Do Virtue Signals Signal Virtue?

Deivis Angeli* Matt Lowe* The Village Team†

Abstract

We study whether tweets about racial justice predict the offline behaviors of nearly 20,000 US academics. In an audit study, academics that tweet about racial justice discriminate more in favor of minority students than academics that do not tweet about racial justice. Racial justice tweets are more predictive of race-related political tweets than political contributions, suggesting that visibility increases informativeness. In contrast, the informativeness of tweets is lower during periods of high social pressure to tweet about racial justice. Finally, most graduate students mispredict informativeness, more often underestimating than overestimating, reducing the welfare benefits of social media.

*University of British Columbia.

†The Village Team is a team of over 100 research assistants led by Carla Colina, Jordan Hutchings, Noor Kumar, Ines Moran, Saloni Sharma, Aurellia Sunarja, Chihiro Tanigawa, Akash Uppal, and Kevin Yu (see Appendix A for a full list). The paper was previously entitled "Do Virtue Signals Signal Virtue?" and was approved by UBC's Behavioural Research Ethics Board (H20-03758) and pre-registered in the AEA RCT Registry (AEARCTR-0009359). This research was undertaken, in part, thanks to funding from the Canada Excellence Research Chairs program awarded to Dr Erik Snowberg in Data-Intensive Methods in Economics. We thank Vincent Pons and Edgard Dewitte for helping with the FEC data. We thank Leo Bursztyn, Claudio Ferraz, Patrick Francois, Alex Frankel, Nick Hagerty, Emir Kamenica, George Loewenstein, Rachael Meager, Jamie McCasland, Sam Norris, Nathan Nunn, Chris Roth, Heather Sarsons, Frank Schilbach, Munir Squires, and Carolyn Stein for useful comments. First draft: November 2022. This draft: May 2023.

1 Introduction

Social signalling is a central aspect of human behavior. Humans signal their ability to the job market (Tyler et al. 2000), their ambition to the marriage market (Bursztyn et al. 2017), and their customs and parenting prowess to their neighbors (Bursztyn et al. 2020a; Karing 2023). A key question about social signalling is whether, and when, signals are informative. This question takes on extra importance when people signal online. With the advent of social media platforms, a huge swathe of signalling is unverifiable, takes place at little cost (Zhuravskaya et al. 2020), and is observed by large audiences – precisely the conditions that theory suggests should reduce the informativeness of communication (Frankel and Kartik 2019). Compounding this, individuals may have particularly strong incentives to misrepresent their values when signalling about politically-charged moral virtues, like opposition to racism or xenophobia (Bursztyn et al. 2020b). How informative are these virtue signals about offline behaviors?

If social media users self-persuade (Schwardmann et al. 2022), have preferences for truth-telling (Abeler et al. 2019), or uniformly inflate their statements, virtue signals online will signal related behaviors offline. Instead, if reputational concerns are sufficiently strong, social media posts may be completely uninformative (Morris 2001; Frankel and Kartik 2019). In fact, if social media users engage in moral licensing or conscience accounting (Mazar and Zhong 2010; Gneezy et al. 2014), virtue signallers may even be the least likely to behave in line with their social media-reported values offline. Which case describes real-world equilibria? And do social media users recognize what equilibrium we are in?

In this paper, we study communication on Twitter, a platform used by 23% of Americans (PRC 2021), and 450 million users worldwide (BOA 2022). We measure the informativeness of the racial justice tweets of 18,514 US-based non-Black academics, with 11,450 of these academics included in an audit study. With our focus on informativeness, we assess the *predictive* value of tweets, rather than the *motives* behind tweets. The informativeness of tweets is not necessarily connected to the motives for tweeting: an informative equilibrium may be consistent with all racial justice-tweeters being motivated by social image concerns or all being motivated by prosociality.

We use three steps to characterize the communication equilibrium. First, we answer the question of whether racial justice tweets are informative of related offline behaviors: discriminatory behavior in the audit study and political contributions. Second, we show how tweet informativeness varies with features predicted by theory (Frankel and Kartik 2019), including social pressure, audience size, and reputational concerns. Third, we explore whether graduate students have sophisticated beliefs about the informativeness of tweets.

We pre-registered a binary measure of racial justice signalling: an academic is "Vocal" if they have at least one tweet from January 2020 to March 2022 that mentions at least one racial justice-related word or phrase (e.g. racism, George Floyd). We classify 62% of the 11,450 audited academics as Vocal. Using a manual validation exercise, we find that almost all racial justice tweets are supportive of racial justice-related efforts. This means that our automated measure of vocality can be thought of as a measure of whether an academic has supported racial justice efforts on Twitter.

We ran the audit experiment in May 2022, sending one email to each academic. Each email came from a fictitious student requesting an online meeting to ask questions about graduate studies. We treat meeting acceptance as a costly, near-private (since it is visible to the student), prosocial behavior. We randomly assigned half of the emails to be sent from a student with a distinctively Black name, and half to be sent from a distinctively White name, with 120 names used overall. We cross-randomized whether the name was distinctively female or distinctively male, and whether the email also mentioned that the sender is a first-generation college student.

Our audit experiment required deception, a clear moral cost. To minimize ethical concerns, (i) we cancelled accepted meetings manually, promptly, and politely, (ii) we did not debrief academics on the fictitious nature of the email after the experiment ended, to reduce the possibility that academics become more suspicious of future emails from genuine students, and (iii) we received formal ethics approval from UBC. To the concern of 'poisoning the well,' we note that correspondence studies with US-based professors are rare – in particular, a recent meta-analysis of correspondence studies measuring racial discrimination in the US since 2000 found only one study with professors – Milkman et al. (2012) – carried out over ten years ago (Gaddis et al. 2021). Even with these efforts, the moral case for our audit study rests on the benefits of the study's results outweighing the costs. Here, our view is that the incremental knowledge from our paper is substantial. In particular, we are not aware of any large-scale measures of discrimination across academia since Milkman et al. (2012), or of any measures of first-generation student discrimination. More importantly, we are aware of no evidence of the usefulness of social media for predicting who discriminates. Each of these findings is valuable to students, and particularly to those from underrepresented groups.

We find no evidence of racial discrimination in the full sample of audited academics. Academics accept 30.6% of meeting requests from distinctively White names, and 29.8% from distinctively Black names ($p = 0.31$ for the difference). Tweeting academics discriminate statistically significantly less than large US employers (Kline et al. 2022) and the broader sample of academics audited in 2010 (Milkman et al. 2012).

The lack of racial discrimination in the full audit sample masks heterogeneity: Silent academics are 5.3 percentage points (18%) less likely to accept a meeting with a Black student than with a White student ($p < 0.001$). Vocal academics are 1.9 percentage points (6%) more likely to accept a meeting with a Black student than with a White student ($p = 0.07$). Virtue signals are highly informative – we firmly reject equality of the racial gaps in meeting acceptance between Vocal and Silent academics ($p < 0.001$).

Tweets also signal broader support for students from underrepresented groups. Vocal academics show more favoritism towards first-generation students than Silent academics ($p < 0.001$), and they are also more likely to favor female students over male students ($p = 0.03$). It follows that Vocality is not a signal of unbiasedness – while Vocal academics show less bias in favor of White students than Silent academics, they show *more* bias in favor of first-generation and female students.

As a way of summarizing the treatment of minority-group students, we pool the 7/8 of the audit emails from any minority group (Black, female, or first-generation) and compare them with the 1/8 of the audit emails from White males with no mention of first-generation status. With the caveat that this analysis has lower statistical power, we find that Vocal academics are 27% less likely to accept a meeting with a White male student than with an underrepresented minority student, while we cannot reject the null that Silent academics treat the two groups equally. Going further, while minority students are more likely to get a positive response from Vocal academics than from Silent academics, the reverse is true for White male students. This gives a force for sorting – with White males advised more by Silent academics, and minority students advised more by Vocal academics – reinforcing offline echo chambers.

Beyond the signal of having ever tweeted about racial justice, we also find informativeness along the intensive margin. We split the Vocal academics into the "Rarely Vocal" and the "Regularly Vocal": those with a percentage of racial justice tweets below versus above the median. Rarely Vocal academics are close to racially unbiased – accepting 30.4% of meeting requests from White students, and 31% from Black students ($p = 0.39$ for the difference). In contrast, Regularly Vocal academics favor Black students by 3.2 percentage points ($p = 0.01$).

Our finding that racial justice tweets are unconditionally informative extends to conditional comparisons. In particular, it is not just the case that Vocal academics behave differently to Silent academics, they also behave differently conditional on their gender, race/ethnicity, academic position, university, department, political views, and how much they tweet. Controlling for these factors, Vocal academics discriminate against Black students roughly five percentage points less than Silent academics ($p < 0.01$). Racial justice tweets provide information about discriminatory behaviors over and above these observables, making tweets informative even for students that know a given professor reasonably well. In fact, Vocality is more predictive of racial discrimination than each of these other observable factors.

Detection is a particular concern when auditing academics. We observed some detection directly – in tweets by economists, as well as 16 email responses explicitly mentioning suspicion that the student was fictitious. Given this, we check whether our core results could be driven by detection. Encouragingly, our findings are very similar when we: (i) drop social scientists (who are more likely to be familiar with the audit methodology), (ii) drop the academics to whom we sent more generic emails, (iii) drop university-departments to which we sent more than five emails (where discussion about the emails may have led to more suspicion), and when we (iv) define the outcome as accepting the meeting within one day (leaving less time to find that other academics had received similar emails). In addition, the 16 emails that mentioned suspicion were almost equally divided between distinctively Black and distinctively White names, suggesting that suspicion may be balanced on race.

Given strong partisan differences in emphasis on racial justice, our next set of analyses explore the informativeness of racial justice tweets for political contributions. Building on Bouton et al. (2022), we link three samples of academics with Federal Election Commission (FEC)-reported political contributions made from January 2020 to March 2022: (i) 18,514 non-Black tweeting academics, (ii) 1,094 Black tweeting academics, and (iii) a random sample of 900 non-Black academics without Twitter accounts. We successfully link 29.9% of these academics with at least one contribution.

Our academics are starkly unrepresentative of the political views of Americans. 0.6% of academics contributed at least once to a Republican-related committee, while 29.5% contributed at least once to Democrats. Academic Twitter is even more left-skewed: while 1.1% of non-tweeting non-Black academics contributed to Republicans, only 0.6% of tweeting non-Black academics did. Vocality signals Democratic support: Vocal academics are 81% more likely than Silent academics to have given to the Democrats, and 55% more likely after adding a rich set of controls.

To parallel the audit, we focus on a case study in which two similar candidates differed on race: the 2021 Georgia Senate runoffs with Jon Ossoff (a White Democratic candidate) and Raphael Warnock (a Black Democratic candidate). Both candidates won close elections and have since voted similarly while in office. In contrast with the audit results, Vocal academics are no more likely to give more to Warnock than Ossoff – instead, both Silent and Vocal academics are slightly less likely to have contributed to Warnock's campaign than to Ossoff's. As one explanation for informativeness in the professional but not the political domain, we note that the Warnock-Ossoff giving gap is not predictive of racial discrimination in the audit – suggesting that the two racial gaps are driven by different underlying preferences and beliefs.

Going further, we use the political setting for a suggestive test of a related hypothesis: does the information value of tweets increase where behavior is more visible? For this, we identify which academics *tweeted* about Jon Ossoff and Raphael Warnock. Vocal academics are more likely to have tweeted about Warnock than about Ossoff, while Silent academics tweet minimally about both candidates. It follows that racial justice tweets predict racial gaps in tweeting more than they predict racial gaps in giving.

Having characterized the informativeness of racial justice tweets, we return to the audit-measured behaviors and ask: when are tweets more informative? Guided by theory (Frankel and Kartik 2019), we hypothesize that higher incentives for (or lower costs of) signalling may reduce informativeness. Consistent with this, we find that informativeness is much lower immediately after the murder of George Floyd, when pressure to tweet about racial justice was higher, than in the five months that followed. In fact, once conditioning on other observables, we cannot reject the null that racial justice tweets are uninformative during this high-pressure period. Otherwise, two findings go against the theoretical prediction. First, informativeness is higher for users with above-median Twitter followers than for those with below-median followers – the tweets of famous academics carry the most information. Second, original tweets are less informative than all other types of tweets, despite original tweets being the most costly to compose. One rationalization of this finding is that original tweets also grant more credit (i.e. social image returns) to the composer.

To maximize the welfare effects of informative social media, the audience must have some awareness of its information value. Do students know that virtue signals are informative about discrimination in academia? In the final part of the paper, we use a survey of 1,752 US-based graduate students to characterize beliefs about discrimination in academia and tweet informativeness (following DellaVigna et al. 2019).

We report four main findings from the student survey, the first two of which establish that our core findings are not ex-ante obvious. First, students tend to overestimate how much academics discriminate against Black students, with 84% predicting anti-Black discrimination to be above the upper bound of our estimate's 95% confidence interval. Second, while almost correct on average about informativeness (a 'wisdom of crowds'), 71% make a prediction outside of our 95% confidence interval. Among these students, underestimation is roughly two times more common than overestimation – with the median student predicting that racial justice tweets are roughly half as informative as they actually are. Third, while Black students believe tweets to be somewhat less informative than non-Black students, their predictions are similarly accurate. Fourth, we find no evidence of learning from experience – Twitter-users and upper-year students are no more likely to make accurate predictions about informativeness.

Drawing together our findings, we conclude that social media posts have considerable information value for discriminatory behavior, despite narratives of pervasive virtue signalling, and despite large audiences heightening social image concerns. We find evidence for two forces affecting informativeness. First, while racial justice tweets predict tweeting about Warnock over Ossoff, they do not predict the more-private giving[1] to Warnock over Ossoff, suggesting that visibility increases informativeness. Second, informativeness is considerably lower when more people are tweeting about racial justice, suggesting a tradeoff between attention and information. As for the audience side, most onlookers misunderstand the information value of posts, reducing the welfare benefits of social media.

Our paper builds on three strands of research. First, in emphasizing an informational benefit of social media, we build on the broader evidence of the welfare effects of social media – whether effects on happiness and depression (Mosquera et al. 2020; Allcott et al. 2020;

[1] Although these FEC-reported political contributions are technically public, individual-level contributions tend not to be publicized, and the public nature of even small contributions has only been documented by academics very recently (Bouton et al. 2022).

Braghieri et al. 2022), or political outcomes (Zhuravskaya et al. 2020; Song 2022).

Second, by characterizing a real-world communication equilibrium, we test models of strategic information transmission (Crawford and Sobel 1982; Austen-Smith 1990; Loury 1994; Morris 2001; Kartik 2009; Frankel and Kartik 2019). While most existing tests of these models use lab experiments (Cai and Wang 2006; Blume et al. 2020), Braghieri (2021) uses an online experiment to explore how visibility affects the self-reported sensitive political attitudes of 304 college students. He finds that students' public statements are less predictive of demographics and incentivized behaviors than their private statements. We carry out a similar test of informativeness, but in the naturalistic context of social media and discriminatory behaviors. Otherwise, while Jelveh et al. (2018) explore the predictiveness of language used in research by economists, we explore the predictiveness of language used on a social media platform.

Third, in showing that posts on social media predict heterogeneity in discrimination, we contribute to the large economics literature on discrimination. Recent contributions emphasize heterogeneity in discriminatory behaviors across firms (Kline et al. 2022), while our student survey relates to work that emphasizes the importance of sorting away from discriminators (Becker 1957; Charles and Guryan 2008). In focusing on academia, we complement Milkman et al. (2012), who find that academics are more likely to discriminate against minorities when students request a meeting in one week rather than that day, and Ajzenman et al. (2023), who find that economists on Twitter discriminate against low-ranked and Black students, and in line with our results, male students. Above all, we build on Pager and Quillian (2005). With a sample of 156 employers, they find that self-reported attitudes toward hiring ex-offenders are not predictive of hiring behavior in an audit study. Some of our analysis is similar in spirit, though we increase power with a sample size that is over 70 times larger, and we focus on statements made on social media, where social signalling concerns are more important. Despite these social signalling motives, we reach the opposite qualitative conclusion to Pager and Quillian (2005): words predict audit-measured behavior.

2 Data and Sample

2.1 Academics

While virtue signalling can be studied in many populations, we focus on academics for three main reasons. First, given that our focus is on *differences* in discrimination between the Vocal and the Silent, we require large sample sizes for statistical power. The large number of professors with independent decision-making power makes this feasible. Second, academics have a well-defined audience for whom signals about discriminatory behaviors are potentially important – graduate students. Many of these students have publicly available email addresses, allowing us to elicit their beliefs about the signalling equilibrium in a survey, and then test whether these beliefs are accurate. Third, academics are conducive to an audit study, given the public availability of their email addresses, and the regularity with which they are cold-contacted by prospective students.

We began assembling the data in early-2021. We first listed all research academics in PhD-granting departments in the top-150 universities according to the 2019 US News University Rankings.[2] We found over 125,000 research academics in 5,113 departments in this step.[3]

In the second step, we found the subset of academics with Twitter accounts. We used a search engine with automated searches to create a shortlist of possible Twitter handles for each academic. Research assistants manually picked the correct handle from the shortlist. If an academic's handle was not shortlisted, the research assistant conducted a manual search for the correct handle. We then recorded the academic's email address, position (Assistant, Associate, or Full Professor), gender (Male, Female, Other), and race or ethnicity (White, Black, East Asian, South Asian, Hispanic, Other, Uncertain).[4] We dropped any academics without an email address available online, and those with an explicit policy of not answering prospective student's

[2] The 2022 rankings can be found here.

[3] We also collected data on a few large universities ranked outside the top-150 to participate in a pilot audit experiment. We describe the pilot experiment and findings in our AEA-registered pre-analysis plan.

[4] When not explicitly stated online, the RA team guessed the gender and race/ethnicity using any hints available (e.g., photos, education history, last name).

emails listed on their website.

In early-2022 the RA team double-checked the entire list of tweeting academics. We used this check to drop duplicates (e.g., professors listed twice because they held multiple positions), and drop any non-research-active or non-professor academic (e.g., those on leave, post-docs, emeritus professors, and adjunct professors). This leaves us with a sample of 28,302 tweeting research academics.

For some of the analysis, we use features of the university and department of each academic. For each of the universities, we linked the Black or African American share of undergraduates in Fall 2020 from the National Center for Education Statistics. The team standardized the coded department of each academic, assigning each academic to one of seven broad categories (e.g. Social Sciences), and to one of 75 narrow categories (e.g. Economics).

We imposed six final eligibility criteria. First, we kept only the academics that joined Twitter on May 1, 2020 (just prior to George Floyd's murder) or before, ensuring that our experimental sample were on Twitter during the height of tweeting about racial justice. Second, we required a minimal level of public Twitter activity, keeping only the academics with at least five public tweets in 2020. Third, we dropped academics with lab-oriented Twitter accounts with no personal tweet content. Fourth, we dropped a few academics in departments for which our email templates do not fit well (e.g. Theater, Education departments oriented only towards practitioners). Fifth, we dropped a handful of academics with whom we had discussed the project. Finally, we dropped Black academics and those with race/ethnicity coded as Uncertain, given that the core research question is about how *non-Black* academics signal support for Black people in America. This leaves us with a final experimental sample of 18,514 academics.

We planned to audit the full set of 18,514 academics. Due to a detection-related concern explained below, we stopped the experiment after emailing 11,450 academics – these 11,450 academics comprise the final audit sample.[5]

[5]The reduction in sample size for the audit is one of two major deviations from our pre-analysis plan (the second involves exploring the informativeness of tweets for political contributions, see full details in Appendix B). More accurately, we stopped after emailing the 11,480th academic, but (i) we drop 23 academics because some

For our analysis of political contributions, we use a broader sample of 20,608 academics – the 18,514 non-Black academics selected for the audit experiment, the 1,094 non-audited Black academics satisfying the same selection criteria, and a random sample of 900 non-Black academics without Twitter accounts, using the same non-Twitter-related selection criteria (including that they are from the top-150 universities).

Among the non-Black tweeting academics, audited academics are similar to non-audited academics on most observables (columns 1 and 2, Table 1), suggesting that our audit results might generalize to the broader population of tweeting academics. The exceptions are that, on average, the audited academics are at slightly lower-ranked universities (59th of 150 vs. 52nd), tweet somewhat less (1,022 tweets vs. 1,207), have fewer Twitter followers (2,781 vs. 3,426), and make smaller political contributions.

Black academics on Twitter are more likely to be female, less likely to be Full Professors, and less likely to work in the Sciences (column 3). They are more active on Twitter, much more likely to tweet about racial justice, and they have more followers. Academics without Twitter accounts are less likely to be female, more likely to be Full Professors, and less likely to have donated at least once to the Democrats (column 4).

2.2 Twitter Data and Political Preferences

We used Twitter's academic API to download user-level and tweet-level data for each of the academics with a Twitter account. Among other variables, the user-level data includes the number of followers, number of accounts following, and the date that the account was created. We mostly use user-level data as of May 10, 2022, just prior to the launch of the audit experiment. We also scraped the full list of accounts followed by each academic during July 2022.

The tweet-level data includes the full text of all original tweets, replies, quote tweets, and

academics in their departments were sent an email addressed to the wrong name, (ii) we skipped three academics because of interruptions in our automated email-sending code, and (iii) after debriefing the 16 academics who sent an email reply indicating suspicion of an audit study, four requested for their data to be withdrawn. This leaves us with 11,450 academics in the final analysis sample.

retweets from January 1, 2020 to March 27, 2022. We use January 1, 2020 as the start date to cover the tweets before and around May 25, 2020, the date of the murder of George Floyd. We use March 27, 2022 as the last date as we began to download the tweets shortly after.

We also collected three measures of political preferences. First, we web-scraped Blindspotter to get a measure of the political slant of the news each user interacts with on Twitter. Second, we used and updated data from UCSD on the Twitter accounts of politicians in the US Senate and House of Representatives. We then linked this data with the full list of accounts followed by each academic to calculate (i) the number of political accounts each academic follows, and (ii) the percentage of political accounts followed that are Democrats.

Third, building on Bouton et al. (2022), we linked each academic with their FEC-reported political contributions made from January 1, 2020 to March 27, 2022, mirroring the period for which we have tweets. Bouton et al. (2022) show that the vast majority of contributions are now made through conduits (particularly ActBlue and WinRed), and that reporting requirements for these online platforms ensure that all contributions are reported, along with the full name of the donor. Since the observability of nearly all contributions has been documented only recently, we consider an academic's contributions to be perceived as private, and certainly to be perceived as more private than an academic's tweets. We use this difference in visibility between contributions and tweets to test for the effects of visibility on informativeness below.

For linking, we make use of the fact that 98.9% of FEC-reported individual contributions also report the occupation and employer of the contributor. We first kept only the contributions that list an employer that could be one of the top-150 universities in our sample, and only those that list an occupation that could be consistent with being a research-active academic. After these steps, we carried out an exact-match on full name (allowing for nicknames) and university.[6] We link 29.9% of our academics with at least one FEC contribution, while Bouton et al. (2022) find that 8.5% of the adult US citizen population contributed in 2019 or 2020.

[6]For more details on the FEC data and the matching procedure, see Appendix C. While we pre-specified the use of FEC-linked contributions as controls in the audit analysis, we did not pre-specify our analysis of the informativeness of tweets for contributions. For full details see Appendix B.

2.3 Measuring Racial Justice Signalling

A key part of our paper is determining which academics tweeted in support of racial justice on Twitter and which did not. Given the large sample size, we automated this classification based on the words and phrases included in tweets. We pre-registered our core measure, Vocal_i (with $\text{Silent}_i = 1 - \text{Vocal}_i$), as a binary variable equal to one if academic i has at least one tweet (of any type) from January 1, 2020 to March 27, 2022 that mentions at least one of these racial justice-related words or phrases:[7]

1. *Racism-related*: racism / racist / racial bias / racial discrimination / racial justice / racial prejudice / anti black / white supremacy
2. *Black Lives Matter movement-related*: BLM / black lives / blackintheivory
3. *References to Black individuals killed*: george floyd / ahmaud arbery / breonna taylor / daunte wright / justiceforgeorgefloyd / justiceforgf / justiceforahmaudarbery / justiceforbreonnataylor / justicefordauntewright / sayhername / sayhisname / nojusticenopeace / icantbreathe

We chose these words and phrases to cover the most popular racial justice-related hashtags and to explicitly reflect racial justice themes.[8]

Our automated approach raises two main concerns. First, an academic auto-classified as a signaller may have tweeted about racial justice, but not necessarily *in support* of racial justice. For example, the auto-classification would consider an academic a racial justice signaller if they have ever tweeted "the racial justice movement has gone too far." This case would be a false

[7]For retweets and quote tweets, we include the text in the tweet being retweeted. Throughout, we also include words and phrases found in any (expanded) hyperlinks included in tweets.

[8]For example, we do not include phrases like 'affirmative action' given that a tweet that references this may not necessarily be referring to affirmative action around race. Otherwise, we allow for slight variants of the above terms (e.g. "breonnataylor"). For "BLM", we require the term to not be part of a longer word. This way we avoid classifying medieval manuscript lovers (who may refer to the account @BLMedieval) as racial justice signallers. For full details, see Appendix D.

positive. Second, there may be academics that have tweeted in support of racial justice without using one of the words or phrases above. These cases would be false negatives.

To test for these concerns, we used a richer manual measure of signalling status for a random subset of our experimental sample ($N = 450$). In this random subsample, we automatically classified 64% as vocal. In the full experimental sample of 18,514 academics, we automatically classify 63% as vocal.

For each academic in the random subsample, one or two team members each spent up to five minutes scrolling through the academic's tweets, beginning around May 25, 2020 (the date of George Floyd's murder). After doing so, they selected as many options that apply from the following:

- I did not find any tweets or retweets related to racial justice for Black people
- I found at least one tweet or retweet **opposing** efforts to promote racial justice for Black people
- I found at least one tweet or retweet **questioning** efforts to promote racial justice for Black people
- I found at least one **neutral** tweet or retweet about racial justice for Black people (e.g., mentioning a neutral statistic or a study)
- I found at least one tweet or retweet **showing some support** of efforts to promote racial justice for Black people (e.g. one short tweet with #BLM)
- I found at least one tweet or retweet **heavily supporting** efforts to promote racial justice for Black people (e.g., a long and thoughtful tweet describing why we should support racial justice for Black people, or problems with how police treat Black people)

The data from this user-level manual classification exercise increases confidence in our cruder automated measure, Vocal_i. In particular, no academic in this random subsample ever tweeted in opposition of racial justice, and only two academics questioned efforts to promote racial justice (Figure A1). It follows that there are practically no false positives – in the sense of academics auto-classified as supporting racial justice that in fact question or oppose the movement.

Second, while 13% of those auto-classified as Silent have shown any support on Twitter for racial justice, the figure is 74% for the Vocal. The difference is even larger for heavy support, with Vocal academics roughly 20 times more likely than Silent academics to have tweeted in heavy support of racial justice efforts. This validation shows that false negatives are not too common, and that our automated measure of signalling has a large first stage for the richer manual measures of signalling.

Vocal_i is our pre-registered measure of whether an academic has signalled support for racial justice on Twitter. Nevertheless, we also use a continuous measure (the percentage of tweets auto-classified as racial justice-related) and measures that treat each type of tweet separately (original tweets, replies, quote tweets, and retweets).

We show what factors predict tweeting behavior in Figure A2, using one multivariate regression. Female academics are almost 10 percentage points more likely to tweet about racial justice, as are academics that gave money to Democrats. Comparing fields, social scientists are the most likely to tweet about racial justice, and engineers the least. Comparing universities, racial justice tweeting does not differ much by rank, nor is it more common in universities with more Black undergraduate students. Unsurprisingly, academics who tweet more in general are far more likely to ever tweet about racial justice. We show below that racial justice tweets are informative even after conditioning on how much academics tweet.

2.4 Graduate Students

To study whether students have sophisticated beliefs about the signalling equilibrium, we used graduate program websites to collect the email addresses of doctoral students at the top-80 universities, again as per the US News Rankings of 2019.[9] To oversample Black students, we collected all email addresses of graduate students with photographs where the team judged the student to be likely to self-identify as Black ($N = 3,502$).[10] Otherwise, we randomly sampled three students per doctoral program, provided email addresses were available ($N = 7,337$).[11]

In late-November 2022, we emailed these students with a survey to elicit their predictions about our experimental results (see Appendix E for survey wording). We included a description of how audit studies can identify discrimination, and we described key experiment details – including the timing and how we categorize academics as Vocal or Silent. We asked each student to predict the meeting acceptance rate for distinctively Black names, separately for the full sample, the Vocal academics, and the Silent academics. In each of the three cases, we gave the true acceptance rate for White students as a benchmark. Next, we told the student the unconditional difference in discrimination implied by their answers. Following this, we asked them to report their prediction of the *conditional* difference, using the following text: "...suppose you know of two professors of the same rank in the same department and university. They also share the same gender and race/ethnicity and tweet the same amount. But one of the professors tweeted about racial justice in the past two years and the other did not. What would you expect the difference in racial discrimination to be between these two professors?" Given that this question is less straightforward, particularly for non-quantitative graduate students, we are more confident in the predictions of unconditional differences in discrimination.

We offered each student a $5 Amazon gift voucher and a chance to win one of ten $100 cash prizes for taking the survey. To incentivize predictions, we randomly assigned half of the students to receive one additional lottery ticket for one of four $250 cash prizes for each accurate guess.[12] A drawback of monetary incentives is that respondents may bake bias into

[9] We opted for the top-80 universities rather than the full top-150 because of research assistance capacity constraints.

[10] Though for our analysis we use the self-identification of each student respondent to measure race and ethnicity.

[11] This number is not divisible by three because some doctoral programs had only one or two students.

[12] More specifically, after each prediction question these students would read "You will get one additional lottery ticket for a $250 cash prize if your answer is within 3 percentage points of the number we found." This approach

their reports – reporting not what they believe to be the truth (Haaland et al. forthcoming), but what they believe a ivory-tower academic to find. To check for this, we did not provide accuracy incentives to the remaining half of the students. 1,752 students (16.2%) completed the survey (515 Black and 1,237 non-Black). We describe their characteristics in Section 4.4.

3 Audit Experiment

3.1 Experiment Design

Name Selection. We chose 120 racially distinctive names largely following the approach of Kessler et al. (2019). For first names, we used data on baby names from New York City and Massachusetts.[13] We kept only those with birth years from 1995 to 2004, making the individuals around late-college age today. We dropped distinctively Jewish and Italian names, any first names used in Bertrand and Mullainathan (2004) (since these names may be distinctively fictitious-sounding to some academics), and the eight first names used in our pilot experiment.[14] We imposed a popularity threshold, keeping only the first names used by at least 0.01% of a gender-race cell (e.g. White men). We then kept the top-36 most distinctive for each gender-race cell. For example, for White men, we kept the 36 first names with the highest probability of being a White man conditional on the first name being used. This leaves us with 144 potential first names.

For last names, we used the 2010 US Census, and a within-race popularity cutoff of 0.1% (exactly as in Kessler et al. 2019). We again dropped the eight last names used in our pilot.[15] We kept the 72 most racially distinctive last names – 36 for Black last names, 36 for White.

In the final step, we randomly matched each first name to a last name, with each last name

is incentive-compatible for eliciting the mode of each respondent's subjective belief distribution (Haaland et al. forthcoming).

[13] For the data for New York City, see here. We received the Massachusetts data from the Massachusetts Registry of Vital Records and Statistics.

[14] Iyanna, Tyra, Latrell, Tyreek, Jaclyn, Molly, Graham, and Jonah.

[15] Washington, Glover, Ware, Clay, Collins, Peterson, Ward, and Phillips.

used twice – once each for a distinctively male and female name. This leaves us with 144 full names. To select the 120 most racially distinctive names from among these, we paid MTurkers to guess the race of the names. We dropped the six names with the least accurate guesses in each gender-race cell, leaving us with 120 full names to use for the full audit experiment (see Figures A4 and A5).

Email Addresses. We created one gmail account for each of the 120 full names. Stratifying by race and gender, we randomly assigned each name to one of four possible email formats: [firstname].[lastname][X]@gmail.com, [firstname][lastname][X]@gmail.com, [lastname].[firstname][X]@gmail.com, or [lastname][firstname][X]@gmail.com, where X is a number.[16]

Main Randomization. We sent one email to each academic in our audit sample, purporting to be an undergraduate student interested in graduate studies at the academic's university. The core randomizations were:

1. Distinctively Black vs. distinctively White name of sender (50:50), stratifying on university-by-department.

2. Distinctively male vs. distinctively female name of sender (50:50), stratifying on university-by-department-by race of sender.

3. Sentence mentioning that the sender is a first-generation college student or not (50:50), stratifying on university-by-department.

We use the first-generation and gender randomizations to test whether support for racial justice

[16]To choose X we used a protocol that ensures that the number of digits in X is balanced between distinctively Black and White names. In particular, we first randomly paired each full name with a different full name from the same gender but different race. We then found the lowest X such that a gmail account with that X was available for both the Black and White full name in a given pair. We then randomly picked two numbers above that number, with the same number of digits, and assigned one to the Black name and one to the White name.

on Twitter is informative of support for underrepresented students in general, beyond support for racial minorities.

After these steps, we know the purported race, gender, and first-generation student status of the sender for each academic recipient. Next we randomly assigned the student's name. We randomly chose one Black-male name, one Black-female name, one White-male name, and one White-female name to be used for each university-department. This ensured that all tweeting academics in the same department at the same university assigned to receive an email from, for example, a Black male, would receive an email from the exact same Black male.

In the final step, we randomized the subject and main text of each email at the level of the sender-by-university-by-department, subject to the constraint that the same email type is not used by more than one sender for the same university-by-department. This constraint minimizes the possibility of academics detecting the deception by comparing emails and seeing two identical-looking emails from different senders.

We chose the main text of the email from 12 possible variants. We then randomly chose a minor variant of the email from three options for each of the 12 main text variants. The minor variants involve small changes to minimize the chances of our emails being detected as spam (e.g. "final year of undergrad" instead of "final-year undergraduate"). We randomized the minor variant at the level of sender-by-university-by-department, meaning that a given fictitious student uses the same minor email variant for all of their emails to a given university-department.[17] For an example email format, see Appendix F.

Minor Randomization Details. For randomization stratified on university-by-department, we made sure that all strata have at least four observations (covering the four race-by-gender treatments) by joining together small strata (usually creating a strata that includes all of the small departments of a given university).[18]

[17]For some departments in which the academic would not work on research per se (e.g. because they compose music), we used a fourth minor variant which replaces the term "research" with "work" throughout.

[18]Since we ultimately only emailed 11,450 of the 18,514 academics, we have some singleton strata in the

We split the universities into nine groups according to the final exam dates for the last semester. We emailed the academics according to this order, with the email order within each of the nine groups randomized.

Since most of our email types mention the undergraduate institution of the fictitious sender, we assigned this institution randomly at the level of the sender-by-university-by-department. For the set of possible institutions, we started with the same top-150 US News ranked institutions as for our sample of academics. We then used NCES data from Fall 2020 to keep the 90 institutions that satisfy these eligibility criteria: (i) at least 4% Black or African American undergraduate enrollment, (ii) at least 20% White undergraduate enrollment, (iii) 20 to 80% female undergraduate enrollment, (iv) undergraduate degrees offered, (v) at least 4,000 undergraduates enrolled, and (vi) no technology focus (i.e. we drop institutions like MIT). For each university covered by our audit sample, we kept the eight of the 90 institutions that are closest in rank to be considered as the institution of the fictitious student.

Ethics. We received full ethics approval for the audit experiment from UBC's Behavioural Research Ethics Board (ID: H20-03758). The audit experiment necessitates deception to give a plausible measure of actual racial discrimination over email. We took several steps to minimize ethical concerns. First, to reduce the burden on academics, we sent the emails during May when most research academics are not teaching. Second, we excluded Black academics from the experiment entirely.[19] This means that we did not take up any time of Black academics who are already underrepresented in academia. Third, whenever an academic accepted a meeting invite, we sent emails manually to cancel Zoom meetings promptly and politely. This limited the time cost on a given academic to writing the reply to one email. Fourth, we did not debrief academics on the fictitious nature of the email after the experiment ended. This reduces the

analysis sample. Whenever our analysis includes strata fixed effects, these singleton observations are dropped, leaving us with 11,393 observations.

[19] With only 1,094 Black academics satisfying the eligibility criteria, and 88% classified as Vocal (Table 1), we would anyway have had little statistical power to estimate tweet informativeness separately for Black academics.

possibility that the audited academics become more suspicious of future emails from genuine students.

Replies. While we automated the sending of emails, the team classified each email reply manually, as either (i) accepting the meeting request, (ii) declining the meeting request, but sending a helpful reply (e.g. by directing the student to other resources), or (iii) declining the meeting request, without a helpful reply.[20]

We pre-registered meeting acceptance as our main outcome, given that meeting acceptance is more welfare-relevant for students than email replies. We consider meeting acceptance to be a costly prosocial behavior – costly because acceptance effectively commits the academic to a 20-minute or so online meeting, and prosocial given that meetings with students outside of an academic's institution tend to benefit the student rather than the academic.

