a Class 3 felony. Respondents in the homogenous conditions were also less comfortable with a prison sentence for White defendants convicted of a Class 3 felony. One peculiar result of this initial exploration of the data is that respondents were more severe toward *White* defendants than to *Black* defendants, especially in the case of Class

2 felonies. Because this is apparent in the control group, we conclude that this peculiarity is not due to any aspect of our chat treatments. Since this finding is incongruous with the real-world behavior of criminal trial court judges, outcomes that calculate the gap between White and Black defendants should be interpreted with caution (Spohn 2013).

Next, we estimate average marginal effects for our four chat treatments compared to the non-chat control on two outcomes: the average appropriateness of prison each for Black defendants and White defendants. Results are shown in Figure 4. The only chat group which significantly decreases punitiveness toward Black defendants is the diverse chat with the criminal justice prompt. This offers support for Hypothesis 2, which expected that respondents assigned to discuss issues related to criminal justice with a diverse group of colleagues would sentence more leniently than the control. We do not observe that being assigned to any diverse group is associated with greater leniency, counter to Hypothesis 1. Notably, none of the chat treatments significantly affect punitiveness toward White defendants compared to control, although all point estimates are negative.⁹

⁹We conduct robustness for the results in Figure 4. In one specification, we adjust for party identification and anti-Black affect (see Figure A.1); in another, we cluster standard errors at the chat level (see Figure A.3). In both cases, the results are substantively similar. We also interact treatment with party identification and anti-Black affect in Table A.2.

Figure 4: Impact of Treatment on Appropriateness of Prison Sentence

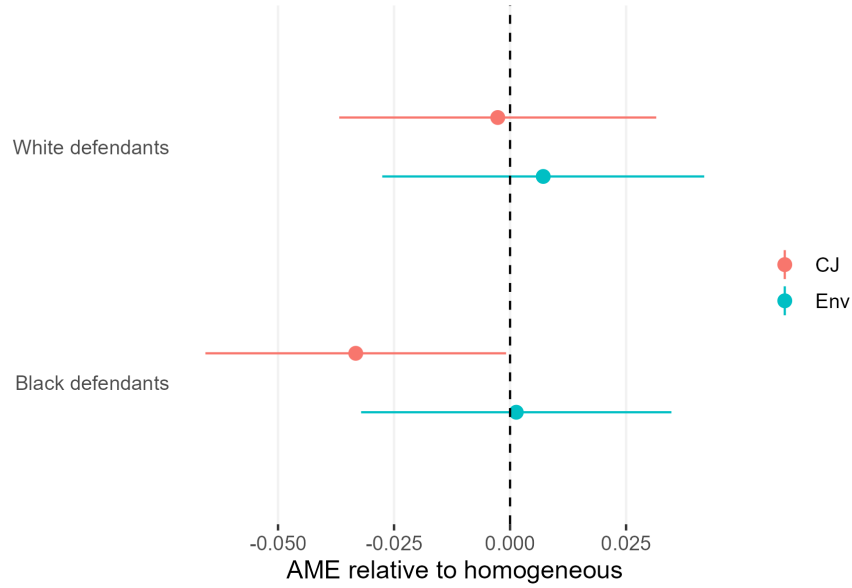


Note: This outcome question is coded numerically as follows: Less Severe = 0, Appropriate = 0.5, More Severe = 1. Lines represent 95% confidence intervals around unadjusted average treatment effects compared to the control condition. The absolute difference is calculated for every individual respondent by taking the absolute value of the difference between the average among Black profiles and the average among white profiles.

Next, we directly test whether the “diverse” chat treatments affect punitive outcomes compared to the “homogeneous” treatments. These analyses isolate the effect of “diverse” group composition compared to “homogeneous” group composition; in other words, these environments most resemble the real-world scenario we wish to replicate. White judges can work in racially homogenous or racially diverse courthouses, and we are interested in how their behavior, when they have a racially diverse group of colleagues, differs from their behavior when they have a racially homogenous group of colleagues. Though, a judge could move from being the only judge in a courthouse to working in a courthouse where other judges also have courtrooms, this is less common. We estimate average marginal effects with an indicator variable for the “diverse” treatment and use the “homogeneous” treatment as the comparison group. We estimate these effects separately for the two chat topics: one related to criminal justice and the other about climate change

(allowing us to test Hypothesis 4). We present these results in Figure 5. For criminal justice chats, diverse group composition decreases punitiveness by 3.3 percentage points, or about 0.15 standard deviations ($p = 0.045$). This supports Hypothesis 4, which anticipated that the impact of diversity (relative to homogeneity) on sentencing outcomes would be especially pronounced when respondents chatted about criminal justice. We do not see this same difference for the environment topic chat groups ($p = 0.937$). This runs counter to Hypothesis 3, which anticipated that diversity would increase leniency across the board. We are not well-powered enough to test the interaction effect between chat topic and group composition, but this evidence is suggestive that the effect of diverse group composition on punitiveness is conditioned by topic-specific deliberation.¹⁰

Figure 5: Impact of Diverse Group Composition on Appropriateness of Prison Sentence



Note: This outcome question is coded numerically as follows: Less Severe = 0, Appropriate = 0.5, More Severe = 1. Lines represent 95% confidence intervals around unadjusted average treatment effects compared to the control condition. The absolute difference is calculated for every individual respondent by taking the absolute value of the difference between the average among Black profiles and the average among white profiles.

¹⁰We conduct robustness for the results in Figure 5). In one specification, we adjust for party identification and anti-Black affect (see Figure A.2); in another, we cluster standard errors at the chat level (see Figure A.4. In both cases, the results are substantively similar.

Text Analysis

Our chat platforms saved and captured the chats respondents had with each other and with confederates, providing a rich source of text data that we use to further understand the results we examined above. To clean the text data, we removed all system-generated messages, messages from chat administrators, and private messages. We also classify messages as *respondent*-generated or *confederate*-generated messages.¹¹ Descriptive statistics of our chat transcripts are included in Table 2. Confederates generally represent about three of every five users in chat rooms, but approximately 50% of messages in the average chat were from actual respondents. More messages were sent in the chats focused on criminal justice than on climate change, and respondents generally sent longer messages than confederates.

