

Myths of Censorship: The Realities and Misperceptions of “Cancel Culture”

Nicholas Dias

University of Pennsylvania

James Druckman

Northwestern University and IPR

Matthew Levendusky

University of Pennsylvania

Version: January 21, 2022

DRAFT

Please do not quote or distribute without permission.

Abstract

Few principles are as central to American democracy as freedom of speech. Yet, some argue “cancel culture”—i.e., censoring offensive speech—undermines this crucial tenet. The authors offer a theory of why people “cancel” others and test it using a conjoint experiment with a representative sample of Americans. They find that when Americans engage in canceling, they do so because of what was said, regardless of the speaker’s identity. Cancellation reflects an attempt to redress speech considered harmful, not punishment borne of partisan or racial animosity. But the researchers also show that the public is significantly misinformed about cancellation: People overestimate the extent to which canceling occurs and they misconstrue why it happens. Even though partisan bias does not cause canceling, (mis-)beliefs about canceling could exacerbate partisan animosity. These findings help to unravel the dynamics of contemporary American free speech.

The authors thank Kirsten Huh, Evan Myers, Uday Tandon, and Anna Wang for research assistance, and the Annenberg Public Policy Center for funding this research.

Few principles are as foundational to American democracy as freedom of speech. As Benjamin Franklin noted, without this right, “a free society is dissolved, and tyranny is erected on its ruins” (Shibley 2016; see also Gibson 1992). Even when confronted with offensive and hateful speech, Americans traditionally opt for “more speech, not enforced silence” (Brandeis 1927). Public opinion data bear this out: much more so than citizens of other democracies, Americans tolerate offensive speech in the name of protecting free speech more generally (Wike 2016).

But is this traditional commitment on the wane? Many argue that the rise of “cancel culture” implies that, rather than counter-arguing offensive speech, Americans now want to censor (i.e., cancel) it (see, e.g., Friesdorf 2015, Mounk 2020, *The Economist* 2021, Applebaum 2021). As Hobbes (2021) notes, cancel culture has escalated to a “moral panic,” particularly among those on the political right. Yet, while media warnings of cancel culture abound, empirical analyses of this supposed phenomenon do not. Consequently, we do not know how many Americans engage in cancel culture or why they do so.

We fill this gap by examining several interconnected issues: (1) how often Americans actually cancel others, (2) what leads them to do so; 3) whether they share an understanding about canceling (that is, is there actually a “cancel culture”); and 4) what motives Americans ascribe to those who cancel. We derive hypotheses about each of these topics, and use a pre-registered conjoint experiment with a representative sample of Americans to test them. Our predictions build from work on affective polarization, given that many argue it motivates canceling (e.g., Romano 2021, Wehner 2021).

Our empirical findings upend the dominant narrative about cancel culture. While many popular accounts argue that Democrats are the ones canceling others, we find only modest

partisan differences. Instead, we find that both Democrats and Republicans alike rarely cancel others. And when they do so, it is a response to speech they consider offensive or harmful, not punishment for members of groups they dislike—cancellation is about what people say, not who they are. Similarly, members of both parties argue they cancel to hold others accountable rather than to punish them. canceling is both uncommon, and just as likely to come from the political left as it is from the political right.

Nevertheless, important misperceptions abound: Americans assume that a plurality of *others*—and especially those from the other political party—actively cancel, and do so for nefarious and discriminatory reasons. Thus, “cancel culture” is no culture at all, but instead reflects fractionalized (mis-)understandings and antagonistic attributions that deepen partisan animosity. While the public’s attitudes toward free speech are shifting (Chong et al. 2021), the prevailing sense that others cancel far more than they do in reality, and for dubious reasons, fuels the aforementioned moral panic. The reality of when and why people do cancel offers sheds new light on contemporary public opinion about free speech.

The Meaning of Free Speech and Cancel Culture

The First Amendment to the U.S. Constitution protects free speech, and politicians and citizens alike view it as a fundamental right.¹ Classic studies show strong support for free speech, at least in the abstract, though that support varies depending on individual attributes, such as values, education, and socialization, as well as contextual circumstances, such as attitudes towards the group speaking, party cues, geography, and so forth (Sullivan et al. 1982, McClosky and Zaller

¹ However, politicians tend to exhibit greater tolerance for offensive speech than the publics they represent (Sullivan et al., 1993).

1984, Peffley et al. 2001, Chong 2006, Gibson 2008, Armstrong and Wronski 2019). While only a few legal rules exist that regulate hateful or offensive speech in the U.S. given first amendment jurisprudence (Schauer 2010), there is a considerable normative debate about the extent to which the U.S. should regulate such offensive speech (Dworkin 1996, Beerbohm 2012). Many argue that, rather than prohibiting hateful speech, a better response entails counter-arguing offensive speech with more speech (Brettschneider 2012, Howard 2019). This debate captures the essence of contemporary free speech controversies—the question no longer concerns just state regulation, but also how *citizens* should react to, and therefore regulate, speech extra-governmentally.

This shift stems from two key developments.² First, with the rise of social media, a large proportion of questionable speech now occurs in forums operating with unparalleled speed and reach. This alters the way speech and counter-speech can, and perhaps even should, work (Guo and Johnson 2020, Waisbord 2020). Second, the polarization of American politics has heightened incivility (Mutz 2015, Sydnor 2019, Druckman et al. 2019) and shifted conversations about how to regulate speech. Concepts of free speech developed in an era of newspapers may be ill-suited to the era of Twitter.

Indeed, the rise of social media is integral to the public debate over cancel culture. The concept of canceling someone originated in Black and queer communities (Clark 2020), with the idea being that those who express offensive ideas should be called out and shunned, or more colloquially, canceled. Social media allowed this idea to spread rapidly and enter mainstream consciousness, making cancel culture functionally symbiotic with social media (Romano 2020, Klein 2021). In recent years, a wide variety of public figures—from Congresswoman Marjorie

² Our focus here is on racist or hate speech. We thus do not consider other types of controversial speech such as leftist or militarist speech (Chong et al. 2021).

Taylor Greene, to Roseanne Barr, to JK Rowling, to Yale faculty members, to Mr. Potato Head—have allegedly been canceled. Of course, whether these individuals actually were canceled, and whether such cancellation is justified, depends on one’s point of view.

The debate over cancel culture centers on whether canceling reflects accountability or punishment (Pew Research Center 2021). Proponents of cancel culture focus on the harm done by offensive speech, especially to stigmatized groups, and the need to hold those who make offensive statements accountable. Opponents, by contrast, equate canceling to punishment or censorship that undermines America’s historic commitment to free speech. This marks a significant shift in the terms of the free speech debate. Rather than passive toleration, or counter-speech, the remedy for offensive speech involves censorship or shunning, at least in some circumstances.

Debates about cancel culture also take on a partisan and ideological cast. Historically, political liberals acted as the staunchest advocates of free-speech protections (e.g., Davis and Silver 2004, Linder and Nosek 2009, Downs and Cowan 2012). However, more recently, progressive liberals have led the charge for countering what they view as harmful language with “safe spaces” where racist, sexist, or other discriminatory words are not tolerated or allowed (Crockett 2016). Along these lines, Chong et al. (2021) report a dramatic shift between liberals and conservatives in free speech attitudes, particularly those concerning racist speech. It is now conservatives, not liberals, who are the stronger proponents of unfettered free expression, resulting in what the authors call “the most significant realignment of political tolerance in the United States in the past half-century” (6).

Indeed, Republicans increasingly bemoan Democrats’ intolerance as the perpetrators of cancel culture. This perspective emerges with particular vehemence on Fox News and other

conservative outlets (see Romano 2021). The 2020 Republican National Convention featured cancel culture as one of its key themes (Gomez 2020), and in a speech at Mt. Rushmore, President Trump inveighed against cancel culture, calling it a “far left fascism” (Trump 2020). Yet, whether there is actually a partisan dimension to canceling behavior is an important, but unanswered, question.

What Drives Canceling?

Canceling someone involves criticizing or censoring their offensive speech in some way. But what determines whether a given statement offends? We argue this depends on who the speaker is, as well as the content of what they say (Gibson and Anderson 1985, Gibson 2006). First, the speaker’s partisanship likely matters. The past quarter century has seen a dramatic rise in affective polarization, with individuals favoring those from their own party and disparaging those from the other party in a wide variety of circumstances (for reviews of these effects, see Iyengar et al. 2019, Finkel et al. 2020). These feelings spillover into censorship decisions: Lelkes and Westwood (2017) show that partisans support outlets publishing news content critical of the opposition, but not content critical of their own party,³ and journalist accounts argue that such animus drives cancel culture more generally (Wehner 2021). Thus, our partisan source hypothesis is *individuals will be more likely to cancel out-party speakers than in-party speakers, all else constant*.

The growth of affective polarization stems, in part, from each party becoming demographically more homogeneous (i.e., social sorting), especially with respect to race (Mason

³ Also see Westwood et al. (2019). Related work shows that partisans are less sensitive to uncivil speech from co-partisans (Mutz 2015, Muddiman 2017, Druckman et al. 2019, Gervais 2019), which should mean they are less sensitive to offensive content from those same individuals.

2018). Given this social sorting, Westwood and Peterson (2020) show that, when people update their affect toward the parties, they also update their affect toward racial groups and vice versa. Race and partisanship are nearly inseparable, with partisans connecting Democrats with Black Americans and Republicans with White Americans (Valentino and Zhirkov 2018). Our race source hypothesis is *individuals will be more likely to cancel speakers from racial groups linked with the out-party (than linked to their party), all else constant*. That is, Democrats will be more likely to cancel White individuals than Black individuals, and Republicans will be more likely to cancel Black individuals than White individuals, all else constant.

A speaker's positionality may also matter, as behavioral expectations accompany positions of power. For instance, constituents expect their elected officials to act on their behalf. As such, if officials speak or take actions deemed offensive to those constituents, they may be more likely to be canceled (as such actions are not representing their constituents). This also aligns with the idea that political elites should follow norms of civility and avoid offensive speech (Uslaner 1993, Jamieson and Hardy 2012). The same can be said for college professors who hold power over their students and are expected to act as arbiters of debate between different perspectives, not to make offensive statements (Daniels et al. 2021).⁴ More generally, we expect any public figure, including celebrities, will be more likely to be canceled. Such canceling carries with it less legal risk since the standards to establish defamation remain much higher for public than private figures. Public figures' speech also has broader reach which could influence many others and set norms of acceptability. Our public figure hypothesis is *individuals*

⁴ The prevalence of cancel culture and safe space discussions on college campuses also may lead people to react more to offensive speech from professors. Isolating the role of faculty is particularly interesting given universities' historic commitment to free speech (Whittington 2018).

will be more likely to cancel public (elected officials, professors, celebrities) versus private figures, all else constant.

Finally, given the aforementioned partisan asymmetry in the volume and tenor of discussion about cancel culture, we expect Republicans will be less likely to engage in canceling behavior. Along those same lines, we expect Republicans to perceive canceling behavior more negatively. We label this the Republican cancellation hypothesis: *Republicans will be less likely to engage in cancel culture and to view it more negatively, than Democrats, all else constant.*

But cancellation does not just depend on who the speaker is, what they say also matters. For instance, Costa (2020) shows that people do *not* prefer representatives who promote affective polarization or partisan slander (e.g., accusing the other party of being corrupt or immoral), but instead want representatives who share their policy views. The same is true of how individuals evaluate other citizens (e.g., Orr and Huber 2020, Mummolo et al. 2021, Dias and Lelkes 2021). Even in today's polarized environment, individuals care what other actors say, and will react negatively when content counters their ideological interests (see Druckman 2022 for a review). This follows because individuals' ideological beliefs reflect their underlying values (Goren et al. 2016), and hence "offensive" speech is therefore speech that violates these core values. Our ideological content hypothesis is *individuals will be more likely to cancel statements that are ideologically inconsistent with their party's positions relative to ideologically consistent ones (i.e., Democrats will punish conservative statements, and Republicans will punish liberal ones), all else constant.*

The political divide around free speech historically, and cancel culture more recently, obfuscates whether source or content will matter more. Consider the 2019 controversy around *Harper's Magazine's* publication of "A Letter on Justice and Open Debate." The letter, signed

by notable public and academic figures, warned that canceling those with offensive views would weaken “our norms of open debate and toleration of differences in favor of ideological conformity.” A response chastised the authors of the initial statement by arguing that it ignored the difficulties faced by stigmatized minorities, particularly Black and transgender people. Yet, at the same time, it also took issue with the original authors, who they purported to all be in positions of power. *The New York Times* pointed out that many felt “that criticism of the *Harper’s* letter centered as much on who signed it as its content” (Schuessler 2020).

This is not simply an academic debate. Obviously, some speakers’ actions matter more because they have a larger platform and reach a broader audience. But there is a clearer, and more direct, link to harm from the content of speech, especially given the framing of cancel culture as countering speech that harms.⁵ If cancel culture debates are about what is being said, it suggests these debates involve legitimate discussions about the boundaries of acceptable speech. If, instead, they focus on who is speaking, then it suggests that they are (relatively) more about silencing figures people dislike, a much less valuable public conversation. Adjudicating the relative importance of speaker’s identity versus what they say is therefore a particularly important task.

What Drives *Perceptions* of Canceling?

Culture refers to a set of shared values or practices among a given group. Whether canceling reflects a set of shared values and hence a culture, is unclear. Many Americans lack familiarity with the term “cancel culture” (Pew Research Center 2021), and those with familiarity may

⁵ That said, we recognize that identical statements from different speakers can connote distinct meanings; our point, all else constant, is that speech content has more harm potential than sources.

vastly misperceive its frequency and nature given the media narrative surrounding the issue. Indeed, individuals tend to misestimate—often grossly—the behaviors of those different from themselves (Robinson et al. 1995). Such patterns are especially likely in a partisan context, given that partisan identity encourages derogation of the out-party (Levendusky and Malhotra 2016). As such, cancel *culture* may be more imagined than concrete.