Meeting acceptance is not a fully private behavior, given that the student observes the academic's response.[21] However, meeting acceptance is substantially less visible than tweets, and individual-level discriminatory behavior is fully private, since discrimination can only be detected at the group-level. Our analysis then characterizes the informativeness of public tweets about a costly, prosocial, and near-private behavior.

Detection on Twitter. We began sending emails in May 2022 with the intention of emailing all 18,514 academics over a two-week period. Following the launch of the experiment, we monitored Twitter for any conversation about the audit study. On May 19th, an economist wrote a tweet thread mentioning their suspicion of our audit study as well as advice on running audit studies. This tweet got some traction among economists, with 46 retweets and 133 likes by May 24th.[22] To minimize the possibility of mass detection (particularly among fields outside

[20] We assigned each email coder to the same number of White and Black email accounts, ensuring that email coder fixed effects are orthogonal to the race of the fictitious student.

[21] One could argue that visibility also increases if an academic decides to tell others about the meeting, but this argument applies to all private behaviors.

[22] As of writing (May 23, 2023), the tweet has 44 retweets and 133 likes.

of economics), we decided on May 19th to not send any further emails. Given this pause, the final audit sample includes 11,450 academics. As reported earlier, audited academics are similar on most observables to the non-audited academics (Table 1). We also use a series of analyses below to show that our findings are unlikely to be explained by academics detecting that the email was part of an audit study.

3.2 Specifications and Outcomes

Racial Discrimination. To estimate overall racial discrimination of academics on Twitter, we use the following specification:

$$\text{Accepted}_i = \alpha_{d(i)} + \alpha_{e(i)} + \beta \text{Black}_i + \varepsilon_i \tag{1}$$

where Accepted_i is a dummy variable equal to one if academic i accepted the meeting invitation,[23] $\alpha_{d(i)}$ are university-by-department of academic i fixed effects (equivalent to randomization strata), and $\alpha_{e(i)}$ are major-by-minor email type fixed effects.

Black_i is a dummy variable equal to one if academic i received an email from a purportedly Black student. We cluster standard errors at the university-by-department-by-sender name-level, the level of treatment, with up to four clusters per university-by-department. Balance checks are consistent with the randomization being carried out successfully (Table A1).

For the more important test of whether discriminatory behavior differs by racial justice

[23] We last checked the email accounts in mid-July, roughly eight weeks after we sent emails. The vast majority of responses came much earlier (Figure A3).

tweeting, we estimate:

$$
\begin{aligned}
\text{Accepted}_i = {} & \alpha_{d(i)} + \alpha_{e(i)} + \gamma_1 \text{Black}_i \\
& + \gamma_2 \left(\text{Black}_i \times \text{Vocal}_i\right) + \gamma_3 \text{Vocal}_i \\
& + \Sigma_j \theta_j \left(\text{Black}_i \times X_i^j\right) + \Sigma_k \eta_k X_i^k + \varepsilon_i
\end{aligned}
\tag{2}
$$

where γ_2 is the key coefficient, Vocal_i is a dummy variable equal to one for those automatically classified as having signalled support for racial justice, and the set of controls X_i^j (with levels and interactions with Black_i) varies across specifications.

The interpretation of γ_2 depends on the set of interacted controls we include in the regression. In particular, the coefficient tells us the signal conveyed by racial justice tweets over and above the information contained in the controls. Without any interacted controls, γ_2 answers the question: what is the unconditional difference in discriminatory behavior between Vocal and Silent academics? In most cases, the more relevant comparison would be conditional. For example, as we have equated Vocal status to having *ever* tweeted about racial justice, academics that tweet more are more likely to be Vocal than those who tweet less (Figure A2). For a student scrolling Twitter, the relevant question may be: knowing that professors X and Y both tweet a similar amount, what additional information about discriminatory behavior is conveyed by the fact that X tweets about racial justice while Y does not? Going further, the student may know the gender and race of X and Y, the institution at which they teach, and perhaps even finer details. What is the signal of a racial justice tweet over and above this information?[24]

To allow for different possible information sets of onlookers, we estimate specification 2 with different sets of interacted controls, with the full set of controls including measures of Twitter activity, basic demographics, university and department fixed effects, and measures of political preferences.

[24] A similar logic applies to the coefficient γ_3, where in this case we are predicting differences in overall response rates to distinctively White names, rather than differences in differential response rates by race.

4 What Do Virtue Signals Signal?

4.1 Discrimination in the Audit Experiment

Overall Discrimination. Audited academics accept 30.2% of the meeting requests overall, suggesting that our fictitious students were generally taken to be making plausible requests. We do not detect racial discrimination in the full sample – academics accept 30.6% of emails from distinctively White names, and 29.8% of emails from distinctively Black names ($p = 0.31$ for the difference, Figure 1). We can reject discrimination against Black students of 2.4 percentage points or more with 95% confidence.

Tweeting academics discriminate against Black individuals less than the 2.1 percentage points found among large US employers (Kline et al. 2022) – we reject the null hypothesis that racial discrimination of academics as a percentage is the same as employers ($p = 0.02$), and we almost reject the null that the percentage *point* discrimination is the same ($p = 0.12$). Otherwise, tweeting academics discriminate far less than a representative sample of over 6,000 academics audited in 2010 (Milkman et al. 2012). Those academics were roughly eight percentage points less likely to accept a request from an African American name, relative to a Caucasian name, for an in-person meeting in one week. We confidently reject this estimate ($p < 0.0001$). We cannot distinguish between two stories for why we find less racial discrimination: academics in 2022 may discriminate less than academics in 2010, or it may be that tweeting academics discriminate less than non-tweeting academics.

Signalling Discrimination. Silent academics are 5.3 percentage points (18%) less likely to accept a meeting with a Black student ($p < 0.001$, Figure 1), whereas Vocal academics are 1.9 percentage points (6%) *more* likely to accept a meeting with a Black student ($p = 0.07$). The difference in discrimination is then 7.2 percentage points ($p < 0.001$). In this setting, academic Twitter is racially unbiased overall only because the pro-Black bias of the Vocal academics almost exactly offsets the anti-Black bias of the Silent academics.

Vocal and Silent academics treat emails from distinctively White names similarly – while the raw White student acceptance rate is 0.8 percentage points higher for Vocal than for Silent academics, we cannot reject the null of no effect (in specification 2, $\hat{\gamma}_3 = 0.6$ percentage points, $p = 0.65$). If these results generalize to other faculty-student interactions, we would conclude that White students have similar experiences with Silent and Vocal academics. In contrast, Black students are 37% more likely to secure a meeting with a Vocal than with a Silent academic. In this world, where Vocal academics treat White students similarly and Black students better than Silent academics, Black students may have a difficult time convincing their White classmates of their experiences – it is as if the two groups of students see the world through different lenses.

While we pre-registered meeting acceptance as our main outcome, our findings are similar if we look at effects on whether the academic accepted the meeting *or* sent a helpful email reply (Figure A6), or on whether the academic replied at all (Figure A7). With both outcomes, Vocal academics are statistically significantly less likely to discriminate against Black students than Silent academics. The one qualitative difference is that for each outcome we cannot reject the null hypothesis that Vocal academics treat Black and White students equally. This suggests that the favoritism towards Black students in Figure 1 comes from the margin of academics accepting a meeting rather than declining while still being helpful.

Supporting Marginalized Students. Racial justice tweets signal support for Black students. Do these tweets also signal more general support for students from underrepresented groups? Or is the signal specific to behaviors with respect to race? For this, we first turn to the first-generation student randomization, given their underrepresentation across academic fields (Schultz et al. 2022).

Overall, tweeting academics are 3.2 percentage points more likely to accept meetings with students that reference their first-generation status ($p < 0.001$, Figure 2). This finding echoes recent audit evidence that minorities benefit from explicitly mentioning their demographic identity when requesting help (Kirgios et al. 2022). The first-generation advantage is driven entirely by the Vocal academics, who favor first-generation students by 5.8 percentage points ($p < 0.001$).

We see the same pattern for gender. Academics overall are 4.3 percentage points more likely to accept meetings from distinctively female names ($p < 0.001$, Figure 3), similar to evidence of gender gaps elsewhere in academia – in the past 20 years, women were more likely to be selected as members of prestigious national academies than men with similar records (Card et al. 2023). Like first-generation student discrimination, Vocal academics discriminate more in favor of women than Silent academics ($p = 0.03$).

When considering race, we have seen that Vocal academics are less biased than Silent academics – the absolute racial gap in meeting acceptance is 1.9 percentage points versus 5.3 percentage points. The first-generation student and gender randomizations show that this is not because Vocal academics are more likely to be unbiased in general. Here, in both cases, Vocal academics discriminate more, in absolute terms. Racial justice tweets then signal support for students from underrepresented groups, but not unbiasedness.

White Males and Sorting. One way to summarize our results across the three dimensions (race, gender, and first-generation status) is to group together the 7/8 of students that belong to any underrepresented group (Black, female, or first-generation) and compare their meeting success rate with the remaining 1/8 of students: White males with no mention of first-generation status. While this exercise unmasks new findings, we caveat that the comparison of 1/8 versus 7/8 of the emails is lower-powered than our earlier comparisons of 1/2 versus 1/2.

Vocal academics are 9.1 percentage points (27%) less likely to accept a meeting with a White male student than with an underrepresented minority student (Figure 4). This large gap was masked in the previous figures, where given the design of the experiment, 75% of the disfavored category (e.g. White) belonged to at least one underrepresented group (with 25% White female first-generation, 25% White female not first-generation, 25% White male first-generation, and 25% White male not first-generation). In contrast, we cannot reject the null that Silent professors treat the two groups equally ($p = 0.64$), although given the lack of statistical power, the 95% confidence interval is large – ranging from 5.1 percentage points discrimination against underrepresented minorities to 3.1 percentage points discrimination against White males.[25]

Next, we compare the overall level of meeting acceptances between Vocal and Silent academics. Recall that Vocal academics accept more meeting requests from White students (Figure 1), regular students (Figure 2), and male students (Figure 3) than Silent academics. These differences are driven by the presence of underrepresented minority students in each of those three groups. In particular, when we look at only White male non-first-generation students, Vocal academics accept 4.4 percentage points fewer meeting requests than Silent academics (Figure 4, $p = 0.06$).

In a world with rational beliefs and students aiming to maximize meeting acceptance, our pattern of results leads to sorting – White male students prefer to request meetings with Silent professors, while underrepresented minority students request meetings with Vocal professors. Sorting can occur even if students have inaccurate beliefs about meeting acceptance rates – provided that they are similarly likely to request meetings from Silent and Vocal professors, and keep meeting with professors that accept.[26] Such sorting may serve to reinforce the views of both sets of academics, to the extent that ingroup and outgroup contact affects attitudes toward race and diversity.

The Intensive Margin. Our results show extensive margin signalling: racial justice ever-tweeters discriminate more in favor of students from underrepresented groups than racial justice never-

[25] We also note that the estimate of the gap from the specification without strata and email type fixed effects ($28.7 - 26.6 = 2.1$) is somewhat larger in this case than the estimate after including strata and email type fixed effects (1 percentage point). As with the previous figures, the p-value of 0.64 comes from the specification with strata and email type fixed effects, while without strata and email type effects, the p-value is 0.3.

[26] While documenting such sorting is beyond the scope of this paper, future research might explore this question by linking academics' tweets with the demographics of their research assistants, advisees, and lab members.

tweeters. To explore signalling on the intensive margin, we split the set of Vocal academics into two groups: those with a percentage of racial justice tweets below versus above the median. We call these two sets of academics the "Rarely Vocal" and the "Regularly Vocal."

Rarely Vocal academics are close to unbiased, accepting 30.4% of meeting requests from White students, and 31% from Black students ($p = 0.39$ for the difference using the specification with strata and email fixed effects, Figure 5). Regularly Vocal academics favor Black students by 3.2 percentage points ($p = 0.01$). Academics who tweet more often about racial justice show more support for Black students than academics who tweet less often about racial justice.

In contrast, the Regularly Vocal are no more likely than the Rarely Vocal to favor first-generation and female students (Figures A8 and A9). So while the extensive margin of racial justice tweets signals support for three types of marginalized students, the intensive margin carries a narrower informativeness: only signalling about racial bias. Put another way, silence speaks louder than words.

Conditional Signalling. Our results so far make unconditional comparisons between Vocal and Silent academics. But as referred to earlier, Vocal and Silent academics differ along many dimensions other than their racial justice tweets: for example, Vocal academics are more likely to be female, Democrats, and less likely to work in Engineering and Technology (Figure A2). More mechanically, they tweet more often. Is vocality merely proxying for these other dimensions, or does vocality predict racial discrimination above and beyond these other observables? We answer this question in Figure 6.

The far-left coefficient replicates our earlier result: Vocal academics discriminate against Black students 7.3 percentage points less than Silent academics (when including strata and email fixed effects). The coefficient falls slightly, to 7 percentage points, when conditioning on the number of tweets. Vocality is then not just capturing the fact that Vocal academics tweet more, and that those that tweet more discriminate less. The coefficient falls gradually as we add more controls: gender, race/ethnicity, position, university, and department fixed effects. With these controls the coefficient falls to 5.3 – smaller, though still statistically significant at the 1% level. These controls likely constitute a common information set of prospective students. Given this, we will refer back to this particular measure of conditional informativeness (the coefficient fourth from the right in Figure 6) below.

While students are perhaps less likely to know the political views of prospective advisors, controlling for our three measures of political views barely changes the informativeness of racial justice tweets. So even students with this richer information set can learn from the tweets of their professors.

There are two possible explanations for why the conditional informativeness of racial justice tweets is similar to the unconditional informativeness. First, it could be that the extra controls are not correlated with whether academics tweet about racial justice, but this possibility is ruled out by Figure A2. Second, it could be that the extra controls do not predict racial discrimination. We explore this explanation in Figures A10 and A11.

While our estimates are imprecise, most observable factors do not predict discriminatory behaviors enough for us to reject the null hypothesis of no effect. This is the case when including each interacted factor one-by-one (Figure A10) or when including all interacted factors at once (Figure A11). Focusing on the latter, only two variables are statistically significant predictors of discrimination – Vocal academics discriminate against Black students 5.6 percentage points less ($p = 0.01$) and academics at universities ranked in the top-50 discriminate 5.2 percentage points less ($p = 0.02$), other things equal. Otherwise, Democratic contributions, fields of study, race/ethnicity, and exposure to Black undergraduates are not predictive, although there is some suggestive evidence that female academics discriminate less (2.5 percentage points, $p = 0.19$). Of these, the non-significance of Democratic contributions is perhaps the most interesting – in this setting, *costly* support for the Democrats is less informative than cheap statements about racial justice on social media.

In terms of the magnitude of the point estimate, Vocality is in fact the single most predictive variable of racial discrimination, whether considering unconditional (Figure A10) or conditional predictiveness (Figure A11). However, given the imprecision of the estimates, we cannot statistically reject equality of the Vocal interaction with several other interaction terms.

In summary, academics' racial justice tweets carry information over and above other observables, and going even further, no other variable in our data outperforms the informativeness of racial justice tweets.

Detecting Detection. Suspicion of the fictitious nature of our emails is more likely in our setting than others – after all, academics essentially invented the audit method. In our case, we also have concrete evidence of detection, including the tweets by economists on Twitter, and 16 email responses from academics that explicitly noted suspicion. These cases give us a lower bound on the extent of detection, but otherwise cases of detection are undetectable – the academic that detects may just ignore our email. Two patterns of detection and behavior would make our results particularly misleading. First, if emails from distinctively Black names raise more suspicion than those from distinctively White names,[27] and if academics respond less when suspicious, racial gaps in responses may be unrelated to actual discriminatory behaviors. Second, even if suspicion is not affected by race, it could be the case that (i) Vocal academics are more likely to be suspicious (perhaps because they are more familiar with audit studies, being more interested in racial justice), and (ii) when suspicious, social signalling concerns drive these academics to reply more to Black names than White names. Either of these two cases could explain our findings, even if Silent and Vocal academics are equally likely to discriminate against Black students in daily interactions. Given this concern, we report a series of checks.

First, of the 16 detectors, we note that seven received an email from a Black name and nine from a White name. This gives some evidence against the first of the two main concerns – that emails from Black names are thought to be more likely to be fictitious.

Second, our core findings are very similar when we use samples and outcomes less subject

[27] Given that Black students are underrepresented in graduate studies, a Bayesian should think that emails from Black students are more likely to be fictitious than emails from White students.

to detection concerns (Figure 7 for race, and Figures A12 and A13 for first-generation status and gender). In particular, the patterns of racial discrimination of the Silent and Vocal are similar when we drop academics in fields more familiar with audit studies: either those in Economics, Political Science, Sociology, and Business (12.4% of the sample), or all of those in the Social Sciences (25% of the sample). The results are also similar if we drop the 7% of academics to whom we sent more generic emails – not mentioning the specific field of the academic, which could arouse more suspicion. Next, assuming that suspicion is more likely when academics see that colleagues have received similar emails, we show that the results are similar when we drop university-departments that received either more than ten, or more than five, emails. Finally, based on the same idea that discussion between academics might increase suspicion, we show that the findings are similar when we define the outcome as accepting the meeting within one day (likely without having the time to discuss the email with other academics), or within three days.

Third, we use an accounting exercise to ask: assuming that there is no true difference in discrimination between Silent and Vocal academics, what percentage of academics would need to be suspicious to fully explain a difference in racial discrimination of 7.3 percentage points? For this exercise, we make the following assumptions: (i) Silent academics never detect the audit (a conservative assumption if detection makes academics weakly more likely to avoid anti-Black discrimination), (ii) some percentage X of the Vocal detect the audit, while the remaining (100-X)% act as they would in real life, (iii) true racial discrimination is the same for Vocal and Silent academics, at 5.3 percentage points, and (iv) Vocal detectors accept meetings only when the student is Black, with 30% overall acceptance (a particularly extreme assumption, giving the Vocal detectors a 60 percentage point discrimination rate). Even under these conservative assumptions, we would need at least 11.2% of the Vocal academics to have detected the audit to fully explain our core unconditional signalling result.[28] We find this number relatively im-

[28]Note that this is a much stronger statement than requiring 11.2% of the Vocal academics to be familiar with the audit study methodology. We are requiring them to recognize that our email was fictitious. This reaction is plausible for those for whom the audit method is top-of-mind, but less so for those who are familiar with the

plausible, especially for fields outside of the Social Sciences. Collectively, we conclude that detection is unlikely to have tainted our findings.

4.2 Signalling Politics

Racial justice tweets are highly predictive of discriminatory behavior in the audit experiment. Given strong partisan differences in the issue importance of racial justice, we now explore the political signal conveyed by racial justice tweets. We use a broader sample of academics for this analysis: the full set of 18,514 non-Black tweeting academics (6,784 Silent and 11,730 Vocal), 1,094 Black tweeting academics, and a random sample of 900 non-Black academics without Twitter accounts.

Our analysis of political signalling has three main advantages over the audit study. First, our broader sample allows us to compare the behavior of non-Black academics with that of Black academics. Here we can see whether the behavior of Vocal academics more closely resembles that of Black academics or Silent academics. Second, our random sample of non-tweeting academics allows us to explore the signal of being on Twitter itself. Third, with two measures of political action that vary in visibility (monetary contributions and political tweets), we can explore whether racial justice tweets are more informative about political behavior that is more visible. Otherwise, our analysis of political signalling has one key limitation relative to the audit: we do not have exogenous variation in the perceived race of political candidates. We use the Ossoff-Warnock case study to approximate such variation, but our conclusions on differences in treatment by race are necessarily more suggestive.

Democratic vs. Republican Contributions. Each sample of academics overwhelmingly supports the Democratic party (Figure 8), consistent with other work on the liberal slant of academia (Langbert et al. 2016). Roughly one in 200 Silent academics and one in 200 Vocal academics

method but work in a field that does not use audit studies.

gave at least once to Republican-related committees, while zero Black academics did.

While vocality is not a signal of support for Republicans, it is a strong signal of support for Democrats – Vocal academics are 81% more likely than Silent academics to have given to Democrats (36.2% vs. 20%, Figure 8), with a similar gap when considering the dollar value of total contributions (Figure A14). If Vocal academics are more likely to have green cards or be citizens, some of this gap may be mechanical – with more Vocal academics legally allowed to make contributions. While we do not have measures of citizenship, we can control for the race/ethnicity categories as coded by the RA team (White, East Asian, South Asian, Hispanic, Other), and richer measures of race and ethnicity based on predictions made by `namsor.app` using the full name of each academic. From namsor we have the first- and second-most likely race (with six categories, e.g. Black, Caucasian) and the first- and second-most likely ethnicity (with 120 categories, e.g. Chinese, British, Jewish). After adding these controls together with our previous set of controls, Vocal academics remain 55% more likely to contribute to Democrats (Figure A15), and contribute 46% more in dollar terms (Figure A16). The Vocal-Silent gap in Democratic support is then unlikely to be fully explained by Vocal-Silent differences in citizenship and green card status.

Non-tweeting academics resemble Silent academics more than Vocal academics, with a similar percentage contributing at least once to Democratic causes (18.7% vs. 20%). On the other hand, non-tweeting academics are roughly twice as likely to have contributed to Republican causes as either Silent or Vocal academics (1.1% vs. 0.6%) and give more to Republicans on average (Figure A14). These findings demonstrate that while academia has a strong bias towards the Democrats, the Democrat-bias of academic Twitter is even stronger.

Ossoff vs. Warnock Contributionsn and Tweets. For an analysis that parallels the audit experiment, we focus on a case study in which two similar political candidates differ on race. Two Democratic candidates ran for the US Senate in Georgia in November 2020: Jon Ossoff (a White candidate), and Raphael Warnock (a Black candidate). Neither candidate won a majority of the vote, forcing runoff elections to be held in January 2021. At this point, the Democrats held 48 seats in the Senate, and the Republicans held 50 seats. Since the Democrats held the vice-presidency with Kamala Harris, they would secure a majority in the Senate by winning both of the Georgia runoffs. This made the runoffs nationally salient, with both Ossoff and Warnock breaking fundraising records. Reflecting this, the FEC committees receiving the most contributions in our data are Biden for President (first), Biden Victory Fund (second), Warnock for Georgia (third), and Jon Ossoff for Senate (fourth). Polling found both races to be similarly close,[29] and both Democrats won their runoffs by slim margins: Ossoff by 0.9 percentage points and Warnock by 2 percentage points. Both have voted similarly since in office (Figure A17).

Consistent with Figure 8, Vocal academics are over twice as likely to give to Ossoff and Warnock than Silent academics, and non-tweeting academics most closely resemble Silent academics (Figure 9). In contrast to the audit results, the Ossoff-Warnock gap in giving is similar for both Silent and Vocal academics – both are slightly less likely to give to Warnock than to Ossoff. We cannot reject the null that the racial gap in giving is the same for Silent and Vocal academics ($p = 0.53$), nor can we reject the null after adding controls (Figure A18). In contrast, Black academics are more likely to give to Warnock than Ossoff (7.6% vs. 5.8%, $p < 0.01$), ruling out the possibility that Democratic contributors almost always support both candidates, or none at all. This pattern of findings is similar when considering total contributions (Figure A19), and when considering only the audited sample (Figure A20), showing that the contrast with the audit results is not due to sample differences.

While racial justice tweets do not predict racial gaps in political contributions, they do predict racial gaps in political tweets. In particular, Vocal academics are 11% more likely to have tweeted at least once about Raphael Warnock than about Jon Ossoff ($p < 0.001$), while Silent academics tweet similarly (and very little) about both candidates (Figure 10). The unconditional Vocal-Silent difference in the racial gap in tweeting is statistically significant ($p < 0.001$,

[29]For example, see https://projects.fivethirtyeight.com/georgia-senate-polls/, accessed May 20, 2023.

Figure 10), and remains so after adding controls (Figure A21), ruling out a simpler explanation that people that tweet more are also more likely to tweet about Warnock than about Ossoff. We consider this finding suggestive evidence that racial justice tweets are more informative of more visible behaviors – suggestive in that visibility is not the only dimension along which political contributions and political tweets differ. All this being said, the racial gap in visible political support is still much greater for Black academics than for Vocal non-Black academics – they tweet about Raphael Warnock 54% more often than they tweet about Jon Ossoff.

Discussion. If racial justice tweets signal support for Black over White students, why do they not signal political contributions for Warnock over Ossoff? One possibility is that the two types of racial gaps are not driven by the same underlying behavioral trait. Consistent with this, we find that the Warnock-Ossoff gap in giving is not predictive of racial discrimination in the audit study.[30] While we lack evidence on exactly why the two racial gaps are uncorrelated, we note two possibilities. First, discrimination in the audit (in both directions), may be statistical. Silent academics predict that Black students are less promising as graduate students, deserving less support, while Vocal academics predict that Black students are more disadvantaged, deserving more support. While academics have little information about our fictitious students, they have rich information about Ossoff and Warnock, making statistical discrimination less likely. Second, race is not randomly assigned in the case study. While Ossoff and Warnock are similar on many dimensions, they differ on others – for example, Warnock has a background as a Baptist pastor, while Ossoff has a background as a journalist. If Vocal academics support racial minorities, but oppose organized religion, they may ultimately support Ossoff and Warnock equally.

[30] In particular, we run specification 2 without the set of controls X_i^j, replacing the Vocal dummy variable with either (i) total contributions made to Warnock minus those made to Ossoff or (ii) a dummy variable for any contributions made to Warnock minus a dummy variable for any contributions made to Ossoff. In the first case the p-value for $\hat{\gamma}_2$ is 0.86, in the second it is 0.11. In both cases, the sign is the opposite of the natural prediction – with those who give relatively more to Warnock discriminating more against Black students.

4.3 When Are Tweets More Informative?

We now turn to theory as a guide for when tweets are more likely to be informative of discriminatory behavior in the audit.[31] Applying Frankel and Kartik (2019) to our setting, let us assume that academics differ along two dimensions: some are more biased for or against Black students than others, which we measure using the audit study, and some have stronger social image concerns than others, motivating them to misrepresent themselves on social media. Frankel and Kartik (2019) show that, under some assumptions, this leads to a problem of muddled information – when we see an academic tweet in support of racial justice, it is unclear whether this public action reflects the fact that the academic has low anti-Black bias, or that they have high social image concerns. The key result of the paper is a comparative static regarding the information extraction problem: as the incentives to signal low anti-Black bias increase, tweets become less informative about bias, and more informative about social image concerns. With this prediction as a guide, in this subsection we explore how tweet informativeness varies with different proxies for signalling incentives.

With our focus now on testing for *differences* in informativeness, our primary tests use a continuous measure of racial justice signalling – the percentage of a user's tweets that are about racial justice – rather than the binary measure Vocal_i. We do this because the binary measure can reveal mechanical differences in informativeness not emphasized by theory. To see this, suppose that academics have pro-Black bias r, and tweet about racial justice whenever $r > k$. Here tweet activity is a simple function of racial justice types – there is no muddled information. As incentives to signal increase (e.g. through a shock that increases the number of Twitter followers), k falls, leading to more academics tweeting about racial justice. Informativeness can change mechanically: $E[r|r > k] - E[r|r \leq k]$, can increase, decrease, or stay unchanged,

[31] We focus on audit-measured behavior rather than political contributions for three reasons. First, this is what we pre-specified, whereas the political contributions analysis was incorporated later on (as explained in Appendix B). Second, the audit experiment gives a cleaner measure of discriminatory behavior. Third, racial justice tweets are more informative of behavior in the audit, giving us more scope for finding correlates of informativeness.

depending on the distribution of r.[32] In this sense, the difference in informativeness using the binary measure may reflect the distribution of underlying racial bias, rather than the muddled information described by Frankel and Kartik (2019).[33] The continuous measure is less subject to this critique, though it is not immune, since it is censored at zero.

The continuous measure has a different drawback: outlying observations can disproportionately affect the results. To address this, we always winsorize continuous measures at the 99th percentile (see Figure A23 for the distribution of the main continuous measures we use, along with the 99th percentile). Despite our preference for the continuous measure, we nevertheless report findings with the binary measure in the Appendix, and note when the findings differ.

Social Pressure. Racial justice tweets among non-Black academics were rare in early-2020, but spiked following the murder of George Floyd by a White police officer on May 25th (Figure A22). After a month or so, racial justice tweets became less common again, though more common than in early-2020. We split 2020 into three periods: the pre-period (January 1 to May 24), the during-period (May 25 to June 30), and the post-period (July 1 to November 30). We think of the during-period as a period of high social pressure to tweet about racial justice, or more generally, as a period of higher incentives to signal low racial bias. Frankel and Kartik (2019) predict that informativeness (i.e. $\hat{\gamma}_2$ from specification 2) should be higher in the post-period than in the during-period, given that the incentives to signal are weaker. The prediction for informativeness in the during-period relative to the pre-period is more ambiguous: while signalling incentives are surely higher in the during-period than the pre-period, the distribution of underlying racial bias of academics is also likely to have changed, with academics getting exposure to many more cases of racial injustice (Reny and Newman 2021). Similar concerns still apply somewhat to the comparison between the during- and post-period – types may still be changing. Given this, we can think of this exercise either as a test of the theory, or less

[32] For example, measured informativeness is constant in k when r is uniformly distributed. When normally distributed, measured informativeness is decreasing in k when $k < E[r]$, but increasing in k when $k > E[r]$.

[33] We thank Alex Frankel for highlighting this point.

ambitiously, as a theory-inspired test of whether informativeness increased as racial justice tweets became less fashionable.

During the pre-period, 17% of audited academics tweeted at least once about racial justice. The low percentage makes this period's estimates the most imprecise, though there is evidence for unconditional and conditional informativeness – a one percentage point increase in racial justice tweets is associated with roughly three percentage points less anti-Black discrimination ($p < 0.02$, Figure 11). The point estimates for the during-period, when 49% of academics are Vocal, are much smaller, though more precisely estimated – an unconditional informativeness of 0.17 percentage points ($p = 0.02$), and a conditional difference in discrimination of 0.11 percentage points ($p = 0.16$). During the height of racial justice tweeting, we cannot reject the null hypothesis that racial justice tweets are uninformative when basic observable characteristics are known.

As predicted, informativeness is substantially higher in the post-period – unconditional informativeness is 1.2 percentage points ($p < 0.01$) and conditional informativeness is 0.94 percentage points ($p = 0.01$). We can reject that during- versus post-period informativeness is equal ($p < 0.01$ for unconditional, $p = 0.02$ for conditional). When using Twitter, rational belief updating about racial discrimination should take into account the current popularity of racial justice tweets. The Bayesian updates much more from seeing one racial justice tweet when such tweets are unpopular than when they are popular.

The key pattern of during- versus post-period informativeness is similar with the binary measure (Figure A24). Unconditional informativeness in the post-period is 57% higher than in the during-period, while conditional informativeness is roughly twice as high, though we have less power to reject the null that during- and post-period informativeness are equal ($p = 0.13$ and $p = 0.19$).

To the extent that the increase in informativeness is due to the fall in social pressure, the results here highlight a tradeoff: while social pressure increases awareness of racial injustice, through prompting more academics to tweet and retweet, social pressure also reduces the informational benefits of social media.

Audience and Reputation. In principle, racial justice signalling incentives increase in audience size, audience left-wing slant, and career incentives. We measure these dimensions with the number of Twitter followers, the Biden 2020 vote share of the academic's university's county, and whether the academic has tenure or not, proxied by whether the academic is an Assistant Professor, or an Associate or Full Professor. For these dimensions we have an even weaker claim of causality – Twitter followers are clearly not randomly assigned, nor is an academic's location, or their career success. Given this, we think of this exercise as more a question of whether theory can guide us toward useful heuristics – i.e. does the causal claim of the theory also show up in correlations in real-world equilibria?

Racial justice tweets are more informative among academics with above-median Twitter followers, going against the theoretical prediction (Figure 12, same direction but smaller magnitudes with the binary measure in Figure A25). The most famous academics have the largest incentives to misrepresent their views (when unverifiable), but their tweets are in fact more informative than those of less-famous academics. One possible rationalization is that academics with more Twitter followers perceive their private behaviors to be more closely monitored, with a greater risk of any hypocrisy being exposed. Otherwise, informativeness is similar in counties with above- versus below-median Biden vote share, and similar for non-tenured and tenured academics (with similar patterns with the binary measure). There is then no compelling evidence for the theory-guided hypothesis that the tweets of tenured professors, who have weaker reputational incentives, should be more informative.

Less guided by theory, we also find that informativeness differs little by university rank and gender, but is much higher for academics that have not made political contributions to the Democrats (Figures A26 and A27). In fact, both unconditional and conditional informativeness are not statistically significant for academics that have made Democratic contributions, whether we use the binary or the continuous measure. In each of those cases, informativeness is statistically significant at the 1% level for non-contributors. Tweets may then provide a way for the apolitical to separate from high anti-Black bias types.

Tweet Type. Some types of tweets are costlier to compose than others. To retweet someone else's racial justice-related tweet, an academic need only tap twice. For the three other types of tweets – original tweets, quote retweets, and replies – an academic will tend to type out their own text. If the cost of typing out racial justice-related text is higher for academics with greater bias against Black students, we might expect costly tweets to be more revealing of types than cheap retweets (Spence 1973).[34] We test this hypothesis in Table 2, decomposing the total percentage of racial justice tweets into its additive components by tweet type.[35] In addition, to ensure that informativeness is driven by the numerator (the number of racial justice tweets), rather than the denominator (the number of total tweets), we include the full set of Twitter activity levels and interactions throughout.

Columns 1 and 2 replicate our earlier result with the continuous measure. For each additional 1% of racial justice tweets (of any type), academics discriminate against Black students 1.62 percentage points less (column 1). This difference falls to 1.38 percentage points after adding the full set of controls and interactions – gender, race/ethnicity, position, university, and department (column 2). Contrary to intuition, racial justice retweets are more informative than other racial justice tweets (columns 3 and 4), although we cannot reject that the two coefficients are equal ($p = 0.48$ and $p = 0.65$).

Breaking up the racial justice tweets into their four components, we see that original tweets are actually the least predictive (columns 5 and 6), so much so that we can almost reject the

[34] Frankel and Kartik (2019) do not make this specific prediction. Their model suggests that a market-wide increase in the cost of signalling would reduce informativeness. But this result may not be true of a setting with *two* actions, one being more costly than the other.

[35] The continuous measure of racial justice signalling has an additional advantage here. In particular, a comparison of the informativeness of a dummy variable for any racial justice retweet with a dummy variable for any racial justice non-retweet is not comparing like-for-like. For example, if the mean number of racial justice retweets among those with at least one was 20, and the equivalent mean for racial justice non-retweets was 10, we would effectively be comparing the informativeness of 20 racial justice retweets with that of 10 racial justice non-retweets.

null hypothesis that original tweets are as informative as retweets ($p = 0.1$ and $p = 0.16$). Quote tweets and replies are the most informative, though the least precisely estimated. Tests of equality with original tweets have $p = 0.1$ for quote tweets and $p = 0.2$ for replies in column 5, and $p = 0.13$ and $p = 0.23$ in column 6.

The results in columns 5 and 6 suggest that cost is not the important distinguishing feature of types of tweets (with retweets being low-cost, and all other types being high-cost). Instead, the more important distinction may be *credit-taking*. Original tweets grant all credit to the tweeter, and these tweets are the least informative. Retweets and quote tweets amplify the thoughts of a different tweeter, while replies join a conversation that another tweeter started. These three types of tweets are more informative. It may then be that post types that have the most potential to boost the reputation of the poster, have the least potential to serve as credible signals of offline discriminator behavior.

4.4 What Do Students Believe?

Theorists tend to assume that communication game receivers have rational beliefs, fully understanding the equilibrium in which they find themselves (Crawford and Sobel 1982; Morris 2001; Kartik 2009; Frankel and Kartik 2019). Some models instead allow for bounded rationality, with some fraction of receivers being naïve, taking signals at face value (Ottaviani and Squintani 2006; Kartik et al. 2007; Chen 2011), or with receivers underestimating the dependence of actions on types (Eyster and Rabin 2005; Jehiel 2005; Jehiel and Koessler 2008). These assumptions about receivers have welfare implications. In particular, if students believe that tweets are uninformative (as in a special case of Eyster and Rabin (2005)), they ignore useful information when choosing advisors. We use our graduate student survey to test these theoretical assumptions.

The 515 Black and 1,237 non-Black graduate students who completed our survey are of similar age (29.5 and 29.2 years on average, Table 3), and at similarly-ranked schools (35.4 and

35.6 on average). A majority of both types of students have a Twitter account, though Black students are more likely to have one (75% versus 60%), and they are also more likely to self-report tweeting about racial justice. Politically, both sets of students skew left-wing – 77% of Black students self-identify as liberal or very liberal, 20% as moderate, and only 4% as conservative or very conservative. For non-Black students, the corresponding numbers are nearly identical: 76%, 20%, and 4%. Among the non-Black students, 58% self-identify as White or European, 34% as Asian, 1% as First Nations or Indigenous, and 11% are of Hispanic, Latino, or Spanish origin.