Table 2: Chat Descriptive Statistics

	Users per Chat	Messages per Chat		Words per Message	
	% Confederates	#	% Confederates	Respondents	Confederates
Diverse, CJ	60.0	41.4	49.7	16.4	10.7
Homogeneous, CJ	61.5	38.4	50.2	15.1	12.0
Diverse, Env	61.2	34.2	50.4	17.4	13.3
Homogeneous, Env	58.2	28.9	50.3	18.0	13.1

One immediate question is whether respondents and confederates were truly talking about the topics randomly assigned to the chat rooms. To investigate this, we estimated a structural topic model (STM) of chat transcripts (Grimmer, Roberts and Stew-

¹¹We use three methods to classify messages as confederate-generated: first, by using our own records of confederates' usernames (available after February 17); second, by classifying usernames which appeared in more than one chat during a session as a confederate (for the period prior to February 17); and third, by comparing the automatically generated User ID to the confederates' manipulated usernames (respondents' user IDs and usernames were identical, while confederates were not). The first two approaches create one indicator, while the last approach creates a second. These classifications match at a rate of 98.8%.

art 2022).¹² The frequency of criminal justice and environmental topics by treatment condition are shown in Table 3. When we regress the frequency of these topics on the conditions, we find that respondents are more likely to talk about climate change in the chat rooms assigned to speak about the environment and that respondents are more likely to talk about criminal justice and drugs in the chat rooms assigned to discuss criminal justice ($ps < 0.001$). We also note that the diverse groups assigned to talk about the environment were more likely to talk about the environment than comparable homogenous groups ($p = 0.041$), but diverse groups assigned to talk about criminal justice were less likely to talk about criminal justice than comparable homogenous groups ($p = 0.002$). Broadly, these results suggest that respondents and confederates were generally talking about the prompted topics.

Table 3: Topic Frequency by Treatment Condition

Chat Condition	Criminal Justice Topics	Environment Topics
Diverse, Criminal Justice	0.676	0.005
Homogeneous, Criminal Justice	0.767	0.005
Diverse, Environment	0.018	0.431
Homogeneous, Environment	0.016	0.378

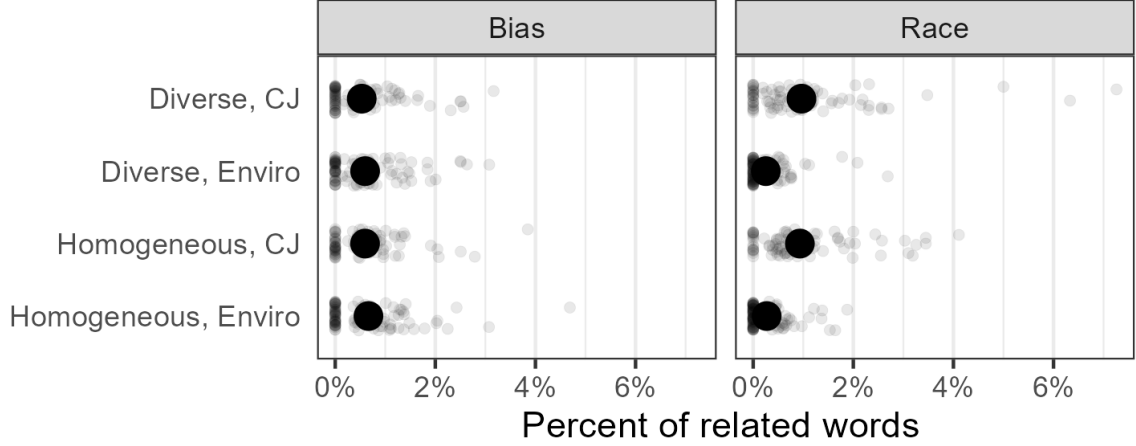
Our theory implies that individuals in diverse settings will be more likely to talk about race, bias, and inequality during interactions. We were therefore interested in the frequency of discussions about race and bias across treatment conditions. To evaluate this, we constructed a dictionary of words related to race and bias.¹³ In total, 2.2% of 196,993 words were coded as relating to bias, while 0.5% were classified as related to race. We plot the distribution of these chats, divided by treatment condition, in Figure 6, with large black dots denoting the average percentage of words related to race/racism or racial bias sent by respondents assigned to that condition. There is significantly

¹²More details about the STM analysis can be found in Appendix B.2. Estimated topic frequency and the top words associated with each topic can be found in Figure B.2.

¹³The full dictionary and details about how we constructed it can be found in Appendix B.3.

more discussion of race in chats focused on criminal justice than those focused on the environment ($ps < 0.001$), and there is no statistically significant difference based on perceived group composition ($ps > 0.7$).

Figure 6: Dictionary Word Frequency, by Treatment Condition



Because our confederates were aware of the study’s hypotheses and the treatment condition of their assigned chat room, we analyzed whether the confederates’ behavior may have been different based on chat condition. First, we examined whether confederates talked about race or racial bias more than respondents in particular conditions. While confederates do talk about race and racial bias more in criminal justice conditions than respondents ($p = 0.006$), there were no statistically significant differences between chats that differed on perceived group diversity ($ps > 0.19$). However, even if confederates were not necessarily sending messages about race or racial bias more often than respondents, it is possible they swayed results by *starting* discussions about race or racial bias. We find that, among chats with at least some discussion of race, confederates began conversations about race (e.g., was the first user in the chat room to use a word coded as race-related) in 278 of 443 chats, or about 63% of the time, across all conditions. This is roughly similar to the average proportion of users in chats who are confederates (60.2%). We find that confederates were more likely to start race-related conversations in chats focused on criminal justice topics ($p < 0.001$), but they were not more likely to start conversations related to race in diverse chat conditions ($p = 0.854$).

Finally, it is plausible that positive experiences in some chats might have motivated our white respondents to be less punitive toward Black defendants. Affective feelings, such as anger or happiness, could affect respondents' judgments, even if such feelings are not directed at a particular group or person (Greifeneder, Bless and Pham 2011). We test this alternative explanation by using a sentiment analysis to approximate the positivity of respondents' "experience."¹⁴ While we find that respondents were more positive in diverse chat rooms when talking about the environment, this diversity effect does not extend to chat rooms focused on criminal justice.

In summary, the preceding text analyses have probed several potential mechanisms that could explain the lower punitiveness of respondents assigned to the diverse chat room focused on criminal justice. First, we tested that our manipulation (e.g., the assignment of topic prompts) worked as expected; according to Table 3, they did. Second, we tested whether our main survey experimental effect was driven by respondents talking more about race or bias in diverse conditions; they did not. Lastly, we tested whether our confederates, who as part of their training were informed of the study's hypotheses and were not blind to treatment assignment, behaved differently in chats where they used Black profile photos. While confederates sent more messages and started more discussions in chat rooms focused on criminal justice, their behavior did not differ based on the group diversity treatment. Based on these results, we argue that our chats focused on criminal justice topics raised the salience of criminal justice as a topic for respondents and often evoked race — but this increased salience translated to less punitiveness toward Black defendants only when respondents were in diverse chat rooms.