We suspect most individuals have not actually canceled anyone, as many Americans remain politically unengaged (Krupnikov and Ryan 2022) and averse to political conflict (Klar and Krupnikov 2016). Even on social media, most American partisans remain apolitical (Wojcieszak et al. 2021). They thus likely believe they cancel less than other partisans, even from their own party. Moreover, if most partisans view cancel culture negatively (as we show later), they will assume out-partisans are even more likely to cancel. Our canceling perception hypothesis is *individuals will predict that in-partisans will be somewhat more likely to cancel than themselves, all else constant. Individuals also will perceive out-partisans to be more likely to cancel than in-partisans, all else constant.* That is, when faced with the same (negatively viewed) behavior, individuals think that they themselves will be the least likely to cancel, in-partisans will be somewhat more likely to cancel, and out-partisans will be the most likely to cancel. We expect this to be true both in how people respond to both hypothetical scenarios as well as their actual behaviors. Again, given the partisan asymmetry in cancel culture attitudes, we expect partisanship will moderate these effects. Our Republican perception hypothesis is: *relative to Democrats, Republicans will perceive a wider gap between their party's propensity to cancel and Democrats' propensity to cancel, all else constant.*

In sum, we predict Americans will have substantial misperceptions about how likely others will engage in, and how often others have engaged in, cancel culture. If true, this would

suggest that, instead of being a shared set of beliefs, cancel culture manifests as a partisan divide where people attribute the other side to embrace the activity more.

What Motives Do People Attribute to cancellation?

Finally, what motives do people attribute to those who cancel? Because humans want to have a positive self-image (Baumeister 2010, Molden and Higgins 2012), they engage in a self-serving attribution bias to make themselves look good (e.g., Kunda 1999, Reeder 2013). When individuals cancel someone, they will be apt to imbue positive motivations to themselves. Our personal attribution hypothesis is *individuals will describe their canceling behaviors as more positive than negative, all else constant.*

Just as positive personal attributions enhance self-esteem, so too do positive attributions for one's group (Reeder 2013, Warner and Villamil 2017, Noor et al. 2019, Zell et al. 2021). In contrast, the intergroup attribution error suggests that people perceive objectionable or anti-social behaviors to be character flaws when it comes from the out-group, including political out-groups (Pettigrew 1979, Munro et al. 2010, Goya-Tocchetto et al. 2022). Our out-party attribution hypothesis is *individuals will describe in-partisans' canceling behaviors as more positive than they will describe out-partisans' canceling behaviors, all else constant. By contrast, they will describe out-partisans' canceling behaviors as more negative than they will describe in-partisans' canceling behaviors, all else constant.*

Finally, in line with our prior discussions of partisan asymmetry, we expect differences in how respondents describe their own canceling behavior, versus the out-party's canceling behavior, to be especially defined among Republicans. That is, our Republican attribution hypothesis is *Republicans will be more likely, than Democrats, to describe their party's*

canceling behaviors positively and out-partisans canceling behaviors negatively, all else constant.

Altogether, our hypotheses suggest that—in addition to cancel culture being entangled in partisan divides—misperceptions surrounding cancel culture could exacerbate polarization by generating attributional tensions. Even if we find that canceling behaviors reflect content more than sources, partisan attributions could turn free speech debates into partisan disputes. We summarize our hypotheses in Table 1.

[Insert Table 1 About Here]

Survey and Experimental Design

To test our hypotheses, we conducted a pre-registered survey that included a set of conjoint experiments (for pre-registration, see https://aspredicted.org/VKJ_DYZ or appendix A). Our data came from a quota-matched sample of 1,752 American adults from Bovitz’s Forthright, a high-quality online panel, collected between September 24 and October 4, 2021. The sample was quota-matched to represent American adults on age, gender, education, Census region, and race; we present sample demographics in appendix B. Forthright participants are recruited via mail campaigns based on addressed-based probability sampling, as well as via online ads, and their data has been used extensively in political science (e.g., Druckman et al. 2019).

The survey began by presenting respondents with six different behaviors that have been labeled “canceling” and asking them which they had done, which they had seen others do, and who (i.e., Democrats or Republicans) they had seen do them. To formulate the list of canceling behaviors, we carefully reviewed the academic and popular literature on this topic, and then pre-

tested open-ended items to ensure that our list reflected what the public thought of as canceling. Throughout our study, our list of these behaviors is: criticizing the speaker on social media; complaining to the speaker's employer; boycotting the speaker's employer or merchandise; boycotting or protesting at an event where the speaker is participating; reporting the speaker to, or trying to ban the speaker from, a social media site; and publishing a speaker's personal information online (i.e., doxing). Thus, these behaviors encompass potentially anodyne activities (e.g., social media criticism), as well more severe actions (e.g., banning someone from social media or doxing them). By capturing both what partisans do and how they perceive what others do, these descriptive questions let us test our canceling perception hypothesis. We did not call these behaviors "canceling" or "cancel culture" in the survey to avoid cueing respondents, but we refer to them as such in the paper in the interest of clarity.

We then presented respondents with four hypothetical scenarios wherein speakers made potentially offensive statements. Each scenario varied either the speaker's partisanship or their race (Democrat/Republican/Black/White),⁶ the speaker's occupation or social role (university student/ professor/elected official/celebrity/voter), and what the speaker said. The statements focused on contemporary debates about history and race, contemporary race and politics, police and protests, and transgender identity. These issues all received substantial attention at the time of our study, as evidenced by internet searches of "cancel culture." We identified controversial liberal and conservative statements about these issues using Google News (see examples in appendix C) and then condensed them down for ease of presentation. For example, on the role of race in the United States' history, the liberal statement is "The Founding Fathers were racist," while the conservative statement is "Confederate statues are about America's heritage and are

⁶ We told respondents only the speaker's race or partisanship. That is, we never provided both pieces of information.

not racist.” In Table 2, we display an overview of our conjoint attributes and levels, including all of the “cancelable” statements. For example, a respondent could have seen the following statement: “A Republican university student posted a comment on social media stating that there is no such thing as transgender, only male and female.”

[Insert Table 2 About Here]

Following each scenario, participants reported their likelihood of engaging in each of our six canceling behaviors on a four-point scale ranging from “Not at all likely” to “Very likely.” We summed respondents’ answers across these behaviors to calculate scores ranging from 0 (i.e., selecting “Not at all likely” for each behavior) to 18 (i.e., selecting “Very likely” for each behavior). These responses allow us to test our five canceling behavior hypotheses (seen in the first section of Table 1).

We then presented respondents with four additional, identically randomized scenarios wherein speakers made potentially offensive statements. However, for the first three scenarios, respondents indicated what canceling behaviors *out-partisans* would likely do, using that same response scale as above (e.g., what a Democratic respondent thought Republicans would do). For the fourth and final scenario, respondents indicated what *in-partisans* would do. We use these scenarios to test our perceptions of canceling hypotheses (seen in second section of Table 1).

When respondents said they themselves, out-partisans, or in-partisans were somewhat or very likely to engage in at least one canceling behavior, we asked them to describe the canceling behavior by choosing from 11 descriptors—e.g., fair, biased, and so forth (we provide a full list below in the results section; the items reflected terms from Pew Research Center 2021 as well as

additional terms used in online debates about this topic). This list also included the term “cancel culture,” the first time we introduced it (quite late in the survey). The descriptors were used to assess our attribution hypotheses (seen in the third section of Table 1) and to see what respondents themselves saw as cancel culture. We also asked respondents how much they had heard of cancel culture and whether they thought it a more positive or negative effect on society. All question wordings are in the appendix D.

Canceling Behavior Results

We begin by considering what drives canceling behavior: is it a function of what gets said, who says it, or both? We use our first-person conjoint results to answer this question. Recall that, for each respondent-scenario observation, we calculated a score between 0 and 18, reflecting the sum of the likelihoods of engaging in our six canceling behaviors. Higher values on this scale indicate a higher likelihood of canceling.⁷

Following Hainmueller et al. (2014), Figure 1 presents the AMCEs (average marginal component effects) estimates for our conjoint features. We conjoined our party-race source and statement ideology features to clearly test the partisan and race source hypotheses, as well as the ideological content hypothesis. This approach puts the relative impact of source and content in stark relief (see appendix E for analyses that treat them separately). We use the prefixes “in-” and “out-” to indicate how the speakers’ identity corresponds to the respondent’s partisanship. For example, “in-party” refers to a Democratic or Black speaker when the respondent is a Democrat. We use the terms “agreeable” and “disagreeable” to indicate statements that cohere with the respondent’s in-party and out-party ideology, respectively. For baselines, we use the

⁷ We get similar results analyzing each behavior separately; see appendix E.

“in-party, agreeable scenario,” and the “voter” social role. We use the former since in that case we expect very little canceling and the latter since it captures the positionality of the respondents.

[Insert Figure 1 about here]

The results in Figure 1 are striking in their consistency. The clearest and most important factor that shapes canceling behavior is *not* the speaker’s partisanship or race. That is, the partisan and race source hypotheses receive no support: people do not censor speech based merely on the speaker’s identity, as one might expect in an era of affective polarization.⁸ Instead, we find strong support for the ideological content hypothesis: respondents are substantially more likely to cancel ideologically disagreeable statements. Democrats are more likely to cancel conservative statements, and Republicans are more likely to cancel liberal ones.⁹ Indeed, respondents are much more likely to cancel someone from their own party (or an associated racial group) who says something disagreeable than to cancel someone from the out-party who says something agreeable! Moreover, in separate pre-test data (presented in appendix G), we show that the more offensive respondents perceive a statement to be, the more likely they are to cancel it. Ideological agreeableness serves as the main driver of offensiveness.

Consistent with the content-over-source pattern, we also find no effect—at all—of the speaker’s occupation on the likelihood of cancellation. Contrary to our public figure hypothesis, the speaker’s job or social role has no effect on how respondents behave. While public officials,

⁸ See appendix F for analysis of affective polarization as a moderator.

⁹ Respondents are slightly more likely to cancel a member of the out-party (or an associated racial group) who makes an ideologically disagreeable statement. However, these differences are dwarfed by the content effects.

celebrities, and professors may get more attention when they make controversial statements, individuals do not hold them to a higher standard.¹⁰

This is a particularly important set of findings because it reveals the essence of why people cancel: it is not about holding some people to a higher standard than others or punishing those whom we dislike, but rather involves calling out disagreeable or offensive speech. Cancellation sets the boundaries of acceptable expression.

A final, notable finding concerns the average likelihood of canceling behavior. While our scale ranges from 0 to 18, these AMCEs are small, only moving respondents one to two points on the scale. Indeed, the modal person scores a 2 on the underlying scale, and the inter-quartile ranges from 0 to 7, suggesting that canceling is rare, even when faced with hateful content. That said, it could be that this low rate of canceling reflects something about our conjoint studies: perhaps we picked the wrong statements, or the wrong actors, etc. To assess this possibility, recall that we asked our respondents if they had *ever* engaged in any of our six canceling behaviors. The left panel of Figure 2 plots the percentage of respondents who report engaging in each of these behaviors.

[Insert Figure 2 About Here]

Consistent with Figure 1, canceling in the real world rarely occurs. No more than one-third of the public has engaged in any of our six canceling behaviors, and many have done none of them. Similarly, in the right-hand panel of Figure 2, we plot how much respondents have

¹⁰ In appendix E, we present results for each specific statement. We find that the exact content matters somewhat, with people having particularly adverse reactions (i.e., they cancel more) in response to the statement that some races are less intelligent than others. All other statements, regardless of topic, are equally likely to be cancelled. These results clarify that canceling depends on ideological tenor more than specific content.

heard about cancel culture. While the modal respondent has heard “a fair amount” about cancel culture, more than a third of respondents (36 percent) have heard “not too much” or “nothing at all” about the topic. Most people themselves do not engage in cancellation, and many have limited familiarity with these debates.

(Mis-)perceptions of Canceling Results

Even though most do not engage in cancel culture, they may think *others*—especially others unlike them—do so. To test this, we asked respondents if they had seen anyone else engage in our six canceling behaviors. The top-left panel of Figure 3 shows how often respondents have seen others engage in these behaviors, with the self-reports of canceling behavior from Figure 2 included as a baseline.

[Insert Figure 3 About Here]

The disparity between perceptions and reality stands out: for every behavior, prevalence perceptions at least *double* actual prevalence. For instance, nearly a third (31 percent) report they have seen others engage in doxing, though only 3 percent have ever done it themselves; the same pattern exists on every item. Clearly, people’s perceptions of cancel culture far outpace the number who engage in these behaviors. Thus, perhaps canceling behavior is rare but made visible by the media coverage or social media attention.¹¹

¹¹ It is theoretically possible that a few people do an enormous amount of canceling. We think this is an unlikely explanation for the vast disparities we observe; it would entail these individuals canceling in fairly extreme quantities and having very broad audiences of that canceling. Moreover, the conjoint estimates we next discuss (in Figure 3C) strongly suggest people think canceling is common among (many) others.

Our argument is that people should think out-partisans cancel more frequently than in-partisans. To test this, we asked those who reported having seen others engage in canceling behavior how often those individuals were Democrats or Republicans, on a five-point scale ranging from “Never” to “Extremely often.” The top-right panel of Figure 3 shows that 50 percent of individuals believe out-partisans cancel more, 30 percent think the parties cancel equally, and 20 percent believe in-partisans cancel more. The other party, more so than one’s own, bears the brunt of the blame for cancel culture. This parallels other findings on individuals’ highly skewed perceptions of the other party (e.g., Ahler and Sood 2018, Druckman et al. Forthcoming).