Predicting Discrimination. Most students overestimate how much academics discriminate against Black students: 84% predict anti-Black discrimination to be above the upper bound of our estimate's 95% confidence interval, 5% predict it to be within our confidence interval, and 11% predict it to be below its lower bound, meaning that they predict that Black students will be favored by at least 0.77 percentage points. We will call these three types of people overestimators, accurate, and underestimators, from now.

When guessing separately for Silent and Vocal academics, on average students correctly anticipate that Vocal academics discriminate less against Black students than Silent academics, but in both cases, they again tend to overestimate anti-Black discrimination. For Silent academics, 72% overestimate discrimination, while for Vocal academics, 74% overestimate. Though we find that Vocal academics discriminate in favor of Black students, 75% of students predict that Vocal academics discriminate in favor of White students.

Non-Black students predict less anti-Black discrimination than Black students, and center- or right-leaning students predict less than left-leaning students (Figure 13). Even so, the median non-Black and the median right-leaning student overestimates anti-Black discrimination.

The overestimation of anti-Black discrimination by Black students is stark given that these are the students that decided to continue with graduate studies. If post-graduation decisions are influenced by perceived discrimination, those that opted against graduate studies may overestimate anti-Black discrimination even more.

Predicting Informativeness. Students make more accurate predictions about informativeness: 29% make a guess for unconditional informativeness within our confidence interval, while 39% do so for conditional informativeness. Among the remaining guesses, students are more likely to underestimate than overestimate informativeness: 2.3 times more likely for unconditional informativeness, and 1.3 times more likely for conditional informativeness. This finding suggests that the fundamental attribution error (Jones and Harris 1967; Andre 2021) – the overattribution of behavior (e.g. tweets) to personality traits (e.g. racial bias) rather than situational factors (e.g. signalling incentives) – is not the key source of biased beliefs in our context.

Following Frankel and Kartik (2019), the higher the fraction of skeptical receivers, the lower the incentives to signal, and the higher informativeness should be.[36] Our finding that a substantial fraction of the audience discounts social media posts is then one possible rationalization of why the equilibrium itself is highly informative about discrimination.[37]

Unlike predictions about discrimination, predictions about informativeness are similar by race (Figure 13), although race does matter for predictions after conditioning on controls, as we discuss below. In addition, despite our prior that 'virtue signaller' is a pejorative used more by the political right, liberals and moderates make similar predictions about informativeness.[38]

Predicting Predictions. We use regressions in Table 4 to explore more carefully what factors

[36] In particular, in Frankel and Kartik (2019), if a fraction of the audience always ignore signals, one can redefine the relevant audience as those that are sophisticated. Making slightly different assumptions, Eyster and Rabin (2005) show that informativeness is impossible when receivers fully neglect the dependence of tweets on types. The key assumption driving this is that the lowest cost action is to not tweet at all. This assumption ensures that no one tweets when the audience is fully incredulous. We find the assumption in Frankel and Kartik (2019) to be more plausible in our setting – that the lowest cost action is to tweet in line with your racial bias type. In this case, a fully incredulous audience ensures a fully separating equilibrium.

[37] This presumes that senders are aware of the skepticism of many receivers. Future research could explore this by eliciting the second order beliefs of senders.

[38] While we explicitly asked the graduate students to make predictions for the top-150 universities (Appendix E), students may have instead substituted an easier question, and made predictions for their own top-80 universities. Encouragingly, our conclusions are unchanged if we instead use the audit results for only the top-80 universities to benchmark the student predictions (Figure A28).

explain the variation in predictions.

On average, students predict meeting requests from distinctively White senders to be 11 percentage points more likely to be accepted than those from Black senders (columns 1 to 3). Black students predict two or three percentage points more anti-Black discrimination, while a one-point increase in conservative political views is associated with a prediction of three percentage points less anti-Black discrimination. The average prediction of conservatives (who score four out of five on a five-point politics scale) is more accurate than the average prediction of each of the four other political groups.

Otherwise, graduate students born in the USA predict two percentage points more anti-Black discrimination (columns 2 and 3), as do those randomly assigned to receive prediction incentives. The latter suggests that students might predict us to be biased toward finding evidence of discrimination. In contrast, prediction incentives do not influence predictions of informativeness (columns 5, 6, 8, and 9), perhaps because it is less obvious what informativeness result we would be biased toward finding.

Four factors are statistically significantly associated with predicted unconditional informativeness: Black students predict one percentage point less informativeness (columns 4 to 6), non-binary students predict three percentage points less informativeness (relative to women), and upper-year students and those from lower-ranked universities predict more informativeness.

While one hypothesis would be that sophistication comes with experience (List 2003), we find no evidence that Twitter users or upper-year students make more accurate predictions about informativeness (columns 8 and 9). And though the average conservative makes more accurate predictions about levels of discrimination, conservatives are less accurate when it comes to predicting informativeness. These findings are for the most part similar when considering predictions about conditional informativeness (Table A2), though as mentioned in Section 4.4, we emphasize the unconditional predictions, given that the conditional prediction question is more difficult to parse.

5 Conclusion and Future Directions

This paper characterizes the informativeness of tweets about racial justice. Academics that tweet about racial justice are less biased against Black students, and more biased in favor of first-generation and female students. These vocal academics are also far more likely to take political action – contributing more often to Democratic candidates – but are no more likely to contribute to Raphael Warnock over Jon Ossoff. However, they do tweet more often about Warnock than Ossoff, suggesting that racial justice tweets are more informative of more visible behaviors.

The predictive value of tweets for discriminatory behaviors suggests an underemphasized informational benefit of social media: students can use tweets to facilitate sorting towards more responsive academics. In practice, we suspect that these sorting gains are not fully realized, given that most of our surveyed students mispredict the informativeness of tweets.

Several questions merit further study. First, informativeness may differ in other populations. In ongoing work, we explore corporate virtue signalling, a context in which employees that post tweets have little broader decision-making power. In preliminary results, we find that UK firms that tweet about International Women's Day have similar gender pay gaps to those that do not – a babbling equilibrium. Second, informativeness may change if informativeness is publicized. For example, if anti-Black Silent academics react to our paper by tweeting about racial justice, and unbiased Silent academics remain silent, informativeness would fall. Future work could explore this by randomizing information about equilibrium informativeness and measuring subsequent social media activity. Third, work could explore the source of student misperceptions. One hypothesis is that hypocrisy is more memorable than consistency, leading people to underestimate the informativeness of the equilibrium. Fourth, work could correct student misperceptions in the spirit of Bursztyn et al. (2020a), and evaluate effects on advisor choice and related behaviors. Finally, while our results speak to the informativeness of racial justice tweets, they do not speak to the motivations for tweeting about racial justice. In particular, our findings do not rule out the possibility that Vocal academics tweet about racial justice to bolster their social image. Experiments could isolate the motives for signalling by creating variation in the visibility of tweets, and by tracking subsequent tweet activity.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for truth-telling," *Econometrica*, 2019, *87* (4), 1115–1153.

Ajzenman, Nicolás, Bruno Ferman, and Pedro C Sant'Anna, "Discrimination in the Formation of Academic Networks: A Field Experiment on# EconTwitter," 2023.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow, "The welfare effects of social media," *American Economic Review*, 2020, *110* (3), 629–76.

Andre, Peter, "Shallow meritocracy: An experiment on fairness views," Technical Report, ECONtribute Discussion Paper 2021.

Austen-Smith, David, "Information transmission in debate," *American Journal of political science*, 1990, pp. 124–152.

Becker, Gary Stanley, *The Economics of Discrimination*, University of Chicago Press, 1957.

Bertrand, Marianne and Sendhil Mullainathan, "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, September 2004, *94* (4), 991–1013.

Blume, Andreas, Ernest K Lai, and Wooyoung Lim, "Strategic information transmission: A survey of experiments and theoretical foundations," *Handbook of experimental game theory*, 2020, pp. 311–347.

BOA, "Twitter Revenue and Usage Statistics 2022," Technical Report, Business of Apps, https://www.businessofapps.com/data/twitter-statistics/ 2022.

Bouton, Laurent, Julia Cagé, Edgard Dewitte, and Vincent Pons, "Small Campaign Donors," Technical Report, National Bureau of Economic Research 2022.

Braghieri, Luca, “Political correctness, social image, and information transmission,” *Work. Pap., Stanford Univ., Stanford, CA*, 2021.

__ **, Ro’ee Levy, and Alexey Makarin**, “Social media and mental health,” *American Economic Review*, 2022, *112* (11), 3660–93.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, “Misperceived social norms: Women working outside the home in saudi arabia,” *American economic review*, 2020, *110* (10), 2997–3029.

__ **, Georgy Egorov, and Stefano Fiorin**, “From extreme to mainstream: The erosion of social norms,” *American economic review*, 2020, *110* (11), 3522–48.

__ **, Thomas Fujiwara, and Amanda Pallais**, “’Acting Wife’: Marriage Market Incentives and Labor Market Investments,” *American Economic Review*, 2017, *107* (11), 3288–3319.

Cai, Hongbin and Joseph Tao-Yi Wang, “Overcommunication in strategic information transmission games,” *Games and Economic Behavior*, 2006, *56* (1), 7–36.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri, “Gender gaps at the academies,” *Proceedings of the National Academy of Sciences*, 2023, *120* (4), e2212421120.

Charles, Kerwin Kofi and Jonathan Guryan, “Prejudice and wages: an empirical assessment of Becker’s The Economics of Discrimination,” *Journal of political economy*, 2008, *116* (5), 773–809.

Chen, Ying, “Perturbed communication games with honest senders and naive receivers,” *Journal of Economic Theory*, 2011, *146* (2), 401–424.

Crawford, Vincent P and Joel Sobel, “Strategic information transmission,” *Econometrica: Journal of the Econometric Society*, 1982, pp. 1431–1451.

DellaVigna, Stefano, Devin Pope, and Eva Vivalt, "Predict science to improve science," *Science*, 2019, *366* (6464), 428–429.

Eyster, Erik and Matthew Rabin, "Cursed equilibrium," *Econometrica*, 2005, *73* (5), 1623–1672.

Frankel, Alex and Navin Kartik, "Muddled information," *Journal of Political Economy*, 2019, *127* (4), 1739–1776.

Gaddis, S Michael, Edvard Larsen, Charles Crabtree, and John Holbein, "Discrimination against black and Hispanic Americans is highest in hiring and housing contexts: A meta-analysis of correspondence audits," *Available at SSRN 3975770*, 2021.

Gneezy, Uri, Alex Imas, and Kristóf Madarász, "Conscience accounting: Emotion dynamics and social behavior," *Management Science*, 2014, *60* (11), 2645–2658.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing Information Provision Experiments," *Journal of Economic Literature*, forthcoming.

Jehiel, Philippe, "Analogy-based expectation equilibrium," *Journal of Economic theory*, 2005, *123* (2), 81–104.

_ **and Frédéric Koessler**, "Revisiting games of incomplete information with analogy-based expectations," *Games and Economic Behavior*, 2008, *62* (2), 533–557.

Jelveh, Zubin, Bruce Kogut, and Suresh Naidu, "Political language in economics," *Columbia Business School Research Paper*, 2018, (14-57).

Jones, Edward E. and Victor A. Harris, "The attribution of attitudes," *Journal of Experimental Social Psychology*, 1967, *3* (1), 1–24.

Karing, Anne, "Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone," *Working Paper, University of California, Berkeley*, 2023.

Kartik, Navin, "Strategic communication with lying costs," *The Review of Economic Studies*, 2009, *76* (4), 1359–1395.

_ , **Marco Ottaviani, and Francesco Squintani**, "Credulity, lies, and costly talk," *Journal of Economic theory*, 2007, *134* (1), 93–116.

Kessler, Judd B, Corinne Low, and Colin D Sullivan, "Incentivized resume rating: Eliciting employer preferences without deception," *American Economic Review*, 2019, *109* (11), 3713–44.

Kirgios, Erika L, Aneesh Rai, Edward H Chang, and Katherine L Milkman, "When seeking help, women and racial/ethnic minorities benefit from explicitly stating their identity," *Nature Human Behaviour*, 2022, *6* (3), 383–391.

Kline, Patrick, Evan K Rose, and Christopher R Walters, "Systemic discrimination among large US employers," *The Quarterly Journal of Economics*, 2022, *137* (4), 1963–2036.

Langbert, Mitchell, Anthony J Quain, Daniel B Klein et al., "Faculty voter registration in economics, history, journalism, law, and psychology," *Econ Journal Watch*, 2016, *13* (3), 422–451.

List, John A, "Does market experience eliminate market anomalies?," *The Quarterly Journal of Economics*, 2003, *118* (1), 41–71.

Loury, Glenn C, "Self-censorship in public discourse: A theory of "political correctness" and related phenomena," *Rationality and Society*, 1994, *6* (4), 428–461.

Mazar, Nina and Chen-Bo Zhong, "Do green products make us better people?," *Psychological science*, 2010, *21* (4), 494–498.

Milkman, Katherine L, Modupe Akinola, and Dolly Chugh, "Temporal distance and discrimination: An audit study in academia," *Psychological science*, 2012, *23* (7), 710–717.

Morris, Stephen, "Political correctness," *Journal of political Economy*, 2001, *109* (2), 231–265.

Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie, "The economic effects of Facebook," *Experimental Economics*, 2020, *23* (2), 575–602.

Ottaviani, Marco and Francesco Squintani, "Naive audience and communication bias," *International Journal of Game Theory*, 2006, *35* (1), 129–150.

Pager, Devah and Lincoln Quillian, "Walking the talk? What employers say versus what they do," *American Sociological Review*, 2005, *70* (3), 355–380.

PRC, "Social Media Use in 2021," Technical Report, Pew Research Center, Washington, D.C., https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ 2021.

Reny, Tyler T and Benjamin J Newman, "The opinion-mobilizing effect of social protest against police violence: Evidence from the 2020 George Floyd protests," *American Political Science Review*, 2021, *115* (4), 1499–1507.

Schultz, Robert, Anna Stansbury et al., "Socioeconomic diversity of economics PhDs," Technical Report 2022.

Schwardmann, Peter, Egon Tripodi, and Joël J Van der Weele, "Self-persuasion: Evidence from field experiments at international debating competitions," *American Economic Review*, 2022, *112* (4), 1118–46.

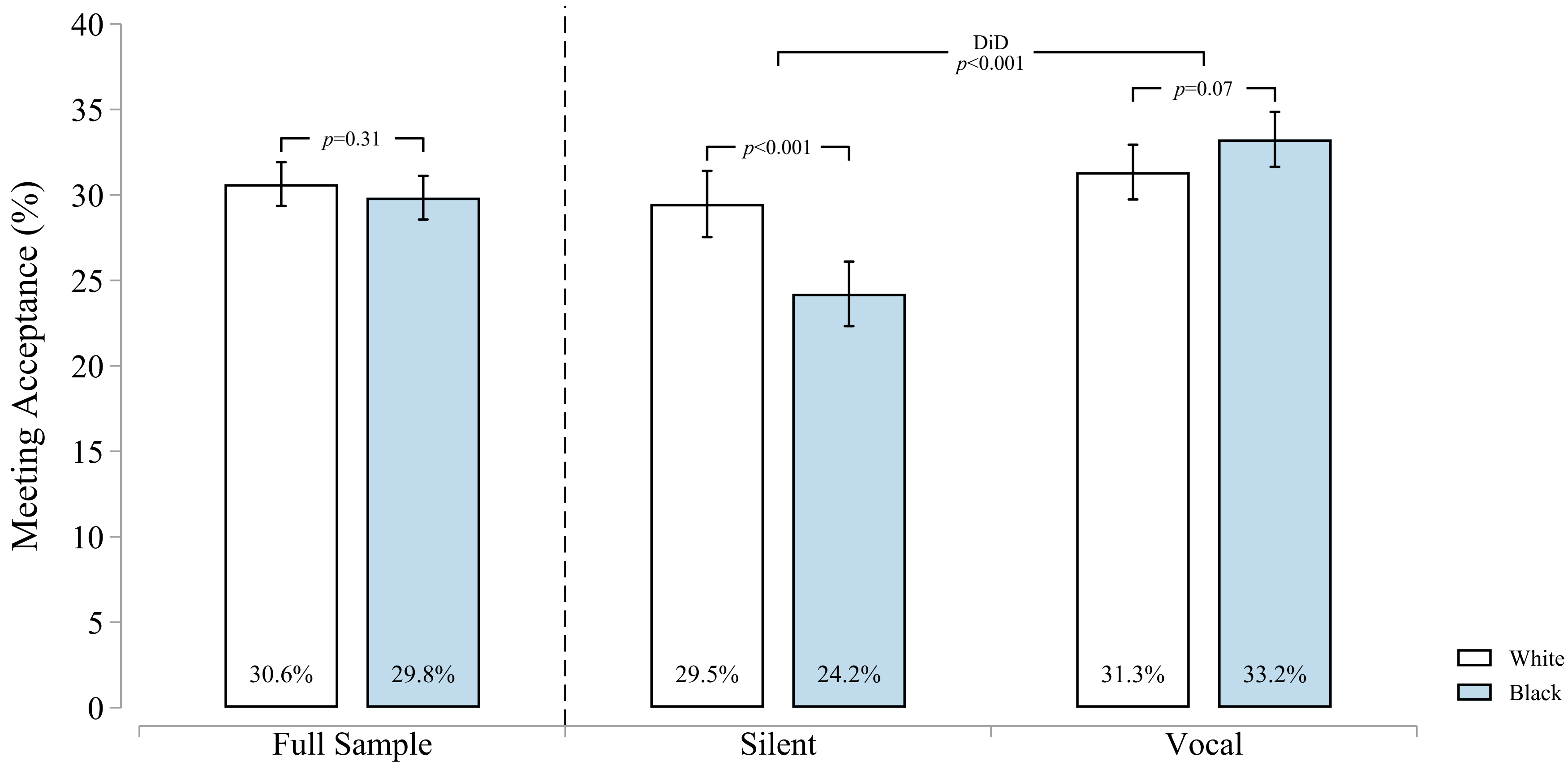
Song, Lena, "Essays on Information Technology and Inequality," 2022.

Spence, Michael, "Job Market Signaling," *The Quarterly Journal of Economics*, 1973, *87* (3), 355–374.

Tyler, John H, Richard J Murnane, and John B Willett, "Estimating the labor market signaling value of the GED," *The Quarterly Journal of Economics*, 2000, *115* (2), 431–468.

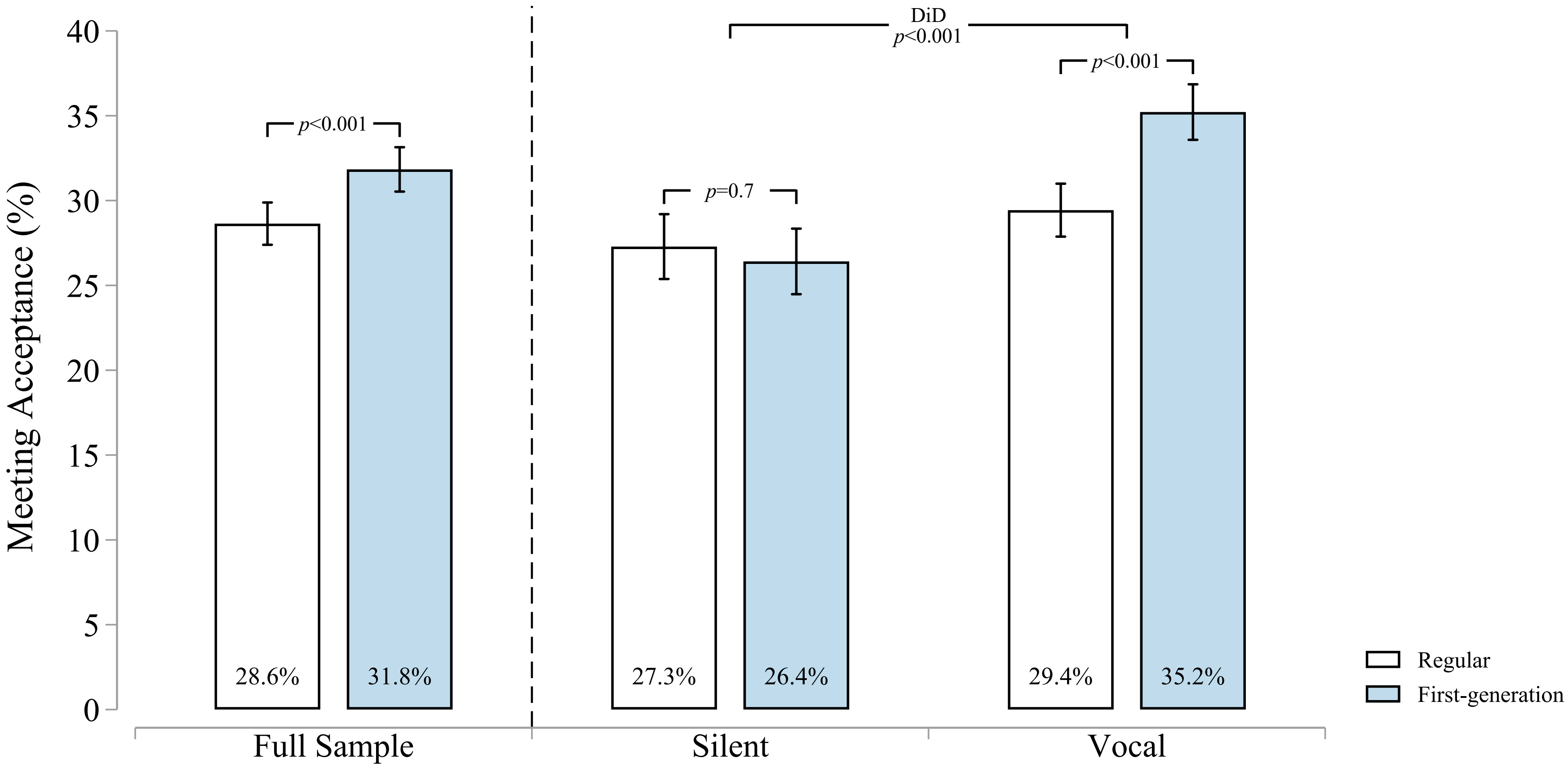
Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov, "Political effects of the internet and social media," *Annual review of economics*, 2020, *12*, 415–438.

Figure 1: Silent Academics Discriminate Against Black Students, Vocal Academics Discriminate (Somewhat) Against White Students



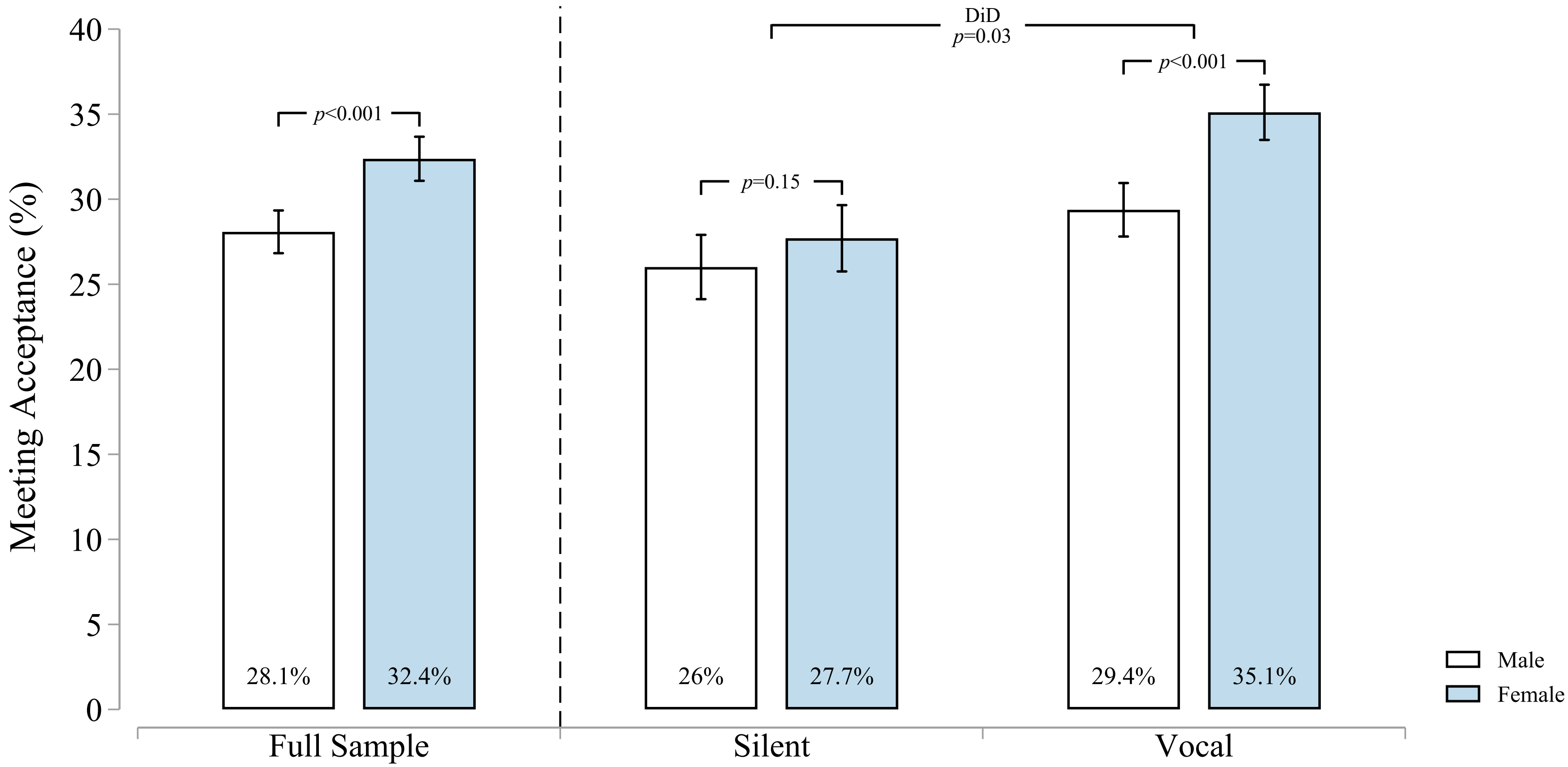
Notes: The bars show what percentage of audited academics accepted meeting requests from distinctively White and distinctively Black names. The Full Sample includes the 11,450 audited academics (4,318 Silent and 7,132 Vocal). Vocal academics are those that tweeted at least once about racial justice from January 2020 to March 2022. Silent academics are those that did not. The raw means and 95% confidence intervals come from a regression of Accepted_i on dummy variables for White and Black email sender (to the left of the vertical dashed line), and a regression on dummy variables for White email sender to Silent academic, White email sender to Vocal academic, and the same for Black email sender (to the right of the vertical dashed line). The p-values come from the specification that also includes strata and email type fixed effects. The DiD (diff-in-diff) p-value is from a test for equal discrimination rates across Vocal and Silent academics (γ_2 in specification 2). Standard errors are clustered at the university-by-department-by-sender name-level.

Figure 2: Vocal Academics Favor First-Generation Students, Silent Academics Show No Bias



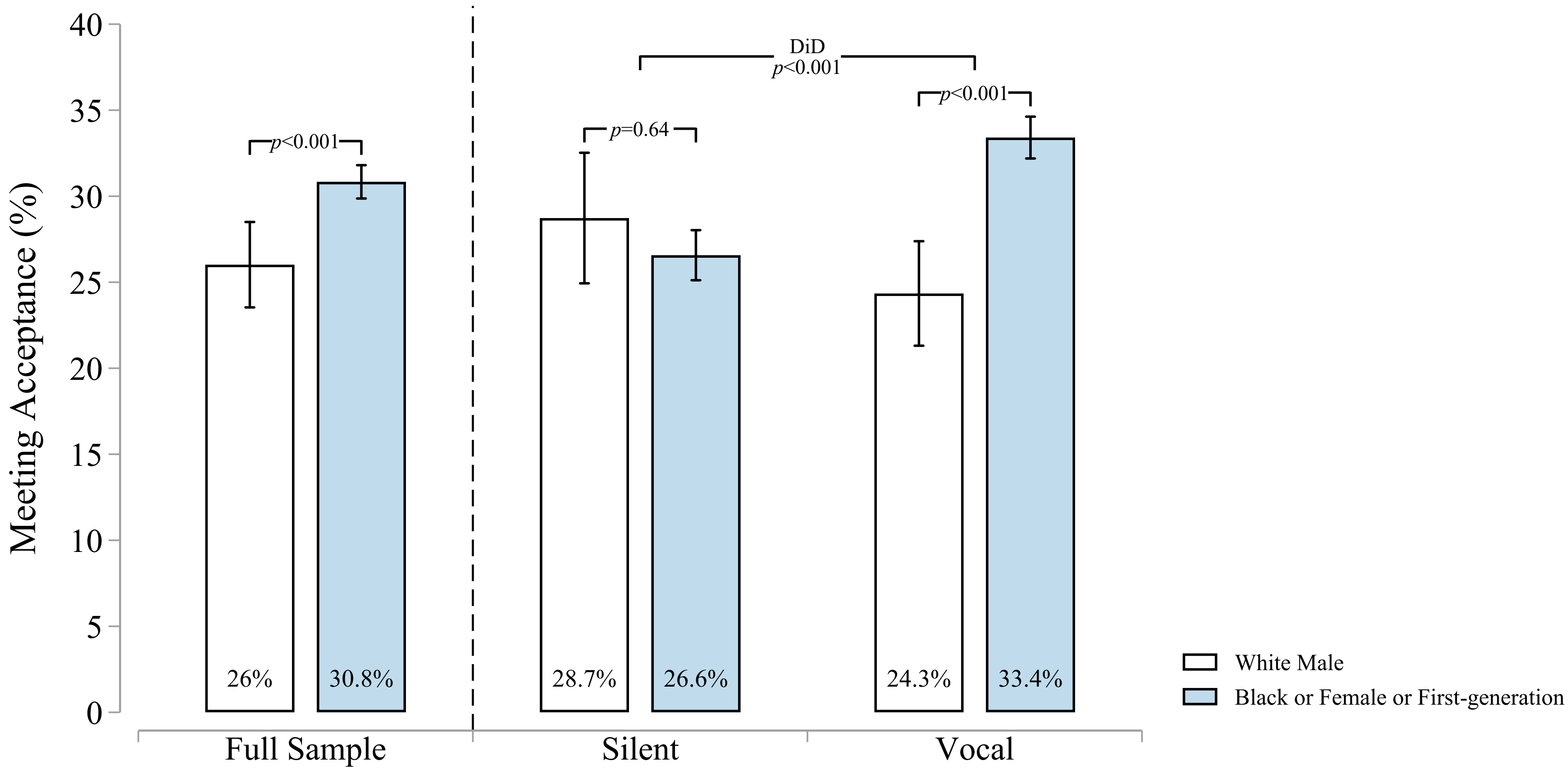
Notes: The bars show what percentage of audited academics accepted meeting requests from students that mention their first-generation status and students that did not ("Regular" students). The Full Sample includes the 11,450 audited academics (4,318 Silent and 7,132 Vocal). Vocal academics are those that tweeted at least once about racial justice from January 2020 to March 2022. Silent academics are those that did not. The raw means and 95% confidence intervals come from a regression of $Accepted_i$ on dummy variables for first-generation student and regular student email sender (to the left of the vertical dashed line), and a regression on dummy variables for first-generation email sender to Silent academic, first-generation email sender to Vocal academic, and the same for Black email sender (to the right of the vertical dashed line). The p-values come from the specification that also includes strata and email type fixed effects. The DiD (diff-in-diff) p-value is from a test for equal discrimination rates across Vocal and Silent academics (γ_2 in specification 2). Standard errors are clustered at the university-by-department-by-sender name-level.

Figure 3: Vocal Academics Favor Female Students More Than Silent Academics



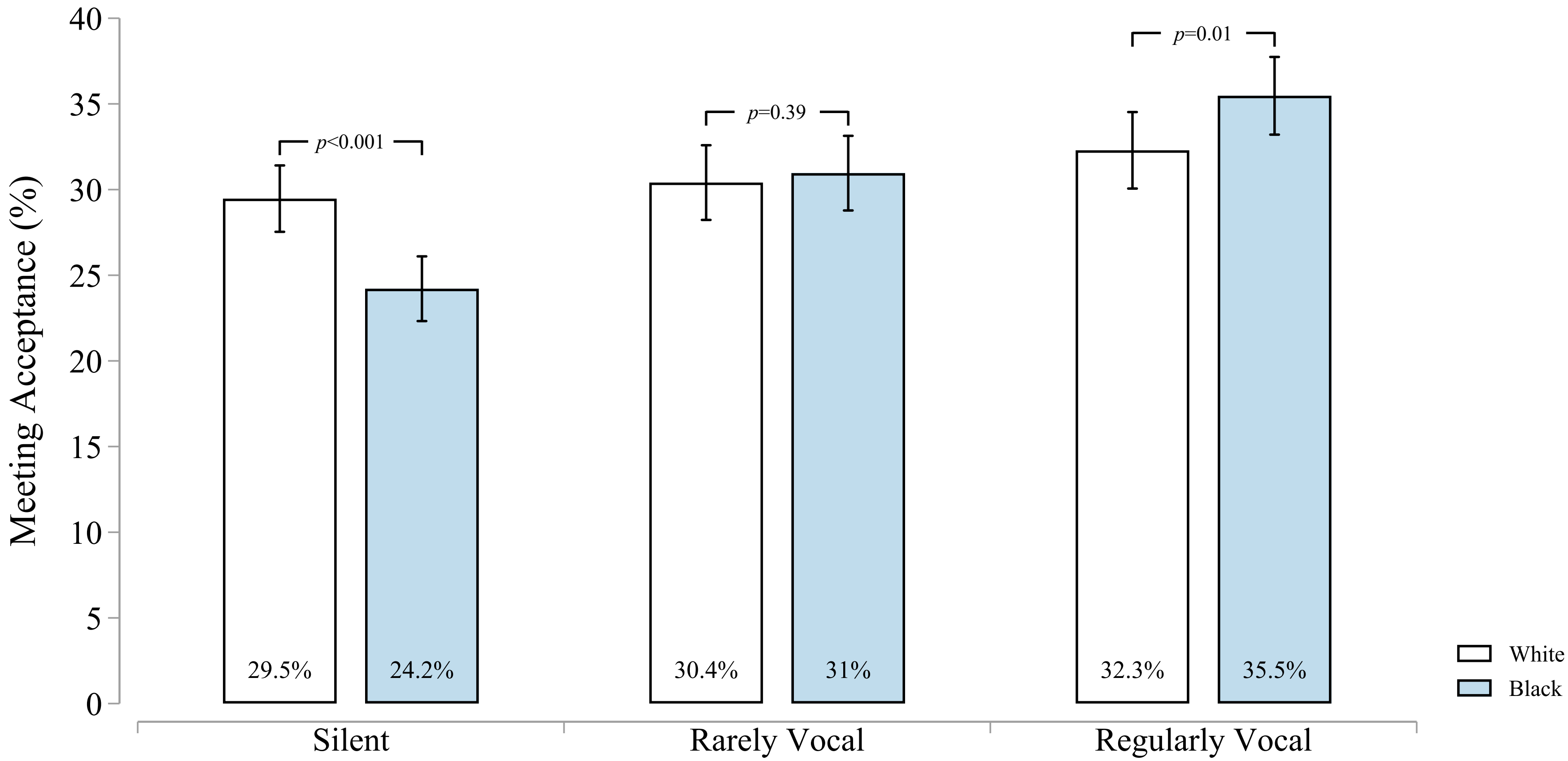
Notes: The bars show what percentage of audited academics accepted meeting requests from distinctively male and distinctively female names. The Full Sample includes the 11,450 audited academics (4,318 Silent and 7,132 Vocal). The raw means and 95% confidence intervals come from a regression of Accepted_i on dummy variables for female and male email sender (to the left of the vertical dashed line), and a regression on dummy variables for female email sender to Silent academic, female email sender to Vocal academic, and the same for Black email sender (to the right of the vertical dashed line). The p-values come from the specification that also includes strata and email type fixed effects. The DiD (diff-in-diff) p-value is from a test for equal discrimination rates across Vocal and Silent academics (γ_2 in specification 2). Standard errors are clustered at the university-by-department-by-sender name-level.