Conclusion

This article examines how engaging with a racially diverse group can influence racial disparities in decision making for members of a dominant racial group. Drawing from recent research about racial diversity among judges and the broader prejudice reduction literature, we focus on criminal sentencing decisions with a novel, interactive survey

¹⁴Full details and analysis can be found in Appendix B.4.

experiment. We randomized individuals into either a diverse conversation group or a homogenous one, and tasked them with chatting with one another in response to a prompt prior to making sentencing decisions about a series of criminal profiles. We further wanted to know whether membership in a racially diverse group was sufficient to yield more similar outcomes for profiles featuring Black or White defendants, or whether talking about crime and punishment promoted these outcomes. Thus, we also varied the prompts to which people responded during their chats. Individuals were therefore randomized into one of five groups: a diverse group tasked with chatting about decriminalizing marijuana (the criminal justice related chat), a homogenous group tasked with chatting about the same topic, a diverse group or homogenous group each tasked with chatting about the environment, or a control group that did not complete any chat.

Chat participants who participated in the diverse criminal justice chat condition were more lenient towards Black defendants than were those who did not complete any chat – and this effect did not hold in any other chat condition. We also observe that individuals in the diverse criminal justice chat condition reduced the sentencing of Black defendants relative to those who talked about the same topic but with a homogenous group of colleagues. We also evaluated the content of the chats. We find that even individuals in the homogenous condition discussed topics related to race, although the prompt did not include race-specific language, and these discussions did not appear to influence post-chat decision making. These findings suggest that even short intergroup interactions featuring relevant discussions can lead members of dominant racial groups to make decisions that are less racially disparate.

This project makes several contributions. First, we offer causal evidence that group-level diversity (among our hypothetical judges) itself can change individual behavior (here, sentencing decisions) and decrease racial disparities (in sentences between White and Black defendants) through relevant discussion that occurs during intergroup interactions. Previous studies of the courts and police forces have relied on administrative records to observe the linkage between criminal legal officials’ racial identities and racial disparities in civilian outcomes. We consider two potential mechanisms underlying those associations and show that descriptive racial representation does not simply change

outcomes because members of racial minority groups behave differently than members of racial minority groups. Instead, diversity itself can change the behavior of all group members.

While we tasked our respondents with pretending to be judges and working through sentencing decisions, these findings have implications for other organizations, and are not limited to the judiciary. We can imagine the type of intergroup contact we simulate here yielding similar results among colleagues whose work arrangements, like judges', resemble the conditions Allport outlined in his theory: equal status, common goals, and institutional supports. With respect to government agencies, these likely include those where a significant number of employees do similar work, share a similar status within the organization, and interact with one another, including assistant prosecutors, teachers, and refugee officers, for example. Taken together, our findings suggest that racial diversity itself can generate more similar outcomes for civilians across different racial groups in their interactions with the state.

This project raises a number of additional questions. We focused here on interactions between and differences in outcomes for Black and White individuals. But what about inequity faced by other racial and ethnic groups in the US? Could diversity on the bench improve outcomes for Hispanic defendants as this group is now experiencing the fastest growth in incarceration in the U.S.? Scholars have likewise examined how gender diversity on the bench can impact outcomes; do we observe similar dynamics with respect to the treatment of women by criminal legal actors? Finally, this experiment was completed among individuals recruited through an online platform. Would we observe similar dynamics among public officials themselves? Finally, as we stated, we think the findings should extend to organizations beyond the courts, but this remains an outstanding question ripe for further exploration by scholars. We hope the methodological tools we have developed here can be used to examine this new line of inquiry, particularly in a political moment when the value of diverse organizations is being actively contested.

Ethical Standards

All authors declare no conflicts of interest or ethical issues in this research. The human subjects research was reviewed and approved by [redacted] IRB.

References

- 10 Reasons Why DEI Is Important for Businesses and How it's Evolving*. 2024. *Forbes* . <https://www.forbes.com/councils/forbescoachescouncil/2024/05/06/10-reasons-why-dei-is-important-for-businesses-and-how-its-evolving/>.
- Abrams, David, Marianne Bertrand and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race?" *Journal of Legal Studies* 41(2):347–383.
- Alimo, Craig John. 2012. "From dialogue to action: The impact of cross-race intergroup dialogue on the development of white college students as racial allies." *Equity & Excellence in Education* 45(1):36–59.
- Allport, Gordon W. 1955. *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley Pub. Co.
- Ba, Bocar A., Dean Knox, Jonathan Mummolo and Roman Rivera. 2021. "The role of officer race and gender in police-civilian interactions in Chicago." *Science* 371(6530):696–702. American Association for the Advancement of Science.
URL: <https://www.science.org/doi/full/10.1126/science.abd8694>
- Bestvater, Samuel E and Burt L Monroe. 2023. "Sentiment is not stance: Target-aware opinion classification for political text analysis." *Political Analysis* 31(2):235–256.
- Butcher, Jonathan. 2023. "DEI Doesn't Work – Taxpayers Shouldn't Pay For It." *The Heritage Foundation* . <https://www.heritage.org/education/commentary/dei-doesnt-work-taxpayers-shouldnt-pay-it>.
- Cameron, A Colin and Douglas L Miller. 2015. "A practitioner's guide to cluster-robust inference." *Journal of human resources* 50(2):317–372.
- Cohen, Alma and Crystal S Yang. 2019. "Judicial Politics and Sentencing Decisions." *American Economic Journal: Economic Policy* 11(1):160–91.
- Dessel, Adrienne B. 2010. "Effects of intergroup dialogue: Public school teachers and sexual orientation prejudice." *Small Group Research* 41(5):556–592.

- Dottolo, Andrea L and Abigail J Stewart. 2013. ““I never think about my race”: Psychological features of White racial identities.” *Qualitative Research in Psychology* 10(1):102–117.
- d’Urso, Amanda Sahar, Tabitha Bonilla and Genni Bogdanowicz. 2025. “Where to place sensitive questions? Experiments on survey response order and measures of discriminatory attitudes.” *Political Science Research and Methods* pp. 1–22.
- Enos, Ryan D. 2014. “Causal Effect of Intergroup Contact on Exclusionary Attitudes.” *Proceedings of the National Academy of Sciences* 111(10):3699–3704.
- Greifeneder, Rainer, Herbert Bless and Michel Tuan Pham. 2011. “When do people rely on affective and cognitive feelings in judgment? A review.” *Personality and Social Psychology Review* 15(2):107–141.
- Grimmer, Justin, Margaret E Roberts and Brandon M Stewart. 2022. *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.
- Gurin, Patricia, Biren Ratnesh A Nagda and Ximena Zúñiga. 2013. *Dialogue across difference: Practice, theory, and research on intergroup dialogue*. Russell Sage Foundation.
- Harris, Allison P. 2024. “Can racial diversity among judges affect sentencing outcomes?” *American Political Science Review* 118(2):940–955.
- Hutto, Clayton and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*. Vol. 8 pp. 216–225.
- Kalla, Joshua L and David E Broockman. 2020. “Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments.” *American Political Science Review* 114(2):410–425.
- Klar, Samara, Thomas Leeper and Joshua Robison. 2020. “Studying identities with