This pattern also manifests in our second-order conjoint data. Recall that after respondents indicated how they themselves would respond to four conjoint scenarios, they indicated how out-partisans and in-partisans would respond to a separate but similar set of conjoint scenarios. In the bottom panel of Figure 3, we show that people think they themselves are the least likely to cancel someone, followed by those from the in-party, and finally those from the out-party.¹² In terms of magnitude, people think out-partisans are two to three times more likely to engage in canceling behavior than they themselves are. While respondents’ views of out-partisans (versus themselves) exhibit the largest gap, they also see members of their own party as being much more likely to cancel compared to themselves. Simply put, canceling is something that “others” do, strongly confirming our canceling perception hypothesis. Indeed, the term “cancel culture” constitutes a misnomer given the lack of widely shared beliefs around canceling behaviors.

¹² For the full analysis of these data a la Figure 1, see appendix E. We find results that are extremely consistent with those in Figure 1: people cancel because of what is said, not who someone is.

Canceling Motivation Results

The findings in Figure 1 underscore that canceling stems from disagreeable speech rather than partisan or racial sentiment. While one cannot impute definitive motivations, these findings suggest canceling behaviors reflect authentic reactions to sentiments seen as offensive or harmful, not examples of discrimination (as mentioned, also see appendix G).

We posited, however, that people would attribute different motivations to cancelers based on those cancelers' partisanship. As mentioned, when respondents said they themselves, out-partisans, or in-partisans were somewhat or very likely to engage in at least one canceling behavior, we asked them to describe those behaviors by choosing from 11 descriptors: four positive words (fair, empathetic, accountability, considerate), four negative words (biased, over-sensitive, punishment, racism), and three somewhat ambiguous words (political correctness, wokeness, cancel culture); respondents could pick as many terms as they liked. We label these last three words as ambiguous because their valence depends on one's point of view. Perhaps not surprisingly, this is often how people have described cancel culture.¹³

[Insert Figure 4 About Here]

Consistent with our personal attribution hypothesis, respondents described their own canceling behavior in positive terms (more than negative terms). Respondents felt their own behavior was justified: they described their own behavior with at least one positive term 73

¹³ Our pre-registration originally designated the terms political correctness, wokeness, and cancel culture as negative. However, given that the significance of these terms varies across the political spectrum, we re-classified these terms as ambiguous. We re-compute our tests of the personal attribution, out-party attribution, and Republican out-party attribution hypotheses in appendix F using our original designations. Our results are not substantively different, with one exception: counting these three terms as negative makes Republicans somewhat more likely to describe the out-party's canceling behavior in negative terms.

percent of the time and with at least one negative term 34 percent of the time (this difference is statistically significant, $p < 0.01$). Interestingly, respondents also tended to avoid describing their own behaviors as “political correctness” (15 percent), “wokeness” (8 percent), or “cancel culture” (8 percent). This last finding provides preliminary evidence that people have an aversion to the term “cancel culture” (we return to this point below).

Figure 4 also provides strong support for our out-party attribution hypothesis. Respondents described in-partisans’ canceling behavior positively 58 percent of the time, but described *identical* behavior by out-partisans positively only 33 percent of the time ($p < .001$). Similarly, respondents described in-partisans’ canceling behavior negatively only 46 percent of the time, but described identical behavior by out-partisans negatively 67 percent of the time, ($p < .001$). Interestingly, respondents used the term “political correctness” slightly more when describing canceling behavior by in-partisans than out-partisans (20 percent versus 15 percent) but reserved “cancel culture” much more for the out-party (13 percent versus 26 percent).¹⁴ They were virtually equivalent for wokeness, with 10 percent and 12 percent describing in-partisans’ and out-partisans’ canceling behavior with this term, respectively.

In short, respondents not only perceive that out-partisans cancel more, but they perceive out-partisans’ cancelations to be malicious, even though our prior results suggest substantive based canceling. This makes it difficult to arrive at a shared cultural understanding of appropriate canceling and what canceling signifies. It also ironically creates a situation where polarization may not cause canceling (e.g., the lack of support for our partisan source hypothesis), but the misperceptions of canceling could cause polarization. While we cannot

¹⁴ As mentioned, in our design, we—purposefully—avoided calling particular behaviors “cancel culture” because we wanted to avoid pushing subjects one way or another. Instead, we can infer what they view as cancel culture by looking at which practices they label that way; see below note and appendix E for these results, which echo what we present in the body of the paper.

directly speak to this possibility, we suspect that perceptions of substantial out-party canceling with bad intent generates additional partisan animosity. The misperceptions underlying cancel culture could therefore generate real harm if they heighten animus between the parties.

Is There Partisan Asymmetry?

So far, we have discussed the broad contours of cancel culture, including the how and why, for the public at large. But we also expect—in light of our Republican cancellation and attribution hypotheses—to find partisan differences in engagement in, and perceptions of, cancel culture. Given Republican Party elites’ focus on describing cancel culture in the most negative terms, we expect Republican voters to be less likely to cancel others (Republican cancellation hypothesis), and to think that cancel culture has a more pernicious effect on the nation (Republican cancellation hypothesis). Finally, we expect Republicans to be especially likely to say that Democrats perpetrate cancel culture (Republican perception hypothesis) and that Democrats act with particularly malicious motivations (Republican attribution hypothesis).

That said, nothing in our hypotheses predict that distinct causal processes underlie Republicans’ and Democrats’ decisions to cancel. That is, when they engage in canceling behavior, Republicans and Democrats should do so for the same reasons. Republican officials’ attention to cancel culture should change how often Republicans engage in it, but not how or why they do so.

Take, first, the question of how much the parties cancel. In the top panel of Figure 5, we see that Democrats are somewhat more likely to cancel than Republicans (Republican cancellation hypothesis). We also see, in the bottom-left panel of Figure 5, that Republicans in particular perceive the other party is more likely to cancel (Republican perception hypothesis).

But the biggest, and most striking, difference is in whether people see cancel culture as having a net-positive or net-negative effect on society, shown in the bottom-right panel of Figure 5. We asked respondents to evaluate cancel culture's effect on society using a seven-point Likert scale, ranging from "Much more negative of an effect" to "Much more positive of an effect." We see that both parties' perceptions are more negative than positive. However, the distribution of Democrats' perceptions is flatter, with many people clustered in the middle. Republicans, by contrast, see cancel culture almost uniformly as very negative. Indeed, nearly a quarter of Republicans use the lowest scale point! This confirms the second part of the Republican canceling hypothesis.

[Insert Figure 5 about here]

But what stays the same across parties is, in some ways, even more fascinating. In the left panel of Figure 6, we show respondents' preferences across the conjoint features do not differ substantially by party, using the method developed by Leeper et al. (2020) for making sub-group comparisons. Thus, to the extent that both parties cancel, they do so for the same reasons. Further, in the right two panels of Figure 6, we show that both Democrats and Republicans describe in-partisans' canceling behavior in positive terms, and the other party's behavior in negative ones, disconfirming the Republican attribution hypothesis.

That said, there are statistically significant differences in Democrats and Republicans' preferences about some conjoint features, $F(12, 1752) = 3.58, p < .001$. Specifically, Democrats express slightly more willingness to cancel disagreeable statements, though not when they come from in-partisans. Moreover, Democrats are somewhat more likely to describe out-partisans'

canceling behaviors using certain negative terms, namely “racism” and “biased.” This runs counter to our Republican attribution hypothesis which suggests Republicans—not Democrats—would view the other party relatively more negatively. That said, Republicans were more likely to use terms like “political correctness,” “wokeness,” and “cancel culture” to describe out-partisans’ canceling behavior (see appendix F).¹⁵ Most importantly, while levels of canceling behavior differ somewhat between parties, the motivations behind that behavior—and perceptions of canceling behavior—do not differ substantially.¹⁶

[Insert Figure 6 about here]

These findings refute the narrative that cancel culture simply occurs among the censorious left (e.g., *The Economist* 2021). Yes, Democrats are more likely to cancel, but this difference is modest. The only large partisan difference is that Republicans perceive cancel culture more negatively. But both parties do, in fact, engage in cancel culture—and do so for similar reasons. Both parties also see their own behavior in positive terms, and the opposition’s behavior in negative terms. So, consistent with the Republican cancellation hypothesis, we do observe real and significant differences between the parties. But focusing only on those differences—and ignoring the many commonalities—would be just as much a mistake as ignoring those differences in the first place.

¹⁵ More generally, respondents are more likely to label something as “cancel culture” when it is by others, and especially out-partisans. Moreover, when a canceler is responding to a statement presumably perceived to be disagreeable to the canceler, their actions are more likely to be considered cancel culture. Remarkably, the particular action taken by the canceler has almost no effect of whether their actions are labeled as cancel culture. See appendix E for analyses.

¹⁶ Another potential source of heterogeneity is that perhaps Fox News viewers are more attentive to this issue, given its extensive coverage on that network and related conservative media outlets. We tested for such heterogeneity, among Republicans, and generally found it to be weaker than the partisan one discussed above; see appendix F.

Conclusion

The United States has historically had an exceptional commitment to free speech and has seen the solution to offensive rhetoric as counter-speech rather than censorship. Is that now changing with the rise of cancel culture—is America’s understanding of free speech evolving? Our data suggest so: while rates of canceling remain low, at least some Americans willingly silence those with whom they disagree. While some differences across party lines emerge, with Democrats being more likely to engage in these behaviors, the gaps remain modest, with members of both parties engaging in these behaviors for the same reasons. Importantly, however, we show enormous perceptual gaps: people think that others are many times more likely to cancel than they are, and that when they do so, they do so for ill, rather than good. While affective polarization does not generate canceling, canceling—given these misperceptions—can, and likely does, fuel animus.

The fact that we found almost no partisan effects was surprising, but important in that it highlights the mechanisms at work underlying our effects. We expected the speaker’s identity to substantially matter. Yet, we found that it does not. Instead, when individuals encounter an offensive statement counter to their ideological belief systems, they are more likely to cancel it. This is in line with a theory of naïve realism where people presume their opinions and beliefs to be objectively accurate, with anyone disagreeing being biased (Ross and Ward 1996). Individuals view certain language as harmful, cancel it, and believe those who cancel from the other side have bad intent.

This suggests two new dimensions to debates over free speech. First, as we mentioned, it signals a shift a shift in the response to offensive speech, away from counter-speech and toward

cancellation.¹⁷ But more subtly, it also suggests a shift in *who* gets to determine the boundaries of acceptable expression. In an earlier age, it would have been gatekeepers who set the terms of the debate, with politicians, journalists, lawyers, and judges largely determining what was, and was not, acceptable. But with the rise of social media, those intermediaries have been disempowered: there is applicable law, but in most cases, what other people think matters more. In effect, the people now help to set the limits of what constitutes acceptable discourse, and those limits evolve rapidly, explaining why speech acceptable at one time becomes unacceptable at another. What this shift means for our understanding of free speech is a fascinating area for future research.

Second, and just as importantly, our findings provide insight into the intensity of cancel culture debates, on two levels. On one level, people see their canceling of others as being driven by positive motives; they view canceling by others as being driven by nefarious intentions. When voters read stories about it, especially stories centered on those from the other party, they see those involved as acting in bad faith. To the extent that partisan media outlets (and partisans on social media) push such stories to disparage the other side, this effect should be especially likely. Motives here really lie in the eye of the beholder.

Finally, and perhaps most importantly, our results lay bare the extent to which these stories are ultimately about political power. Cancel culture, at its heart, democratizes power away from elites and toward ordinary voters, and ensuring that the voices of the marginalized become enfranchised (Romano 2021).¹⁸ For centuries, elites largely policed themselves, but now, they

¹⁷ Of course, one could point out that the most common activity—criticizing someone on social media—constitutes a form of counter-speech, at least in theory. That is true, and future work should explore the boundary between what is counter-speech and what is canceling in greater detail.

¹⁸ From this perspective, one can see that the debates over political correctness in the 1990s are the forerunners of cancel culture today (see [Shapiro et al. 2021](#)). Indeed, such debates are centuries old (Mishan 2020). This also

can be—and are being—called out by the masses in a highly visible fashion. Our results highlight that voters do not judge public figures more harshly than anyone else, so why do those cases get so much attention? We imagine it is because these sorts of powerful people otherwise rarely have to account for their actions. Cancel culture makes them as vulnerable as others. For this reason, debates about cancel culture will likely remain heated for years to come, with much for scholars to study moving forward.

explains why Republicans express more concern about cancel culture as it can be seen as a form of status threat or a challenge to the system (Jost 2020).

References

- Ahler, Douglas J, and Gaurav Sood. 2018. "The Parties in Our Heads: Misperceptions about Party Composition and Their Consequences." *The Journal of Politics* 80 (3): 964–81. doi:10.1086/697253.
- Applebaum, Anne. 2021. "The New Puritans." *The Atlantic*, 31 August.
- Armstrong, Grant M., and Julie Wronski. 2019. "Framing Hate: Moral Foundations, Party Cues, and (In)Tolerance of Offensive Speech." *Journal of Social and Political Psychology* 7 (2): 695–725. doi:10.5964/jspp.v7i2.1006.
- Baumeister, Roy F. 2010. "The Self." In Roy F. Baumeister, and Eli J. Finkel, eds., *Advanced Social Psychology: The State of the Science*. New York, NY: Oxford University Press.
- Beerbohm, Eric. 2012. *In Our Name: The Ethics of Democracy*. Princeton University Press.
- Brandeis, Louis. 1927. Concurring Opinion, *Whitney v. California*. Available online at: <https://bit.ly/3cge7N4>.
- Brettschneider Corey. 2012. *When the State Speaks, What Should It Say?: How Democracies Can Protect Expression and Promote Equality*. Princeton, NJ: Princeton Univ. Press
- Chong, Dennis. 2006. "Free Speech and Multiculturalism In and Out of the Academy." *Political Psychology* 27 (10): 29-54. <https://doi.org/10.1111/j.1467-9221.2006.00448.x>
- Chong, Dennis, Jack Citrin, and Morris Levy. 2021. "The Realignment of Political Tolerance in the United States." *SSRN Electronic Journal*. doi:10.2139/ssrn.3951377.
- Clark, Meredith. 2020. "Drag Them: A Brief Etymology of So-Called 'Cancel Culture.'" *Communication and the Public* 5(3-4): 88-92.