Figure 4: Vocal Academics Discriminate Against White Males



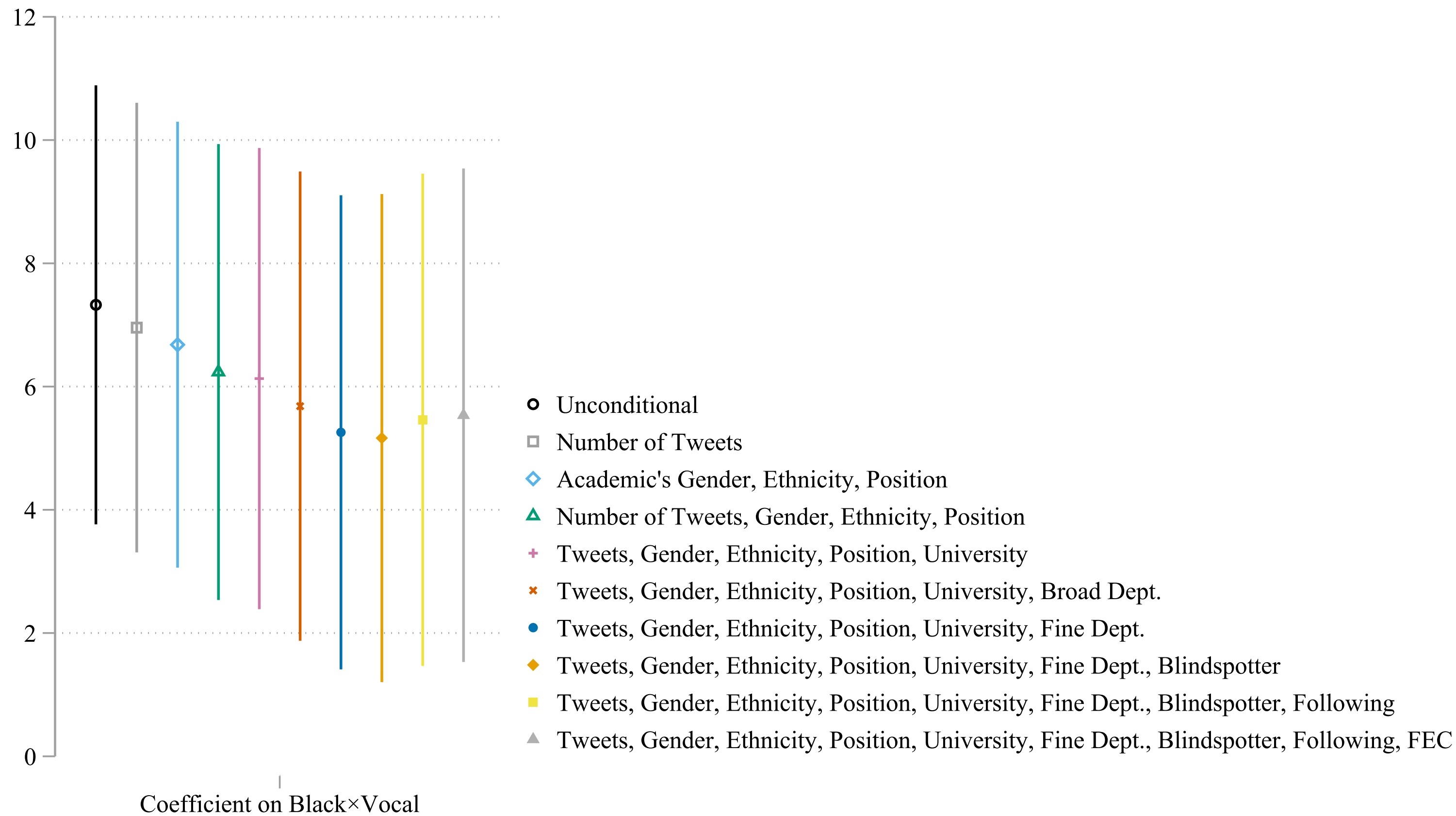
Notes: The bars show what percentage of audited academics accepted meeting requests from distinctively White male names with no mention of first-generation status (1/8 of emails) vs. emails from distinctively Black names or female names, or emails that mention first-generation status (7/8 of emails). The Full Sample includes the 11,450 audited academics (4,318 Silent and 7,132 Vocal). The raw means and 95% confidence intervals come from a regression of Accepted_i on dummy variables for White male and Black/female/first-generation email sender (to the left of the vertical dashed line), and a regression on dummy variables for White male email sender to Silent academic, White male email sender to Vocal academic, and the same for Black/female/first-generation email sender (to the right of the vertical dashed line). The p-values come from the specification that also includes strata and email type fixed effects. The DiD (diff-in-diff) p-value is from a test for equal discrimination rates across Vocal and Silent academics (γ_2 in specification 2). Standard errors are clustered at the university-by-department-by-sender name-level.

Figure 5: The Rarely Vocal Are Unbiased, the Regularly Vocal Favor Black Students



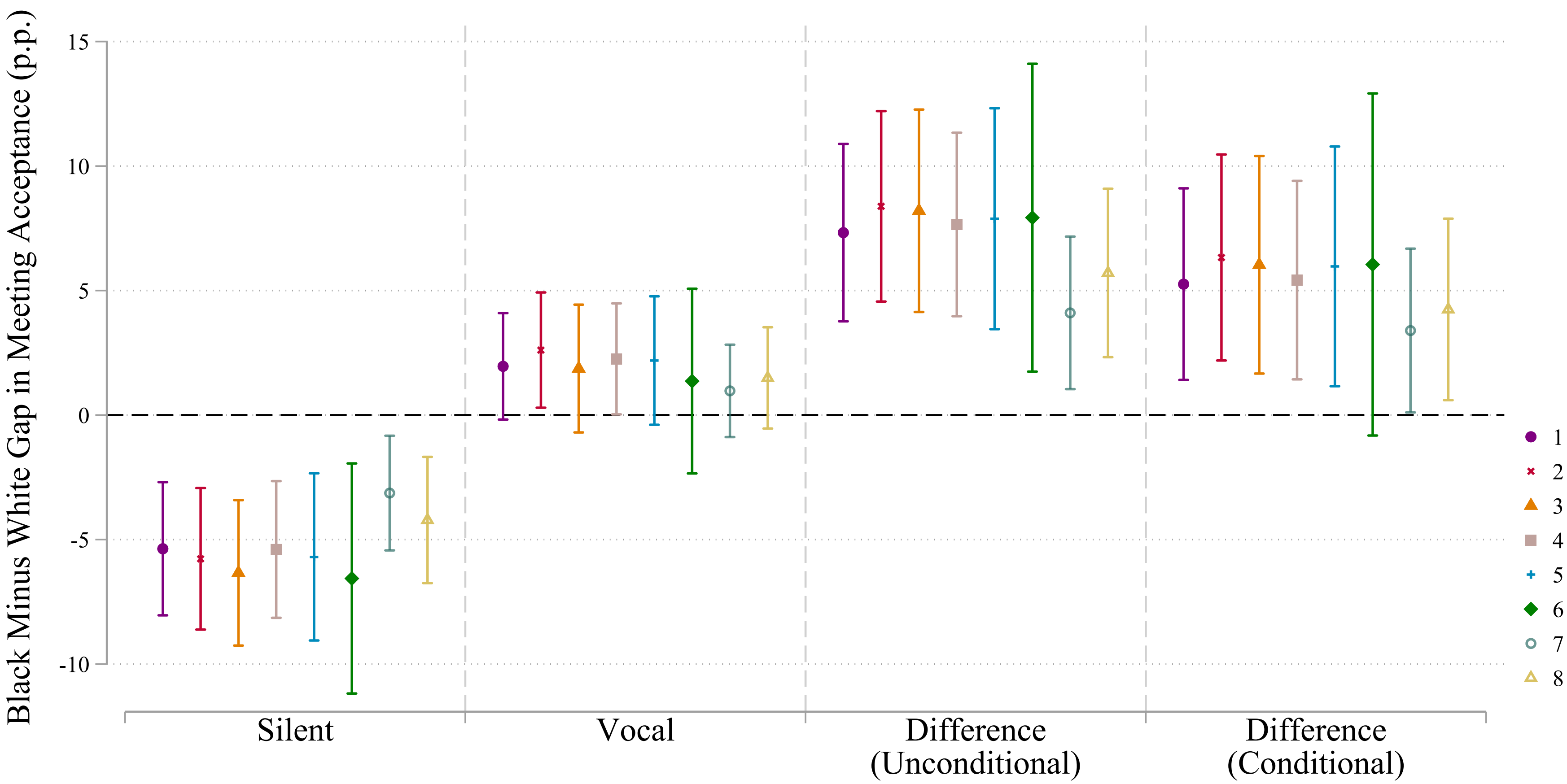
Notes: The bars show what percentage of audited academics accepted meeting requests from distinctively White and distinctively Black names. Silent academics are those that did not tweet about racial justice from January 2020 to March 2022. The bars for the Silent academics replicate the findings in Figure 1. Among the Vocal, the Rarely Vocal are those with below-median percentage of tweets from January 2020 to March 2020 that are about racial justice (0.6% on average), while the Regularly Vocal academics are above-median (3.9% on average). The raw means and 95% confidence intervals come from a regression of Accepted_i on dummy variables for Black email sender to Silent academic, White email sender to Silent academic, and the same for emailed to the Rarely Vocal and to the Regularly Vocal. The p-values come from the specification that also includes strata and email type fixed effects. Standard errors are clustered at the university-by-department-by-sender name-level.

Figure 6: Vocal Academics Discriminate Less Against Black Students, Even After Adding Controls



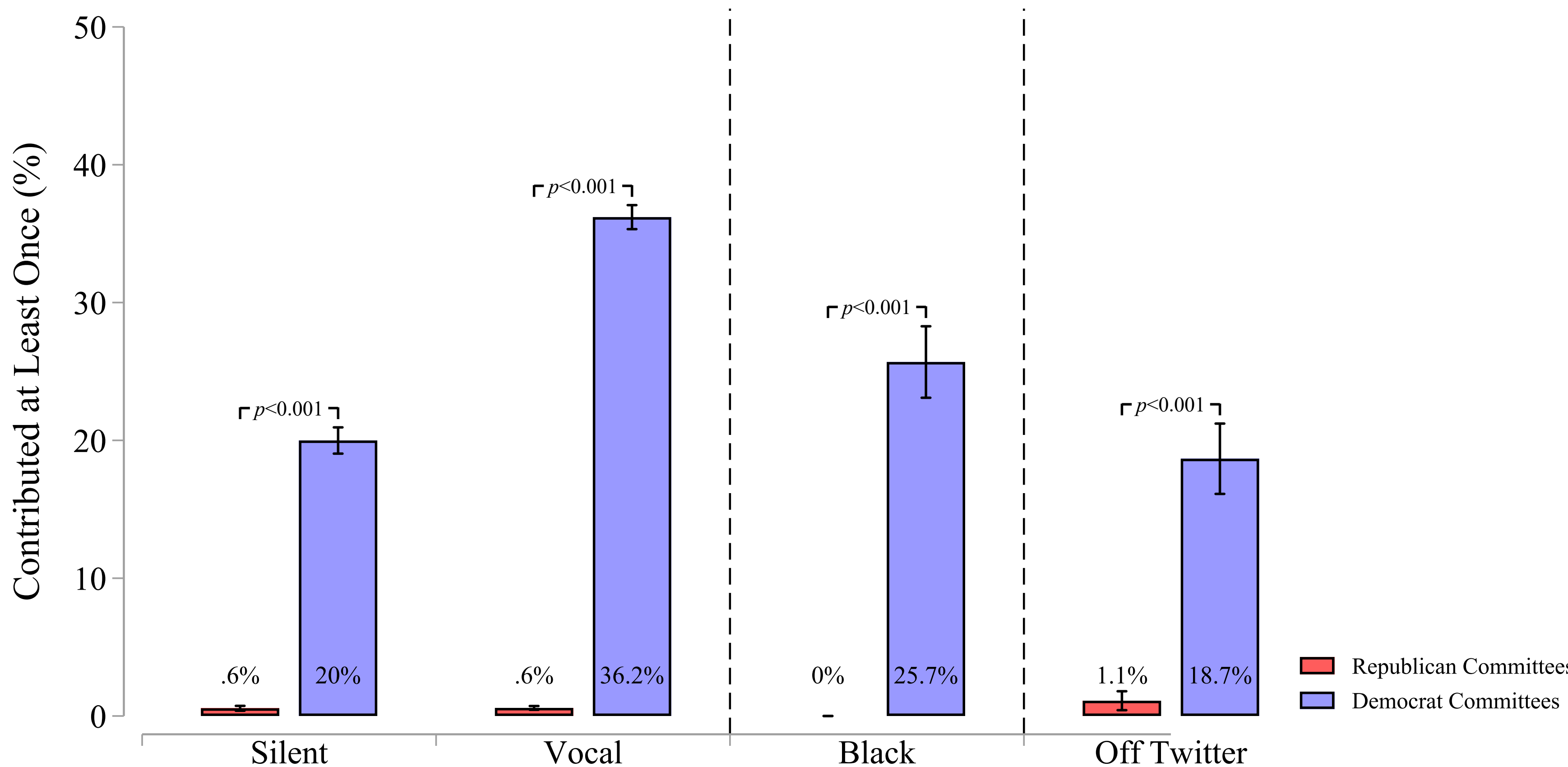
Notes: The figure shows how the difference in discrimination between Vocal and Silent academics ($\hat{\gamma}_2$ from specification 2) changes as the set of X_i^j covariates change. $\hat{\gamma}_2 > 0$ indicates that Vocal academics discriminate against Black students less than Silent academics. The covariates are: (1) Number of Tweets: the number of original tweets, reply tweets, retweets, quote tweets, and quote reply tweets, all for the period January 1, 2020 to March 27, 2022; (2) dummy variable for female, dummy variables for Assistant Professor and Associate Professor, dummy variables for race/ethnicity; (3) includes both the tweet variables from (1) and the demographics variables from (2); (4) adds university fixed effects; (5) adds broad department dummy variables (seven departments, e.g. Social Sciences); (6) replaces broad departments with narrowly defined department dummy variables (75 departments, e.g. Economics); (7) adds Blindspotter-measured percentage left, percentage center (percentage right is omitted), dummy variable for missing because profile is private, dummy variable for missing because of insufficient content; (8) adds number of political accounts followed, percentage of political accounts followed that are Democrats, dummy variable for at least one political account followed, dummy variable for follow zero accounts, and dummy variable for missing following data (e.g. because profile is private); and (9) adds dummy variable for contributed to Democrats, dummy variable for contributed to Republicans, total contributions to Democrats, and total contributions to Republicans, all for the period January 1, 2020 to March 27, 2022. Standard errors are clustered at the university-by-department-by-sender name-level. 95% confidence intervals are shown.

Figure 7: Detection Is Unlikely to Explain the Results



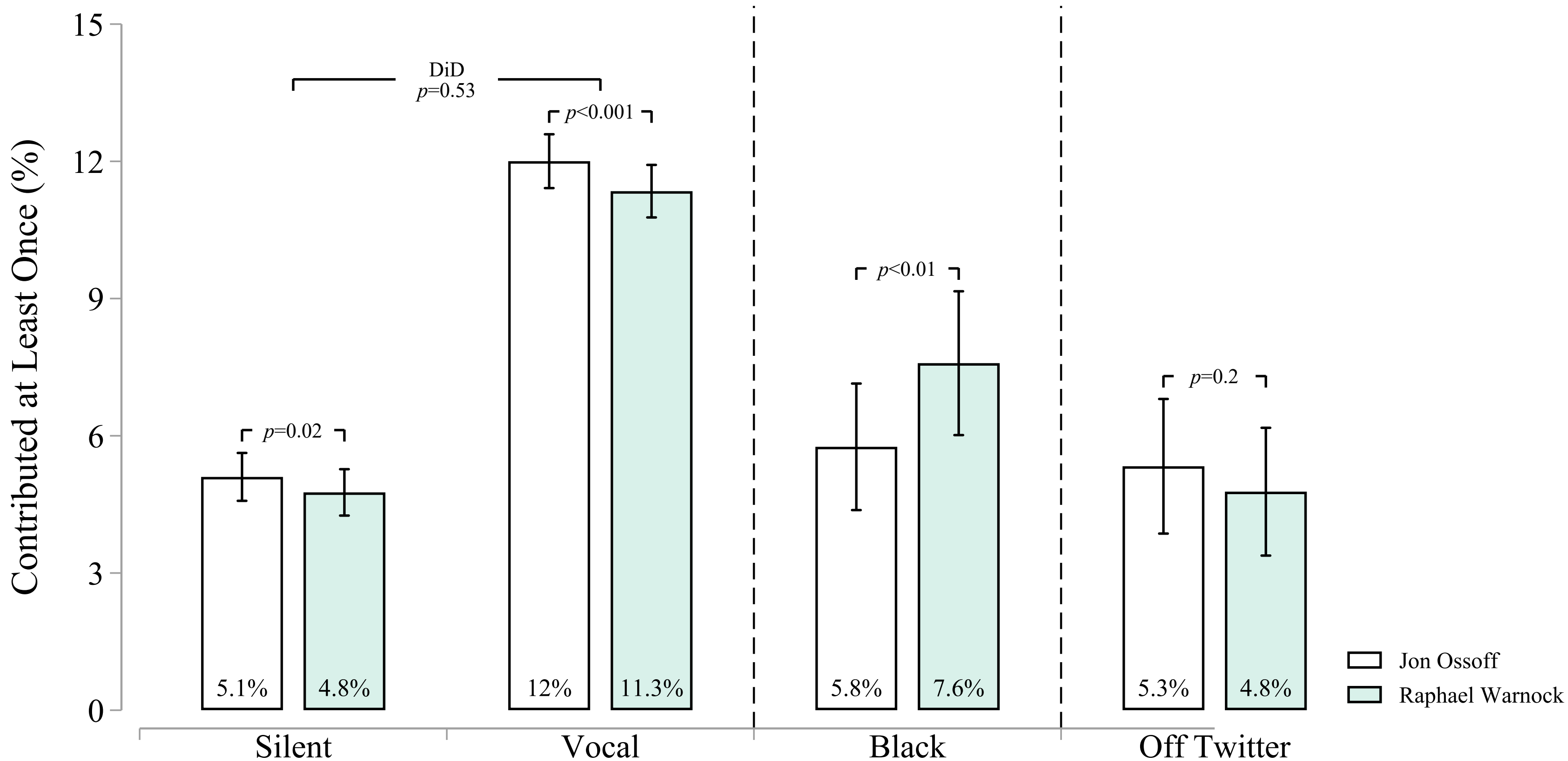
Notes: The figure shows the robustness of our main results (coefficient set 1) to seven alternative samples or outcomes where we expect audit detection to be less likely (coefficients sets 2 to 8). Specifically, for each of the eight we show from left to right: the racial gap in meeting acceptance for Silent academics (negative for discrimination against Black students), the racial gap for Vocal academics, the unconditional difference in the racial gap ($\hat{\gamma}_2$ from specification 2 without any X_i^j covariates, positive if Vocal academics discriminate against Black students less than Silent academics), and the conditional difference using the fourth-from-the-right specification from Figure 6. The specification and sample variants are (percentage of the sample dropped in parentheses): (2) drop academics in Economics, Political Science, Sociology, and Business (12.4%), (3) drop academics in the Social Sciences (25%), (4) drop academics to whom we sent more generic emails (7%), (5) drop university-departments to which we sent more than ten emails (27%), (6) drop university-departments to which we sent more than five emails (61%), (7) outcome is meeting accepted within one day, and (8) outcome is meeting accepted within three days. Standard errors are clustered at the university-by-department-by-sender name-level. 95% confidence intervals are shown.

Figure 8: Academics Almost Never Contribute to Republicans, Vocal Academics Are More Active



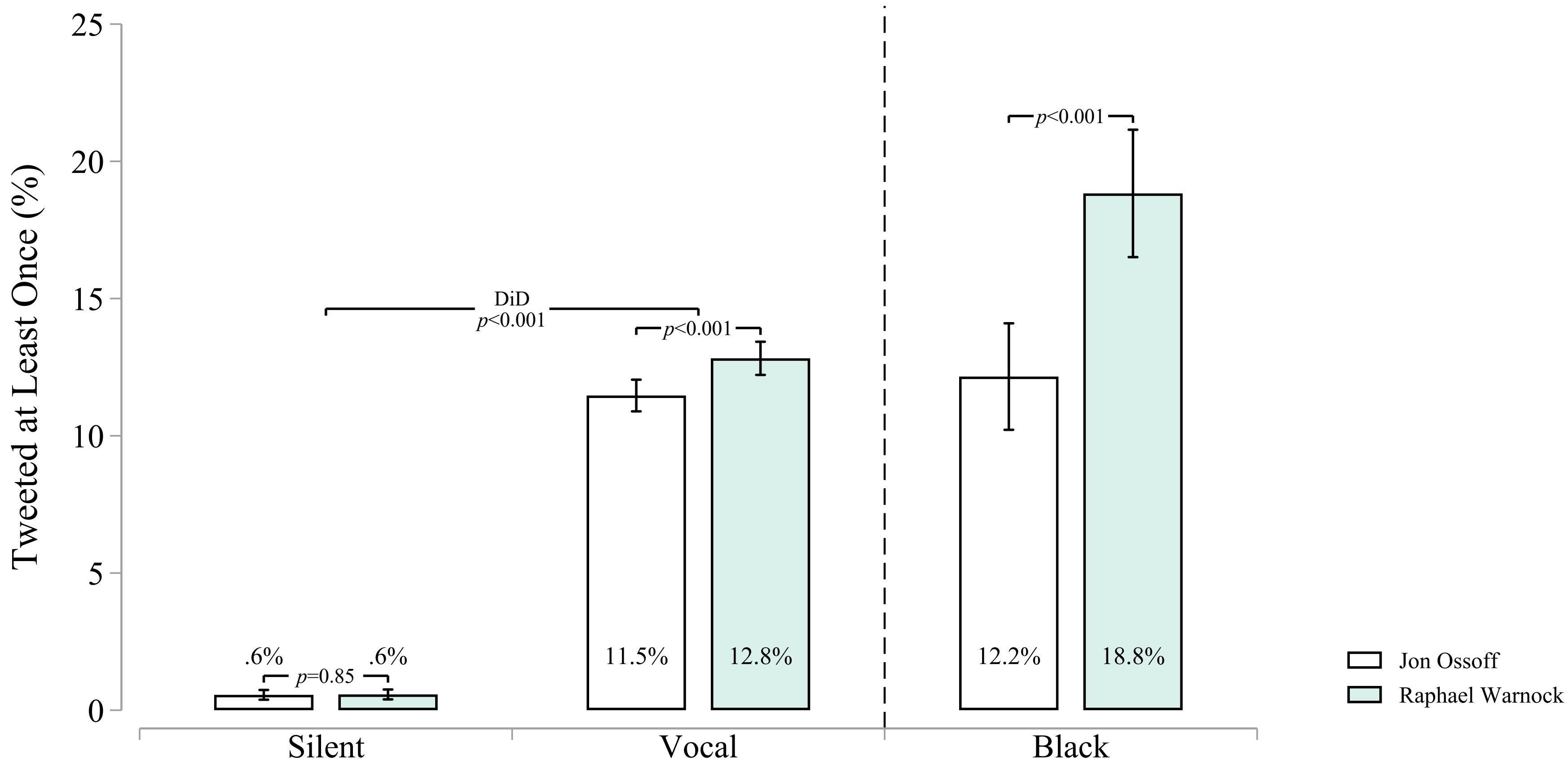
Notes: The bars show what percentage of academics made FEC-reported political contributions to Republican- and Democrat-related committees from January 2020 to March 2022. Silent includes the 6,784 non-Black academics that did not tweet about racial justice during the same time period, Vocal includes the 11,730 non-Black academics that did tweet about racial justice, Black includes the 1,094 tweeting Black academics, and Off Twitter includes the random sample of 900 non-Black academics without Twitter accounts. Unconditional raw means with 95% confidence intervals are shown.

Figure 9: Vocal Academics Are No More Likely to Give to Warnock Over Ossoff Than Silent Academics



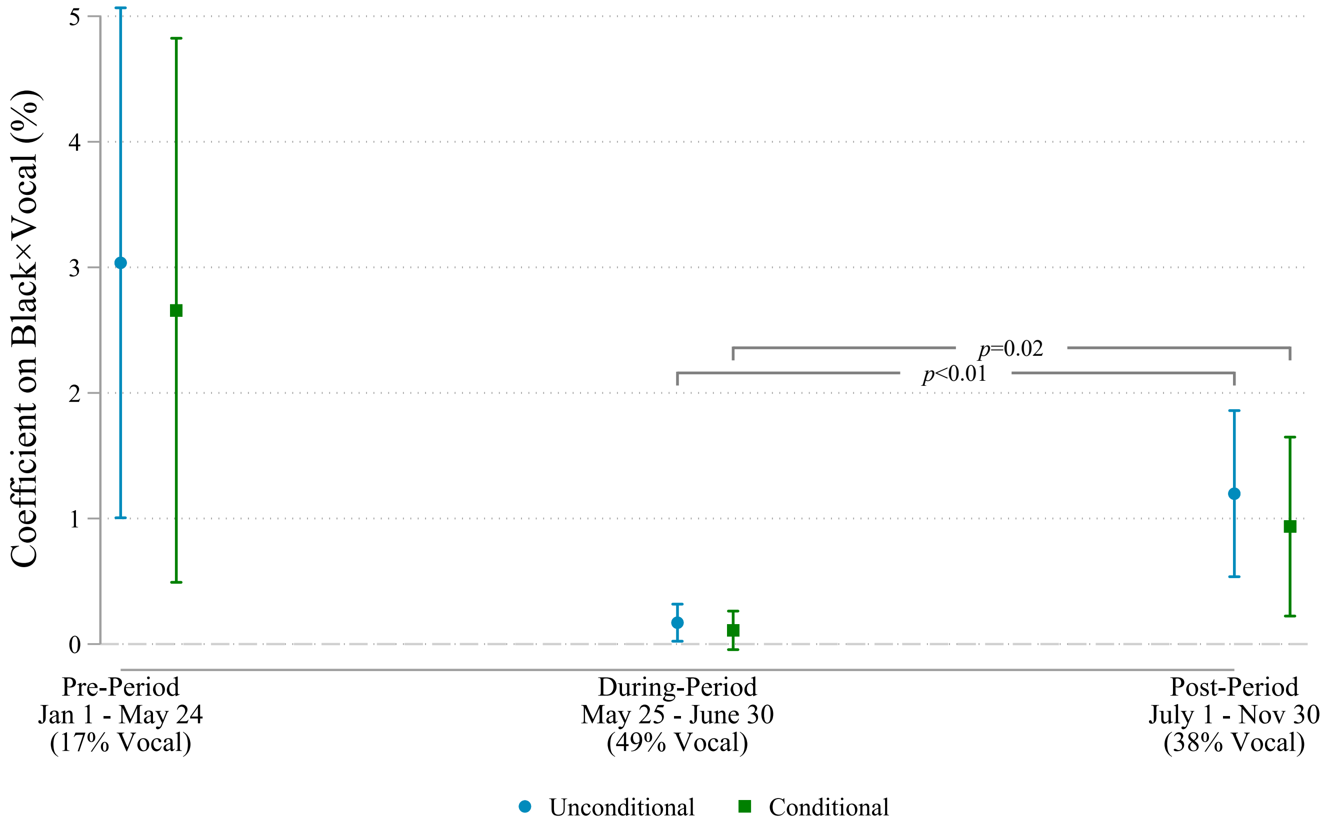
Notes: The bars show what percentage of academics made FEC-reported political contributions to the Senate campaigns of Jon Ossoff (a White Democratic candidate in Georgia) and Raphael Warnock (a Black Democratic candidate in Georgia) from January 2020 to March 2022. Silent includes the 6,784 non-Black academics that did not tweet about racial justice during the same time period, Vocal includes the 11,730 non-Black academics that did tweet about racial justice, Black includes the 1,094 tweeting Black academics, and Off Twitter includes the random sample of 900 non-Black academics without Twitter accounts. Unconditional raw means with 95% confidence intervals are shown.

Figure 10: Vocal Academics Are More Likely to Tweet About Warnock Over Ossoff Than Silent Academics



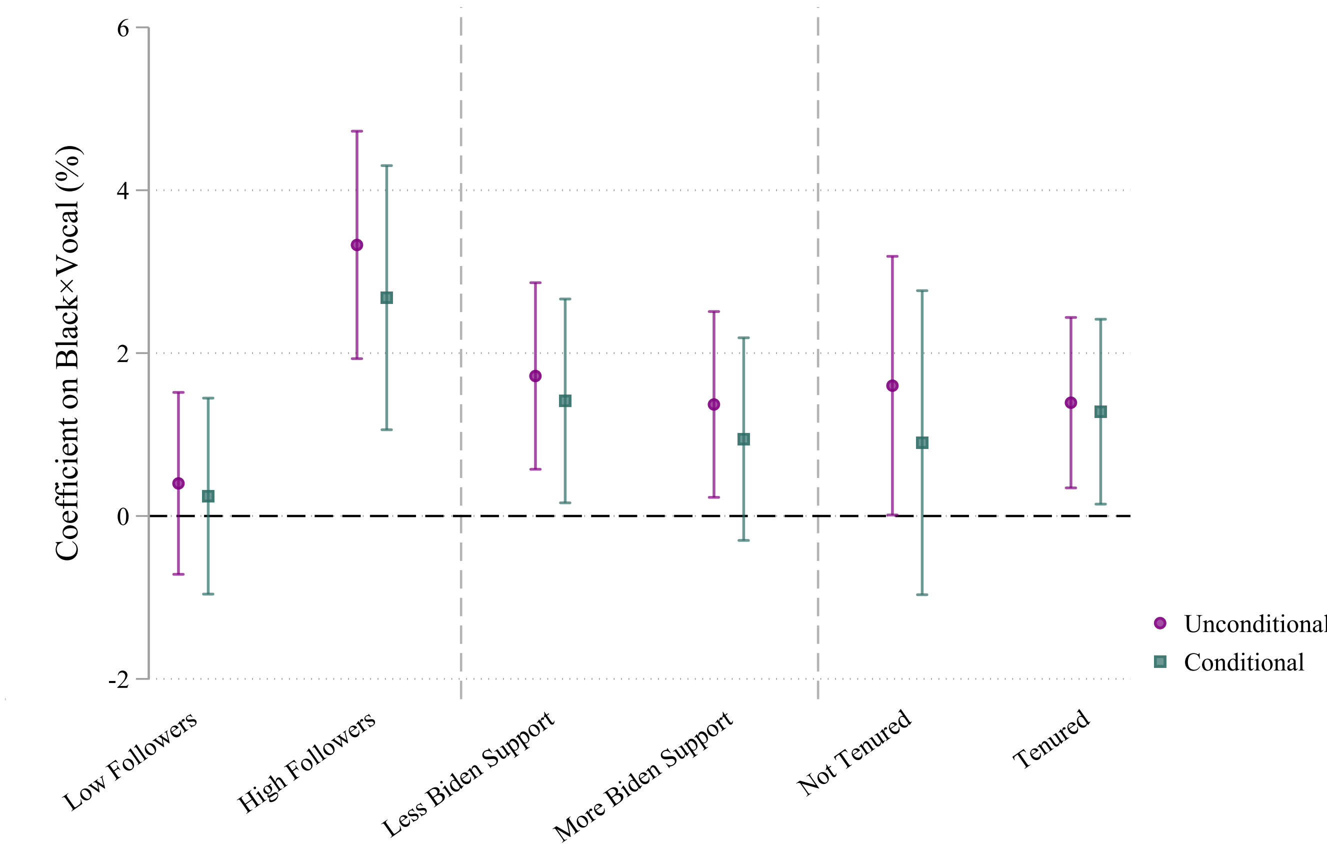
Notes: The bars show what percentage of academics tweeted at least once about Jon Ossoff (any tweet converted to lower-case containing "jon ossoff", "@ossoff", or "@SenOssoff") or Raphael Warnock (any tweet with "raphael warnock", "@ReverendWarnock", or "@SenatorWarnock") from January 1, 2020 to March 27, 2022. Silent includes the 6,784 non-Black academics that did not tweet about racial justice during the same time period, Vocal includes the 11,730 non-Black academics that did tweet about racial justice, and Black includes the 1,094 tweeting Black academics. Unconditional raw means with 95% confidence intervals are shown.

Figure 11: Informativeness Is Higher When Fewer People Are Tweeting About Racial Justice



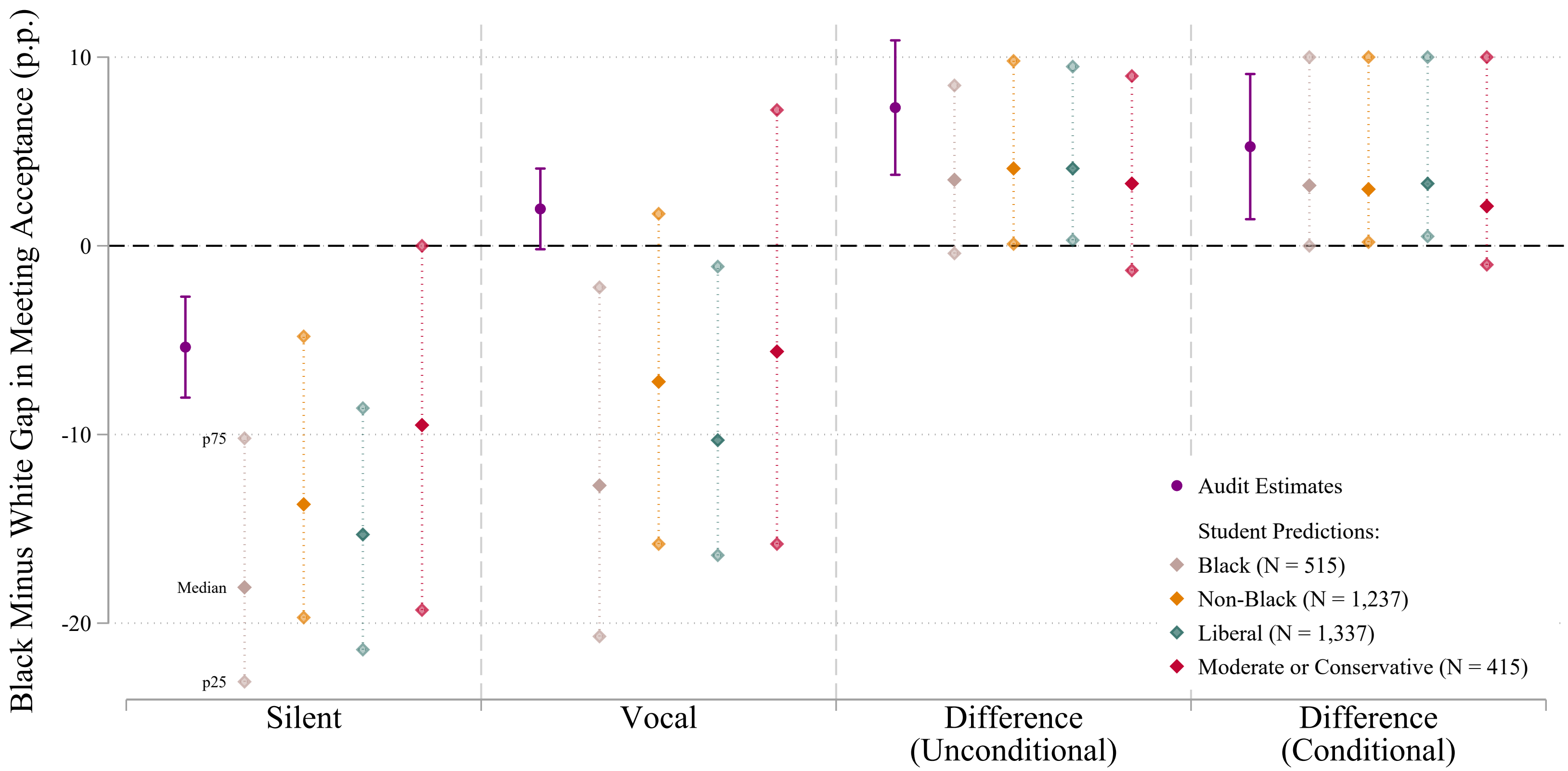
Notes: The figure shows how tweet informativeness ($\hat{\gamma}_2$ from specification 2, though with Vocal (%)$_i$ replacing Vocal$_i$) changes before, during, and after the murder of George Floyd on May 25, 2020. Vocal (%)$_i$ is the percentage of academic i's tweets that are about racial justice, winsorized at the 99th percentile, and set to zero (rather than undefined) for academics that did not tweet at all in a given period. 17% of the 11,450 audited academics tweeted about racial justice at least once during January 1 to May 24, rising to 49% during May 25 to June 30, and falling to 38% during July 1 to November 30. The unconditional estimate denotes the reduction in audit-measured anti-Black discrimination associated with a one percentage point increase in the percentage of racial justice tweets (Vocal (%)$_i$). The conditional estimate denotes the conditional difference in discrimination, using the fourth-from-the-right specification from Figure 6. Standard errors are clustered at the university-by-department-by-sender name-level. 95% confidence intervals are shown.

Figure 12: Informativeness by User Type



Notes: The figure shows how racial justice tweet informativeness ($\hat{\gamma}_2$ from specification 2, though with Vocal $(\%)_i$ replacing Vocal$_i$) differs for different sets of our 11,450 audited academics: (i) below- versus above-median number of Twitter followers, (ii) below- versus above-median Biden 2020 vote share in the university's county, and (iii) Assistant Professors versus Associate and Full Professors. The unconditional estimate denotes the reduction in audit-measured anti-Black discrimination associated with a one percentage point increase in the percentage of racial justice tweets (Vocal $(\%)_i$). The conditional estimate denotes the conditional difference in discrimination, using the fourth-from-the-right specification from Figure 6. Vocal $(\%)_i$ is winsorized at the 99th percentile. Standard errors are clustered at the university-by-department-by-sender name-level. 95% confidence intervals are shown.