- experiments: Weighing the risk of posttreatment bias against priming effects.” *Journal of Experimental Political Science* 7(1):56–60.
- Legewie, Joscha and Jeffrey Fagan. 2016. “Group Threat, Police Officer Diversity and the Deadly Use of Police Force.” p. 64. Columbia Public Law Research Paper No. 14-512.
- Majumdar, Rajeshwari. 2026. “Reducing prejudice and support for religious nationalism through conversations on Whatsapp.” *Comparative Political Studies* p. 00104140261418627.
- Montgomery, Jacob M, Brendan Nyhan and Michelle Torres. 2018. “How conditioning on posttreatment variables can ruin your experiment and what to do about it.” *American Journal of Political Science* 62(3):760–775.
- Nathan, Noah L and Melissa L Sands. 2023. “Context and contact: Unifying the study of environmental effects on politics.” *Annual Review of Political Science* 26(1):233–252.
- NCSC Annual Report 2023: Inspiring Future-Ready Courts.* 2023. *National Conference of State Courts* . <https://www.ncsc.org/sites/default/files/media/document/ncsc-annual-report-2023.pdf>.
- Paluck, Elizabeth Levy and Donald P Green. 2009. “Prejudice reduction: What works? A review and assessment of research and practice.” *Annual review of psychology* 60(1):339–367.
- Paluck, Elizabeth Levy, Roni Porat, Chelsey S Clark and Donald P Green. 2021. “Prejudice reduction: Progress and challenges.” *Annual review of psychology* 72(1):533–560.
- Paluck, Elizabeth Levy, Seth A Green and Donald P Green. 2019. “The contact hypothesis re-evaluated.” *Behavioural Public Policy* 3(2):129–158.
- Pettigrew, Thomas F. 1998. “Intergroup contact theory.” *Annual review of psychology* 49(1):65–85.
- Pettigrew, Thomas F and Linda R Tropp. 2006. “A meta-analytic test of intergroup contact theory.” *Journal of personality and social psychology* 90(5):751.

- Schanzenbach, Max. 2005. "Racial and Sex Disparities in Prison Sentences: The Effect of District-level Judicial Demographics." *The Journal of Legal Studies* 34(1):57–92.
- Scherer, Nancy. 2004. "Blacks on the Bench." *Political Science Quarterly* 119(4):655–675.
- Selvanathan, Hema Preya and et al. 2018. "Whites for Racial Justice: How Contact with Black Americans Predicts Support for Collective Action among White Americans." *Group Processes & Intergroup Relations* 21(6):893–912.
- Sharp, Elaine B. 2014. "Minority Representation and Order Maintenance Policing: Toward a Contingent View." *Social Science Quarterly* 95(4):1155–1171.
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ssqu.12078>
- Shoub, Kelsey and Leah Christiani. 2023. "Context matters: The conditional effect of black police chiefs on policing outcomes." *Urban Affairs Review* 59(6):2043–2056.
- Sommers, Samuel R. 2006. "On racial diversity and group decision making: identifying multiple effects of racial composition on jury deliberations." *Journal of personality and social psychology* 90(4):597.
- Spohn, Cassia. 1990. "The Sentencing Decisions of Black and White Judges: Expected and Unexpected Similarities." *Law and Society Review* 24(5):1197–1216.
- Spohn, Cassia. 2013. "Racial disparities in prosecution, sentencing, and punishment." *The Oxford handbook of ethnicity, crime, and immigration* pp. 166–193.
- Steffensmeier, Darrell and Chester L Britt. 2001. "Judges' Race and Judicial Decision Making: Do Black Judges Sentence Differently?" *Social Science Quarterly* 82(4):749–764.
- Welch, Susan, Michael Combs and John Gruhl. 1988. "Do Black Judges Make a Difference?" *American Journal of Political Science* 32(1):126–136.
- Wu, Stephen. 2021. "Leadership matters: Police chief race and fatal shootings by police officers." *Social Science Quarterly* 102(1):407–419.

Appendix

A Alternative Specifications for Survey Results

In this section, we present average treatment effects for our survey results estimated using alternative specifications. In Figures A.1 and A.2, we replicate Figures 4 and 5 using specifications that control for party identification, measured using the standard branching question and seven-point scale; anti-Black affect, measured using a feeling thermometer; and demographic characteristics, including gender, income, education, and age. As shown in Table A.1, our treatment groups are slightly unbalanced on gender, education, income, and anti-Black affect. While some researchers urge that using covariates measured post-treatment can induce bias (see Montgomery, Nyhan and Torres 2018), others argue that survey questions that could prime respondents, such as demographics or prejudice measures, should be placed after treatment, as we did (see d’Urso, Bonilla and Bogdanowicz 2025; Klar, Leeper and Robison 2020). As seen in the below figures, results are substantively similar to those presented in the main manuscript: the diverse, criminal justice reduces punitiveness among Black defendants compared to the control condition and the homogeneous criminal justice condition.

Table A.1: Balance Table

	Criminal Justice			Environment		
	Control	Homogenous	Diverse	Homogenous	Diverse	P-value
Female	0.52	0.60	0.61	0.56	0.53	0.03
Party ID	3.50	3.33	3.55	3.62	3.40	0.47
Education	4.62	4.71	4.67	4.55	4.51	0.04
Age	40.06	40.89	39.26	42.96	39.62	0.34
Income	5.88	6.04	6.10	6.09	5.34	0.02
Anti-Black	0.03	0.12	-0.23	-0.08	-0.00	0.02
N	551	323	315	312	341	-

In Figures A.3 and A.4, we replicate Figures 4 and 5 using specifications with

Figure A.1: Impact of Treatment on Appropriateness of Prison Sentence (with controls)