- Costa, Mia. 2021. "Ideology, Not Affect: What Americans Want from Political Representation." *American Journal of Political Science* 65 (2): 342–58. doi:10.1111/ajps.12571.
- Crockett, Emily. 2016. "Safe Spaces, Explained." *Vox*, 25 August. Available online at: <https://bit.ly/3tkNMXU>.
- Daniels, Ronald J., with Grant Shreve and Phillip Spector. 2021. *What Universities Owe Democracy*. Baltimore, MD: Johns Hopkins University Press.
- Davis, Darren W., and Brian D. Silver. 2004. "Civil Liberties vs. Security: Public Opinion in the Context of the Terrorist Attacks on America." *American Journal of Political Science* 48 (1): 28–46. doi:10.1111/j.0092-5853.2004.00054.x.
- Dias, Nicholas, and Yphtach Lelkes. 2021. "The Nature of Affective Polarization: Disentangling Policy Disagreement from Partisan Identity." *American Journal of Political Science*. doi:10.1111/ajps.12628.
- Downs, Daniel M., and Gloria Cowan. 2012. "Predicting the Importance of Freedom of Speech and the Perceived Harm of Hate Speech." *Journal of Applied Social Psychology* 42 (6): 1353–75. doi:10.1111/j.1559-1816.2012.00902.x.
- Druckman, James N. 2022. "A Framework for the Study of Persuasion," *Annual Review of Political Science*, Forthcoming.
- Druckman, James N., S.R. Gubitz, Matthew S. Levendusky, and Ashley Lloyd. 2019. "How Incivility On Partisan Media (De-)Polarizes the Electorate." *The Journal of Politics* 81 (1): 291-295. <https://doi.org/10.1086/699912>.
- Druckman, James N., Samara Klar, Yanna Krupnikov, Matthew Levendusky, and John Barry Ryan. Forthcoming. "(Mis-)Estimating Affective Polarization." *The Journal of Politics*

- Dworkin Ronald. 1996. *Freedom's Law: The Moral Reading of the American Constitution*.
Oxford, UK: Oxford University Press.
- Finkel, Eli J, *et al.* 2020. "Political Sectarianism in America: A Poisonous Cocktail of Othering, Aversion, and Moralization." *Science* 370 (6516): 533-536. DOI: 10.1126/science.abe1715.
- Friesdord, Conor. 2015. "The New Intolerance of Student Activism." *The Atlantic*, 9 November.
- Gervais, Bryan T. 2019. "Rousing the Partisan Combatant: Elite Incivility, Anger, and Anti-deliberative Attitudes." *Political Psychology* 40 (3): 637-655.
<https://doi.org/10.1111/pops.12532>
- Gibson, James L. 1992. "The Political Consequences of Intolerance: Cultural Conformity and Political Freedom." *American Political Science Review* 86 (2): 338–56.
doi:10.2307/1964224.
- Gibson, James L. 2006. "Do Strong Group Identities Fuel Intolerance? Evidence from the South African Case." *Political Psychology* 27 (5): 665–705. doi:10.1111/j.1467-9221.2006.00528.x.
- Gibson, James L. 2008. "Intolerance and Political Repression in the United States: A Half Century after McCarthyism." *American Journal of Political Science* 52 (1): 96–108.
doi:10.1111/j.1540-5907.2007.00301.x.
- Gibson, James L., and Arthur J. Anderson. 1985. "The Political Implications of Elite and Mass Tolerance." *Political Behavior* 7 (2): 118–46. doi:10.1007/bf00988795.
- Gomez, Henry. 2020. "Donald Trump And Republicans Have Seen the Future and It Is Canceled." *Buzzfeed News*, Available online at: <https://bit.ly/3Hkq2b4>.

- Goren, Paul, Harald Schoen, Jason Reifler, Thomas Scotto, and William Chittick. 2016. "A Unified Theory of Value-Based Reasoning and U.S. Public Opinion." *Political Behavior* 38: 977–997. <https://doi.org/10.1007/s11109-016-9344-x>
- Goya-Tocchetto, Daniela, Aaron C. Kay, Heidi Vuletich, Andrew Vonasch, and Keith Payne. 2002. "The Partisan Trade-Off Bias: When Political Polarization Meets Policy Trade-Offs." *Journal of Experimental Social Psychology* 98: 104231. <https://doi.org/10.1016/j.jesp.2021.104231>.
- Guo, Lei, and Brett G Johnson. 2020. "Third-Person Effect and Hate Speech Censorship on Facebook." *Social Media + Society* 6 (2): 1–12. doi:10.1177/2056305120923003.
- Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22 (1): 1–30. doi:10.1093/pan/mpt024.
- Hobbes, Michael. 2021. "The Methods of Moral Panic Journalism." *Confirm My Choices*, Available online at: <https://bit.ly/3HumEdK>.
- Howard, Jeffrey W. 2019. "Free Speech and Hate Speech." *Annual Review of Political Science* 22 (1): 93-109. <https://doi.org/10.1146/annurev-polisci-051517-012343>,
- Jamieson, Kathleen Hall and Bruce Hardy. 2012. "What Is Civil Engaged Argument and Why Does Aspiring to It Matter?" *PS: Political Science and Politics* 45(3): 412-5.
- Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. 2018. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22 (1): 1–18. doi:10.1146/annurev-polisci-051117-073034.
- Jost, John T. 2020. *A Theory of System Justification*. Cambridge, MA: Harvard University Press.

- Klar, Samara, and Yanna Krupnikov. 2016. *Independent Politics: How American Disdain for Parties Leads to Political Inaction*. Cambridge: Cambridge University Press.
- Klein, Ezra. 2021. "A Different Way of Thinking About Cancel Culture." *New York Times* 18 April.
- Krupnikov, Yanna, and John Barry Ryan. 2022. *The Other Divide: Polarization and Disengagement In American Politics*. Cambridge: Cambridge University Press.
- Kunda, Ziva. 1999. *Social Cognition: Making Sense of People*. Cambridge, MA: The MIT Press.
- Leeper, Thomas J., Sara B. Hobolt, and James Tilley. 2020. "Measuring Subgroup Preferences in Conjoint Experiments." *Political Analysis* 28 (2): 207–21. doi:10.1017/pan.2019.30.
- Lelkes, Yphtach, and Sean J Westwood. 2017. "The Limits of Partisan Prejudice." *The Journal of Politics* 79 (2): 485–501. doi:10.1086/688223.
- Levendusky, Matthew S., and Neil Malhotra. 2016. "(Mis)perceptions of Partisan Polarization in the American Public." *Public Opinion Quarterly* 80 (S1): 378–391.
<https://doi.org/10.1093/poq/nfv045>
- Lindner, Nicole M., and Brian A. Nosek. 2009. "Alienable Speech: Ideological Variations in the Application of Free-speech Principles." *Political Psychology* 30 (1): 67–92.
<https://doi.org/10.1111/j.1467-9221.2008.00681.x>
- Mason, Lilliana. 2018. *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press.
- McClosky, Herbert, and John Zaller. 1985. *The American Ethos: Public Attitudes toward Capitalism and Democracy*. Harvard University Press.

- Mishan, Ligaya. 2020. "The Long and Tortured History of Cancel Culture." *New York Times*, 3 December.
- Molden, Daniel C., and E. Tory Higgins. 2012. "Motivated Thinking." In Keith J. Holyoak, and Robert G. Morrison, eds., *The Oxford Handbook of Thinking and Reasoning*. New York, NY: Oxford University Press.
- Mounk, Yasha. 2020. "Stop Firing the Innocent." *The Atlantic*, 27 June.
- Muddiman, Ashley. 2017. "Personal and Public Levels of Political Incivility." *International Journal of Communication* 11: 3182–3202.
<https://ijoc.org/index.php/ijoc/article/view/6137/2106>
- Mummolo, Jonathan, Erik Peterson, and Sean Westwood. 2021. "The Limits of Partisan Loyalty." *Political Behavior* 43: 949-972. <https://doi.org/10.1007/s11109-019-09576-3>.
- Munro, Geoffrey D., Carrie Weih, and Jeffrey Tsai. 2010. "Motivated Suspicion: Asymmetrical Attributions of the Behavior of Political Ingroup and Outgroup Members." *Basic and Applied Social Psychology* 32 (2): 173-184, DOI: 10.1080/01973531003738551.
- Mutz, Diana C. 2015. *In-Your-Face Politics: The Consequences of Uncivil Media*. Princeton University Press.
- Noor, Masi, Nour Kteily, Birte Siem, and Agostino Mazziotta. 2019. "'Terrorist' or 'Mentally Ill': Motivated Biases Rooted in Partisanship Shape Attributions About Violent Actors." *Social Psychological and Personality Science* 10(4): 485-493 DOI: 10.1177/1948550618764808
- Orr, Lilla V., and Gregory A. Huber. 2020. "The Policy Basis of Measured Partisan Animosity in the United States." *American Journal of Political Science* 64 (3): 569–86.
doi:10.1111/ajps.12498.

- Peffley, Mark, Pia Knigge, and Jon Hurwitz. 2001. "A Multiple Values Model of Political Tolerance." *Political Research Quarterly* 54 (2): 379. doi:10.2307/449162.
- Pettigrew, Thomas F. 1979. "The Ultimate Attribution Error: Extending Allport's Cognitive Analysis of Prejudice." *Personality and Social Psychology Bulletin* 5 (4): 461–476. <https://doi.org/10.1177/014616727900500407>.
- Pew Research Center. 2021. "Americans and 'Cancel Culture': Where Some See Calls for Accountability, Others See Censorship, Punishment." Available online at: <https://pewrsr.ch/3oxA00m>.
- Reeder, Glenn D. 2013. "Attribution as a Gateway to Social Cognition. In Donal E. Carlston, ed., *The Oxford Handbook of Social Cognition*. New York, NY: Oxford University Press.
- Robinson, Robert, Dacher Keltner, Andrew Ward, and Lee Ross. 1995. "Actual versus assumed differences in construal: 'Naive realism' in intergroup perception and conflict." *Journal of Personality and Social Psychology*, 68(3): 404-17.
- Romano, Aja. 2020. "Why We Can't Stop Fighting about Cancel Culture." *Vox*, Available online at: <https://bit.ly/3kHdS2a>.
- Romano, Aja. 2021. "The Second Wave of 'Cancel Culture.'" *Vox*, Available online at: <https://bit.ly/3DpnNRr>.
- Ross, Lee, and Andrew Ward. 1996. "Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding," In Edward S. Reed, Elliot Turiel, and Terrance Brown, eds., *Values and Knowledge*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Schauer, Frederick. 2010. "Facts and the First Amendment." *UCLA Law Review* 57 (4): 897–919.

- Schuessler, Jennifer. 2020. "An Open Letter on Free Expression Draws a Counterblast." *New York Times*, 10 July.
- Shapiro, Ari, Alejandra Marquez Janse, Mia Venkat, Noah Caldwell, and Patrick Jarenwattananon. 2021. "How Cancel Culture Became Politicized — Just Like Political Correctness." *NPR*, July 26. <https://n.pr/333n1Nf>.
- Shibley, Robert. 2016. "For the Fourth: Benjamin Franklin on Freedom of Speech—50 Years Before the Constitution." Fire.org, Available online at: <https://bit.ly/3FfN5lw>.
- Sullivan, John L., James Pierson, and George E. Marcus. 1982. *Political Tolerance and American Democracy*. Chicago, IL: The University of Chicago Press.
- Sullivan, John L., Pat Walsh, Michal Shamir, David G. Barnum, and James L. Gibson. 1993. "Why Politicians Are More Tolerant: Selective Recruitment and Socialization Among Political Elites in Britain, Israel, New Zealand and the United States." *British Journal of Political Science* 23 (1): 51–76. doi:10.1017/s0007123400006566.
- Sydnor, Emily. 2019. *Disrespectful Democracy: The Interaction of Political Incivility and Psychological Conflict Orientation*. New York, NY: Columbia University Press.
- The Economist*. 2021. "The Threat from the Illiberal Left." 4 September.
- Trump, Donald. 2020. "Remarks by President Trump at South Dakota's 2020 Mount Rushmore Fireworks Celebration." *The Archived Trump White House*. <https://bit.ly/3uCckt3>.
- Uslaner, Eric. 1993. *The Decline of Comity in Congress*. Ann Arbor, MI: University of Michigan Press.
- Valentino, Nicholas, and Kirill Zhirkov. 2018. "Blue Is Black and Red Is White? Affective Polarization and the Racialized Schemes of U.s. Party Coalitions." Available online at: <https://bit.ly/3IjU3rp>.

- Waisbord, Silvio. 2020. "Mob Censorship: Online Harassment of US Journalists in Times of Digital Hate and Populism." *Digital Journalism* 8 (8): 1030-1046, DOI: 10.1080/21670811.2020.1818111.
- Warner, Benjamin R., and Astrid Villamil. 2017. "A Test of Imagined Contact as a Means to Improve Cross-Partisan Feelings and Reduce Attribution of Malevolence and Acceptance of Political Violence." *Communication Monographs* 84 (4): 1–19. doi:10.1080/03637751.2017.1336779.
- Wehner, Peter. 2021. "You're Being Manipulated." *The Atlantic*, Available online at: <https://bit.ly/3r20CrA>.
- Westwood, Sean J., and Erik Peterson. 2020. "The Inseparability of Race and Partisanship in the United States." *Political Behavior*, 1–23. doi:10.1007/s11109-020-09648-9.
- Westwood, Sean J, Erik Peterson, and Yphtach Lelkes. 2019. "Are There Still Limits on Partisan Prejudice?" *Public Opinion Quarterly* 83 (3): 584–97. doi:10.1093/poq/nfz034.
- Whittington, Keith E. 2018. *Speak Freely: Why Universities Must Defend Free Speech*. Princeton, NJ: Princeton University Press.
- Wike, Richard. 2016. "Americans More Tolerant of Offensive Speech than Others in the World." *Pew Research Center*, Available online at: <https://pewrsr.ch/3FkNzXx>.
- Wojcieszak, Magdalena, Andreu Casas, Xudong Yu, Jonathan Nagler, and Joshua A. Tucker. 2021. "Echo Chambers Revisited: The (overwhelming) Sharing of In-group Politicians, Pundits and Media on Twitter." OSF Preprints. February 4. doi:10.31219/osf.io/xwc79.
- Zell, Ethan, Christopher A. Stockus, and Michael J. Bernstein. 2021. "It's Their Fault: Partisan Attribution Bias and Its Association with Voting Intentions." *Group Processes &*

Intergroup Relations, 1–18 sagepub.com/journals-permissions DOI:
10.1177/1368430221990084.