Figure 13: Students Overestimate Discrimination and Tend to Underestimate Informativeness



Notes: The figure shows our audit study estimates and 95% confidence intervals in purple (following Figure 7). The diamonds denote the 25th, 50th, and 75th percentile of student predictions, separately by (i) students that self-identify as Black or African American versus students that do not, and (ii) students that describe their political views as liberal or very liberal versus those that describe their political views as moderate, conservative, or very conservative. Before making predictions, students were informed of the meeting acceptance rate for distinctively White student names, separately for Silent and Vocal academics.

Table 1: Audited Academics Are Similar on Most Observables to Non-Audited Academics

	Audited *N = 11,450*	Not Audited *N = 7,064*	Black *N = 1,094*	Off Twitter *N = 900*
	Mean	Mean	Mean	Mean
Female	0.42	0.42	0.57	0.32
Rank of University	59.08	51.50	53.80	61.71
Full Professor	0.36	0.40	0.26	0.56
Assistant Professor	0.35	0.32	0.40	0.18
Associate Professor	0.29	0.28	0.34	0.26
Business	0.04	0.03	0.02	0.06
Engineering and Technology	0.15	0.14	0.10	0.19
Humanities	0.14	0.16	0.26	0.14
Life Sciences	0.23	0.21	0.13	0.20
Physical Sciences	0.07	0.08	0.03	0.12
Professional Schools	0.12	0.11	0.09	0.13
Social Sciences	0.25	0.27	0.38	0.16
Black	0.00	0.00	1.00	0.00
White	0.79	0.78	0.00	0.74
East Asian	0.08	0.09	0.00	0.14
Hispanic	0.04	0.05	0.00	0.02
South Asian	0.07	0.07	0.00	0.08
Other Race/Ethnicity	0.01	0.01	0.00	0.02
Years On Twitter	8.60	8.71	8.58	
Number of Tweets	1,022.01	1,206.86	2,008.33	
Number of Twitter Followers	2,781.09	3,425.69	4,501.41	
Number of Accounts Following	754.98	775.73	1,038.94	
Vocal	0.62	0.65	0.88	
Percentage of Racial Justice Tweets	1.41	1.50	5.29	
Blindspotter % Left-Wing Score	58.46	59.01	71.69	
Republican Accounts Following	0.23	0.23	0.17	
Democratic Accounts Following	1.76	1.98	3.05	
Any Republican Contributions	0.01	0.01	0.00	0.01
Any Democratic Contributions	0.30	0.31	0.26	0.19
Total Republican Contributions (USD)	1.55	2.72	0.00	38.65
Total Democratic Contributions (USD)	244.59	332.95	132.67	258.09

Notes: Columns 1 to 3 summarise our data on tweeting academics. Column 1 shows variable means for the audited non-Black academics, Column 2 covers the non-audited non-Black academics, and Column 3 covers Black academics. Column 4 summarises our random sample of non-Black academics that do not have a Twitter account. The following are dummy variables: Female, Full Professor, Assistant Professor, Associate Professor, Business, Engineering and Technology, Humanities, Life Sciences, Physical Sciences, Professional Schools, Social Sciences, White, East Asian, South Asian, Hispanic, Other Race/Ethnicity, Vocal, and Any Republican/Democratic Contributions. Rank of University ranges from 1 to 150. Years On Twitter is the number of years since the academic joined Twitter, as of May 10, 2022. Number of Tweets is the number of tweets of any type made from January 1, 2020 to March 27, 2022. Number of Twitter Followers and Accounts Following is as of May 10, 2022. Vocal is equal to one for academics that tweeted at least once about racial justice from January 1, 2020 to March 27, 2022. Percentage of Racial Justice Tweets is the number of racial justice-related tweets as a percentage of the total number of tweets, over the same time period. The Blindspotter score is a measure of the left-wing slant of the news the academic engages with on Twitter. Republican Accounts Followed is the number of Republican Senators and House Representatives followed as of July 2022 (similar for Democrats). The Blindspotter score is missing for 4% of the full sample, while the Republican and Democratic Accounts Followed are missing for 0.5% of the full sample. The table shows the means for the academics with non-missing data. Any Republican Contributions is equal to one if the academic is linked to at least one FEC-reported political contribution to a Republican FEC Committee from January 1, 2020 to March 27, 2022 (similar for Democrats). Total Contributions are for the same period.

Table 2: Informativeness by Tweet Type

	Meeting Accepted (%)					
	(1)	(2)	(3)	(4)	(5)	(6)
Black × Racial Justice Tweets (%)	1.62***	1.38***				
	(0.41)	(0.44)				
Black × Racial Justice Retweets (%)			2.30***	1.97***	2.21***	1.93***
			(0.65)	(0.68)	(0.66)	(0.68)
Black × Racial Justice Non-Retweets (%)			1.38	1.37		
			(0.91)	(0.94)		
Black × Original Racial Justice Tweets (%)					-0.89	-0.75
					(1.66)	(1.68)
Black × Quoted Racial Justice Tweets (%)					4.68*	4.31*
					(2.55)	(2.59)
Black × Replied Racial Justice Tweets (%)					4.29	4.17
					(3.35)	(3.40)
Observations	11,393	11,393	11,393	11,393	11,393	11,393
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Email Type Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Twitter	All	Twitter	All	Twitter	All
p-value			0.48	0.65	0.10	0.16

Notes: Each variable shown in the table interacted with Black is winsorized at the 99th percentile. Sample mean of Racial Justice Retweets (%) = 0.79; Racial Justice Non-Retweets (%) = 0.52; Original Racial Justice Tweets (%) = 0.23; Quoted Racial Justice Tweets (%) = 0.16; Replied Racial Justice Tweets (%) = 0.09. The denominator for each of these measures is the total number of tweets made from January 1, 2020 to March 27, 2022. Before winsorizing, Racial Justice Tweets is equal to the sum of Racial Justice Retweets and Racial Justice Non-Retweets; while Racial Justice Non-Retweets is equal to the sum of Original, Quoted, and Replied Racial Justice Tweets. All regressions also include level variables for the interaction term (winsorized for those shown in the table). Odd columns additionally include the number of original tweets, reply tweets, retweets, and quote reply tweets, all for the period January 1, 2020 to March 27, 2022, and each of these variables interacted with Black. Even columns include the additional variables used in the fourth-from-the-right coefficient in Figure 6. The bottom row p-value tests for equality of the Retweets and Non-Retweets interactions in columns 3 and 4, and equality of Retweets and Original interactions in columns 5 and 6. Standard errors are clustered at the university-by-department-by-sender name-level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Graduate Students Summary Statistics

	Black Students				Non-Black Students			
	N	Min	Mean	Max	N	Min	Mean	Max
Age (Years)	515	23.50	29.52	54.50	1,237	19.50	29.24	54.50
Female/Woman	515	0.00	0.60	1.00	1,237	0.00	0.58	1.00
Male/Man	515	0.00	0.34	1.00	1,237	0.00	0.36	1.00
Non-binary/Genderfluid/Genderqueer	515	0.00	0.05	1.00	1,237	0.00	0.05	1.00
Black or African American	515	1.00	1.00	1.00	1,237	0.00	0.00	0.00
White or European	515	0.00	0.07	1.00	1,237	0.00	0.58	1.00
Asian	515	0.00	0.03	1.00	1,237	0.00	0.34	1.00
First Nations or Indigenous	515	0.00	0.01	1.00	1,237	0.00	0.01	1.00
Native Hawaiian or Other Pacific Islander	515	0.00	0.00	0.00	1,237	0.00	0.00	1.00
Hispanic, Latino, or Spanish origin	515	0.00	0.07	1.00	1,237	0.00	0.11	1.00
Born in the USA	515	0.00	0.70	1.00	1,237	0.00	0.57	1.00
Studying for a PhD	515	0.00	0.98	1.00	1,237	0.00	0.97	1.00
Rank of University (1 to 80)	515	1.00	35.42	78.00	1,237	1.00	35.55	80.00
Year in Graduate Program	515	1.00	3.67	7.00	1,237	1.00	3.86	7.00
Political Views (1 = v. liberal, 5 = v. conservative)	515	1.00	1.86	5.00	1,237	1.00	1.90	5.00
Has Twitter	515	0.00	0.75	1.00	1,237	0.00	0.60	1.00
How Often Tweets About Racial Justice (1 to 4)	384	1.00	2.45	4.00	737	1.00	1.94	4.00

Notes: Columns 1 to 4 show summary statistics for the surveyed students that self-identify as Black or African American, while columns 5 to 8 show summary statistics for the remaining surveyed students. Age (Years) is the mid-point of three-year categorical answers, with the exception of the category "Over 53" where we code Age as 54.5. Female/Woman, Male/Man, and Non-binary/Genderfluid/Genderqueer are binary measures of gender identity. The handful of respondents who are not studying for a PhD are either Master's students or Postdoctoral Fellows. Political views are self-reported as either 1 = Very Liberal, 2 = Liberal, 3 = Moderate, 4 = Conservative, and 5 = Very Conservative. How Often Tweets About Racial Justice was only asked to students that have Twitter, with 1 = Never, 2 = Rarely, 3 = Sometimes, and 4 = Often.

Table 4: What Predicts Student Predictions?

	Predicted Discrimination			Predicted Informativeness (Unconditional)			Accuracy (Unconditional)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Black Student (0/1)	3.25***	2.69***	2.80***	-1.22*	-1.32**	-1.26*	0.18	-0.11	-0.08
	(0.74)	(0.73)	(0.76)	(0.63)	(0.64)	(0.65)	(0.48)	(0.48)	(0.50)
Political Views (1 = v. liberal, 5 = v. conservative)		-3.48***	-3.42***		-0.44	-0.43		-0.80***	-0.67**
		(0.47)	(0.48)		(0.38)	(0.39)		(0.28)	(0.29)
Man (0/1)		-0.76	-0.75		-0.84	-0.73		0.86*	0.69
		(0.71)	(0.75)		(0.62)	(0.63)		(0.45)	(0.47)
Non-binary/Genderfluid/Genderqueer (0/1)		1.22	1.00		-2.89***	-2.97***		2.17***	1.79***
		(1.16)	(1.29)		(0.82)	(0.89)		(0.60)	(0.64)
Has Twitter (0/1)		0.94	0.96		0.71	0.77		0.37	0.51
		(0.70)	(0.73)		(0.63)	(0.65)		(0.46)	(0.49)
Year in Graduate Program (1 to 7)		-0.07	-0.10		0.36*	0.38*		-0.08	-0.06
		(0.20)	(0.21)		(0.20)	(0.21)		(0.15)	(0.16)
Rank of University (1 to 80)		0.00			0.02*			-0.01	
		(0.01)			(0.01)			(0.01)	
Born in the USA (0/1)		2.33***	2.14***		0.06	0.16		1.74***	1.86***
		(0.72)	(0.74)		(0.66)	(0.68)		(0.49)	(0.50)
Prediction Incentives (0/1)		1.51**	1.66**		-0.52	-0.59		0.58	0.69
		(0.65)	(0.69)		(0.58)	(0.61)		(0.43)	(0.47)
Observations	1,752	1,752	1,752	1,752	1,752	1,752	1,752	1,752	1,752
Full Sample Outcome Mean	11	11	11	6.1	6.1	6.1	-8.2	-8.2	-8.2
University Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Predicted Discrimination is the student's prediction of racial discrimination in the full sample of audited example (positive if Black students are discriminated against). Predicted Informativeness (Unconditional) is the student's prediction of the unconditional difference in racial discrimination between Vocal and Silent academics (positive if Vocal academics discriminate against Black students less than Silent academics). Accuracy (Unconditional) is the negative of the absolute percentage point difference between a student's Predicted Informativeness (Unconditional) and the unconditional informativeness we estimate in the audit study (7.3). Political Views are either 1 = very liberal, 2= liberal, 3 = moderate, 4 = conservative, or 5 = very conservative. The omitted gender category is Female/Woman, and a dummy variable for students selecting "Prefer not to say" is included in the regression but omitted from the table. We code Year in Graduate Program as 7 when a student selects "7th or more." Rank of University is from the US News 2019 University Rankings. Prediction Incentives is equal to one for the half of the sample randomly assigned to receive additional lottery tickets for accurate predictions. Standard errors are HC3 robust. *** p<0.01, ** p<0.05, * p<0.1.

Online Appendix

Table A1: Audit Study Balance Check

	Female	Assistant Professor	Associate Professor	White	Number Of Tweets	Vocal	Number Of Followers	Any Democrat Contributions
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Balance for Full Sample				
Black	-0.01*	0.00	-0.01*	-0.00	34.75	-0.00	-119.21	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(53.80)	(0.01)	(610.15)	(0.01)
				Panel B: Balance by Vocality				
Black × Vocal	-0.01	0.01	-0.01	0.00	8.79		-954.58	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(82.09)		(751.60)	(0.01)
Black × Silent	-0.01	-0.00	-0.02	-0.02	86.60*		1260.24	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(45.85)		(1321.36)	(0.01)
Observations	11,393	11,393	11,393	11,393	11,393	11,393	11,393	11,393
Full Sample Outcome Mean	.42	.35	.29	.79	1,023	.62	2,789	.3
p-value (Black × Vocal = Black × Silent)	.97	.63	.64	.25	.42		.18	.24
Vocal Dummy (Panel B only)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Strata Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Email Type Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

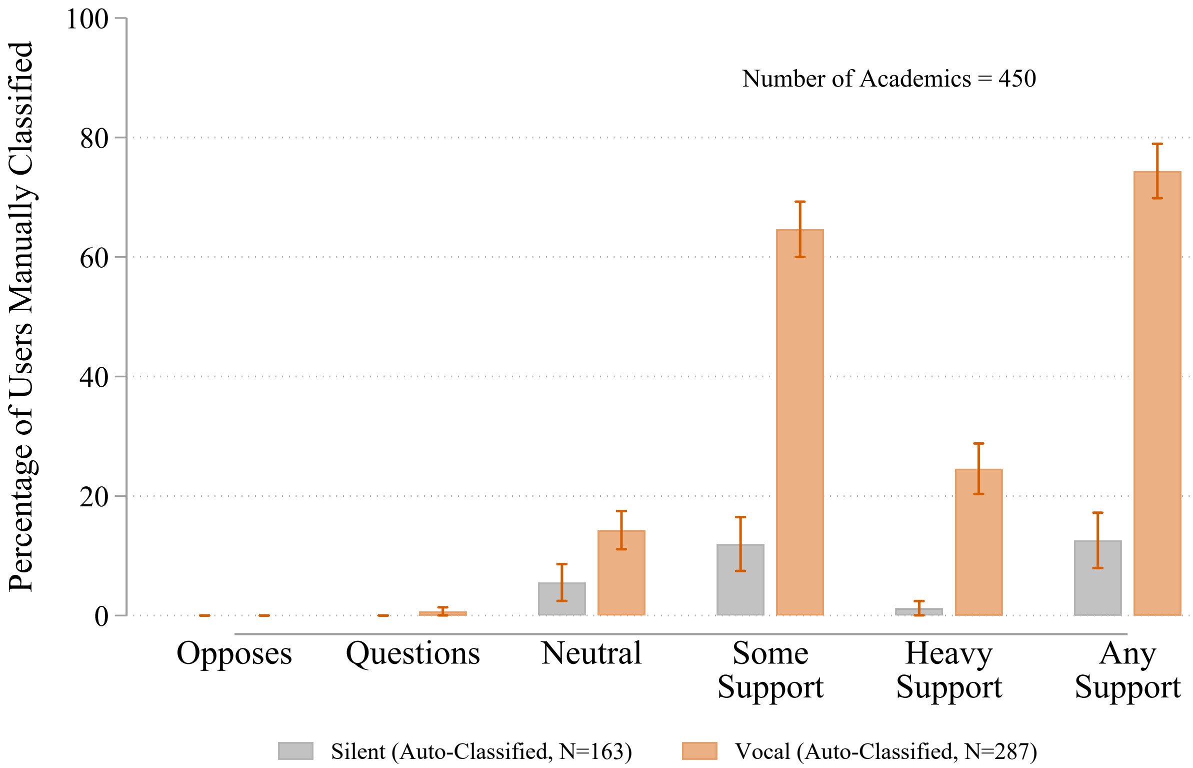
Notes: Standard errors are clustered at university-department-sender name-level. Black is a dummy variable equal to one for receiving an email from a distinctively Black name. Outcome variables are: (1) dummy variable for female academic, (2) dummy variable for Assistant Professor, (3) dummy variable for Associate Professor, (4) dummy variable for White academic, (5) number of tweets (Jan 1, 2020 to Mar 27, 2022), (6) dummy variable for whether tweeted about racial justice (Jan 1, 2020 to Mar 27, 2022), (7) number of Twitter followers as of May 10, 2022, and (8) dummy variable for any FEC-linked contributions to Democrat-related committees (Jan 1, 2020 to Mar 27, 2022). *** p<0.01, ** p<0.05, * p<0.1.

Table A2: What Predicts Student Predictions? (Conditional)

	Predicted Informativeness (Conditional)			Accuracy (Conditional)		
	(1)	(2)	(3)	(4)	(5)	(6)
Black Student (0/1)	-0.93	-1.15	-1.33	-0.74	-1.13*	-1.07
	(0.75)	(0.77)	(0.81)	(0.60)	(0.61)	(0.66)
Political Views (1 = v. liberal, 5 = v. conservative)		-1.07**	-1.04**		-0.70**	-0.68**
		(0.42)	(0.43)		(0.33)	(0.34)
Man (0/1)		0.11	0.35		0.22	0.10
		(0.72)	(0.76)		(0.56)	(0.61)
Non-binary/Genderfluid/Genderqueer (0/1)		-2.07**	-3.02***		2.02***	2.21***
		(0.95)	(1.07)		(0.75)	(0.85)
Has Twitter (0/1)		1.37*	1.44*		0.56	0.56
		(0.72)	(0.74)		(0.57)	(0.59)
Year in Graduate Program (1 to 7)		-0.13	-0.12		0.27*	0.24
		(0.21)	(0.22)		(0.16)	(0.17)
Rank of University (1 to 80)		0.01			-0.02**	
		(0.01)			(0.01)	
Born in the USA (0/1)		-0.62	-0.52		2.47***	2.35***
		(0.71)	(0.72)		(0.56)	(0.58)
Prediction Incentives (0/1)		-0.35	-0.66		-0.31	-0.14
		(0.66)	(0.69)		(0.53)	(0.56)
Observations	1,752	1,752	1,752	1,752	1,752	1,752
Full Sample Outcome Mean	11	11	11	6.1	6.1	6.1
University Fixed Effects	No	No	Yes	No	No	Yes

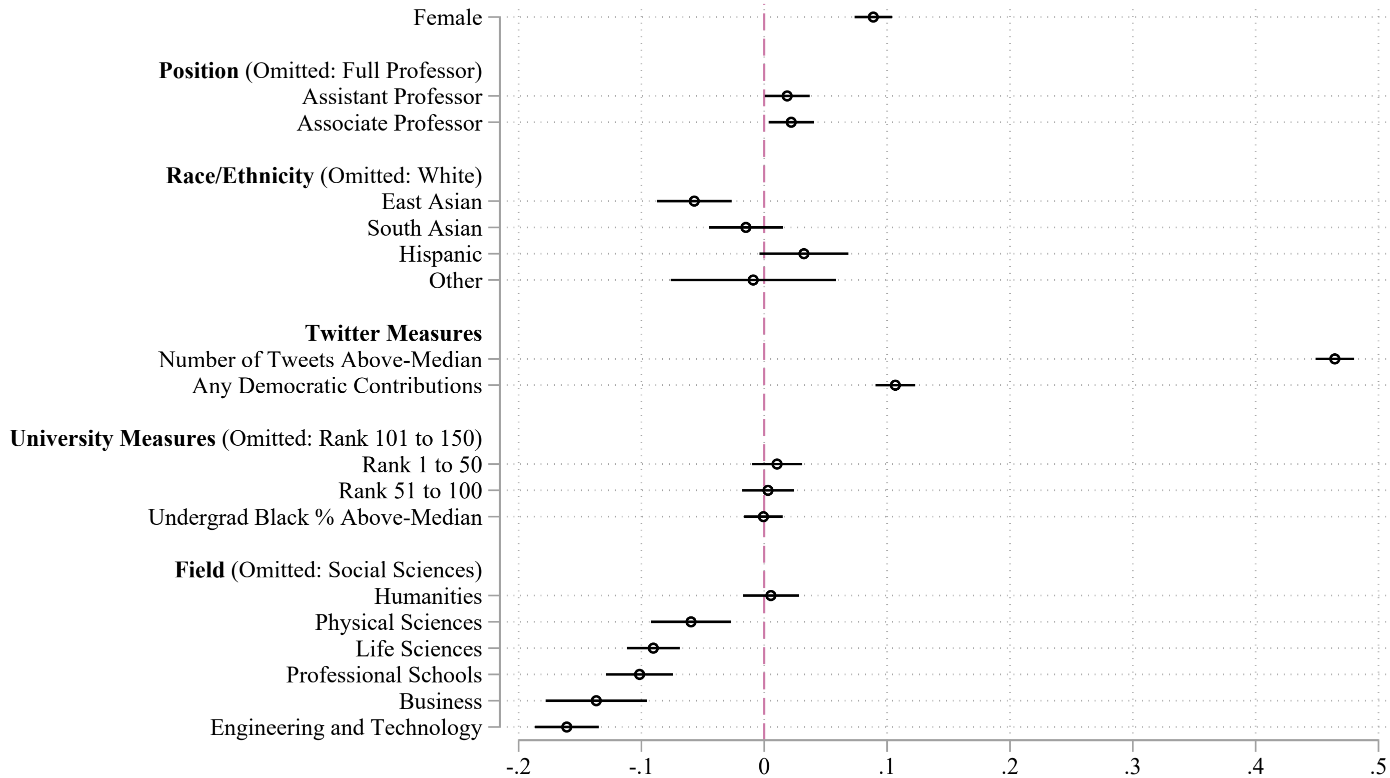
Notes: Predicted Informativeness (Conditional) is the student's prediction of the conditional difference in racial discrimination between Vocal and Silent academics (positive if Vocal academics discriminate against Black students less than Silent academics). Accuracy (Conditional) is the negative of the absolute percentage point difference between a student's Predicted Informativeness (Conditional) and the conditional informativeness we estimate in the audit study (5.3). Political Views are either 1 = very liberal, 2= liberal, 3 = moderate, 4 = conservative, or 5 = very conservative. The omitted gender category is Female/Woman, and a dummy variable for students selecting "Prefer not to say" is included in the regression but omitted from the table. We code Year in Graduate Program as 7 when a student selects "7th or more." Rank of University is from the US News 2019 University Rankings. Prediction Incentives is equal to one for the half of the sample randomly assigned to receive additional lottery tickets for accurate predictions. Standard errors are HC3 robust. *** p<0.01, ** p<0.05, * p<0.1.

Figure A1: Validating the User-level Measure of Vocality



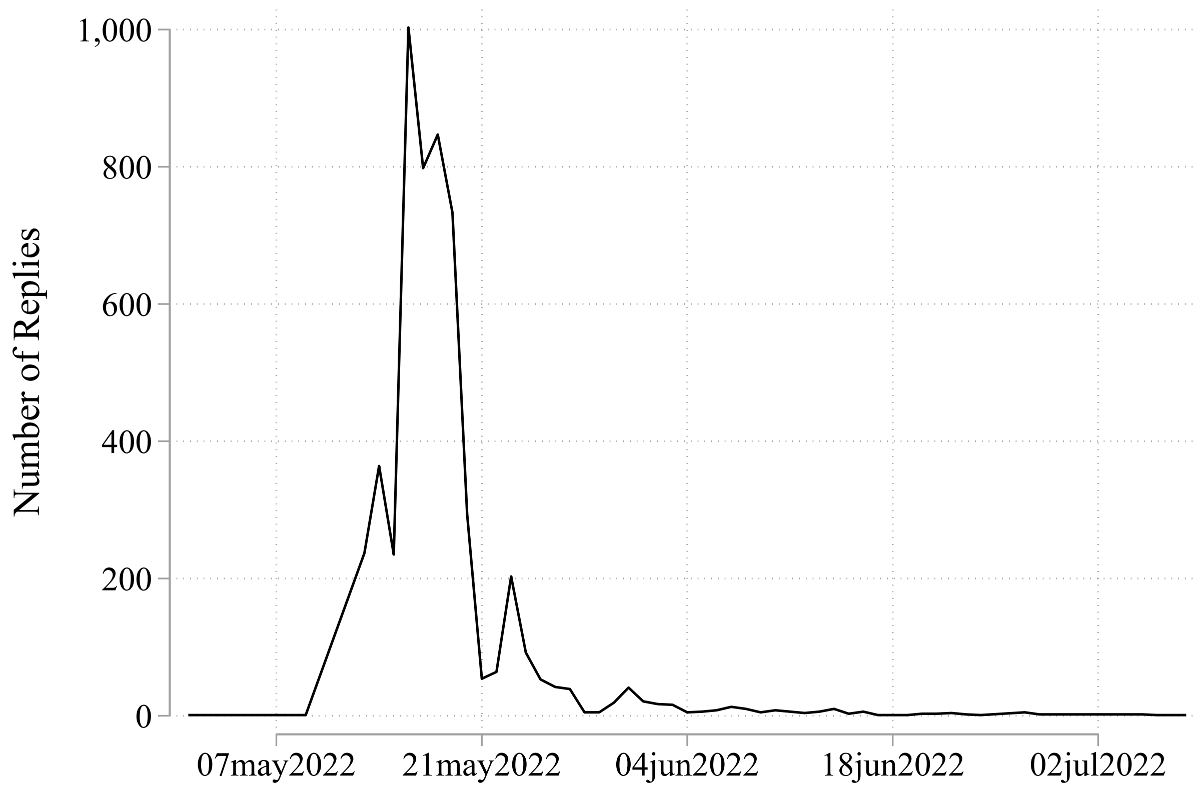
Notes: This figure validates our automated signalling algorithm. The team scrolled through the post-May 2020 Twitter feeds of a random subset of our experimental sample ($N = 450$), recording for each user whether they ever: (i) opposed racial justice, (ii) questioned racial justice, (iii) tweeted neutrally about racial justice, (iv) tweeted some support for racial justice, or (v) tweeted heavy support for racial justice. The orange bars include data for the 287 academics we automatically classify as Vocal. The grey bars include data for the 163 academics we automatically classify as Silent. As an example, the orange "Neutral" bar shows the percentage of users that the team *manually* found to have ever tweeted neutrally about racial justice, only among the academics that we *automatically* classify as Vocal. The "Any Support" category shows the percentage of users that ever tweeted some or heavy support. 95% confidence intervals are shown. [Note: The figure differs slightly from the corresponding one in our pre-analysis plan because we updated our automated measure of vocality to (i) incorporate racial justice-related words and phrases in tweeted hyperlinks (as promised in our pre-analysis plan), and (ii) include the full text of all retweets, correcting an error in our earlier measure. See Appendix B for more details.]

Figure A2: Who Tweets About Racial Justice?



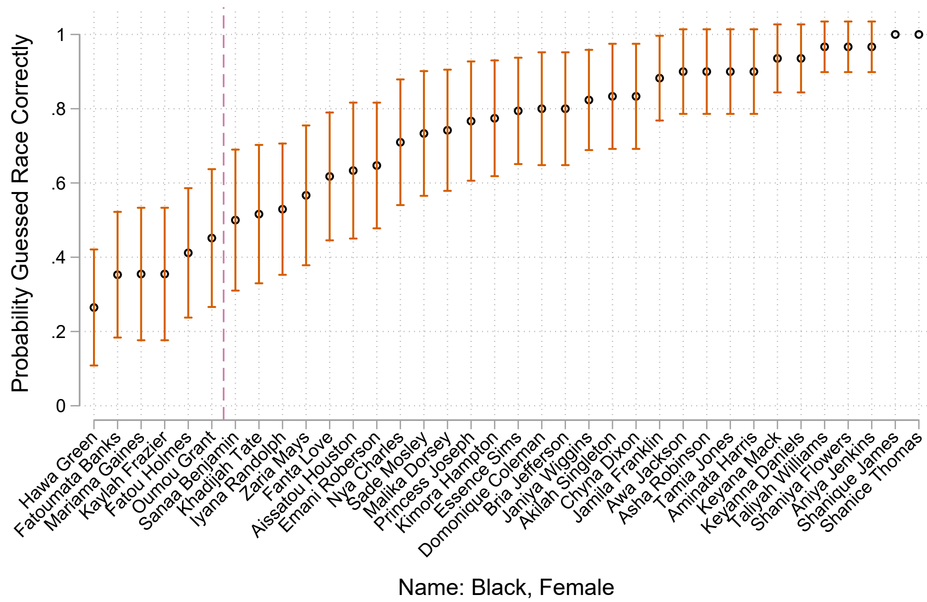
Notes: The figure displays coefficients and 95% confidence intervals from a regression of Vocal_i on a set of covariates, for the sample of 11,450 non-Black academics included in the audit study. The covariates are: dummy variable for female, dummy variables for Assistant and Associate Professor, race/ethnicity dummy variables, dummy variable for above-median total tweets from January 1, 2020 to March 27, 2022, dummy variable for any contributions to Democratic-related FEC committees from January 1, 2020 to March 27, 2022, dummy variables for university ranked 1 to 50 and 51 to 100, dummy variable for undergraduate Black student share above-median, and dummy variables for broad academic fields. Standard errors are HC3 robust.

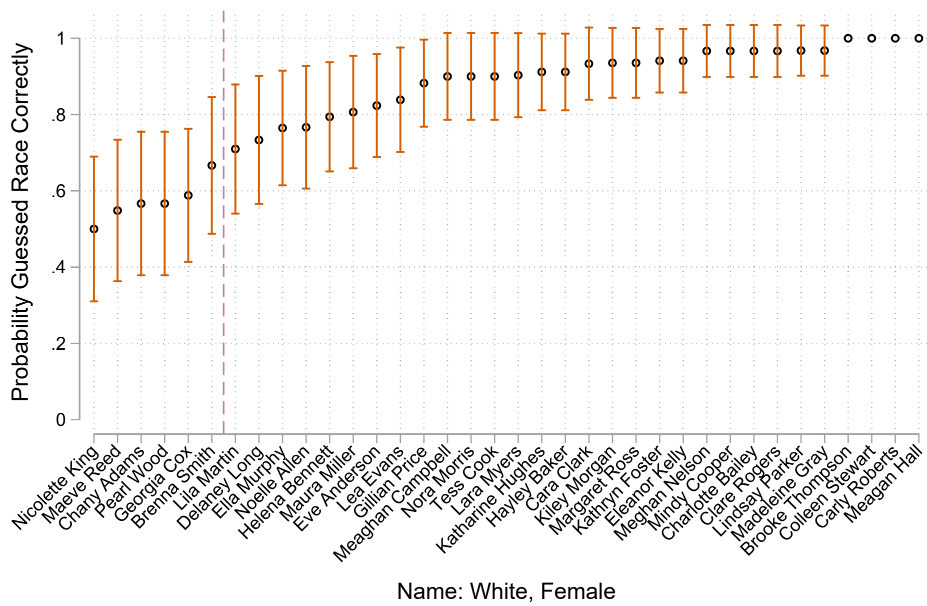
Figure A3: Most Audited Academics Replied by Late-May 2022



Notes: The figure shows the number of email replies we received each day from the audited academics.

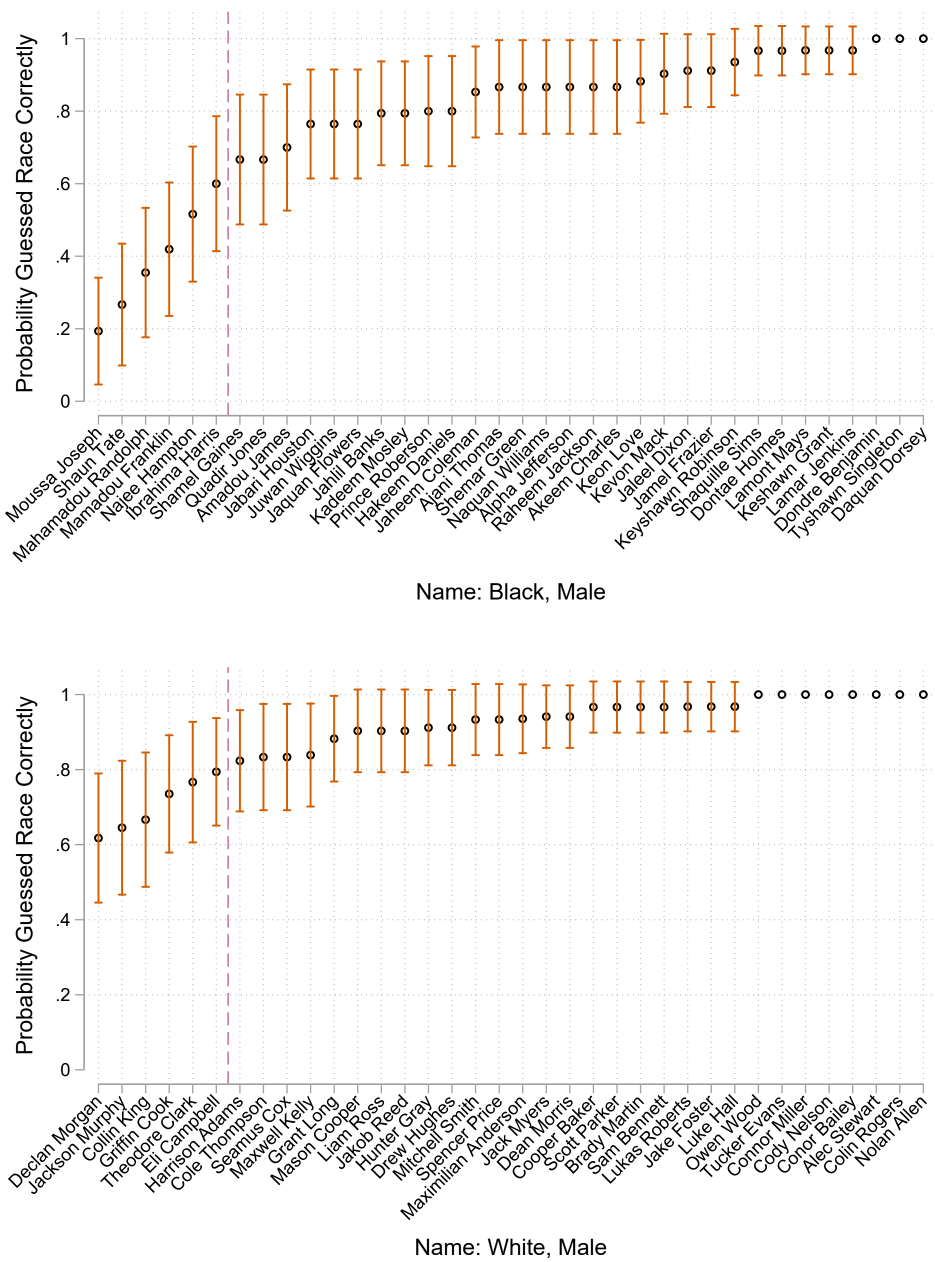
Figure A4: MTurk Validation of Racial Distinctiveness: Female Names





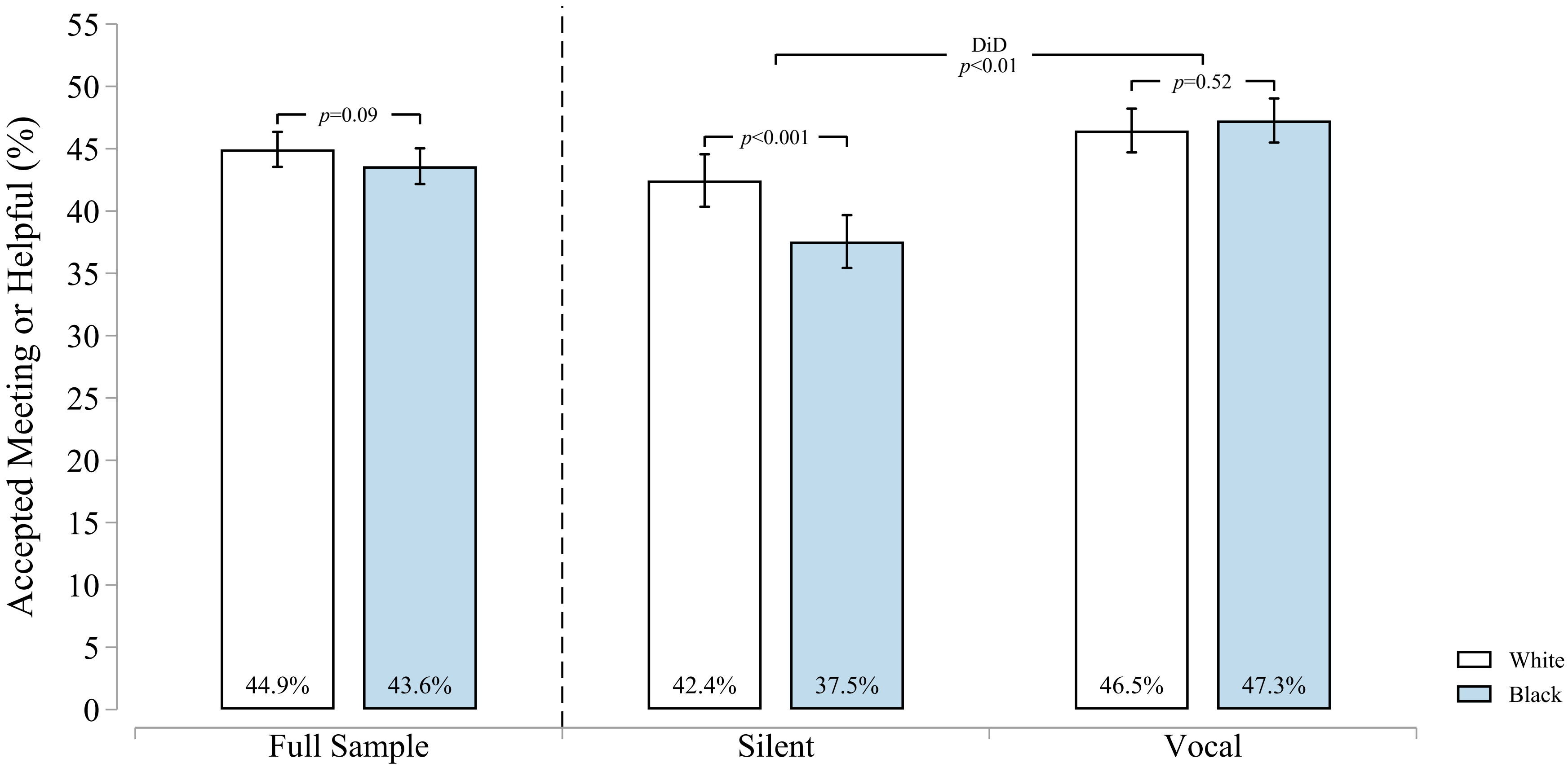
Notes: Each full name was guessed by at least 30 MTurkers. The figure shows the fraction of MTurkers that guessed the race correctly for the female names, along with 95% confidence intervals. Each MTurker chose one answer from Black, White, Hispanic, or Other. They did not know that no names were chosen to be Hispanic- or Other-sounding. We dropped the six names to the left of the vertical dashed line.