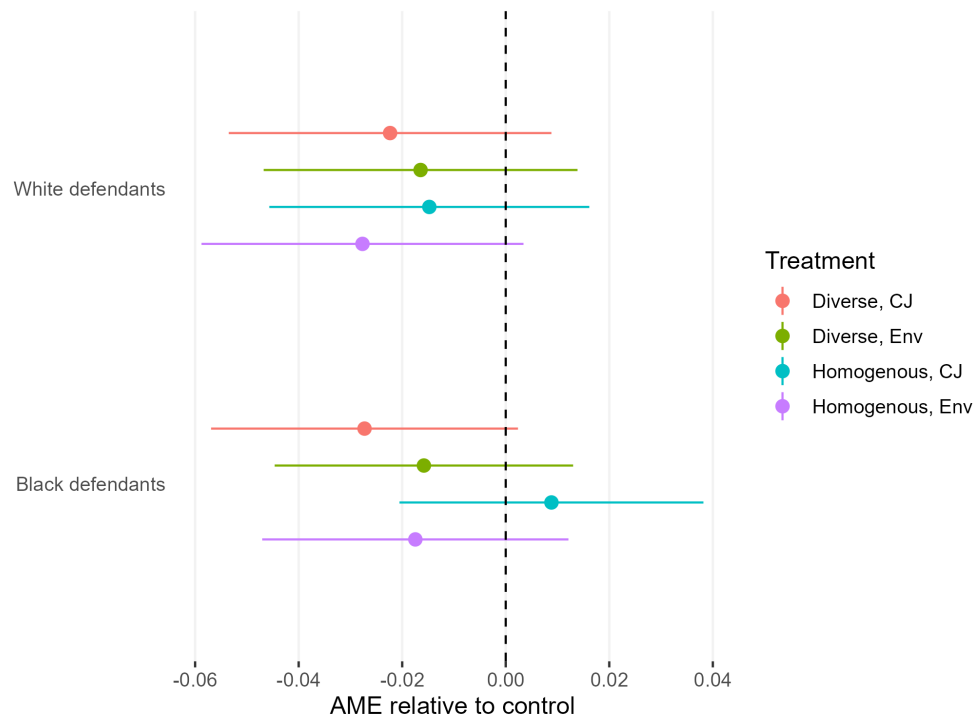
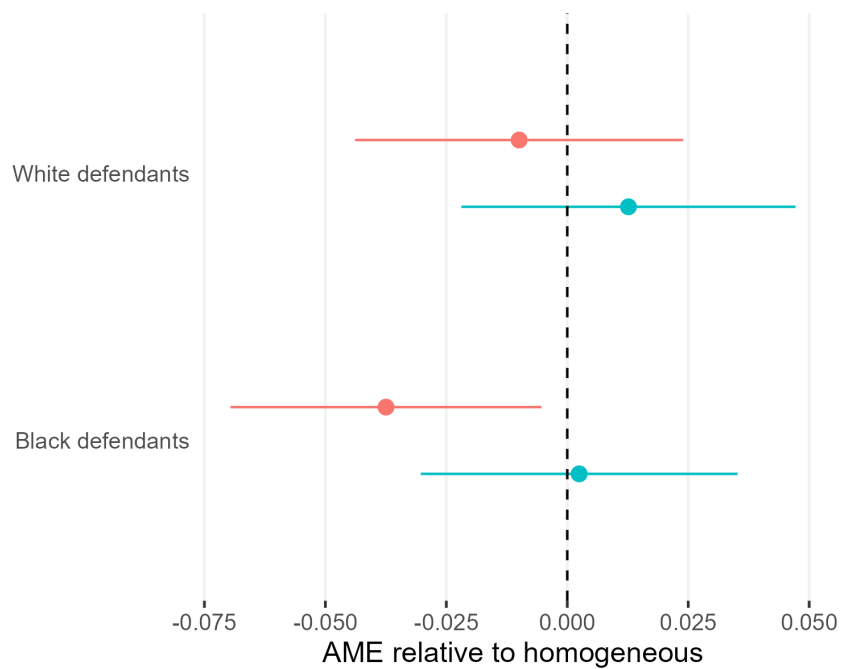
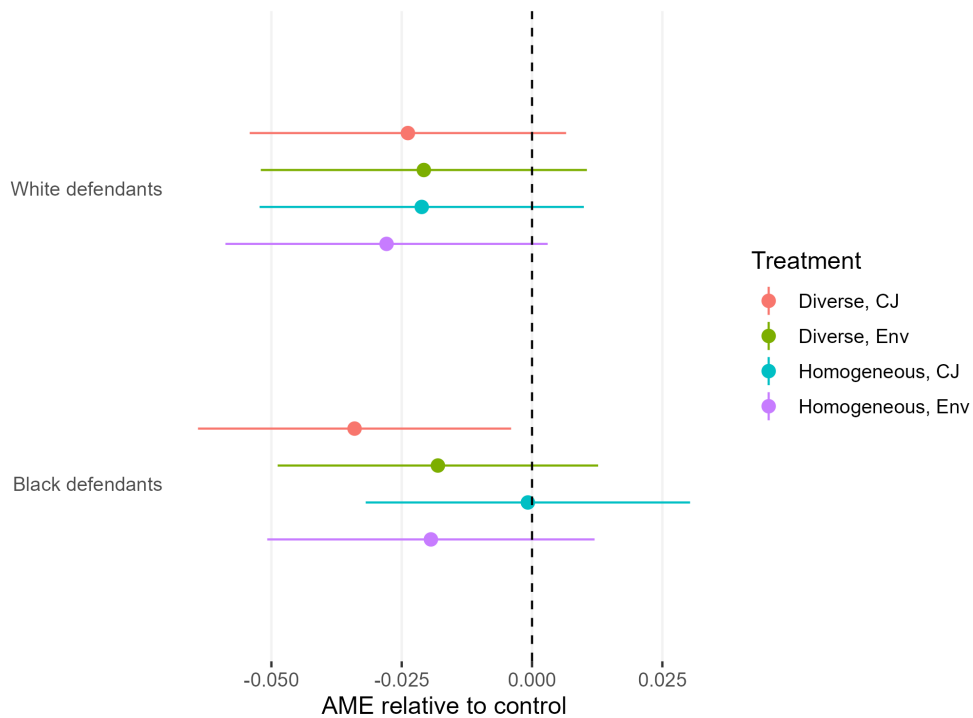


Figure A.2: Impact of Diverse Group Composition on Appropriateness of Prison Sentence (with controls)



cluster-robust standard errors. Specifically, we use CR0 standard errors with clustering at the chat level. We motivate this analysis by noting that although respondents are individually randomly assigned to treatment arms through the Qualtrics survey, treatment is administered at the group level. Therefore, clustering standard errors at the chat level is appropriate (Cameron and Miller 2015). Results are substantively similar to those presented in the manuscript and in the covariate-adjusted analyses above.

Figure A.3: Impact of Treatment on Appropriateness of Prison Sentence (with clustered standard errors)



Lastly, we interacted our treatments with party identification and anti-Black affect. These interactions were pre-registered. As shown in Tables A.2 and A.3, few interaction effects are statistically significant. We also analyzed for treatment effect heterogeneity on other collected covariates (gender, education, income, age) and find no statistically significant treatment effect heterogeneity.