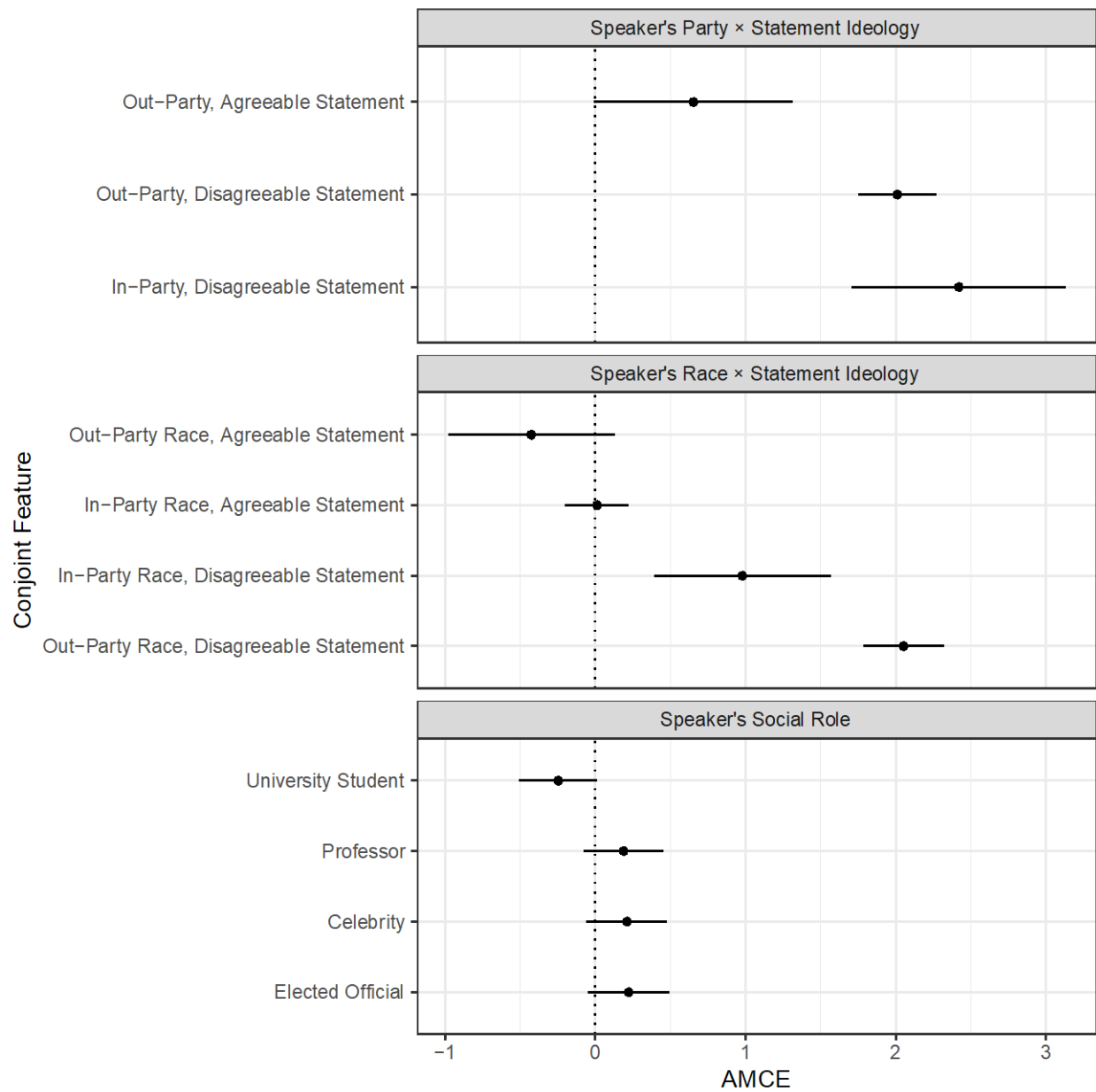
Table 1: Hypotheses

Canceling Behavior Hypotheses	
Partisan source hypothesis	Individuals will be more likely to cancel out-party speakers than in-party speakers, all else constant.
Race source hypothesis	Individuals will be more likely to cancel speakers from racial groups linked with the out-party (than linked to their party), all else constant. (That is, Democrats will be more likely to cancel White individuals than Black individuals, and Republicans will be more likely to cancel Black individuals than White individuals, all else constant.)
Public figure hypothesis	Individuals will be more likely to cancel public (elected officials, professors, celebrities) versus private figures, all else constant.
Republican cancellation hypothesis	Republicans will be less likely to engage in cancel culture and to view it more negatively, than Democrats, all else constant.
Ideological content hypothesis	Individuals will be more likely to cancel statements that are ideologically inconsistent with their party's positions relative to ideologically consistent ones (i.e., Democrats will punish conservative statements, and Republicans will punish liberal ones), all else constant.
Perceptions of Canceling Hypotheses	
Canceling perception hypothesis	Individuals will predict that in-partisans will be somewhat more likely to cancel than themselves, all else constant. Individuals also will perceive out-partisans to be more likely to cancel than in-partisans, all else constant.
Republican perception hypothesis	Relative to Democrats, Republicans will perceive a wider gap between their party's propensity to cancel and Democrats' propensity to cancel, all else constant.
Attribution Hypotheses	
Personal attribution hypothesis	Individuals will describe their canceling behaviors as more positive than negative, all else constant.
Out-party attribution hypothesis	Individuals will describe in-partisans' canceling behaviors as more positive than they will describe out-partisans' canceling behaviors, all else constant. By contrast, they will describe out-partisans' canceling behaviors as more negative than they will describe in-partisans' canceling behaviors, all else constant.
Republican attribution hypothesis	Republicans will be more likely, than Democrats, to describe their party's canceling behaviors positively and out-partisans canceling behaviors negatively, all else constant.

Table 2: Conjoint Attributes and Levels

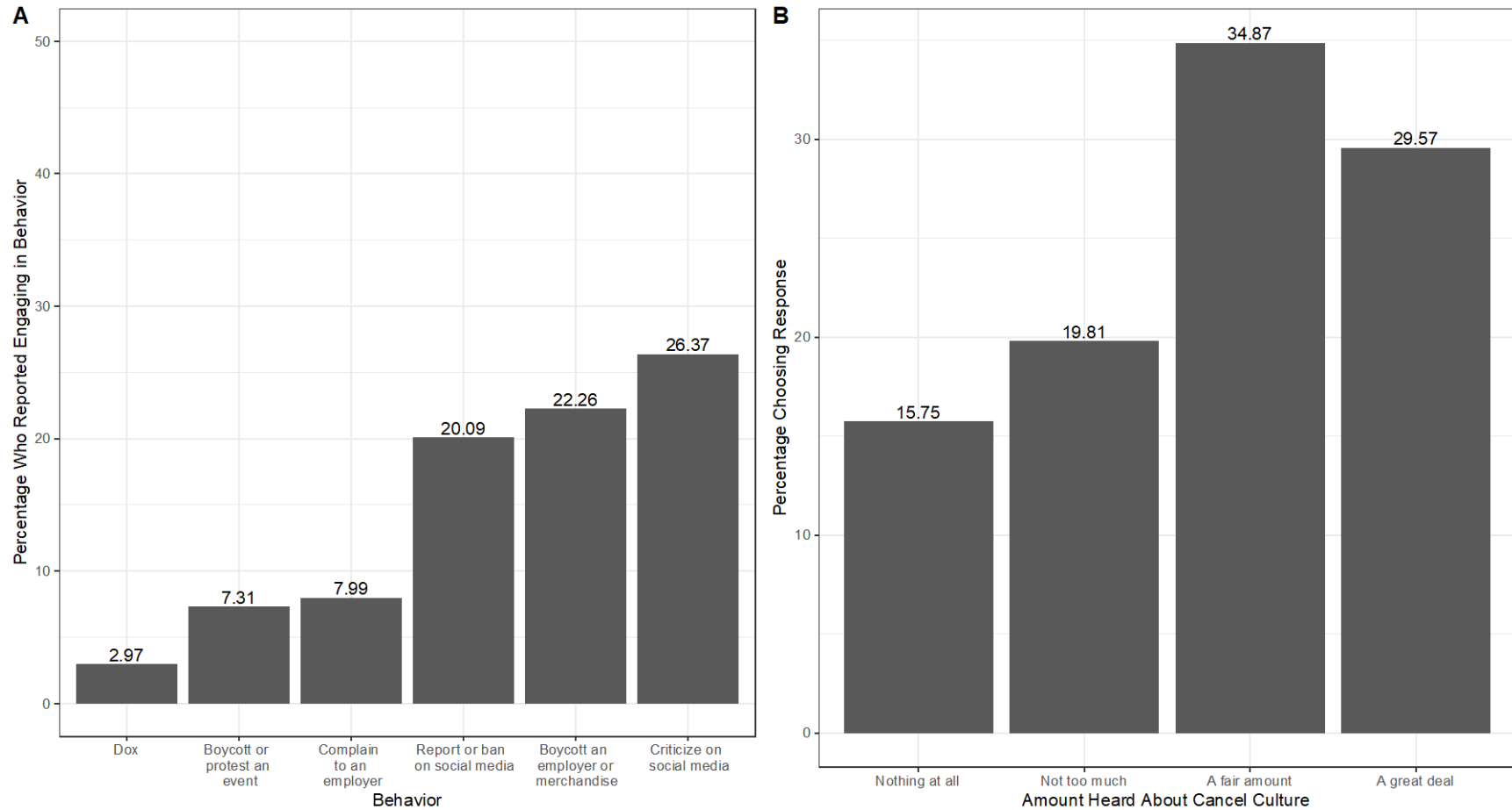
Speaker Partisanship or Race	Democratic, Republican, Black, White
Speaker Occupation or Social Role	Elected Official, Professor, Celebrity, Voter, University Student
Statement	<p>Liberal:</p> <ul style="list-style-type: none">• The Founding Fathers were racist.• America is a racist nation.• All police are bad.• Schools should require students to learn about transgender life and why it is normal. <p>Conservative:</p> <ul style="list-style-type: none">• Confederate statues are about America's heritage and are not racist.• Some races are less intelligent than others.• Athletes who kneel during the National Anthem should be kicked off their teams.• There is no such thing as transgender, only male and female.

Figure 1: Americans Cancel Disagreeable Statements, Not Disliked Speakers



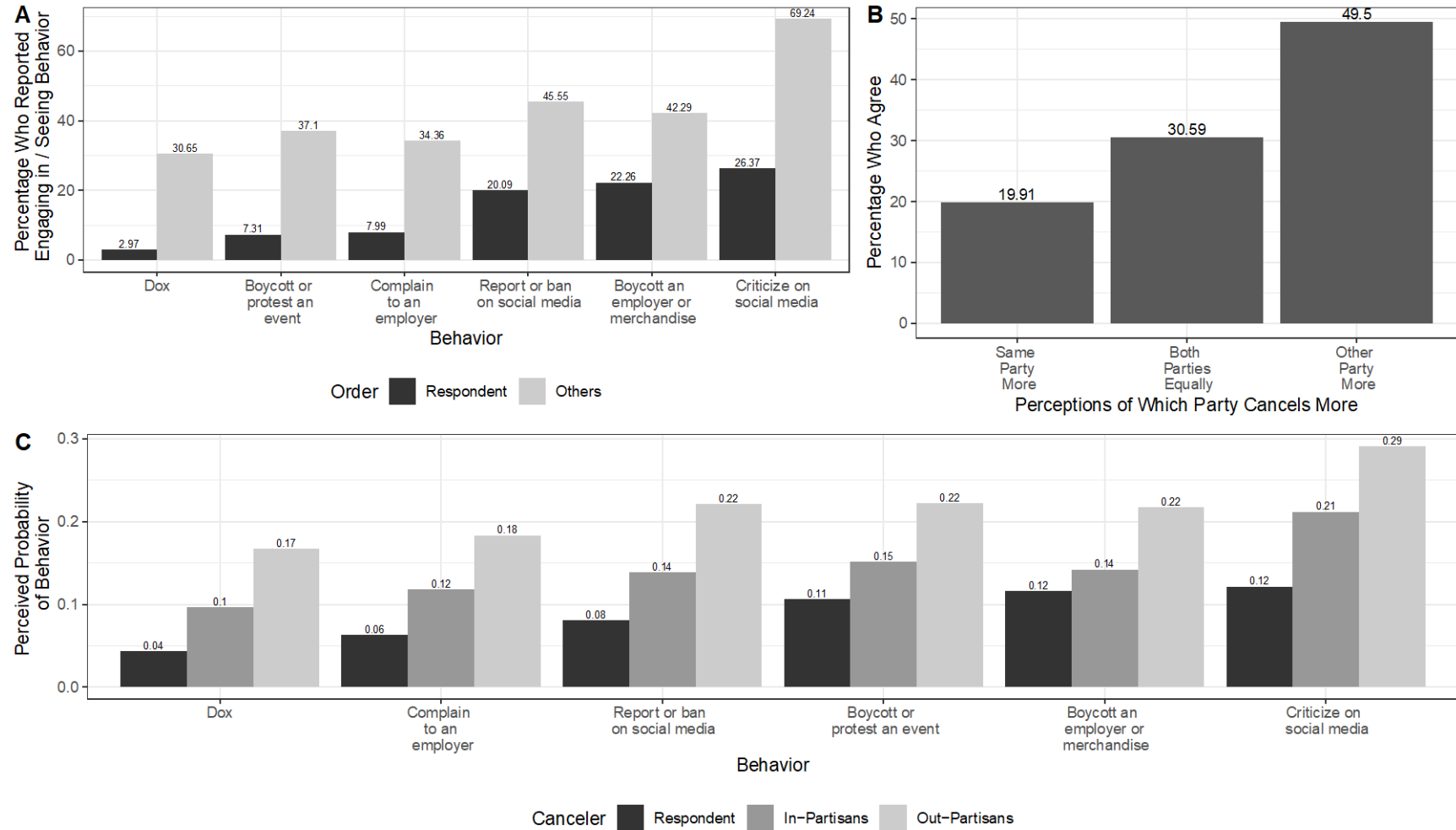
Note: Points are the AMCEs with bars representing 95 percent confidence intervals. The reference category for the Speaker's Party and Speaker's Race levels is the In-Party, Agreeable Statement condition.