Figure A5: MTurk Validation of Racial Distinctiveness: Male Names



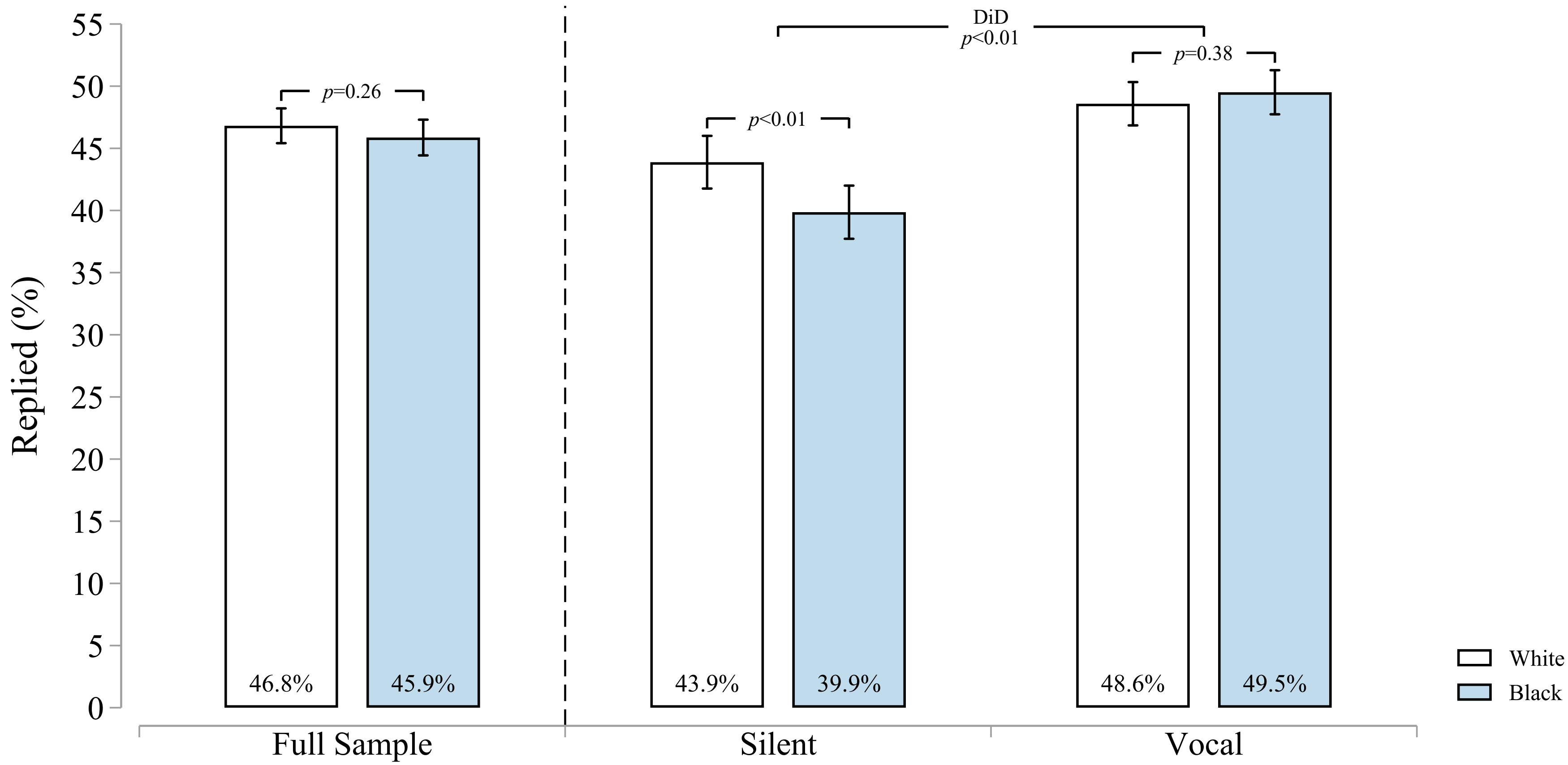
Notes: Each full name was guessed by at least 30 MTurkers. The figure shows the fraction of MTurkers that guessed the race correctly for the male names, along with 95% confidence intervals. Each MTurker chose one answer from Black, White, Hispanic, or Other. They did not know that no names were chosen to be Hispanic- or Other-sounding. We dropped the six names to the left of the vertical dashed line.

Figure A6: Accepted or Helpful: Vocal Professors Discriminate Against Black Students 5.7 Percentage Points Less



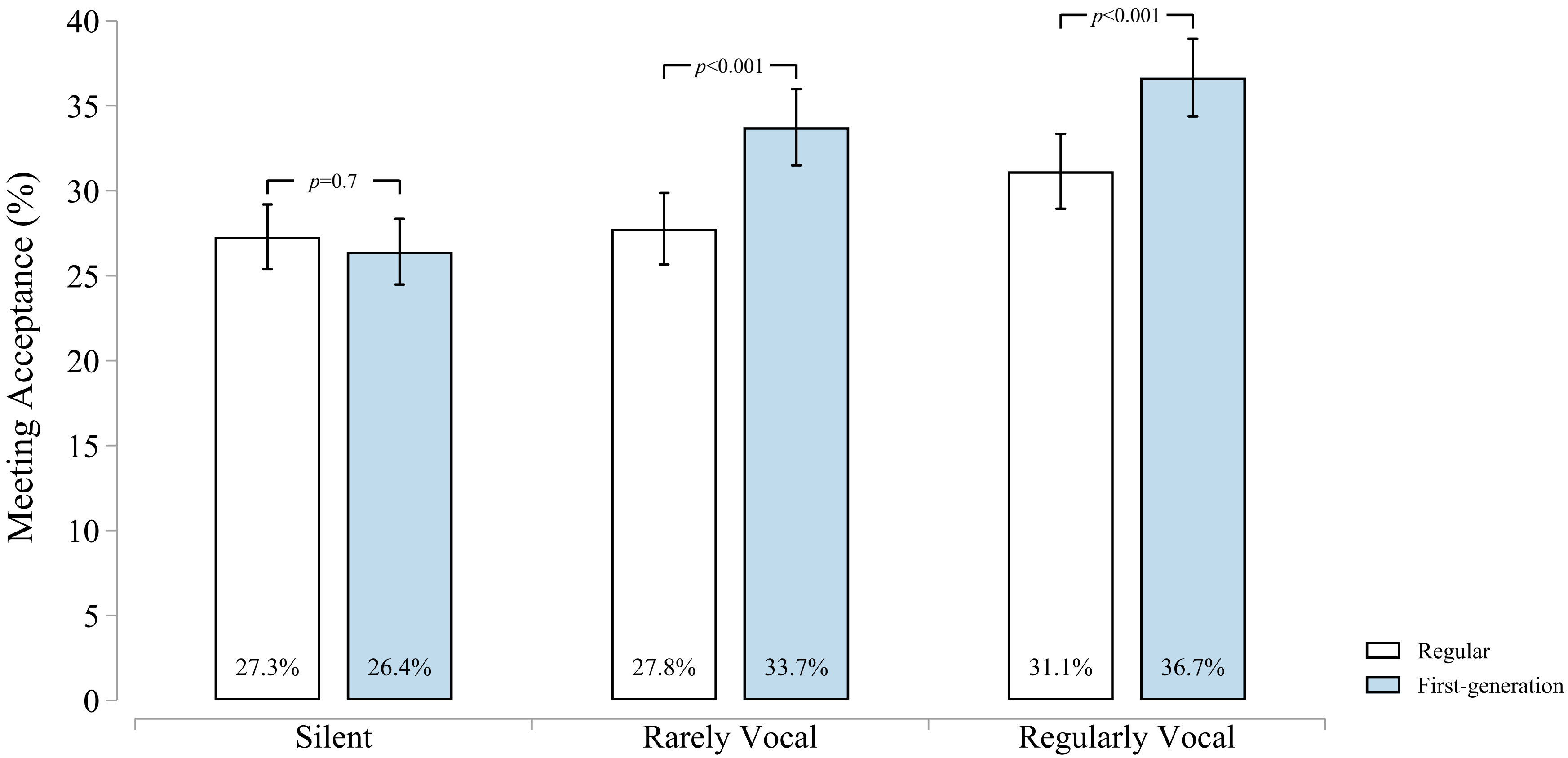
Notes: The bars show what percentage of audited academics accepted, or replied helpfully to, meeting requests from distinctively White and distinctively Black names. The Full Sample includes the 11,450 audited academics (4,318 Silent and 7,132 Vocal). Vocal academics are those that tweeted at least once about racial justice from January 2020 to March 2022. Silent academics are those that did not. The raw means and 95% confidence intervals come from a regression of Helpful_i on dummy variables for White and Black email sender (to the left of the vertical dashed line), and a regression on dummy variables for White email sender to Silent academic, White email sender to Vocal academic, and the same for Black email sender (to the right of the vertical dashed line). The p-values come from the specification that also includes strata and email type fixed effects. The DiD (diff-in-diff) p-value is from a test for equal discrimination rates across Vocal and Silent academics (γ_2 in specification 2). Standard errors are clustered at the university-by-department-by-sender name-level.

Figure A7: Replied: Vocal Professors Discriminate Against Black Students 4.9 Percentage Points Less



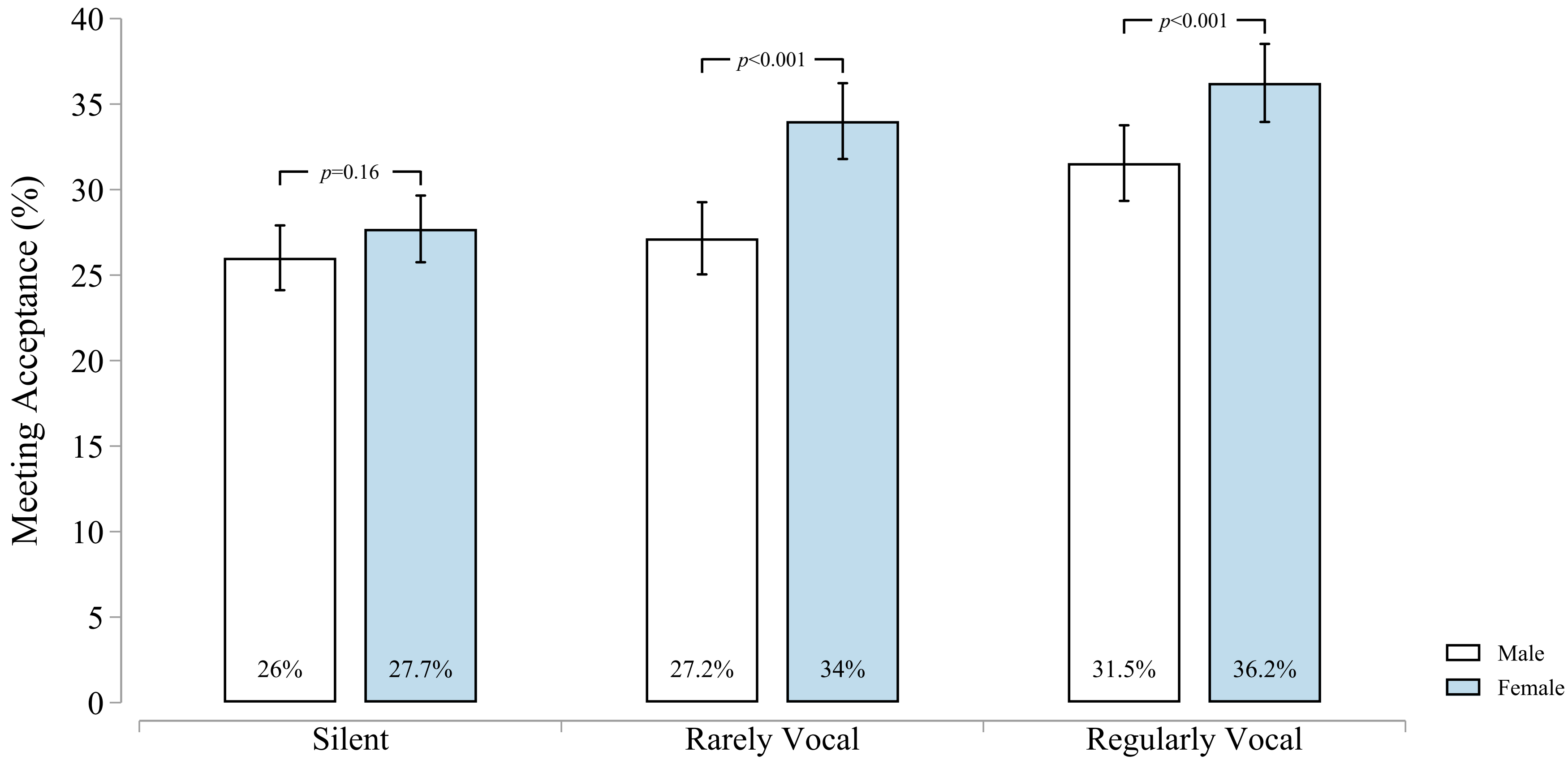
Notes: The bars show what percentage of audited academics replied to meeting requests from distinctively White and distinctively Black names. The Full Sample includes the 11,450 audited academics (4,318 Silent and 7,132 Vocal). Vocal academics are those that tweeted at least once about racial justice from January 2020 to March 2022. Silent academics are those that did not. The raw means and 95% confidence intervals come from a regression of Replied_i on dummy variables for White and Black email sender (to the left of the vertical dashed line), and a regression on dummy variables for White email sender to Silent academic, White email sender to Vocal academic, and the same for Black email sender (to the right of the vertical dashed line). The p-values come from the specification that also includes strata and email type fixed effects. The DiD (diff-in-diff) p-value is from a test for equal discrimination rates across Vocal and Silent academics (γ_2 in specification 2). Standard errors are clustered at the university-by-department-by-sender name-level.

Figure A8: The Rarely and Regularly Vocal Favor First-Generation Students Similarly



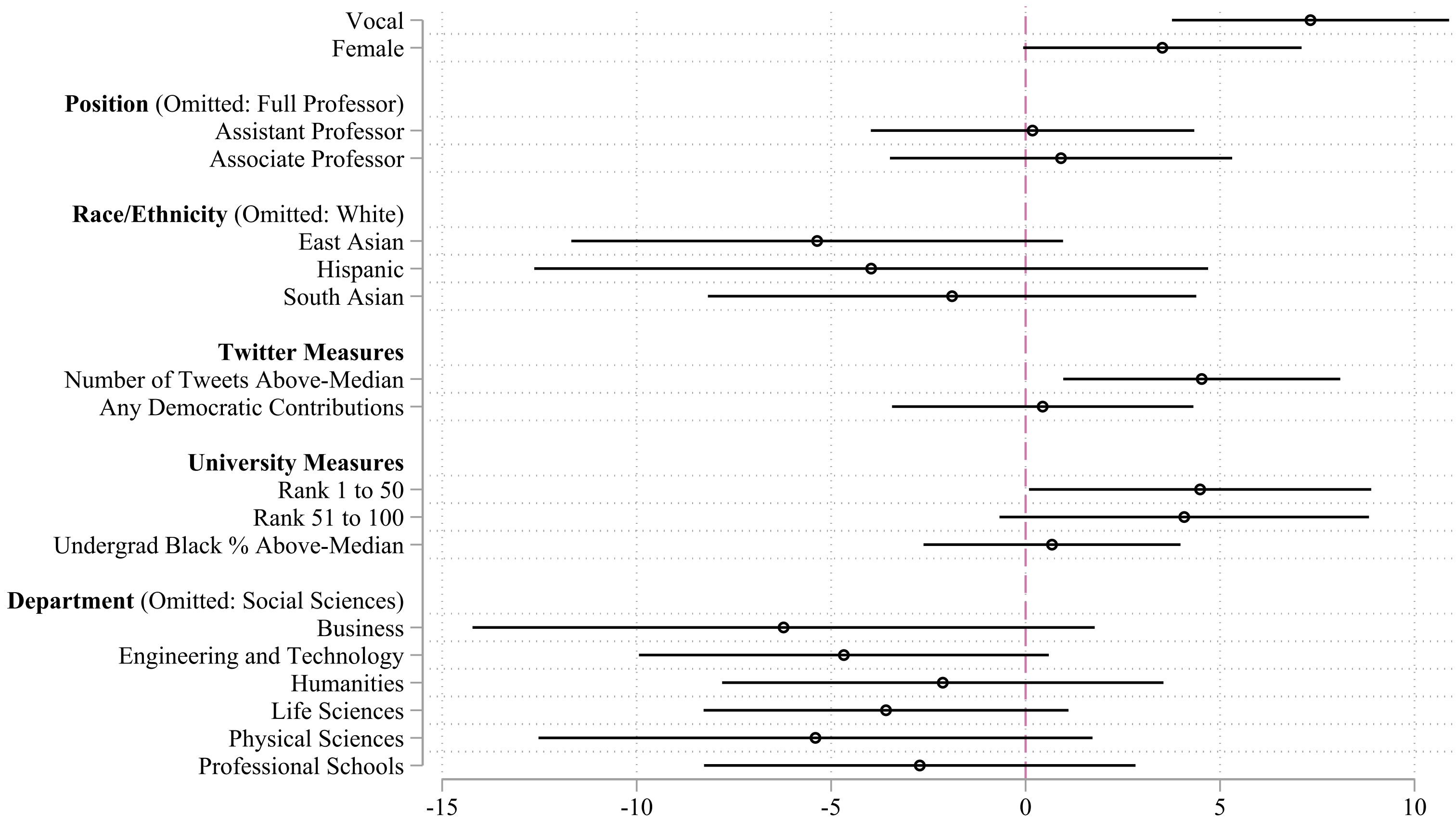
Notes: Silent academics are those that have not tweeted about racial justice. The bars for the Silent academics replicate the findings in Figure 2. Among the Vocal, the Rarely Vocal are those with below-median percentage of tweets from January 1, 2020 to March 27, 2020 that are about racial justice (0.6% on average), while the Regularly Vocal academics are above-median (3.9% on average). The raw means and 95% confidence intervals come from a regression of Accepted_i on dummy variables for first-generation student email sender to Silent academic, regular student email sender to Silent academic, and the same for emailed to the Rarely Vocal and to the Regularly Vocal. The p-values come from the specification that also includes strata and email type fixed effects. Standard errors are clustered at the university-by-department-by-sender name-level.

Figure A9: The Rarely and Regularly Vocal Favor Female Students Similarly



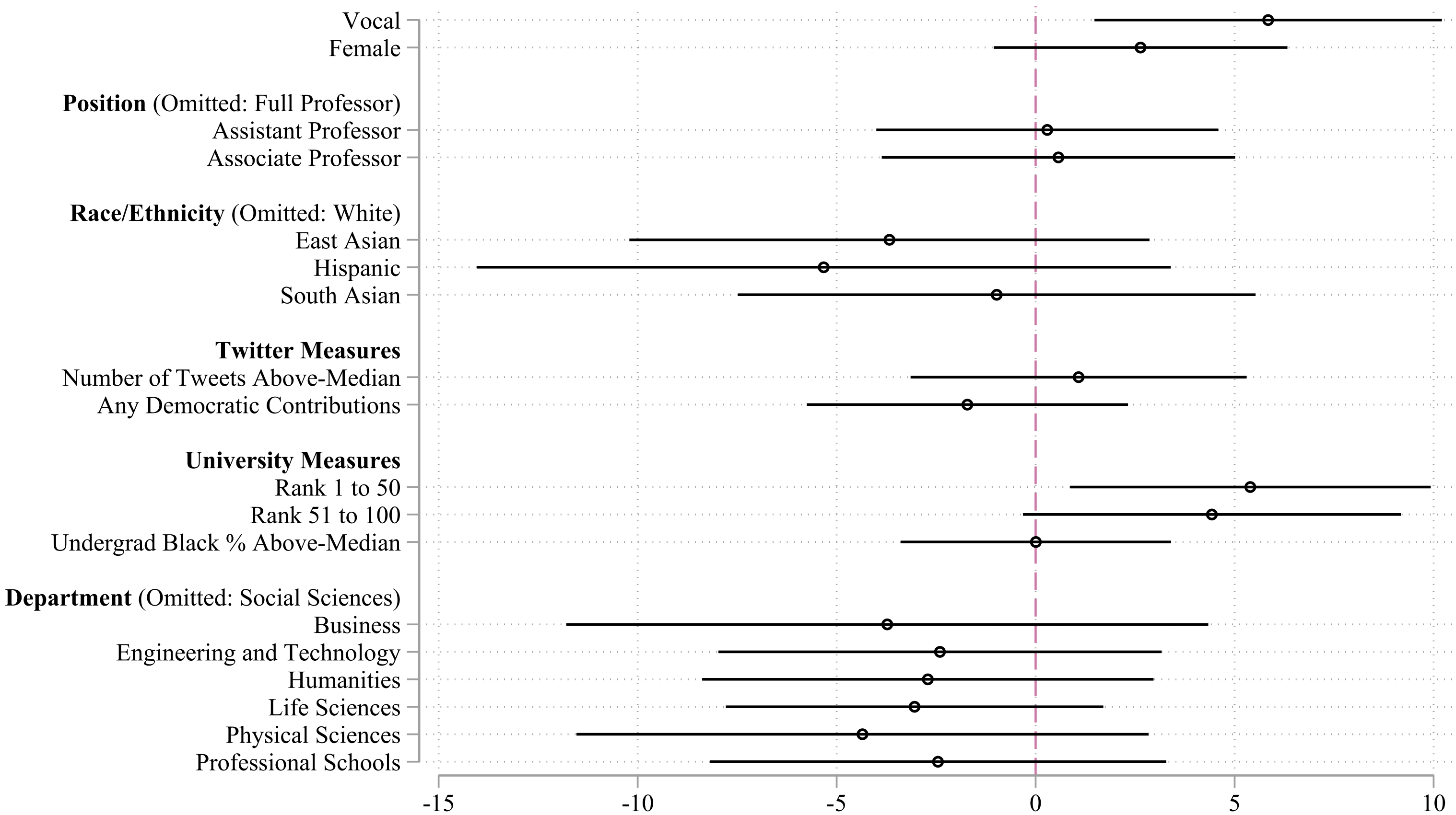
Notes: Silent academics are those that have not tweeted about racial justice. The bars for the Silent academics replicate the findings in Figure 3. Among the Vocal, the Rarely Vocal are those with below-median percentage of tweets from January 1, 2020 to March 27, 2020 that are about racial justice (0.6% on average), while the Regularly Vocal academics are above-median (3.9% on average). The raw means and 95% confidence intervals come from a regression of Accepted_i on dummy variables for female student email sender to Silent academic, male student email sender to Silent academic, and the same for emailed to the Rarely Vocal and to the Regularly Vocal. The p-values come from the specification that also includes strata and email type fixed effects. Standard errors are clustered at the university-by-department-by-sender name-level.

Figure A10: What Predicts Racial Discrimination? (Factor-By-Factor)



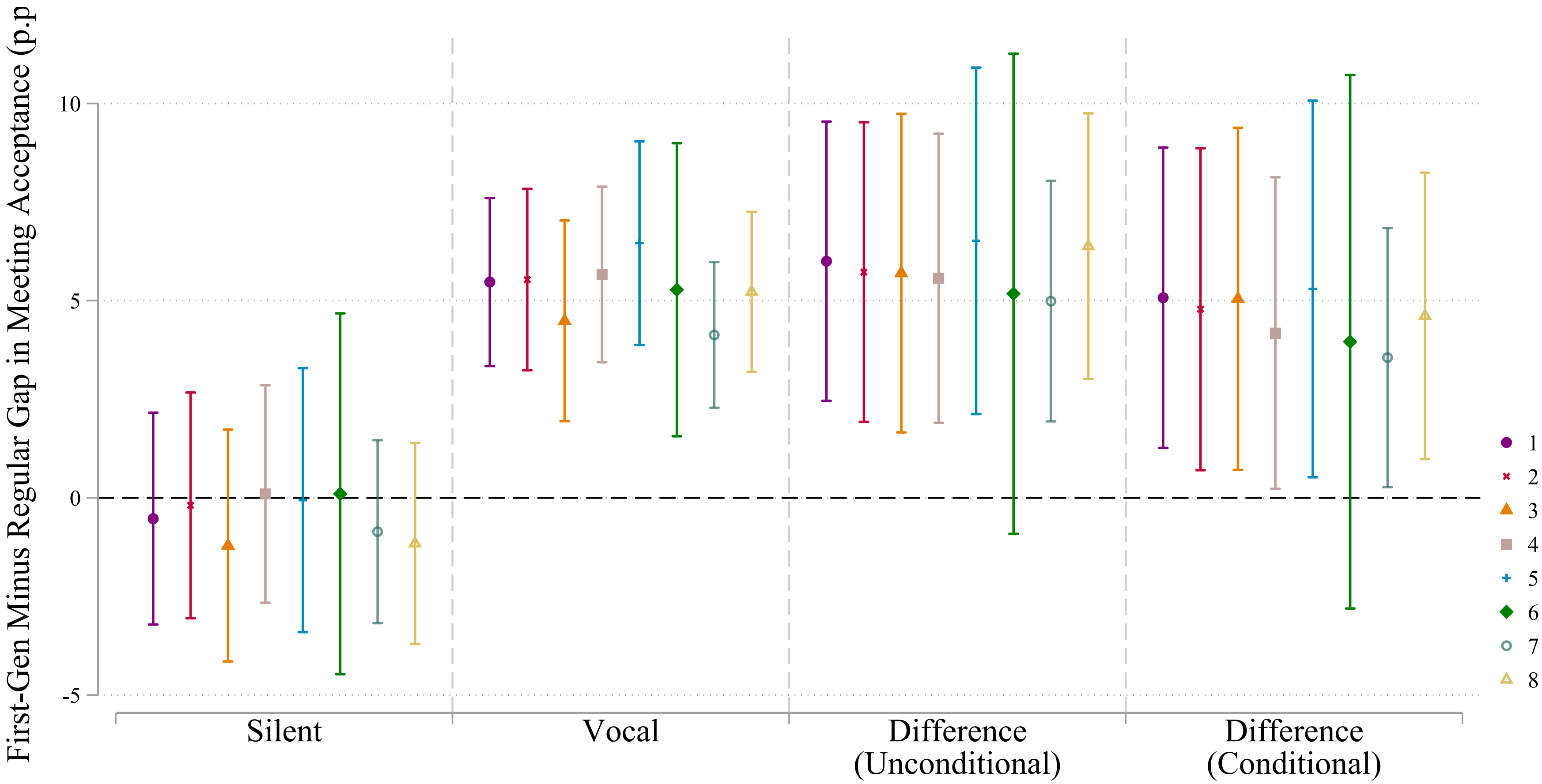
Notes: The figure shows which factors predict racial discrimination. Specifically, the figure shows the $\hat{\gamma}_2$ and $\hat{\theta}_j$s (along with 95% confidence intervals) estimated from specification 2, where the X^js are the same as those used in Figure A2. The coefficients come from nine regressions using specification 2, with the following variables interacted with Black in each: (1) Vocal, (2) Female, (3) Position dummies, (4) Race/Ethnicity dummies, (5) Number of Tweets Above-Median, (6) Any Democratic Contributions, (7) Rank dummies, (8) Undergrad Black % Above-Median, and (9) Department dummies. As an example, the first coefficient tells us that Vocal academics discriminate against Black students 7.3 percentage points less than Silent academics.

Figure A11: What Predicts Racial Discrimination? (All Factors Together)



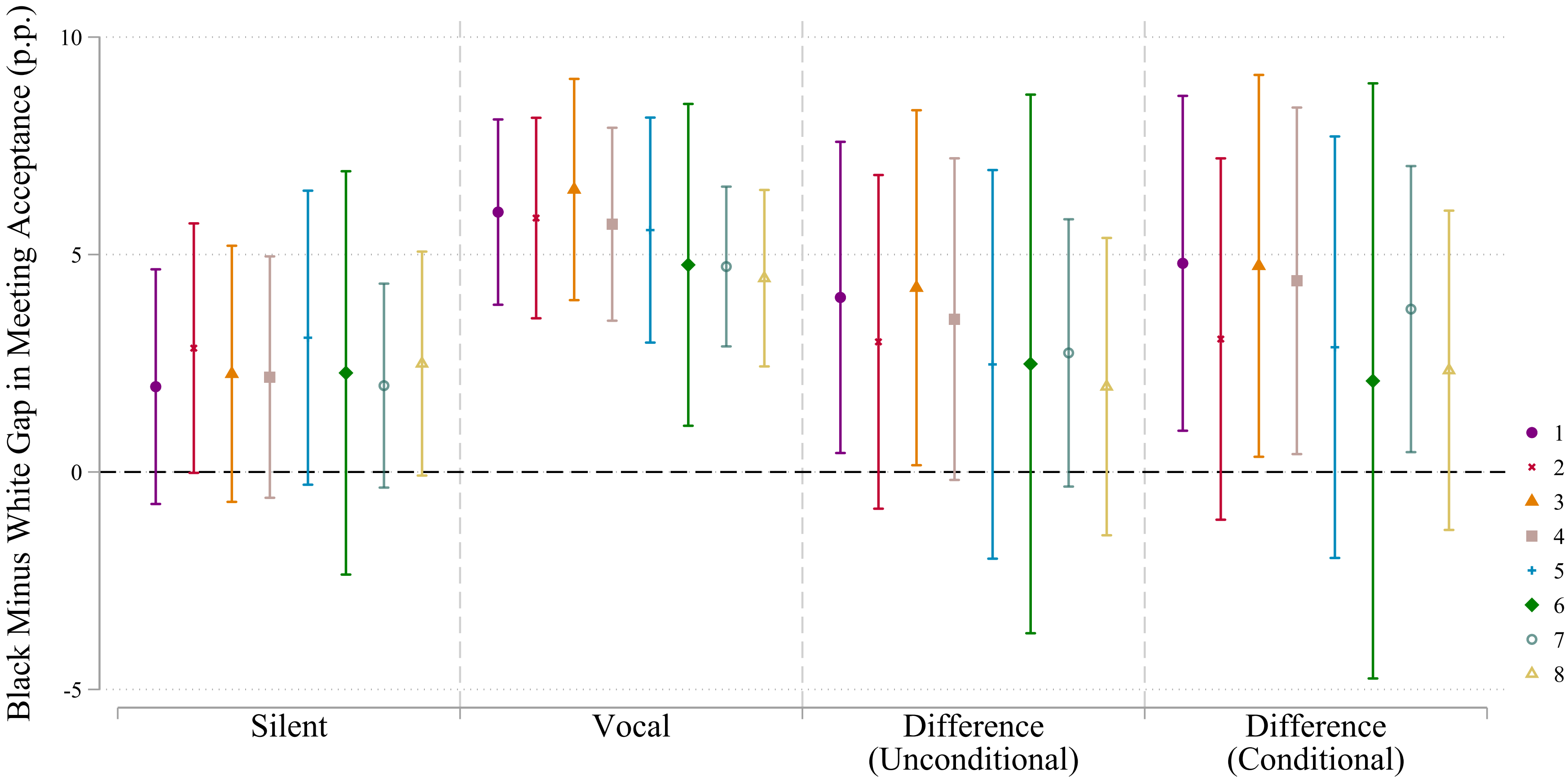
Notes: The figure shows which factors predict racial discrimination, based on one regression with all interactions included at once. Specifically, the figure shows the $\hat{\gamma}_2$ and $\hat{\theta}_j$s (along with 95% confidence intervals) estimated from specification 2, where the X^js are the same as those used in Figure A2. As an example, the first coefficient tells us that Vocal academics discriminate against Black students over five percentage points less than Silent academics, holding the other variables in the figure constant.

Figure A12: Detection Is Unlikely to Explain the First-Generation Status Results



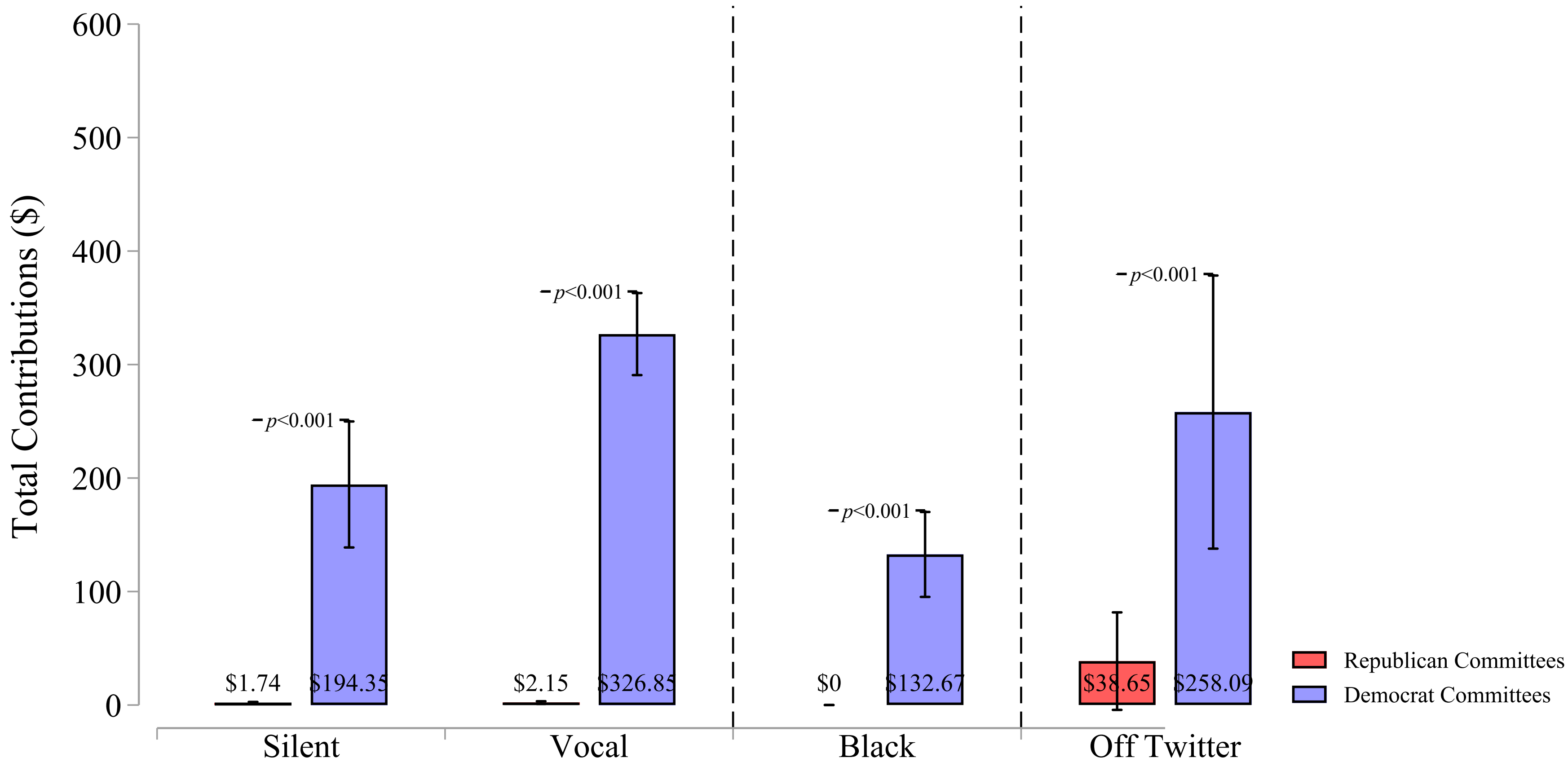
Notes: The figure shows the robustness of our main results (coefficient set 1) to seven alternative samples or outcomes where we expect audit detection to be less likely (coefficients sets 2 to 8). Specifically, for each of the eight we show from left to right: the first-generation versus regular gap in meeting acceptance for Silent academics (negative for discrimination against first-generation students), the same gap for Vocal academics, the unconditional difference in the gap ($\hat{\gamma}_2$ from specification 2 without any X_i^j covariates, positive if Vocal academics favor first-generation students more than Silent academics), and the conditional difference using the fourth-from-the-right specification from Figure 6. The specification and sample variants are (percentage of the sample dropped in parentheses): (2) drop academics in Economics, Political Science, Sociology, and Business (12.4%), (3) drop academics in the Social Sciences (25%), (4) drop academics to whom we sent more generic emails (7%), (5) drop university-departments to which we sent more than ten emails (27%), (6) drop university-departments to which we sent more than five emails (61%), (7) outcome is meeting accepted within one day, and (8) outcome is meeting accepted within three days. Standard errors are clustered at the university-by-department-by-sender name-level. 95% confidence intervals are shown.