Figure A.4: Impact of Diverse Group Composition on Appropriateness of Prison Sentence
(with clustered standard errors)

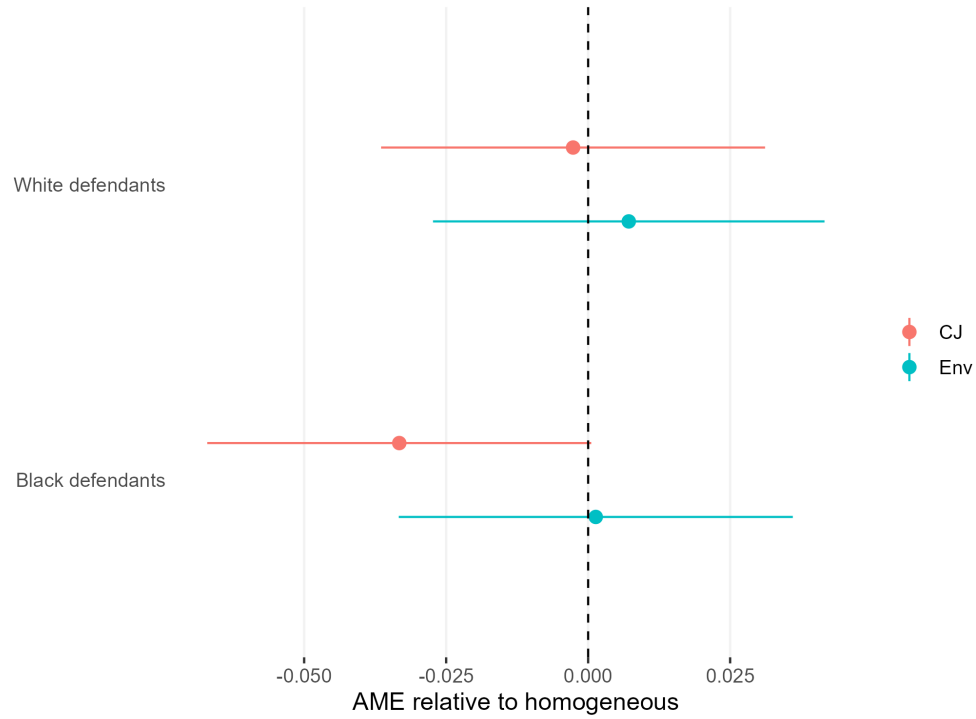


Table A.2: Coefficients for Treatments Interacted with Covariates

Interacted with ...	DV: Black defendants		DV: White defendants	
	<i>Party ID</i>	<i>Anti-Black Affect</i>	<i>Party ID</i>	<i>Anti-Black Affect</i>
Diverse, CJ	0.002	-0.023	0.01	-0.003
Homogenous, CJ	0.007	-0.003	0.011	-0.005
Diverse, Env	0.003	-0.017	0.008	-0.007
Homogenous, Env	0.015*	0.003	0.016*	0.011

Notes: * denotes $p < 0.05$. All regressions use the non-chat control as the omitted category. Each column represents a single regression model.

Table A.3: Coefficients for Diversity Interacted with Covariates

	Topic	DV: Black defendants		DV: White defendants	
		<i>Party ID</i>	<i>Anti-Black Affect</i>	<i>Party ID</i>	<i>Anti-Black Affect</i>
Interacted with ...					
Diverse	CJ	-0.005	0.02	-0.001	-0.002
Diverse	Env	-0.012	0.02	-0.008	0.018

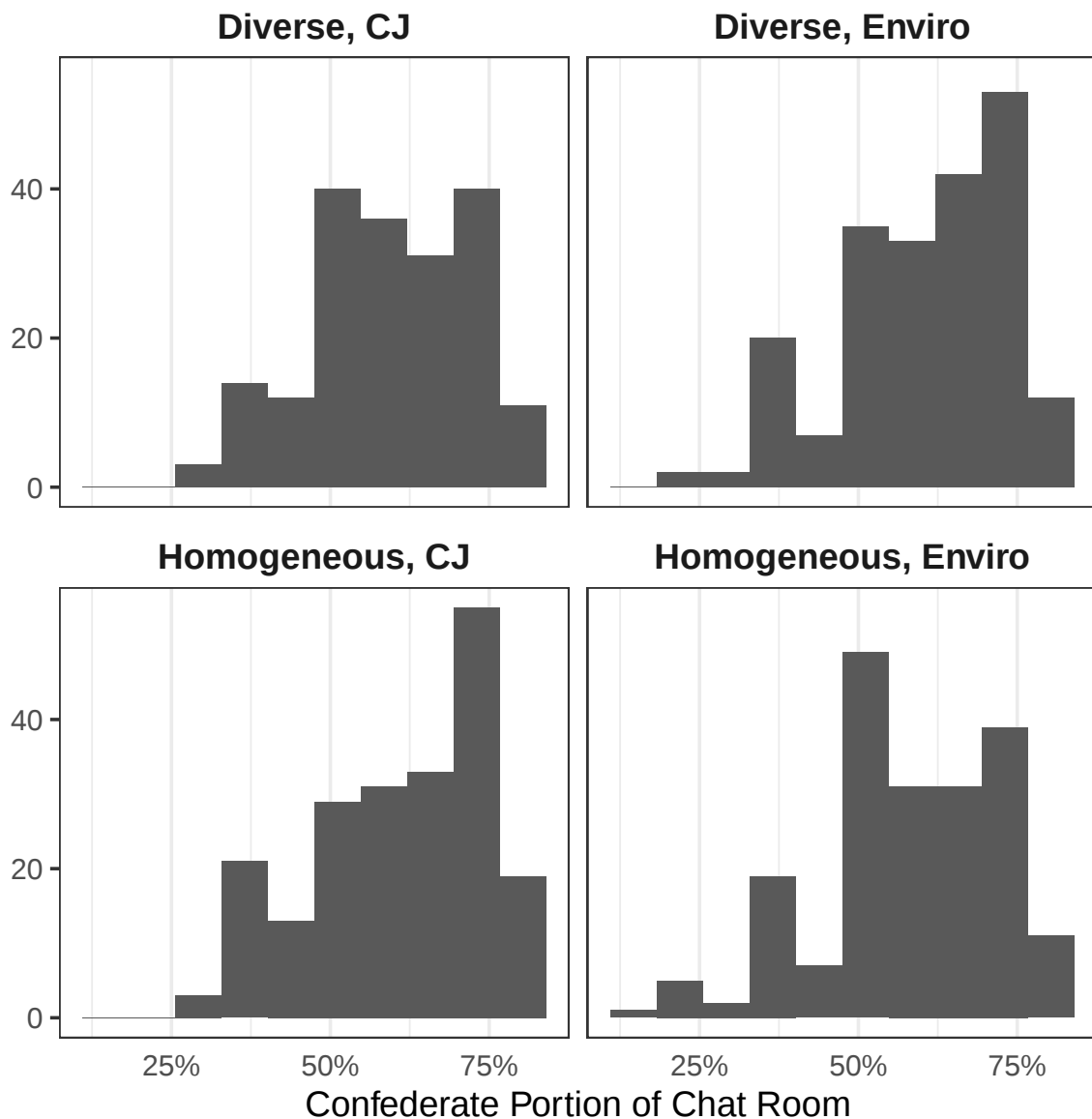
*Notes: * denotes $p < 0.05$. All regressions use the homogenous chat condition for the respective topic as the omitted category.*

B Additional Chat Analyses

B.1 Chat Diversity

To expand upon our descriptive analysis of chat room composition started in Table 2, we show the percentage of respondents for each chat room, broken out by chat condition, in Figure B.1.