Figure 2: Canceling in the Real World is Rare, though Americans have Heard of Cancel Culture



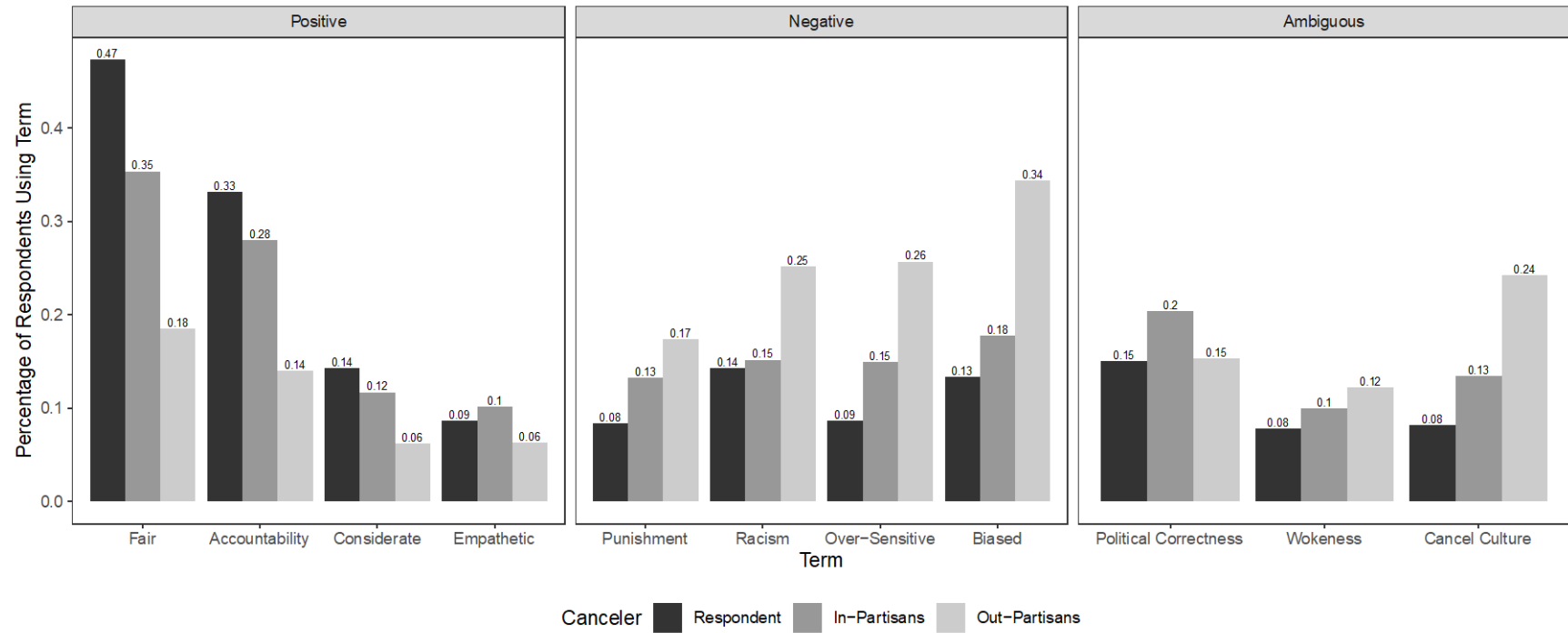
Note: The left panel shows how often respondents engage in the different cancellation behaviors, and the right panel illustrates how much people have heard about cancel culture.

Figure 3: Americans Perceive that Others, Especially Out-Partisans, Engage in Cancel Culture



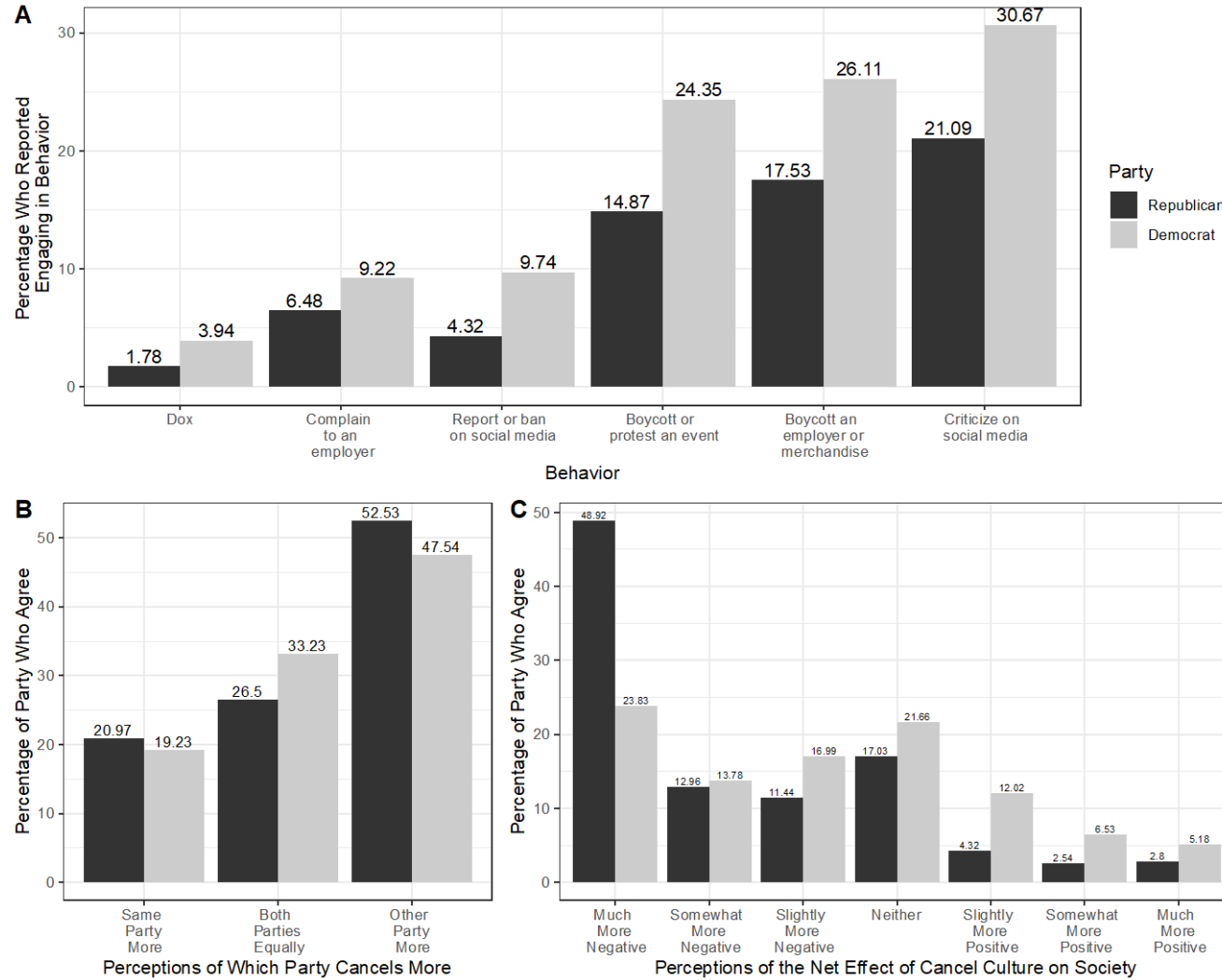
Note: The top-left panel shows how often individuals have seen others engage in cancel culture, the top-right panel shows how often people think canceling behavior comes from the other-party versus the respondent's own party, and the bottom panel shows perceptions of the likelihood of cancellation by the respondent, versus co-partisans, versus out-partisans.

Figure 4: People Attribute Positive Motives to their Own Cancellations, But Not to Others' Cancellations



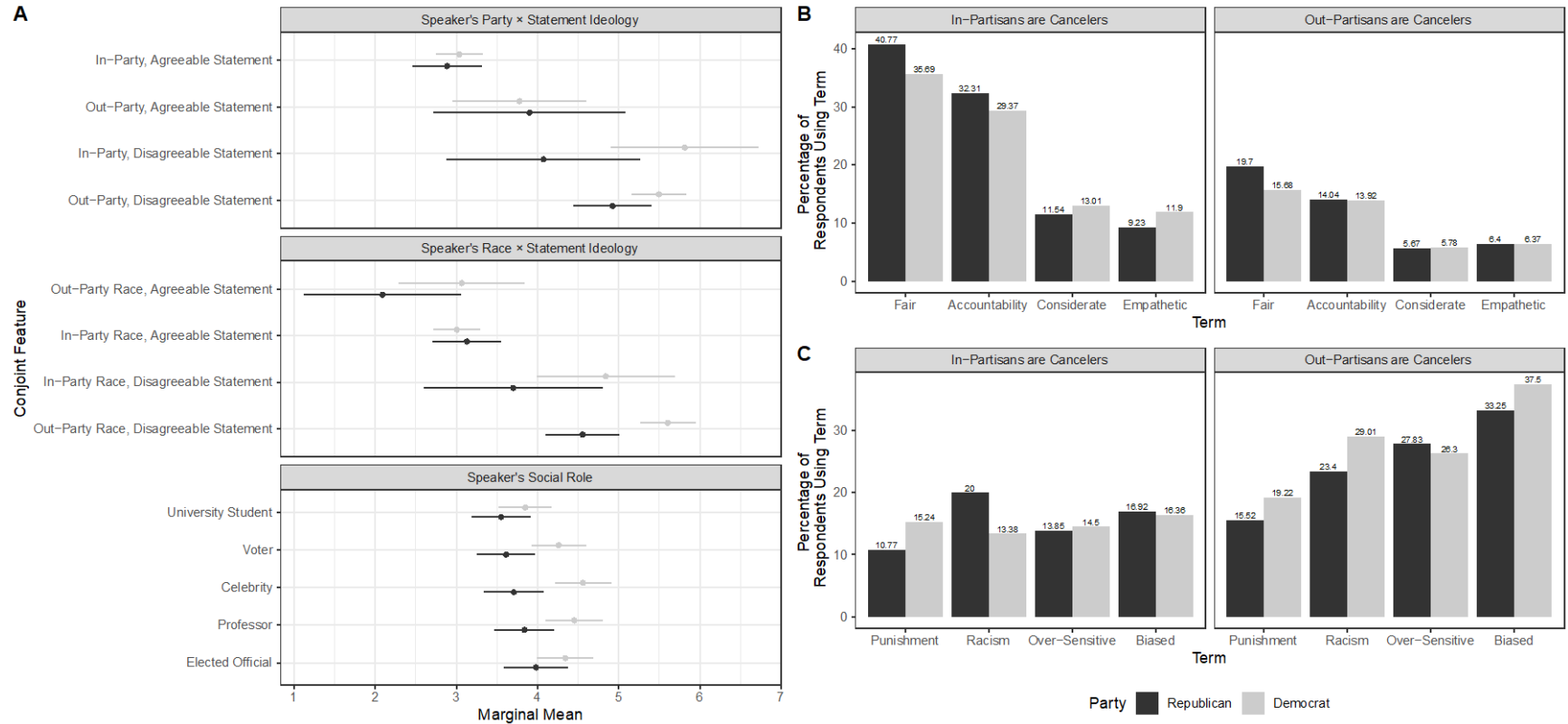
Note: Cell entries are the percentage of times individuals describe their own vs. others cancellation behaviors with various terms.

Figure 5: Democrats Cancel Somewhat More, and Republicans View Cancel Culture More Negatively



Note: The top panel shows partisan differences in actual cancellation behavior, the bottom-left panel shows partisan differences in perceptions of whether the in-party or out-party cancels more, and the bottom-right panel shows partisan differences in assessments of whether cancel culture is a positive or negative force in society.

Figure 6: Cancellations by Democrats and Republicans are Similarly Motivated and Perceived



Note: The left panel shows the marginal means by party, and the right two panels show how often members of each party use positive (top-right panel) and negative (bottom-right) terms to describe in-partisans' and out-partisans' canceling behaviors. Points are the AMCEs with bars representing 95 percent confidence intervals

Supplemental Appendix for: “Myths of Censorship: The Realities and Misperceptions of ‘Cancel Culture’”

Table of Contents:

Appendix A: Pre-Analysis Plan	1–6
Appendix B: Sample Demographics	7–8
Appendix C: Examples of Cancellable Statements from Internet Searches	9–13
Appendix D: Question Wording	14–18
Appendix E: Supplemental Analyses	19–23
Appendix F: Deviations from the Pre-Analysis Plan	24–25
Appendix G: Pre-Test Analyses	27

Appendix A: Pre-Analysis Plan

Our pre-analysis plan is included below and is available at:
https://aspredicted.org/blind.php?x=VKJ_DYZ.