Figure A13: Detection Is Unlikely to Explain the Gender Results



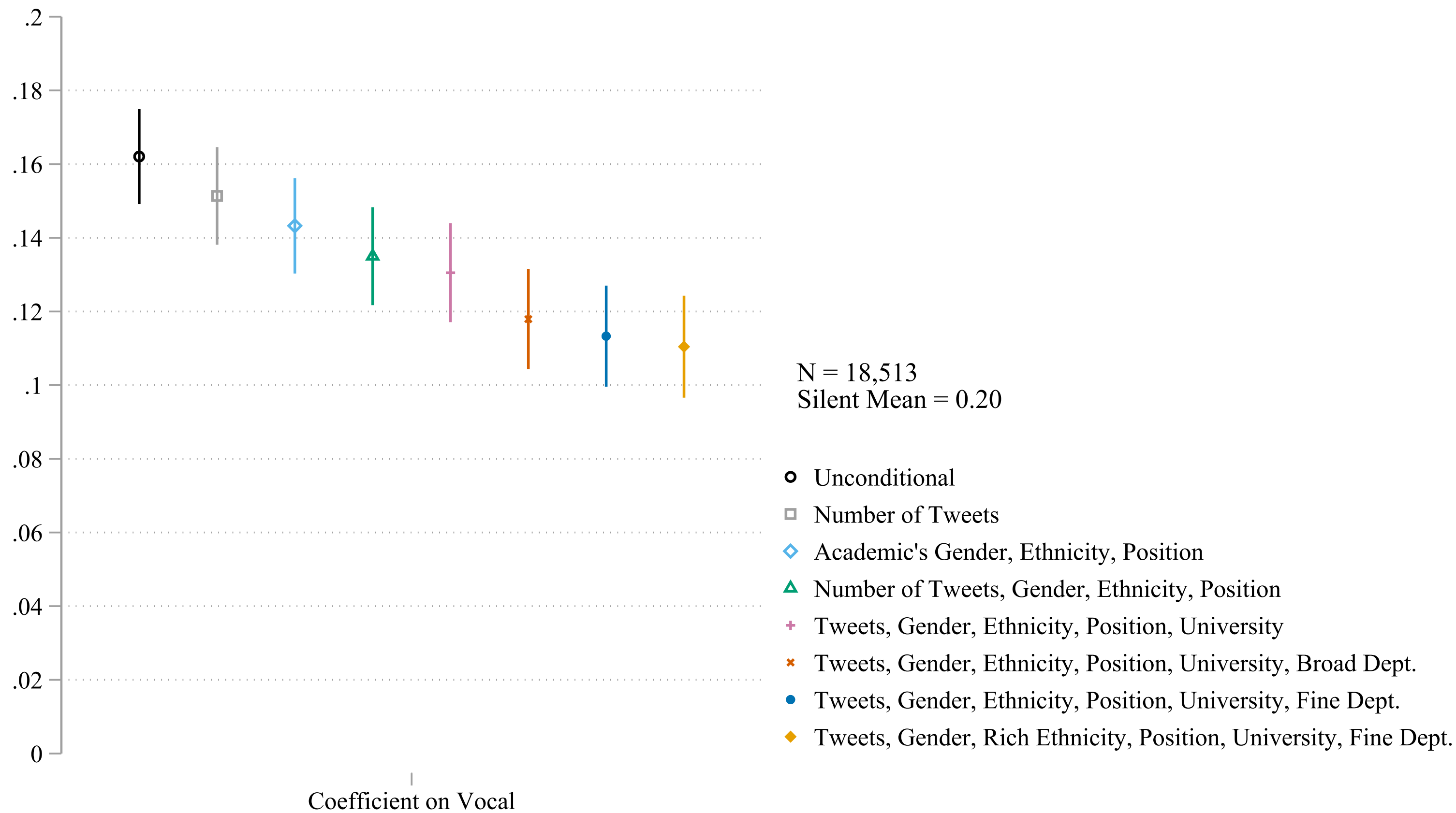
Notes: The figure shows the robustness of our main results (coefficient set 1) to seven alternative samples or outcomes where we expect audit detection to be less likely (coefficients sets 2 to 8). Specifically, for each of the eight we show from left to right: the gender gap in meeting acceptance for Silent academics (positive for discrimination against male students), the same gap for Vocal academics, the unconditional difference in the gap ($\hat{\gamma}_2$ from specification 2 without any X_i^j covariates, positive if Vocal academics favor female students more than Silent academics), and the conditional difference using the fourth-from-the-right specification from Figure 6. The specification and sample variants are (percentage of the sample dropped in parentheses): (2) drop academics in Economics, Political Science, Sociology, and Business (12.4%), (3) drop academics in the Social Sciences (25%), (4) drop academics to whom we sent more generic emails (7%), (5) drop university-departments to which we sent more than ten emails (27%), (6) drop university-departments to which we sent more than five emails (61%), (7) outcome is meeting accepted within one day, and (8) outcome is meeting accepted within three days. Standard errors are clustered at the university-by-department-by-sender name-level. 95% confidence intervals are shown.

Figure A14: Vocal Academics Contribute More in Total to Democrat-Related Committees



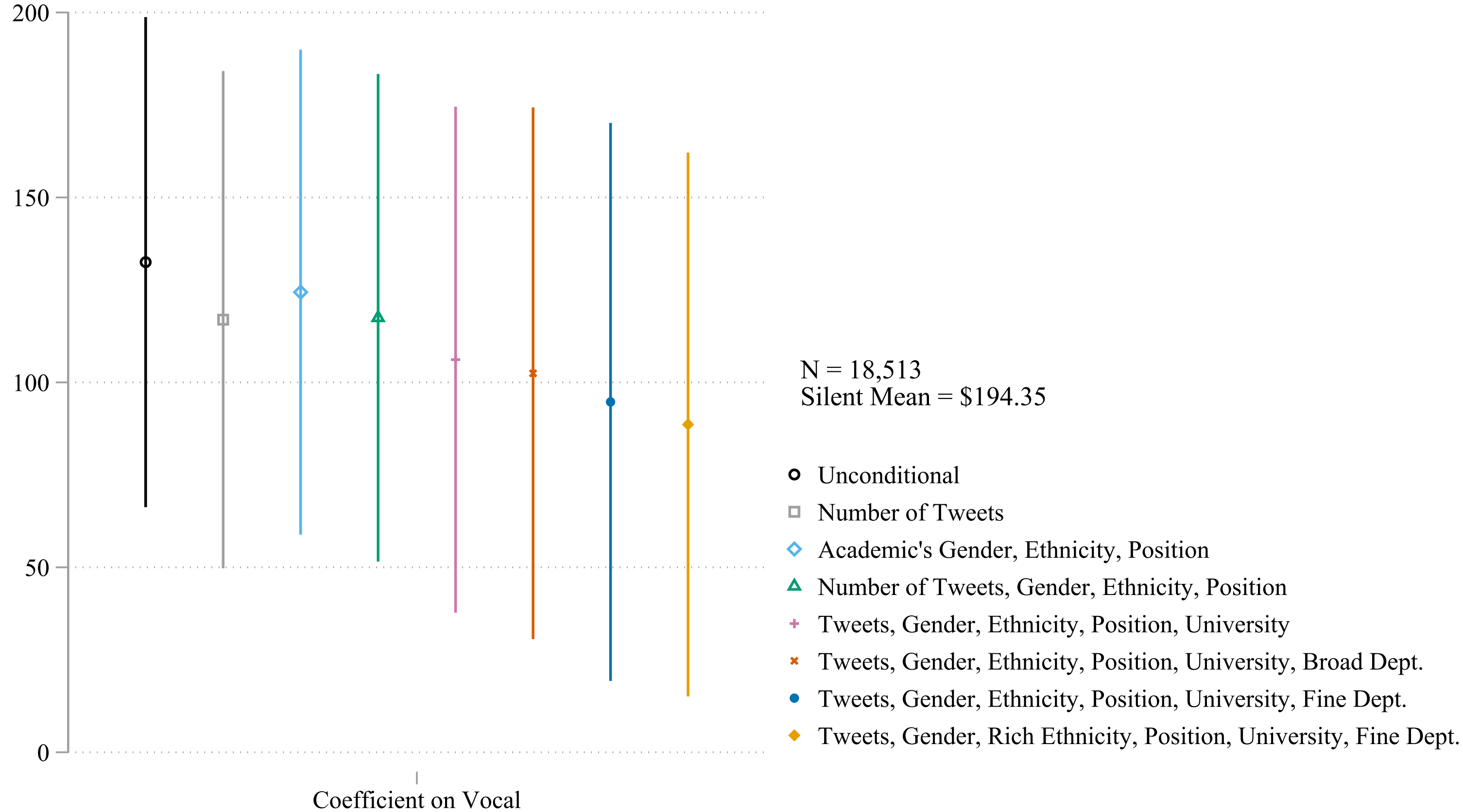
Notes: The bars show mean FEC-reported political contributions to Republican- and Democrat-related committees from January 2020 to March 2022. Silent includes the 6,784 non-Black academics that did not tweet about racial justice during the same time period, Vocal includes the 11,730 non-Black academics that did tweet about racial justice, Black includes the 1,094 tweeting Black Professors, and and Off Twitter includes the random sample of 900 non-Black academics without Twitter accounts. Unconditional raw means with 95% confidence intervals are shown.

Figure A15: Vocal Academics Contribute More Often to Democrats, Even After Adding Controls



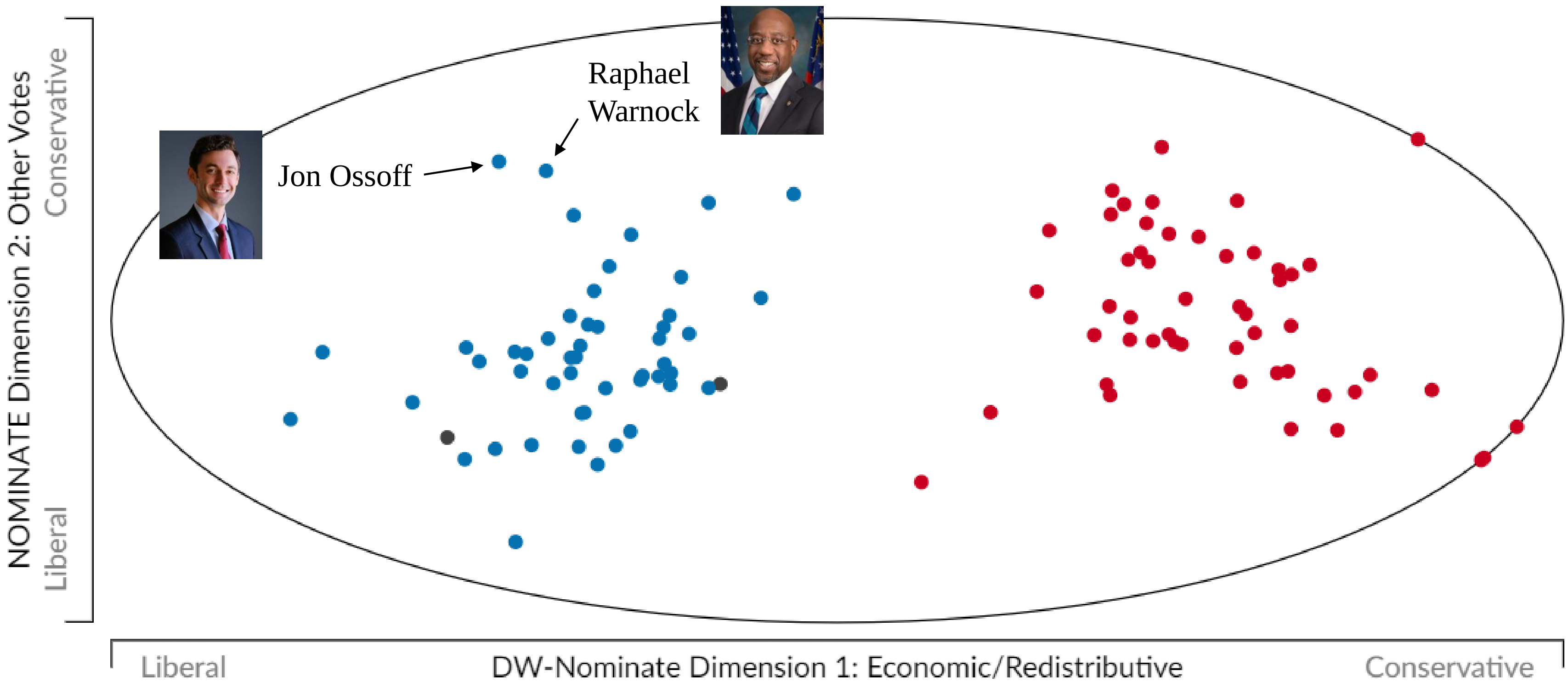
Notes: The figure shows the Vocal-Silent difference in whether contributed at least once to Democrat-related committees from January 2020 to March 2022. The far-left coefficient shows the unconditional difference, while the remaining coefficients add controls: (1) Number of Tweets: the number of original tweets, reply tweets, retweets, quote tweets, and quote reply tweets, all for the period January 1, 2020 to March 27, 2022; (2) dummy variable for female, dummy variables for Assistant Professor and Associate Professor, dummy variables for race/ethnicity, coded by RAs; (3) includes both the tweet variables from (1) and the demographics variables from (2); (4) adds university fixed effects; (5) adds broad department dummy variables (seven departments, e.g. Social Sciences); (6) replaces broad departments with narrowly defined department dummy variables (75 departments, e.g. Economics); (7) adds fixed effects for the first and second most-likely race predicted by `namsor.app` using the academic's full name, and the same fixed effects for first and second most-likely ethnicity. Standard errors are robust and 95% confidence intervals are shown.

Figure A16: Vocal Academics Contribute More in Total to Democrats, Even After Adding Controls



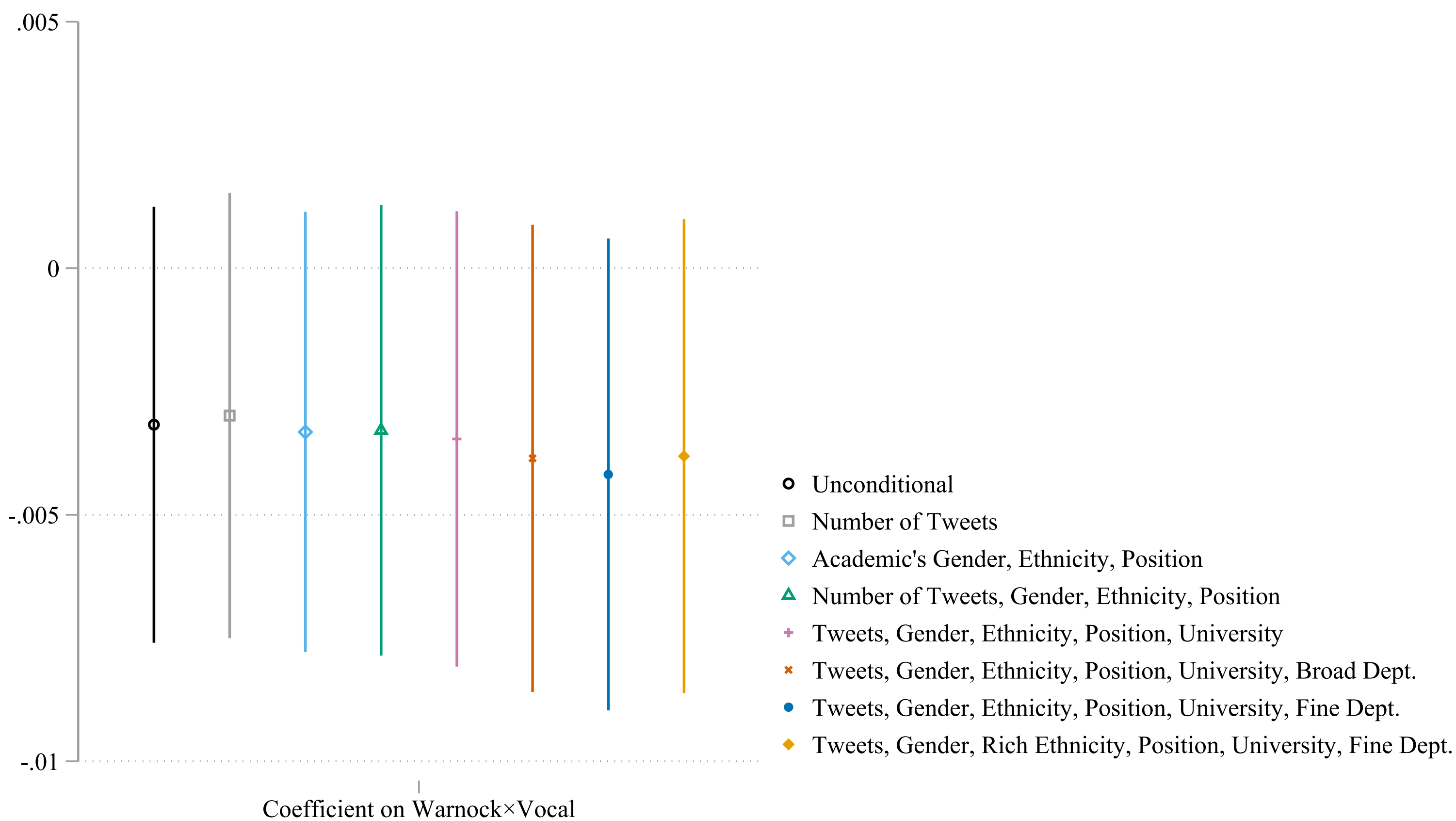
Notes: The figure shows the Vocal-Silent difference in total contributions to Democrat-related committees from January 2020 to March 2022. The far-left coefficient shows the unconditional difference, while the remaining coefficients add controls: (1) Number of Tweets: the number of original tweets, reply tweets, retweets, quote tweets, and quote reply tweets, all for the period January 1, 2020 to March 27, 2022; (2) dummy variable for female, dummy variables for Assistant Professor and Associate Professor, dummy variables for race/ethnicity, coded by RAs; (3) includes both the tweet variables from (1) and the demographics variables from (2); (4) adds university fixed effects; (5) adds broad department dummy variables (seven departments, e.g. Social Sciences); (6) replaces broad departments with narrowly defined department dummy variables (75 departments, e.g. Economics); (7) adds fixed effects for the first and second most-likely race predicted by `namsor.app` using the academic's full name, and the same fixed effects for first and second most-likely ethnicity. Standard errors are robust and 95% confidence intervals are shown.

Figure A17: Ossoff and Warnock Vote Similarly Once in Office



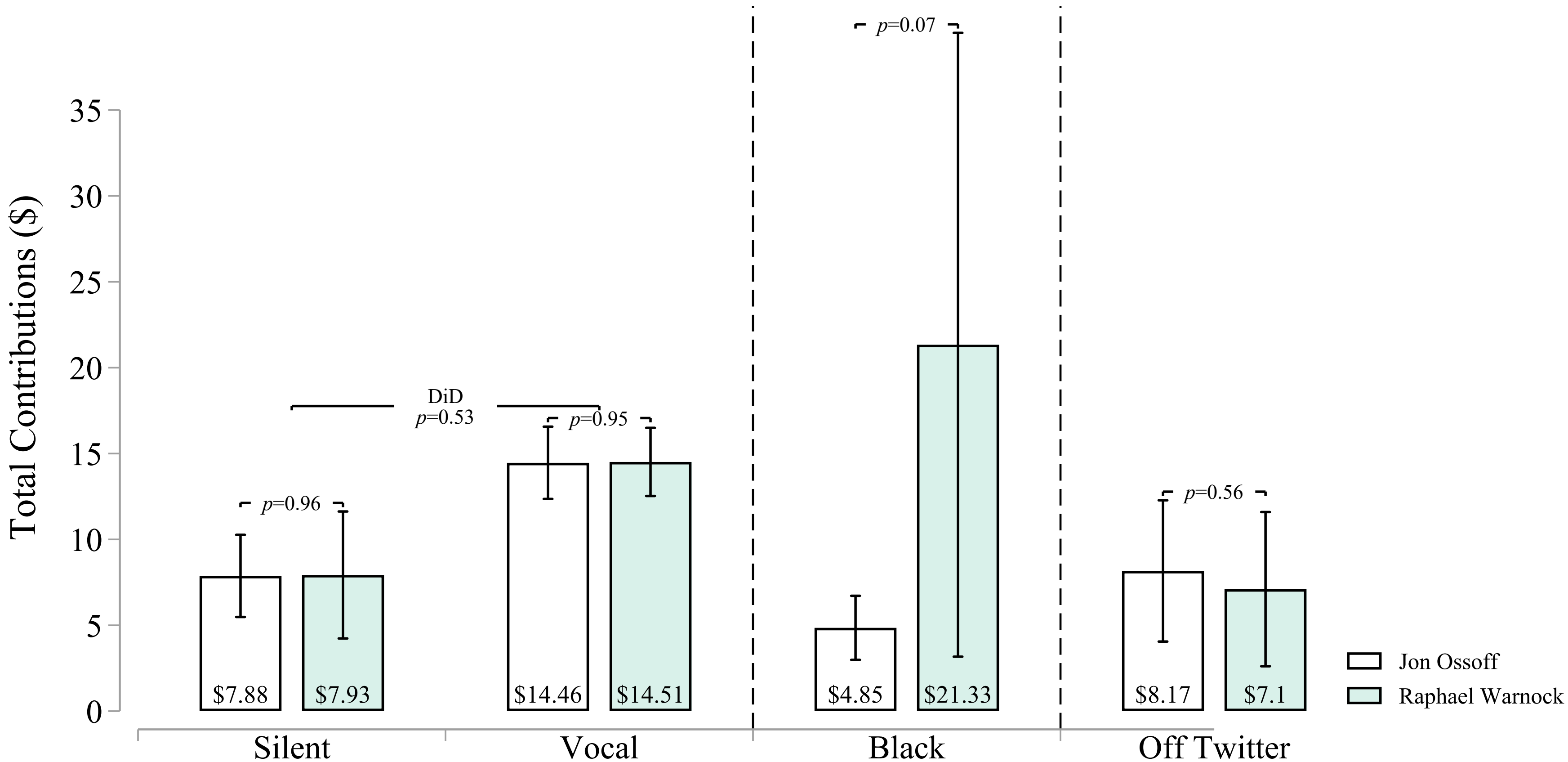
Notes: The figure shows the DW-Nominate scores of US Senators during the 117th Congress (2021-23). Blue dots are Democrats, red dots are Republicans, and grey dots are Independents. Downloaded from voteview.com on November 22, 2022.

Figure A18: Vocal Academics Are No More Likely to Give to Warnock Over Ossoff, Conditional on Controls



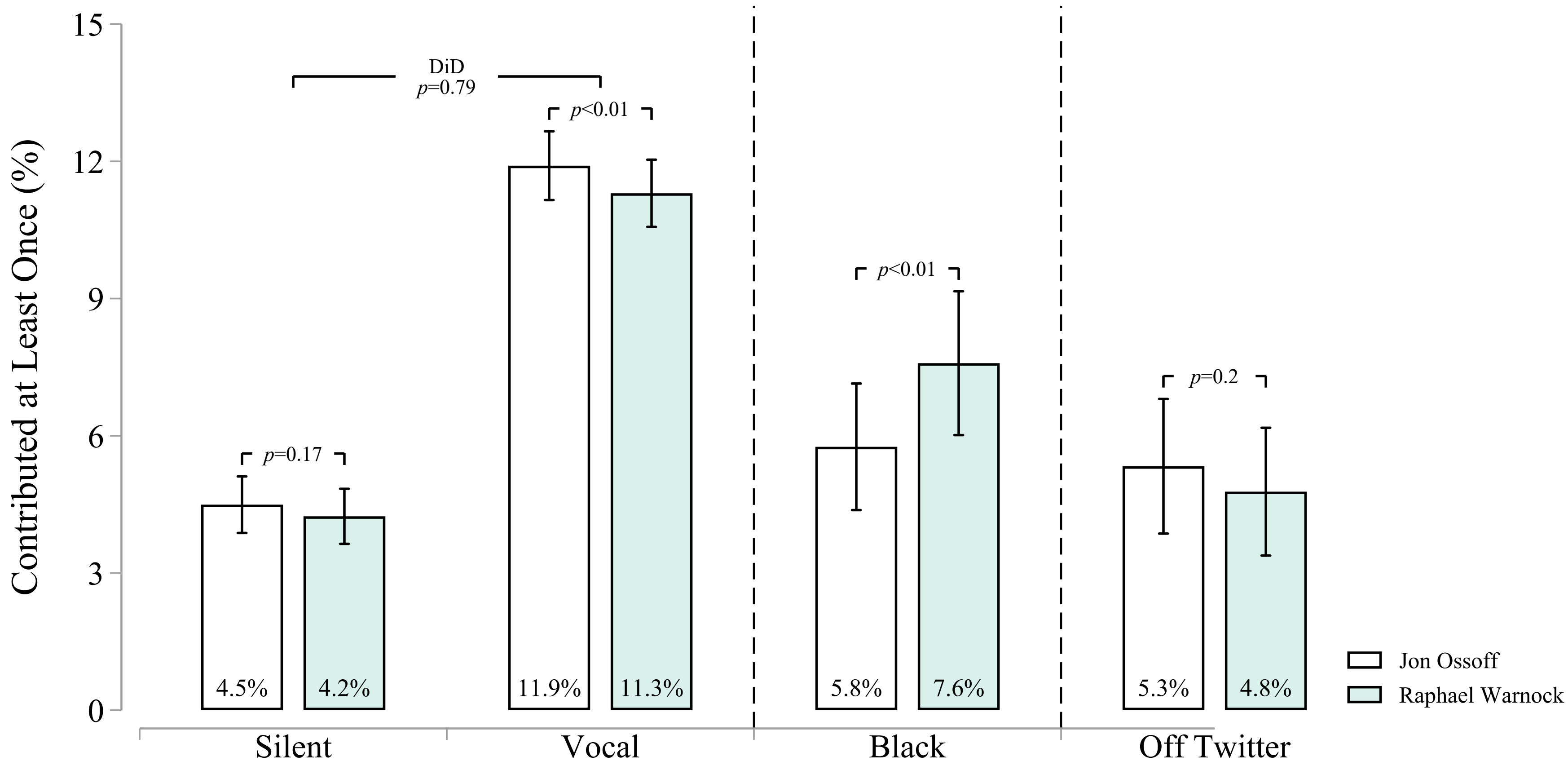
Notes: The figure shows how the difference-in-difference (DiD) coefficient from Figure 9 changes after adding controls, paralleling Figure 6. The far-left coefficient shows the unconditional difference, while the remaining coefficients add controls: (1) Number of Tweets: the number of original tweets, reply tweets, retweets, quote tweets, and quote reply tweets, all for the period January 1, 2020 to March 27, 2022; (2) dummy variable for female, dummy variables for Assistant Professor and Associate Professor, dummy variables for race/ethnicity, coded by RAs; (3) includes both the tweet variables from (1) and the demographics variables from (2); (4) adds university fixed effects; (5) adds broad department dummy variables (seven departments, e.g. Social Sciences); (6) replaces broad departments with narrowly defined department dummy variables (75 departments, e.g. Economics); (7) adds fixed effects for the first and second most-likely race predicted by `namsor.app` using the academic's full name, and the same fixed effects for first and second most-likely ethnicity. Standard errors are robust and 95% confidence intervals are shown.

Figure A19: Vocal Academics Give No More to Warnock Over Ossoff



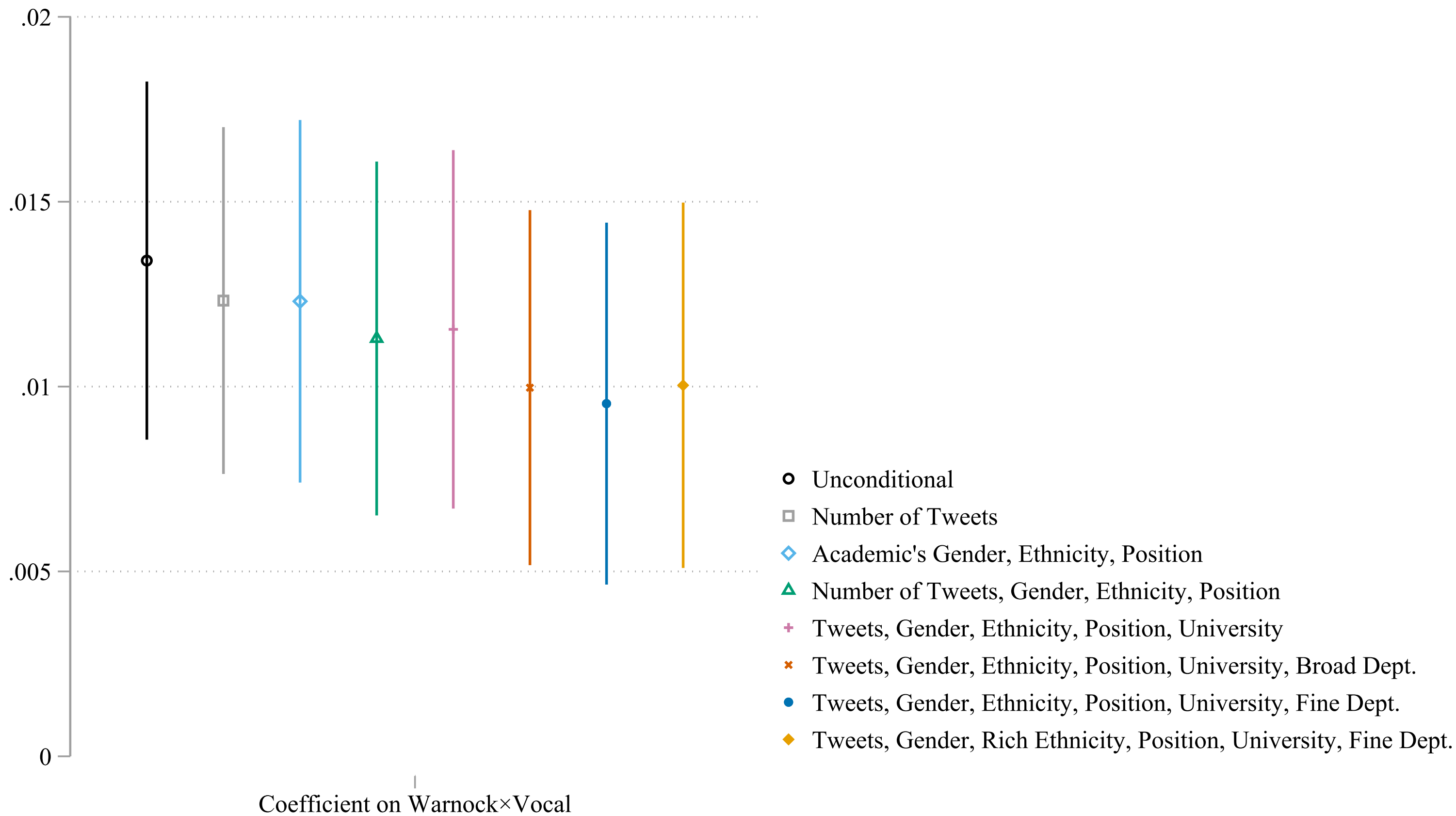
Notes: The bars show mean FEC-reported political contributions to the Senate campaigns of Jon Ossoff (a White Democratic candidate in Georgia) and Raphael Warnock (a Black Democratic candidate in Georgia) from January 2020 to March 2022. Silent includes the 6,784 non-Black academics that did not tweet about racial justice during the same time period, Vocal includes the 11,730 non-Black academics that did tweet about racial justice, Black includes the 1,094 tweeting Black Professors, and and Off Twitter includes the random sample of 900 non-Black academics without Twitter accounts. Unconditional raw means with 95% confidence intervals are shown.

Figure A20: Vocal Academics Are No More Likely to Give to Warnock Over Ossoff (Audited Sample)



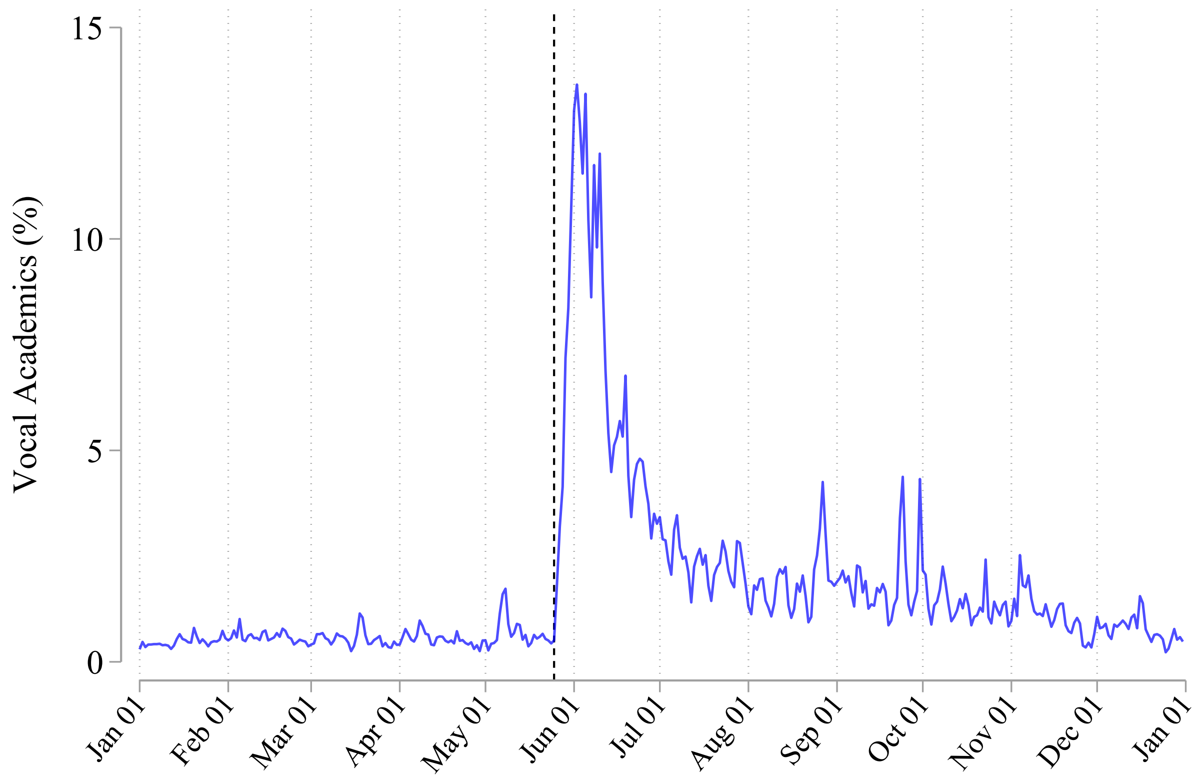
Notes: The bars show what percentage of audited academics made FEC-reported political contributions to the Senate campaigns of Jon Ossoff (a White Democratic candidate in Georgia) and Raphael Warnock (a Black Democratic candidate in Georgia) from January 2020 to March 2022. Silent includes the 4,318 audited non-Black academics that did not tweet about racial justice during the same time period, Vocal includes the 7,132 audited non-Black academics that did tweet about racial justice, Black includes the 1,094 tweeting Black Professors, and and Off Twitter includes the random sample of 900 non-Black academics without Twitter accounts. Unconditional raw means with 95% confidence intervals are shown.