Figure B.1: Chat Room Composition, By Treatment Condition



B.2 Structural Topic Model (STM)

In order to analyze the topics under discussion in our chats, we used a Structural Topic Model (STM). Using the *stm* package in R (Grimmer, Roberts and Stewart 2022), we treated each chat room transcript as an individual document for the purposes of training the model. This allows us to holistically understand the topics discussed by both respondents and confederates in chat rooms. We removed messages containing researcher-provided prompts, removed stopwords (e.g., and, but, the) as well as words occurring fewer than ten times. We set the parameter of K, the number of topics to estimate, at 10. The estimated topics can be found in Figure B.2, along with their estimated frequency and the top three words associated with each topic. In the manuscript, we combine topics 1, 2, 3, 7, and 10 as “criminal justice topics.” Only Topic 9 has words strongly associated with the environment, so we define that as the “environmental topic.”

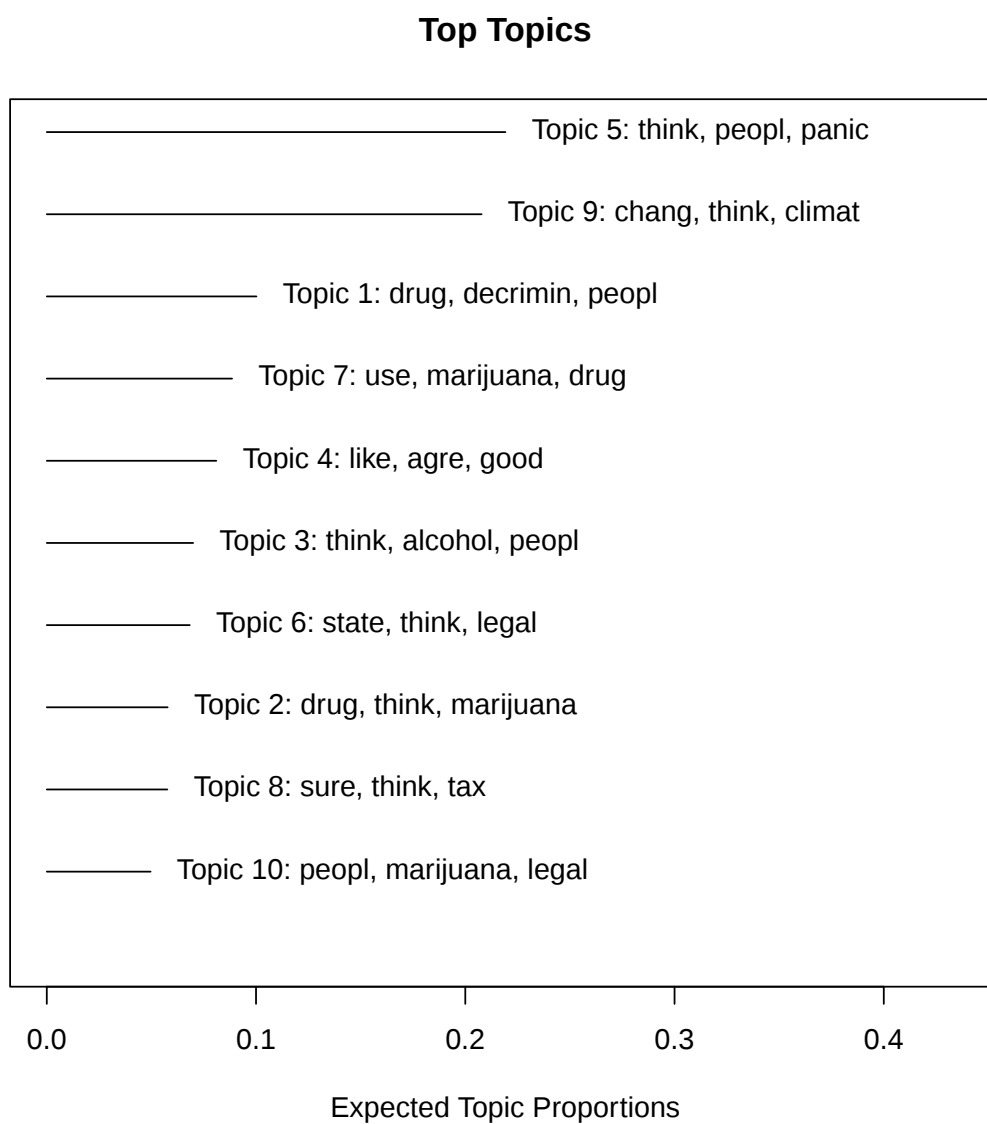
B.3 Dictionary-Based Analysis

In order to investigate whether respondents used words related to race or bias more frequently in certain conditions, we conducted a dictionary-based analysis. We constructed dictionaries of words that denote race or bias using a two-step process. In Step 1, we drew a random sample of twelve chat transcripts, stratified by treatment condition, and asked two research assistants who had no previous involvement with the project to identify words and phrases associated with our topics of interest. In Step 2, we combined the manually coded words with an additional list of words used to denote race and bias. See below for complete dictionaries; words in italics were added in Step 2.

Race: systematic racism, racism, skin color, minorities, singled out, systematically racist, *black, white, asian, latino, hispanic, latina, latinx, people of color, race, racial, racism, african american, colorblind, systemic racism, minority, bipoc, communities of color, inner city, segregation, apartheid, prejudice, bigotry, intolerance, tolerance, slavery, Jim Crow, diversity*

Bias: singled out, unbiased, systematically racist, disparate, targeted, unfairly, equitable, disenfranchised, elites, power, capitalism, capitalistic, low-level, heavy, profits, *dis-*

Figure B.2: Structural Topic Model (STM) Topics



Note: Results from a structural topic model using chat room transcripts as documents and the parameter of ten topics.

crimination, inequality, equality, equity, deserve, deserving, fair, justice, unjust, unfair, equal, disenfranchised, bias, biased, stratify, status, second class