1. Have any data been collected for this study already?

No, we have not collected any data for our main study yet.

2. What's the main question being asked or hypothesis being tested in this study?

Our study seeks to answer three questions related to American's perceptions of cancel culture:

1. How do people respond to various sorts of offensive statements when they encounter it under different circumstances?
 - a. What factors lead them to be more likely to "cancel" someone?
 - b. When do people perceive their actions as constituting "cancel culture"—if ever?
2. How do people perceive others would respond to the same comments, under the same circumstances?
 - a. How do they conceive of the motives for others' canceling behavior, and when do they call it "cancel culture"?
3. What actions have people actually taken, or have seen others take, in response to offensive speech or actions? How do these realities compare to what people think others would do?

Hypotheses about What Predicts cancellation:

H0: Individuals will punish statements that are ideologically inconsistent with their party's position. That is, Democrats will punish conservative statements, and Republicans will punish liberal ones, regardless of the speaker's party.

H1: Individuals will be more likely to cancel public versus private figures. For our purposes, we classify politicians, celebrities, and university faculty as public figures, and voters and college students as private figures.

H2: Individuals will be more likely to cancel out-party speakers than in-party speakers.¹⁹

H3: Individuals will be more likely to cancel speakers from racial groups linked with the out-party. This means Democrats will be more likely to cancel White individuals, and Republicans will be more likely to cancel Black individuals.

H4: Individuals will predominantly describe their canceling behaviors positively—as fair, empathetic, accountability, and considerate.

¹⁹ For our purposes, we classify Independents who lean toward the Republican (Democratic) Party as Republicans (Democrats). Pure Independents (those who do not lean toward either party) are excluded from our main analyses.

Hypotheses about Perceptions of Others' Behavior & Own Behavior:

H5: Individuals will predict that in-partisans will be somewhat more likely to cancel than themselves. Individuals will perceive out-partisans be more likely to cancel than in-partisans. That is, when faced with the same behavior, individuals think that they themselves will be the least likely to cancel, in-partisans will be somewhat more likely to cancel, and out-partisans will be the most likely to cancel.

- a. Given conservative media's extensive coverage of cancel culture, Republican respondents will perceive a wider gap between their own propensity to cancel and Democrats' propensity to cancel.
- b. This will be especially pronounced among those who watch a great deal of conservative media.

H6: Individuals will predominantly describe in-partisans' canceling behaviors positively—as fair, empathetic, promoting accountability, and considerate. By contrast, they will describe out-partisans canceling behaviors negatively—as biased, over-sensitive, punishment, cancel culture, political correctness, racism, and wokeness.

- a. Given conservative media's extensive coverage of cancel culture, descriptions of in-partisans and out-partisans canceling behaviors will be more distinct for Republican respondents.

H7: Individuals will report having seen others cancel far more frequently than they will report having engaged in canceling behavior themselves. We also expect them to think those from the other party are more likely to have engaged in cancellation than those from their own party (with the same partisan asymmetry in H5 above).

3. Describe the key dependent variable(s) specifying how they will be measured.

To measure actual canceling behavior, we will ask respondents them if they themselves have ever engaged in any of the following behaviors when they encountered offensive behavior; we then repeat the question asking them what they have seen others do to measure their perceptions of others' behavior. We based this list off of media reports, and our pre-test suggests this are the main actions that people perform:

- Criticize the person on social media
- Complained to the person's employer
- Boycotted the person's employer, merchandise, or body of work (e.g., books, TV shows)
- Boycotted or protested an event where the person is speaking
- Posted the person's personal information online (i.e., dox them)
- Try to get that person banned from social media
- Other (with a text box for respondents to describe what they had done)

To measure what respondents would do (or what they think others would do) in response to various types of potentially "cancelable" behaviors (see part 4 below), we will ask respondents how likely they would be to engage in each of the behaviors listed above on a four-point Likert

scale (response options: Not at all likely, Not too likely, Somewhat likely, Very likely). To produce a dependent variable for our regressions, we will recode responses to a 1–4 scale and sum these scores across the canceling behaviors.

To measure motivations for hypothetical canceling behavior, we will ask respondents to pick as many of the following terms to describe the canceling behavior in question:

- Accountability
- Biased
- Cancel culture
- Considerate
- Empathetic
- Fair
- Overly sensitive
- Political correctness
- Punishment
- Racism
- Wokeness

Use of each descriptor will be coded as a dummy variable.

We will analyze scenarios with different hypothetical reactors—the respondent themselves, in-partisans, out-partisans—separately.

4. How many and which conditions will participants be assigned to?

We will use a conjoint design wherein the speaker’s racial or partisan identity, social role, and statement are randomized (see the table below). Statements will vary in their ideological content, such that half will be conservative and half will be liberal. In 7/8 scenarios, the speaker will make a statement consistent with her racial or partisan identity, according to well-established partisan stereotypes (Ahler and Sood 2018; e.g., Republican and conservative, Black and liberal). In 1/8 scenarios, the speaker will make a statement *inconsistent* with her racial or partisan identity (e.g., Republican and liberal, Black and conservative).

Respondents will be asked how they would themselves react to four scenarios, how in-partisans would react to one scenario, and how out-partisans would react to three scenarios.

Speaker Race or Partisanship	None, White, Black, Democrat, Republican
Speaker Social Role	Elected Official, Professor, Celebrity, Voter, College Student.
Statement	<p>Liberal:</p> <ul style="list-style-type: none"> -The Founding Fathers were racist. -America is a racist nation. -All police are bad. -Schools should require students to learn about transgender life and why it is normal. <p>Conservative:</p> <ul style="list-style-type: none"> - Confederate statues are about America's heritage and are not racist. -Some races are less intelligent than others. -Athletes who kneel during the National Anthem should be kicked off their teams. -There is no such thing as transgender, only male and female.

We had RAs confirm that these statements reflect actual statements made by various actors. We condensed down those real-world statements here in the interest of simplicity.

5. Specify exactly which analyses you will conduct to examine the main question/hypothesis.

For our conjoint analyses, we will calculate average marginal composite effects (AMCE) and examine marginal means, consistent with the recommendation in Hainmueller, Hopkins, and Yamamoto (2014). To test for differences between conditions, we will concatenate respondents' responses across scenarios, estimate respondent random effects and statement fixed effects, and run the following model:

$$CancelingBehavior_{i,s} = \beta_1 * SpeakerRaceParty_{i,s} + \beta_2 * SpeakerRole_{i,s} + \beta_3 * StatementIdeology_{i,s}$$

To test for statistically significant differences in how groups react (e.g., in the terms Republicans and Democrats use to label actions) across conditions, we add to the previous model an interaction between the condition in question and the grouping variable (e.g., respondent political party).

For continuous dependent variables (e.g., canceling behavior), we will run linear regressions. For binary dependent variables (e.g., use of a particular term to describe canceling behavior), we will run linear regressions with appropriate corrections to the standard errors.

6. Any secondary analyses?

There has been much more focus on cancel culture on the right than on the left. As such, we expect that Republicans will be less likely than Democrats to engage in or report having engaged in canceling behavior. Republicans will also be more likely than Democrats to think that out-partisans are canceling / would cancel various types of behavior.

We also expect that Republicans will be especially likely to describe out-partisan canceling behavior with politically charged terms like “cancel culture”, “political correctness”, and “wokeness”.

Given the extensive coverage of cancel culture on conservative media outlets, especially Fox News, we expect that Republicans who consume more conservative media will be especially likely to display these trends. Here, we measure conservative media exposure using the approach from Dilliplane, Goldman, and Mutz (2013) and ask them which shows they watch on Fox News from a list of choices.

We will also examine the moderating effect of affective polarization. Given our theory, we expect that those who are more affectively polarized will be more likely to cancel those from the out-party, more likely to think out-partisans will engage in canceling behavior (relative to themselves), and more likely to describe out-partisans’ canceling behavior in negative terms.

7. How many observations will be collected or what will determine the sample size? No need to justify decision, but be precise about exactly how the number will be determined.

To ensure a power level of 80 power on an effect size of 4 percent from a five-level variable on three tasks (i.e., conjoints per respondent), our goal is to collect data from 1,750 respondents. We used Lukac and Stefanelli’s (2020) power calculator for conjoint analyses to estimate power.

8. Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)

We conducted a pre-test to determine the feasibility of this approach, and may bring in some of the data from it to the ultimate paper/analysis. When we do so, we will clearly indicate that we are doing so and will flag that the pre-test was not pre-registered.

References

- Ahler, Douglas and Sood, Guarav. 2018. “The Parties in Our Heads: Misperceptions about Party Composition and Their Consequences.” *The Journal of Politics*, 80(3), 964–981.
- Dilliplane, Susanna, Seth Goldman, and Diana Mutz. 2013. “Televised Exposure to Politics: New Measures for a Fragmented Media Environment.” *American Journal of Political Science* 57(1): 236-48.

- Hainmueller, Jens, Daniel Hopkins, and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22(1): 1-30.
- Lukac, Martin and Alberto Stefanelli. 2020. Conjoint Experiments: Power Analysis Tool. Retrieved from <https://mblukac.shinyapps.io/conjoints-power-shiny/>. Accessed August 2021.

Appendix B: Sample Demographics

	Demographic Category	Percentage of Sample
<i>Age</i>	18–24 years old	12.2%
	25–34 years old	17.2%
	35–44 years old	17.5%
	45–54 years old	17.6%
	55–64 years old	17.4%
	65–74 years old	15.0%
	75 years or older	3.1%
<i>Gender</i>	Male	48.5%
	Female	50.2%
	Other	1.3%
<i>Race & Ethnicity</i>	White	73.5%
	Black	13.8%
	American Indian or Alaska Native	1.1%
	Asian American	5.7%
	Native Hawaiian or Pacific Islander	0.4%
	Mixed Race	5.5%
<i>Hispanic</i>	Yes	18.9%
	No	81.1%
<i>Education</i>	Less than high school degree	3.0%
	High school graduate (diploma or GED)	19.1%
	Some college but no degree	24.3%
	Associate degree or vocational training	18.1%
	Bachelor’s degree	24.2%
	Master’s degree	8.6%
	Professional degree (e.g., JD, MD)	1.3%
	Doctoral degree	1.4%
<i>Census Region</i>	Northeast	17.9%
	Midwest	19.7%
	South	37.7%
	West	24.8%

We compare our sample to 2018 benchmarks from the U.S. Census Bureau, via the American Community Survey (ACS). The relevant ACS numbers are as follows.

Age: 18-24: 12.08%; 25-34: 17.87%; 35-50: 24.54%; 51-65: 24.88%; Over 65: 20.65% (Notice we use slightly different categories.)

Gender: Male: 49.2%; Female: 50.8%; (they do not ask “other” but Flores et al. (2016) estimate less than 1 percent identify as transgender).

Race: White: 72.2%; Black: 12.7%; Hispanic: 18.3%; Asian American: 5.6%; Native American: <1%; Other: 5%

Education: Less than high school: 12%; High school: 27.1%; Some college: 28.9%; 4-year college degree: 19.7%; Advanced Degree: 12.3%

Census Region: Northeast: 17.2%; Midwest: 20.9%; South:38.1%; 23.8%

Across categories, our sample matches the ACS benchmarks fairly well. Our biggest discrepancy is that we under-estimate the least well-educated (and over-estimate those with some college or a bachelor's degree). These are well-known limitations of any survey sampling procedure, not just our own—this is linked in that this population is less likely to be online,

Reference

Flores, Andrew, Jody Herman, Gary Gates, and Taylor Brown. 2016. “How Many Adults Identify as Transgender in the United States?” *The Williams Institute*.
<https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf>.

Appendix C: Examples of Cancellable Statements from Internet Searches

In the experiment, we used four liberal statements and four conservative statements that were designed to be “cancelable”—that is, offensive enough that someone might cancel them. But how to pick those statements? We had a set of research assistants do a Google News search potentially offensive topics and document that these were, in fact, sensitive statements that had appeared in the news.

The Founding Fathers Were Racist:

Only one citizen spoke against adoption of the HB 1775 regulations—Sapphira Lloyd, a public-school student who is black... “Native American voices are not heard, because we’re still on their land,” Lloyd said. “Latinx/Hispanic communities are never cared for. Black voices have never been heard, yet here we are still trying. Are we going to forget the fact that Thomas Jefferson and all of our famous Founding Fathers were slave owners?” (<https://www.ocpathink.org/post/teachers-face-loss-of-license-for-racist-instruction>)

Thomas Jefferson, author of the Declaration of Independence, wrote in his 1789 book “Notes on the State of Virginia” that Blacks were a “distinct and inferior species” when compared with Whites. One year later, Jefferson joined the other Founding Fathers in supporting the Naturalization Act of 1790, which made citizenship in the United States available only to “free, white persons” (<https://www.washingtonpost.com/politics/2021/06/30/would-founding-fathers-support-critical-race-theory/>).

These restrictions create quite a few concerns for, say, a Florida social studies teacher. Now, such a teacher must figure out how to tell students what the Founding Fathers really meant when they wrote “We the people” in the U.S. Constitution, without saying the Founding Fathers were racist for excluding Black people from the meaning of that phrase (<https://theconversation.com/bans-on-critical-race-theory-could-have-a-chilling-effect-on-how-educators-teach-about-racism-163236>).

This was written in pure ignorance. The founding fathers were racist slave owners and there is no sugar coating that, which is what you are trying to do. You can't handle the truth (<https://medium.com/truth-in-between/the-alternative-1619-project-reading-challenge-introduction-2021-245e80231ad6>).

Geiss responded by implying the Founding Fathers were racist and sexist, and that Theis shouldn't “get all weepy and nostalgic” about men that wouldn't have allowed women or minorities to serve with them. “There's a whole bunch of us serving right now that the Founding Fathers wouldn't have wanted in this chamber, or any other for that matter,” Geiss said (<https://www.mlive.com/public-interest/2021/03/michigan-senate-reaffirms-second-amendment-in-resolution-democrats-call-insurrectionist.html>).