Figure A21: Vocal Academics Are More Likely to Tweet About Warnock Over Ossoff, Conditional on Controls



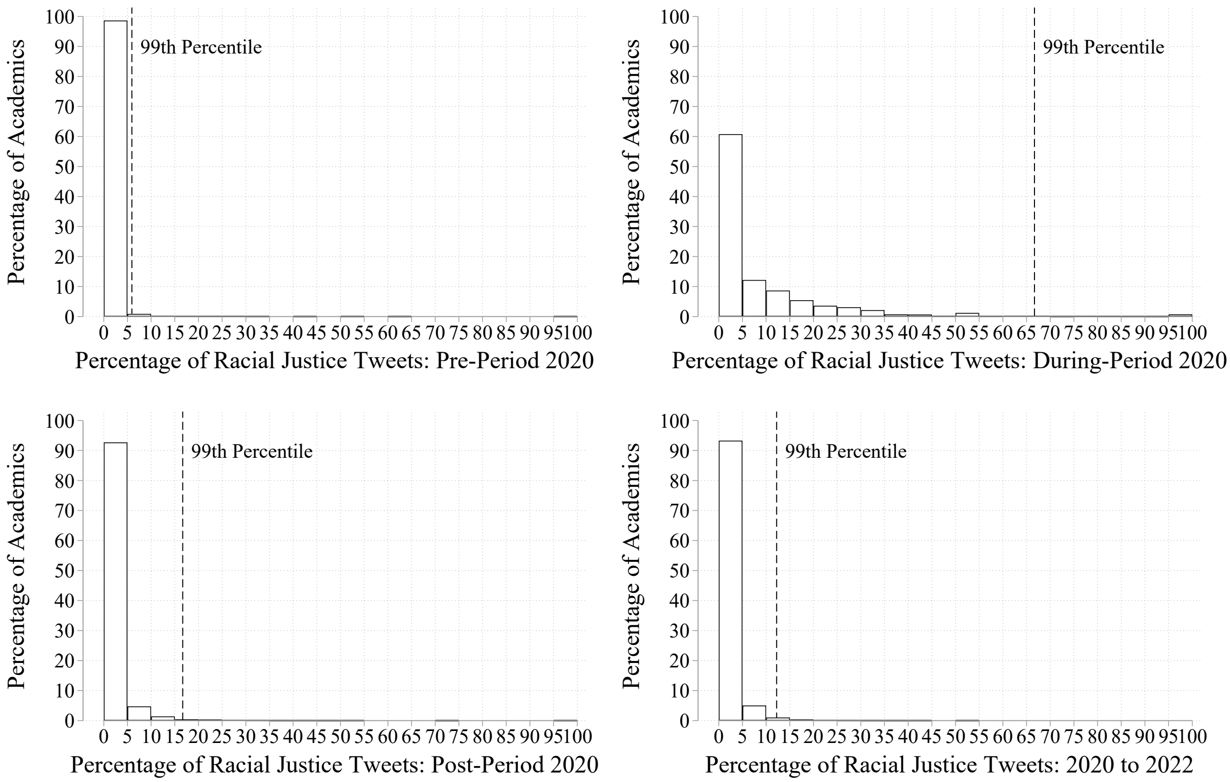
Notes: The figure shows how the difference-in-difference (DiD) coefficient from Figure 10 changes after adding controls, paralleling Figure 6. The far-left coefficient shows the unconditional difference, while the remaining coefficients add controls: (1) Number of Tweets: the number of original tweets, reply tweets, retweets, quote tweets, and quote reply tweets, all for the period January 1, 2020 to March 27, 2022; (2) dummy variable for female, dummy variables for Assistant Professor and Associate Professor, dummy variables for race/ethnicity, coded by RAs; (3) includes both the tweet variables from (1) and the demographics variables from (2); (4) adds university fixed effects; (5) adds broad department dummy variables (seven departments, e.g. Social Sciences); (6) replaces broad departments with narrowly defined department dummy variables (75 departments, e.g. Economics); (7) adds fixed effects for the first and second most-likely race predicted by `namsor.app` using the academic's full name, and the same fixed effects for first and second most-likely ethnicity. Standard errors are robust and 95% confidence intervals are shown.

Figure A22: Racial Justice Tweeting Spiked After the Murder of George Floyd



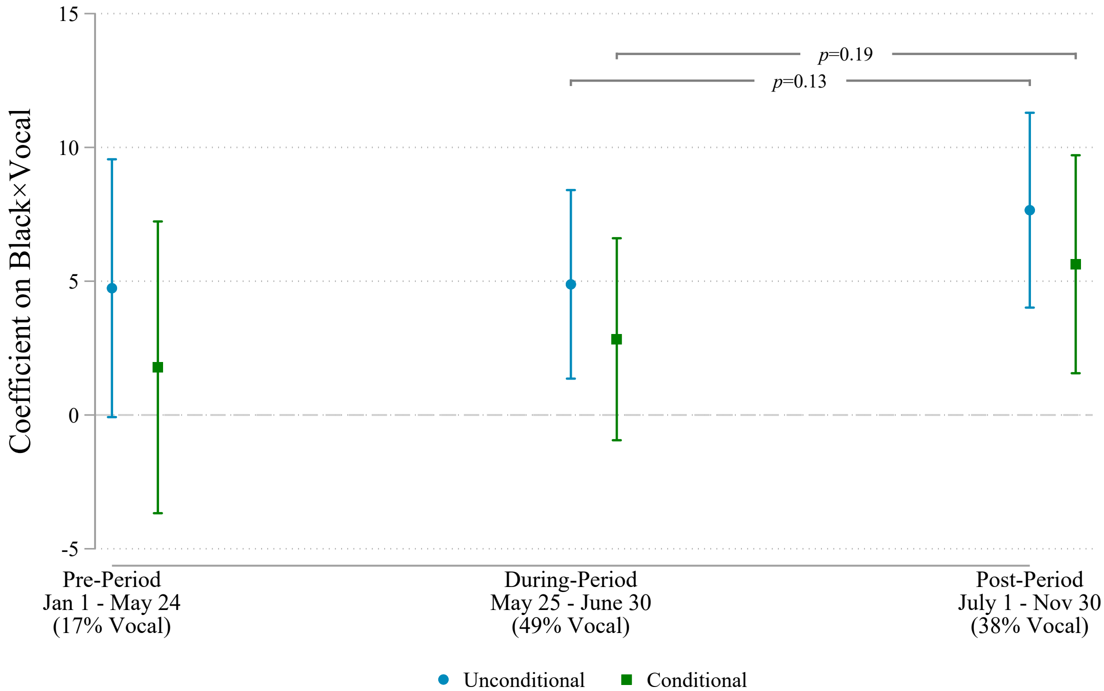
Notes: The figure shows the percentage of audited academics ($N = 11,450$) that tweeted about racial justice each day in 2020. See Section 2.3 for details on how we identify racial justice tweets. The vertical dashed line denotes the murder of George Floyd on May 25th.

Figure A23: Variation in Percentage of Racial Justice Tweets



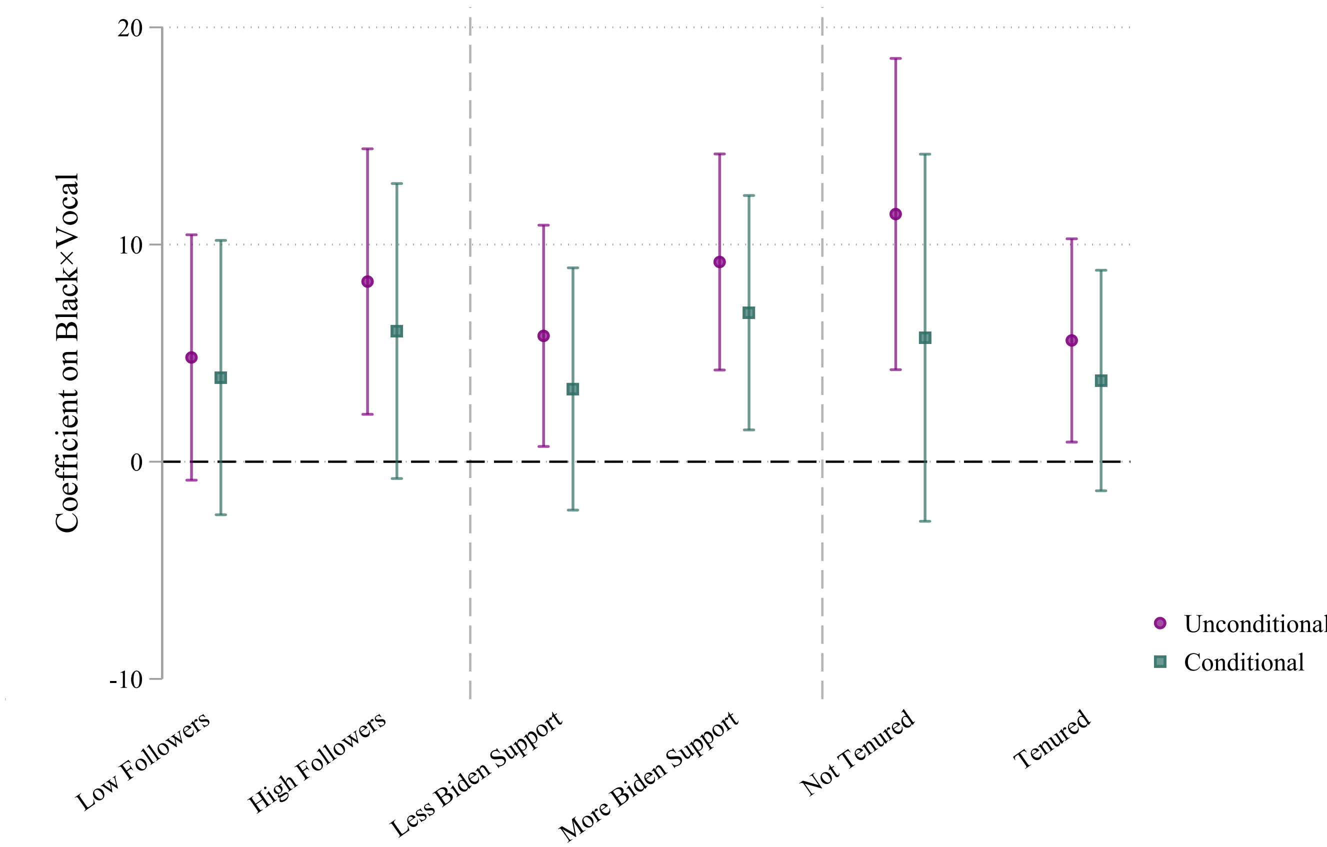
Notes: The figure plots histograms for the 11,450 audited academics for four variables: the percentage of each academic's tweets that are about racial justice (i) during January 1 to May 24, 2020 (the pre-period), (ii) May 25 to June 30, 2020 (the during-period), (iii) July 1 to November 30 (the post-period), and (iv) January 1, 2020 to March 27, 2022 (the full period). The vertical dashed lines denote the 99th percentile, which we use for winsorizing. The percentage of racial justice tweets is set to zero (rather than undefined) for those that did not tweet at all during a given period.

Figure A24: Informativeness Is Higher When Fewer People Are Tweeting About Racial Justice (Binary Measure)



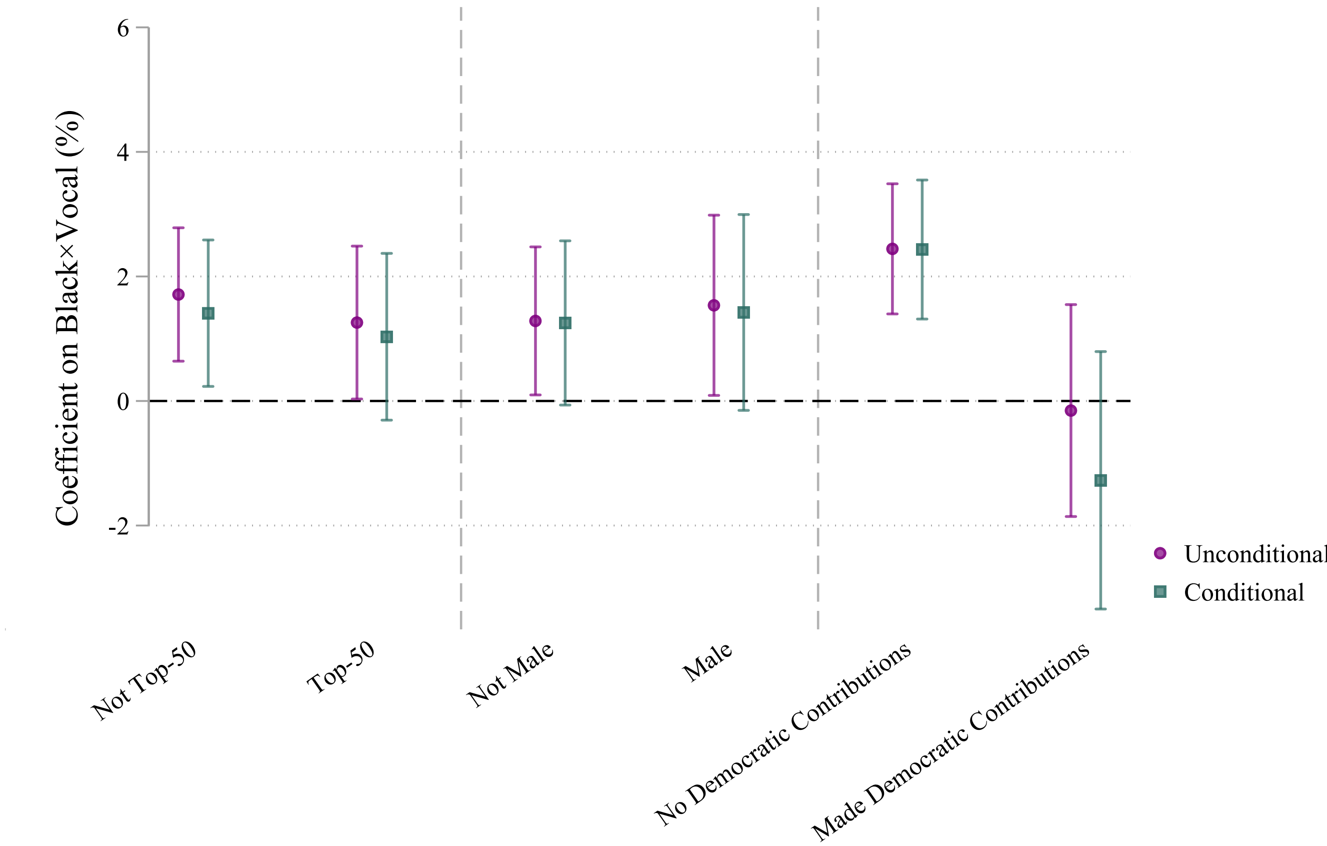
Notes: The figure shows how tweet informativeness ($\hat{\gamma}_2$ from specification 2) changes before, during, and after the murder of George Floyd on May 25, 2020. 17% of the 11,450 audited academics tweeted about racial justice at least once during January 1 to May 24, rising to 49% during May 25 to June 30, and falling to 38% during July 1 to November 30. The unconditional estimate denotes the unconditional difference in audit-measured racial discrimination between the Vocal and Silent during each period. The conditional estimate denotes the conditional difference in discrimination, using the fourth-from-the-right specification from Figure 6. These estimates are positive when Vocal academics discriminate against Black students less than Silent academics. Standard errors are clustered at the university-by-department-by-sender name-level. 95% confidence intervals are shown.

Figure A25: Informativeness by User Type (Binary Measure)



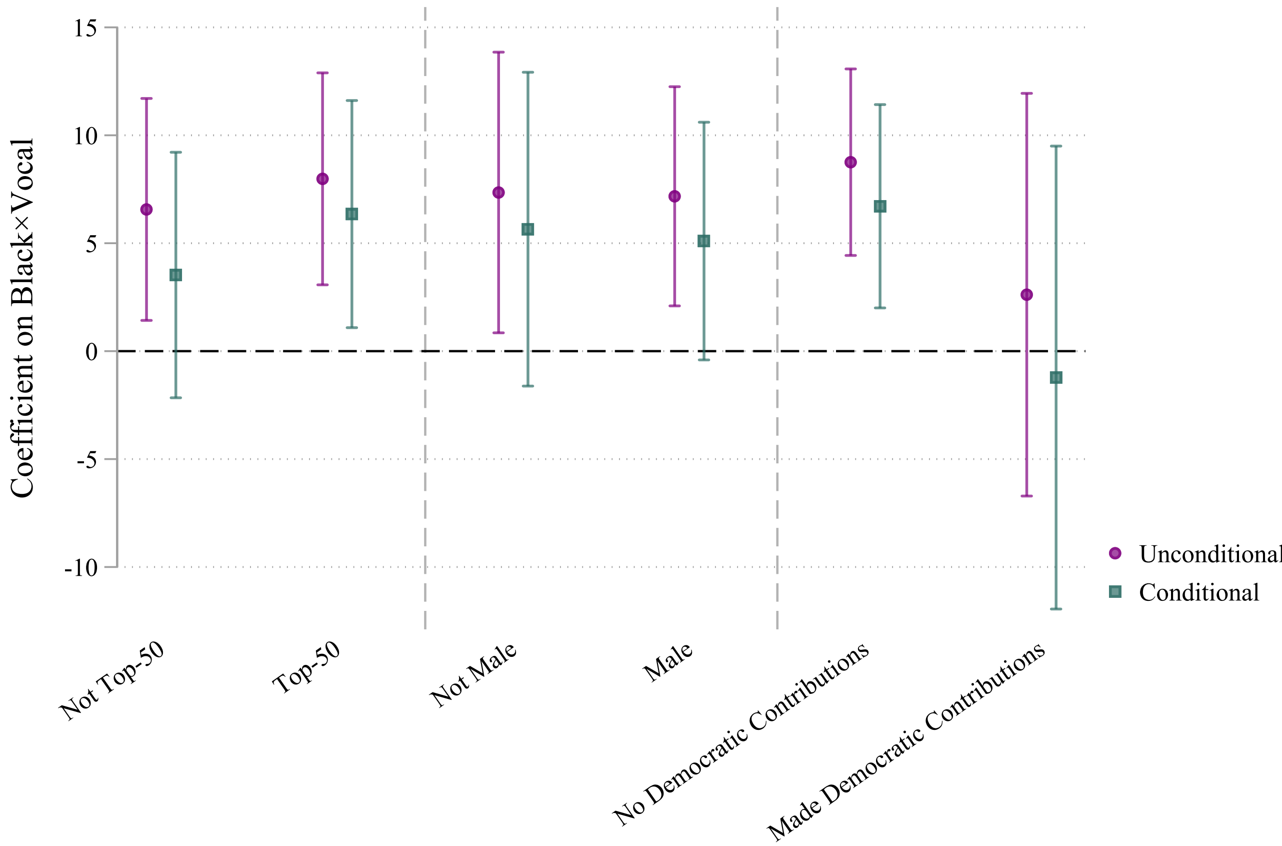
Notes: The figure shows how racial justice tweet informativeness ($\hat{\gamma}_2$ from specification 2, i.e. the difference in discrimination rates between Vocal and Silent academics) differs for different sets of our 11,450 audited academics: (i) below- versus above-median number of Twitter followers, (ii) below- versus above-median Biden 2020 vote share in the university's county, and (iii) Assistant Professors versus Associate and Full Professors. The unconditional estimate denotes the unconditional difference in audit-measured racial discrimination between the Vocal and Silent academics. The conditional estimate denotes the conditional difference in discrimination, using the fourth-from-the-right specification from Figure 6. These estimates are positive when Vocal academics discriminate against Black students less than Silent academics. Standard errors are clustered at the university-by-department-by-sender name-level. 95% confidence intervals are shown.

Figure A26: Informativeness by User Type: Rank, Gender, Politics



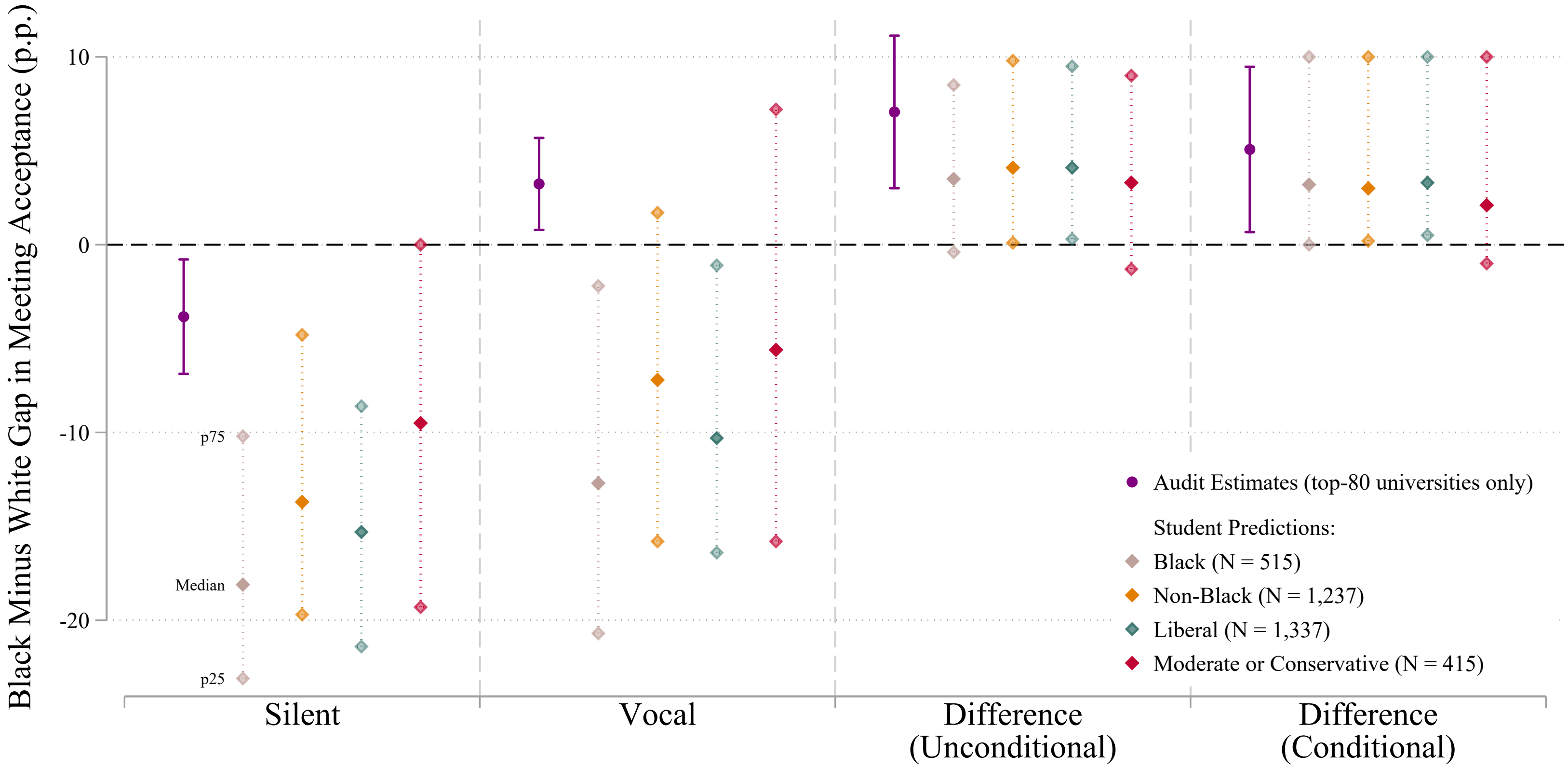
Notes: The figure shows how racial justice tweet informativeness ($\hat{\gamma}_2$ from specification 2, though with Vocal $(\%)_i$ replacing Vocal$_i$) differs for different sets of our 11,450 audited academics: (i) top-50 versus non-top-50 universities, (ii) male versus non-male academics, and (iii) no contributions versus at least one contribution to Democrat-related FEC committees from January 1, 2020 to March 27, 2022. The unconditional estimate denotes the reduction in audit-measured anti-Black discrimination associated with a one percentage point increase in the percentage of racial justice tweets (Vocal $(\%)_i$). The conditional estimate denotes the conditional difference in discrimination, using the fourth-from-the-right specification from Figure 6. Vocal $(\%)_i$ is winsorized at the 99th percentile. Standard errors are clustered at the university-by-department-by-sender name-level. 95% confidence intervals are shown.

Figure A27: Informativeness by User Type: Rank, Gender, Politics (Binary Measure)



Notes: The figure shows how tweet informativeness ($\hat{\gamma}_2$ from specification 2, i.e. the difference in discrimination rates between Vocal and Silent academics) differs for different sets of our 11,450 audited academics: (i) top-50 versus non-top-50 universities, (ii) male versus non-male academics, and (iii) no contributions versus at least one contribution to Democrat-related FEC committees from January 1, 2020 to March 27, 2022. The unconditional estimate denotes the unconditional difference in audit-measured racial discrimination between the Vocal and Silent academics. The conditional estimate denotes the conditional difference in discrimination, using the fourth-from-the-right specification from Figure 6. These estimates are positive when Vocal academics discriminate against Black students less than Silent academics. Standard errors are clustered at the university-by-department-by-sender name-level. 95% confidence intervals are shown.

Figure A28: Students Overestimate Discrimination and Tend to Underestimate Informativeness (Top-80 Universities)



Notes: The figure shows our audit study estimates and 95% confidence intervals in purple (following Figure 7), including only professors in the top-80 universities (the same universities covered by the graduate student survey). The diamonds denote the 25th, 50th, and 75th percentile of student predictions, separately by (i) students that self-identify as Black or African American versus students that do not, and (ii) students that describe their political views as liberal or very liberal versus those that describe their political views as moderate, conservative, or very conservative. Before making predictions, students were informed of the meeting acceptance rate for distinctively White student names, separately for Silent and Vocal academics.

A The Village Team Members

This project was only possible because of the dedication of the following research assistants (most of whom are UBC undergraduates):

G.O.A.T.:

Akash Uppal
Albena Vassileva
Aurellia Sunarja
Carla Colina
Carlos Perez Cavero
Chihiro Tanigawa
Colby Chambers
Conor McCaffrey
Daniella Rolle
Esha Vaze
Eugene Kwok
Jiayu Li
Jordan Hutchings
Kevin Yu
Laura Truong
Louise Cheng
Maria Ines Moran
Noor Kumar
Saloni Sharma
Shardha Nayar
Tierra Habedus-Sorensen
Vinayak Kalra
Yash Ahlawat

Tier 1:

Angela Fan
Bryanna Li
Jaida Smith
Nicholas Latimer
Olivia Klaassen
Angela Lee
Ty Stevenson
Alex Dyky
Amir Ala'a
Avreeta Sandhu
Billy Lam
Cynthia Cui
Gabriel Odeyemi
Jennifer Tran
Jenny Qin
Jeremy Singer
Julianne Nina Marie Uy
Karman Phuong
Kaye Thinh-To
Keshikaa Suthaaharan
Kevin Li
Kevin Tan
Mahrukh Khan
Marianne Sigouin
Maxwell Martel
Nela Radecki
Nikita Gautham
Paulino Tan
Robin Jhatu
Sabine Villaroman
Shahed Salahi
Sophia Huang
Sophia Samilski
Trang Truong
Vanessa Cheung
Yu Fei
Harpreet Khattar
Lulu Wang
Minh Anh Pham
Rayan Aich
Uddhav Kalra

Tier 2:

Odmaa Bayartsogt

Ahana Thakur

Alanna Man

Alejandra Mercadillo

Alejandro Solano Romero

Amit Biswas

Ana Beatriz Pereira

Anahat Kaur Chahal

Anastasia Mishina

Andrea Bartolome

Angela Sequeira

Angela Villavicencio

Becky Zhu

Bhakthie Senanayake

Chris Cheuk Yin Lam

Chris Haun

Connor Wiesner

Deniz Sagnak

Dhairya Chaudhri

Dhruv Bhatia

Emi Oyakawa

Emma Borhi

Erik Bruendl

Fariha Sultana

Florent Dusenge

Francesca Tamberi

Genevieve Moody

Huiqi Liu

Jacqueline Lee

James Brewster

Jane Platt

Jiayi Zhu

Jingyi (Olivia) Cao

Jinmeng Xu

Joaquin Glinoga

Joey Dolayba

Judith Sofiana Haryanto

Julian Kwan

Julianne Louie

Junye Xu

Kaitlin Khu

Karam Kanwar

Karen Liu

Kunwar Modi

Lauren Snow

Liri Zou

Liugu Tan

Lucas Mehling

Marco Lanfranchi

Matteo Tan Zheng Hao

Mridul Manas

Naoki Sakura

Navkiran Takhar

Nicholas Harterre

Percy Chen

Ravi Rinarco

Rohan Prasad

Ruby Taylor

Ruolin Mo

Sanchita Sannigrahi

Sarah Mejia

Sarir Parvizi

Shady Abo El Kasim

Shreya Iyer

Siddhant Kumar

Sirui Bi

Smriti Sukhani

Surotama (Suri) Banerjee

Syra Dhaliwal

Tim Qiao

Tosya Khodarkovsky

Valeria Zolla

Vivian Liu

Vivian Wei

Vyomesh Daga

Xiaoyan Wu

Xixi Xu

Yasin A Zahir

Yawen Zhang

Yuxuan Deng

B Deviations from the Pre-Analysis Plan

We posted our pre-analysis plan to the AEA registry on May 12, 2022, prior to the launch of the experiment. We updated the pre-analysis plan on May 24th to explain the detection-related logic for stopping the experiment before sending emails to the full set of 18,514 academics. We also explained a minor issue where we addressed a handful of emails to the wrong professors, leading us to drop 23 academics from the audit sample. We list other deviations from the pre-analysis plan here:

- In the pre-analysis plan we briefly described the graduate student survey, signalling that we hoped to collect a "third-party report of the behavior of academics at their institution." Given low response rates to our student survey in piloting, we decided not to ask for third-party reports on professors, helping us to keep the survey to roughly 10 minutes.

- We added analysis of informativeness of political contributions and tweets (i) to complement the audit analysis with a non-audit-related behavior, fulfilling somewhat our initially planned role for the third-party student reports, and (ii) to explore the political signalling of racial justice tweets. In the pre-analysis plan we only specified that we would use political contributions as controls in the audit analysis.

- In the pre-analysis plan we signalled that we would omit the analysis of gender discrimination to use instead in a companion paper. We have followed the advice of a previous editor in adding the gender analysis to the current paper.

- Also following editorial advice, we have added analysis in which we compare the emails sent from any minority group (Black, female, or first-generation) to those sent by White males with no mention of first-generation status. This analysis has the advantage of giving a clean measure of minority vs. non-minority treatment, as opposed to our other comparisons (e.g. male vs. female), where by design, 3/4 of the 'majority' group (i.e. male) belongs to one of the other two minority groups (i.e. Black or first-generation).

- While we calculated our key measure Vocal_i prior to running the audit experiment, we since realised that we were unintentionally using text from truncated retweets for roughly half of the academics. For the current paper, we have updated the measures of vocality to include mentions of racial justice-related words and phrases in the full text of the retweets.

C Procedure for FEC Contributions

Overview. To download and link FEC-reported political contributions, we follow the detailed data appendix of Bouton et al. (2022), with some adaptations to fit our context. In particular, while Bouton et al. (2022) aim to describe the donation patterns of *all* donors in the US (requiring them to assign each and every donation to a given donor), we only need to identify the donations of the academics in our dataset. This simplifies the matching process, since the occupation and employer variables in the FEC data are particularly useful for cases in which the target population all have a similar occupation.

Targeting Academics. First, we web-scraped all FEC-reported individual contributions from January 1, 2020 to March 27, 2022 from here. We carried out basic cleaning checks as in Bouton et al. (2022) – dropping duplicates and dropping those from "lines" other than 11A(i) and 17A(i) (these lines denote contributions from individuals).

Second, we used the employer variable to keep only contributions from individuals employed by the top-150 universities, allowing for abbreviations (e.g. UCSB instead of University of California, Santa Barbara), other major name variants, and common misspellings.

Third, we used the occupation variable to keep only contributions from individuals that might be research-active academics (e.g. Professor, Scientist, Historian, etc.). We then carried out basic cleaning checks of the first and last names of contributors in this smaller dataset of contributions.

Matching. To link with our dataset of academics, we looked for perfect matches on first name, last name, and university. We allowed for nickname variants of each first name using the American English Nickname Collection from the Linguistic Data Consortium at UPenn.

Contribution Characteristics. Each contribution in the data has a committee ID, though many contributions go to conduits (especially ActBlue and WinRed) that then channel the donation to a final committee. As much as possible, we identified the final committee of a donation using the memo_text and the receipt_type_full variables. We then used FEC-provided crosswalks from here to merge on the characteristics of each committee – including any linked candidates and their political parties. All of our measures of contribution type then denote features of the final committee to which the contribution is directed.[39]

To determine the political party of each contribution, we first used the political party of the connected candidates, if they exist. For the committees without connected candidates, we used online sources to establish which party the committee is primarily raising funding for. We identify contributions going to Raphael Warnock as those with final committee ID C00736876 or C00740597; and for Ossoff, IDs CC00718866 or C00750919. There were three contributions in our matched data to the Ossoff-Warnock Victory Fund (ID:C00761163) – we count these three contributions as contributions to both candidates.

[39] As explained in Bouton et al. (2022), this also requires us to drop duplicate contributions in cases where the conduit and the final committee both reported the same contribution.

D Measuring Racial Justice Signalling: Minor Details

We classify an academic as Vocal if they have at least one tweet from January 1, 2020 to March 27, 2022 that mentions at least one of the racial justice-related words or phrases listed in Section 2.3. To identify words or phrases, we use the following approach. We call the following characters 'spacing characters' ; : + ? . - _

We make all tweet text lower-case.

For the word "blm", to avoid false positives (e.g. larger words unrelated to racial justice that include the adjacent letters "blm"), we require the blm letters to be:

- Preceded by: nothing, hashtag, spaces, open parentheses, or a spacing character, **AND**
- Followed by: nothing, spaces, close parentheses, or a spacing character.

For all other words and phrases, we disregard preceding and following letters. For phrases, we allow for spaces between each word in the phrase, or for one of the spacing characters (this is particularly relevant for URLs, where spaces are not possible).

E Graduate Student Survey: Prediction Questions

Overall Discrimination Prediction:

We ran an experiment in May to measure racial discrimination in academia. **We sent emails from fictitious students to roughly 11,000 non-Black academics** at top-150 US universities. As we wanted to see how Twitter activity predicts email responses, **we included only academics with Twitter accounts**. Each academic received one email, and **each email requested a Zoom meeting to discuss the possibility of graduate studies**.

Half of the emails were from typically White-sounding names like Owen Wood and Helena Bennett. The remaining emails were from typically Black-sounding names like Lamar Jenkins and Taliyah Williams. Emails from Black-sounding names had similar content to those from White-sounding names. This means that we can measure racial discrimination by comparing the meeting acceptance rate for the two types of names.

We found that 30.6% of meeting requests sent from White-sounding names were accepted.

What percentage of meeting requests sent from Black-sounding names would you guess were accepted?

Predicting Unconditional Differences in Discrimination:

Among the academics we emailed, 62% posted at least one racial justice-related tweet in the two years prior to the experiment. These tweets were almost always **in support of racial justice-related efforts**. We would now like you to guess the email response rates separately for those that tweeted about racial justice and those that did not.

31.3% of meeting requests sent from White-sounding names were accepted by **academics that tweeted about racial justice**.

What percentage of meeting requests sent from Black-sounding names would you guess were accepted by these academics?

29.5% of meeting requests sent from White-sounding names were accepted by **academics that did NOT tweet about racial justice.**

What percentage of meeting requests sent from Black-sounding names would you guess were accepted by these academics?

Predicting Conditional Differences in Discrimination:

Note: the wording of this question differs slightly based on whether the respondent previously guessed that academics that tweet about racial justice discriminate against Black students less, more, or the same as those that don't tweet about racial justice. The version here is for those that think they discriminate less.

Combining your last two answers, **your guess is that academics that have tweeted about racial justice discriminate against Black students ${e://Field/discrimDiff} percentage points less than the academics that did not**.

There can be many differences between the racial justice tweeters and the rest – for example, those that tweet about racial justice may be more likely to be women, or more likely to be junior academics. We now want to ask you what difference in discrimination you would expect after controlling for these factors.

Specifically, suppose you know of two professors of the same rank in the same department and university. They also share the same gender and race/ethnicity and tweet the same amount. But one of the professors tweeted about racial justice in the past two years and the other did not. What would you expect the difference in racial discrimination to be between these two professors?

[Note: enter a negative number if you expect the racial justice tweeter to discriminate against Black students more than the other academic. If you think that accounting for the academics' characteristics does not matter, enter $e://Field/discrimDiff (your answer from before).]

F Email Example

Subject: Some questions about «**recipientDegree**»s

Main Body:

Dear Professor «**recipientFullName**»,

I came across your academic work since I am considering applying this fall to «**recipientDegree**»s «**recipientField**».

While I have done a fair amount of research online, I still feel quite unsure about what «**recipientDegree**» life is like exactly, and whether I would be well-suited for it. «**This is probably partly due to me being a first generation college student.**»

Though I am sure you are very busy, would you have any spare time in the next week or so to answer some of my questions over a call?

I would be very grateful for the help, though I understand if it is not possible.

Thanks, «**senderName**»