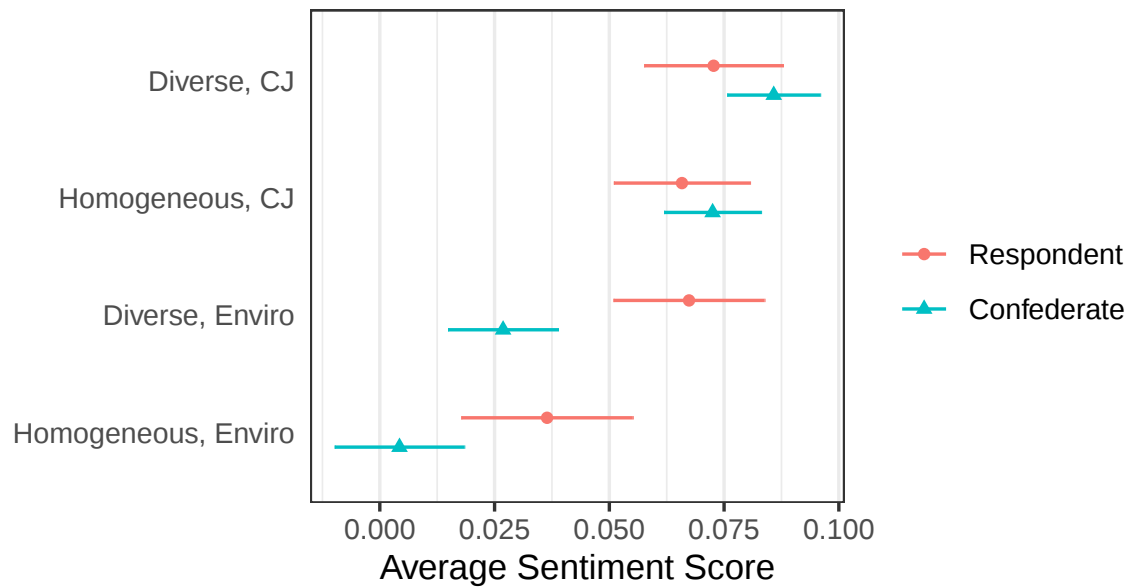
After we constructed this dictionary, we stemmed every word in our chat transcripts and coded whether each stem was included in our dictionary. Because “people” is one of the most common words in our transcripts, and because it is included in the *race* dictionary only incidentally (by way of “people of color”), we exclude the word “people” from our main dictionary-based analyses, including Figure 6 and the p-values cited in the main manuscript.

B.4 Sentiment Analysis

In order to conduct our sentiment analysis, we used the Valence Aware Dictionary and sEntiment Reasoner (VADER) and its pre-existing dictionary to classify the sentiment of our messages and chats (Hutto and Gilbert 2014). Because our chat rooms were online, we believe that VADER, designed for analysis of social media messages, is an appropriate tool. VADER produces three scores of interest: a score of “positive” sentiment, a score of “negative” sentiment, and a *composite* score intended to reflect the overall sentiment of a document. We emphasize that we interpret these scores not as indicative of respondents’ stances toward, for example, those who are convicted of drug possession, but rather as simply indicative of the sentiment of the chat (Bestvater and Monroe 2023). Sentiment analysis is helpful in determining whether respondents’ contact with perceived out-groups in our “diverse” chat conditions were more negative than “homogeneous” chat conditions.

We plot the average composite sentiment score, separated by treatment condition, for messages sent by respondents and confederates in Figure B.3. We highlight two differences: first, while confederates and respondents send messages of similar sentiment in criminal justice conditions, confederates are much more negative in environmental conditions; and second, the diverse condition is more positive than the homogenous condition when talking about the environment. We further examine these patterns with formal statistical analysis in Table B.1. While we present statistical analysis of sentiment at the message-level, we also conducted analysis at the chat level. The results are substantively similar.

Figure B.3: Average Composite Sentiment Score of Messages, by Treatment Condition



Most importantly for our analyses, confederates and respondents sent messages of similar sentiment in criminal justice conditions *across diverse and homogenous chat rooms*, as shown in the statistically insignificant coefficients for the interaction term for perceived group diversity and criminal justice in Table B.1. While we omit a three-way interaction between confederate, criminal justice topic, and perceived group diversity in these regressions for clarity, the results are substantively similar if we include the term.

Table B.1: Sentiment Analysis of Chat Messages

	<i>Composite Score</i>	<i>Positive Score</i>	<i>Negative Score</i>
	(1)	(2)	(3)
Diverse	0.027** (0.013)	−0.0003 (0.007)	−0.011*** (0.004)
Criminal Justice	0.020 (0.012)	−0.013* (0.007)	−0.018*** (0.004)
Word Count	−0.003*** (0.0003)	−0.004*** (0.0002)	0.001*** (0.0001)
Confederate	−0.047*** (0.011)	−0.028*** (0.006)	0.019*** (0.004)
Diverse X Criminal Justice	−0.015 (0.014)	0.010 (0.007)	0.008* (0.005)
Diverse X Confederate	0.047*** (0.011)	0.009 (0.006)	−0.017*** (0.004)
Confederate X Criminal Justice	−0.003 (0.011)	−0.002 (0.006)	−0.001 (0.004)
Constant	0.084*** (0.011)	0.227*** (0.006)	0.084*** (0.004)
Level of Analysis	Message	Message	Message
Observations	28,943	28,943	28,943
R ²	0.010	0.038	0.015

Note:

*p<0.1; **p<0.05; ***p<0.01

C Full Model Results

Table C.1: Full Model Results for Figure 4

	White defendants	Black defendants
	(1)	(2)
Homogenous, CJ	−0.021 (0.016)	−0.001 (0.015)
Diverse, CJ	−0.024 (0.016)	−0.034** (0.015)
Homogenous, Env	−0.028* (0.016)	−0.019 (0.015)
Diverse, Env	−0.021 (0.016)	−0.018 (0.015)
Constant	0.408*** (0.010)	0.333*** (0.009)
Observations	1,807	1,807
R ²	0.002	0.004
Adjusted R ²	0.0002	0.001

Note:

*p<0.1; **p<0.05; ***p<0.01

Table C.2: Full Model Results for Figure 5

	Black defendants		White defendants	
	(1)	(2)	(3)	(4)
Group Diversity	−0.033** (0.017)	0.001 (0.017)	−0.003 (0.017)	0.007 (0.018)
Constant	0.332*** (0.012)	0.314*** (0.012)	0.387*** (0.012)	0.381*** (0.013)
Topic	CJ	Env	CJ	Env
Observations	631	648	632	649
R ²	0.006	0.00001	0.00004	0.0003
Adjusted R ²	0.005	−0.002	−0.002	−0.001

Note:

*p<0.1; **p<0.05; ***p<0.01

Table C.3: P-values for Tables C.1 and C.2

Treatment	White defendants	Black defendants
Table C.1		
Homogenous, CJ	0.182	0.959
Diverse, CJ	0.136	0.028
Homogenous, Env	0.081	0.210
Diverse, Env	0.183	0.230
Table C.2		
Diversity, within CJ	0.879	0.045
Diversity, within Env	0.686	0.937