The point is that the founding fathers were racist horrible people and this is the America they created (<https://variety.com/2020/film/reviews/hamilton-review-lin-manuel-miranda-disney-plus-1234694098/>).

Confederate Statues Are About America's Heritage and Are Not Racist:

The document outlines a range of suggested methods for protecting Confederate monuments, flags, school dedications and mascots from what it describes as “heritage attacks” from those seeking to remove them (<https://www.theguardian.com/us-news/2021/jul/04/sons-of-confederate-veterans-manual-statues-symbols>).

“What I do think is clearly a bridge too far is this nonsense that we need to airbrush the Capitol and scrub out everybody from years ago who had any connection to slavery,” McConnell told reporters when asked about an unrelated provision in a defense bill that would change Confederate-named bases (<https://thehill.com/homenews/senate/503045-mcconnell-rejects-push-to-airbrush-the-capitol-of-confederate-statues>).

These statues, however, still retain cultural value as part of the historic fabric of American communities (<https://thefederalist.com/2020/07/01/why-we-should-keep-confederate-statues-standing/>).

In my opinion, removing historical monuments only removes history facts. If we eliminate historical monuments, what is next? History books that mention Confederate soldiers? All facts about how Africans were brought into this country unwilling? No mention of how our country was developed? The book Tom Sawyer is banned because of the N-word, then the Confederate flag, now all statues of the Confederacy... (<https://www.star-telegram.com/opinion/letters-to-the-editor/article169447232.html>).

Defenders of these monuments, though, say they are benign markers of Southern heritage and culture, important historical markers that pay tribute to the hundreds of thousands of Southern men killed in America's bloodiest conflict. Removing these memorials, they say, is an offensive erasure of history (<https://www.kqed.org/lowdown/27855/heritage-or-hate-a-map-of-confederate-monuments-around-the-country>).

America Is a Racist Nation:

As Soledad O'Brien said, “It sounds like a racist country to me,” in response to Scott—but it seemingly applies to comments by Harris as well (<https://www.brookings.edu/blog/how-we-rise/2021/05/04/is-the-united-states-a-racist-country/>).

Since then, conservative lawmakers, commentators and parents have raised alarm that critical race theory is being used to teach children that they are racist, and that the U.S. is a racist country with irredeemable roots (<https://www.texastribune.org/2021/06/22/texas-critical-race-theory-explained/>).

In Iowa and Tennessee, the laws say teachers cannot teach that the United States is “fundamentally racist” (<https://theconversation.com/bans-on-critical-race-theory-could-have-a-chilling-effect-on-how-educators-teach-about-racism-163236>).

Only then can we truly begin to heal from the deep-rooted pain, dehumanization and prejudice that is embedded in the systems of this country

<https://www.cnn.com/2020/06/18/politics/perspective-american-black-journalist-race/index.html>).

Saying that America is racist is not a radical statement. If that requires a longer explanation or definition, so be it. The fact, in the end, is not altered
(<https://www.nytimes.com/2021/05/02/opinion/america-racism.html>).

That early micro-aggression forewarned me that America may be the land of opportunity for many, but it would still reduce me to the colour of my skin and find me unworthy
(<https://www.bbc.com/news/world-africa-52895490>).

That was certainly the message of progressive Democrat and “Squad” member Rep. Cori Bush of Missouri: “When they say that the 4th of July is about American freedom, remember this: the freedom they’re referring to is for White people,” Bush tweeted. “This land is stolen land and Black people still aren’t free.” (https://wacotrib.com/opinion/columnists/michael-graham-badmouthing-the-flag-a-good-way-to-lose-elections/article_1c7bf35a-df48-11eb-9734-4b355f656485.html).

Some Races Are Less Intelligent than Others:

The idea that certain races are inherently more intelligent than others is being trumpeted by a small group of anthropologists, IQ researchers, psychologists and pundits who portray themselves as noble dissidents, standing up for inconvenient facts
(<https://www.theguardian.com/news/2018/mar/02/the-unwelcome-revival-of-race-science>).

Cofnas’s paper “disingenuously argues that the best explanation of differences in IQ scores between racial and ethnic groups is genetics,” reads a petition posted by Mark Alfano, associate professor of philosophy at Delft University of Technology in the Netherlands and associate professor of philosophy at Macquarie University.
(<https://www.insidehighered.com/news/2020/01/23/intelligent-argument-race>).

In that dissertation Richwine had argued, among other things, that American “Hispanics” are less intelligent than native-born whites as evidenced by their lower average scores on IQ tests. Richwine then attributed Hispanics’ alleged intellectual inferiority at least partly to genetic factors (<https://www.theatlantic.com/national/archive/2013/05/why-people-keep-misunderstanding-the-connection-between-race-and-iq/275876/>).

Denial of any genetic component in human variation, including between groups, is not only poor science, it is likely to be injurious both to unique individuals and to the complex structure of societies (<https://www1.udel.edu/educ/gottfredson/30years/Rushton-Jensen30years.pdf>).

The reasons for the differences in IQ within any racial-ethnic group appear to be the same for all groups: a mixture of genetic and environmental influences (<https://dana.org/article/pretending-that-intelligence-doesnt-matter/>).

All our social policies are based on the fact that their intelligence is the same as ours - whereas

all the testing says not really (<https://www.independent.co.uk/news/science/fury-dna-pioneer-s-theory-africans-are-less-intelligent-westerners-394898.html>).

That year, the scientist told Britain's Sunday Times that he was "gloomy about the prospect of Africa" because he believed African intelligence was genetically lower than that of Europeans (<https://www.washingtonpost.com/nation/2019/01/14/father-dna-says-he-still-believes-link-between-race-intelligence-his-lab-just-stripped-him-his-titles/>).

All Police Are Bad:

At Cypress College in California, there was a professor who berated a student for saying that "police are heroes" and went on to talk about how "police officers have committed atrocious crimes and have gotten away with it, and have never been convicted of any of it" and goes on to state that she doesn't trust the police and that her life is in more danger in their presence. The professor has been put on a leave of absence (<https://www.latimes.com/california/story/2021-05-03/oc-professor-berates-student-who-says-police-are-heroes>).

Athletes Who Kneel During the National Anthem Should Be Kicked Off Their Teams:

In response to Gwen Berry turning her back on the American flag on the podium, Dan Crenshaw has demanded her removal from the U.S. Olympic team and released statements such as "We don't need any more activist athletes" as a campaign against the "anti-racism industry that is incentivizing victimhood" (<https://www.washingtonpost.com/sports/olympics/2021/06/29/gwen-berry-dan-crenshaw/>).

In response to the NFL's new policy where teams will be subject to a fine if a player does not comply with the rules about kneeling, President Trump stated that they did the right thing, while also stating that taking a knee during the national anthem during a National Football League game should maybe be a deportable offense (<https://www.nbcnews.com/politics/donald-trump/trump-says-nfl-players-who-kneel-during-national-anthem-maybe-n876996>).

Schools Should Require Students to Learn about Transgender Life and Why It Is Normal:

School activities and practices should be gender neutral. Transgender students will be allowed to use restrooms, locker rooms, and changing facilities that correspond to their gender identity... Atkinson said the next steps would be to incorporate regular education about transgender students into staff professional development and training and create campaigns about the new policy similar to anti-bullying campaigns (<https://www.wtkr.com/news/portsmouth-public-schools-adopt-policy-supporting-transgender-students>).

Schools must be proactive in creating a culture and practices that respect and value all students and foster understanding of gender identity and expression within the school community. Creating such an inclusive culture will greatly affect what research shows: that transgender and gender expansive students are at higher risk for being marginalized, victimized, or bullied (<https://www.schools.nyc.gov/school-life/school-environment/guidelines-on-gender/guidelines-to-support-transgender-and-gender-expansive-students>).

There Is No Such Thing As Transgender, Only Male or Female:

There is no such thing as transgender

([https://www.realclearpolitics.com/video/2021/06/28/woman outraged at spa for letting trans in womens room are you ok with a man showing his penis around girls.html](https://www.realclearpolitics.com/video/2021/06/28/woman_outraged_at_spa_for_letting_trans_in_womens_room_are_you_ok_with_a_man_showing_his_penis_around_girls.html);

<https://www.washingtonpost.com/dc-md-va/2021/07/05/transgender-woman-los-angeles-spa/>;

Riot police used tear gas and rubber bullets to disrupt the annual Pride parade, intensifying a crackdown on the march at a time of rising government hostility toward LGBTQ individuals in Turkey, advocacy groups say. At least 20 people were detained, local media reported

(<https://www.washingtonpost.com/world/2021/06/26/istanbul-pride/>).

Appendix D: Question Wording

Partisanship

Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what?

- Democrat
- Republican
- Independent
- Other _____
- No preference

(The question below was only shown to respondents who indicated a party.)

Would you call yourself a strong [PARTY] or a not very strong [PARTY]?

- Strong [PARTY]
- Not very strong [PARTY]

(The question below was only shown to respondents who did not indicate a party.)

Do you think of yourself as closer to the Republican or Democratic party?

- Republican Party
- Democratic Party
- Neither

Affective Polarization

We'd like you to rate how you feel towards the Republican and Democratic parties on a scale from 0 to 100. Ratings between 0 and 49 degrees mean that you feel unfavorable and cold (with 0 being the most unfavorable/coldest). Ratings between 51 and 100 degrees mean that you feel favorable and warm (with 100 being the most favorable/warmest). A rating of 50 means you have no feelings one way or the other. How would you rate your feelings toward the Republican and Democratic parties?

	0	10	20	30	40	50	60	70	80	90	100
Republican Party											
Democratic Party											

Partisan News Consumption

Which of the following programs do you watch regularly on television or online at least once a month? Check all that apply. If none apply, simply skip the question.

- Tucker Carlson Tonight
- The Five
- Hannity
- The Ingraham Angle
- Special Report with Bret Baier
- Fox News Primetime
- Gutfeld!
- America's Newsroom
- Outnumbered
- The Faulkner Focus

Which of the following programs do you watch regularly on television or online at least once a month? Check all that apply. If none apply, simply skip the question.

- The Rachel Maddow Show
- The Last Word with Lawrence O'Donnell
- Deadline: White House with Nicolle Wallace
- All In with Chris Hayes
- The Beat With Ari Melber
- The ReidOut
- The 11th Hour with Brian Williams
- The Lead with Jake Tapper
- Cuomo Prime Time
- Anderson Cooper 360

Actual Canceling Behavior

There has recently been a discussion about how people should respond to others who say or do offensive things. Which of the following, if any, have you done in response to someone else's offensive speech or actions? Check all that apply. If none apply, simply skip the question.

- Criticized a person on social media
- Complained to a person's employer
- Boycotted a person's employer, merchandise, or body of work (e.g., books, TV shows)
- Boycotted or protested events where a person is speaking
- Posted a person's personal information online (i.e., doxxed them)
- Reported or tried to ban a person from a social media site
- Other _____

Which of the following, if any, have you seen others do in response to someone else's offensive speech or actions? Check all that apply. If none apply, simply skip the question.

- Criticized a person on social media
- Complained to a person's employer
- Boycotted a person's employer, merchandise, or body of work (e.g., books, TV shows)
- Boycotted or protested events where a person is speaking
- Posted a person's personal information online (i.e., doxxed them)
- Reported or tried to ban a person from a social media site
- Other _____

(The question below was only shown to respondents who indicated they had seen others engage in at least one canceling behavior.)

You indicated that you have seen others do the following in response to offensive speech or actions: [CANCELING BEHAVIORS]. How often were the people who engaged in these actions Republicans?

- Never
- Not too often
- Somewhat often
- Very often
- Extremely often
- Don't know

(The question below was only shown to respondents who indicated they had seen others engage in at least one canceling behavior.)

You indicated that you have seen others do the following in response to offensive speech or actions: [CANCELING BEHAVIORS]. How often were the people who engaged in these actions Democrats?

- Never
- Not too often
- Somewhat often
- Very often
- Extremely often
- Don't know

Conjoint Dependent Variable: Canceling Behavior

(In your opinion,) [h]ow likely would (you/Republican voters/Democratic voters) be to do each of the following in response to the [SPEAKER JOB OR SOCIAL ROLE]'s statement?

	Not at all likely	Not too likely	Somewhat likely	Very likely
Criticize the speaker on social media	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Complain to the speaker's employer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Boycott the speaker's employer, merchandise, or body of work (e.g., books, TV shows)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Boycott or protest events where that person is speaking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Post the speaker's personal information online (i.e., dox them)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Report or try to ban the speaker from a social media site	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

(The question below was only shown to respondents who indicated the canceler in question would be at least "somewhat" likely to engage in at least one canceling behavior.)

You indicated that (you/Republican voters/Democratic voters) would be likely to do the following in response to the [SPEAKER JOB OR SOCIAL ROLE]'s statement: [CANCELING BEHAVIORS]. Which of the following words, if any, would you use to describe (your/Republican voters'/ Democratic voters') response?

- Fair
- Biased
- Empathetic
- Over-sensitive
- Accountability
- Punishment
- Cancel culture
- Considerate
- Political correctness
- Racism
- Wokeness
- Other _____

General Questions about Cancel Culture

How much have you heard about the term “cancel culture”?

- Nothing at all
- Not too much
- A fair amount
- A great deal

As you may know, “cancel culture” is the practice of censoring, shaming, or ostracizing (“canceling”) people who say or do things considered to be offensive. In your opinion, does cancel culture have a more positive or a more negative effect on society?

- Much more negative of an effect
- Somewhat more negative of an effect
- Slightly more negative of an effect
- Neither a positive nor negative effect
- Slightly more positive of an effect
- Somewhat more positive of an effect
- Much more positive of an effect

Appendix E: Supplemental Analyses

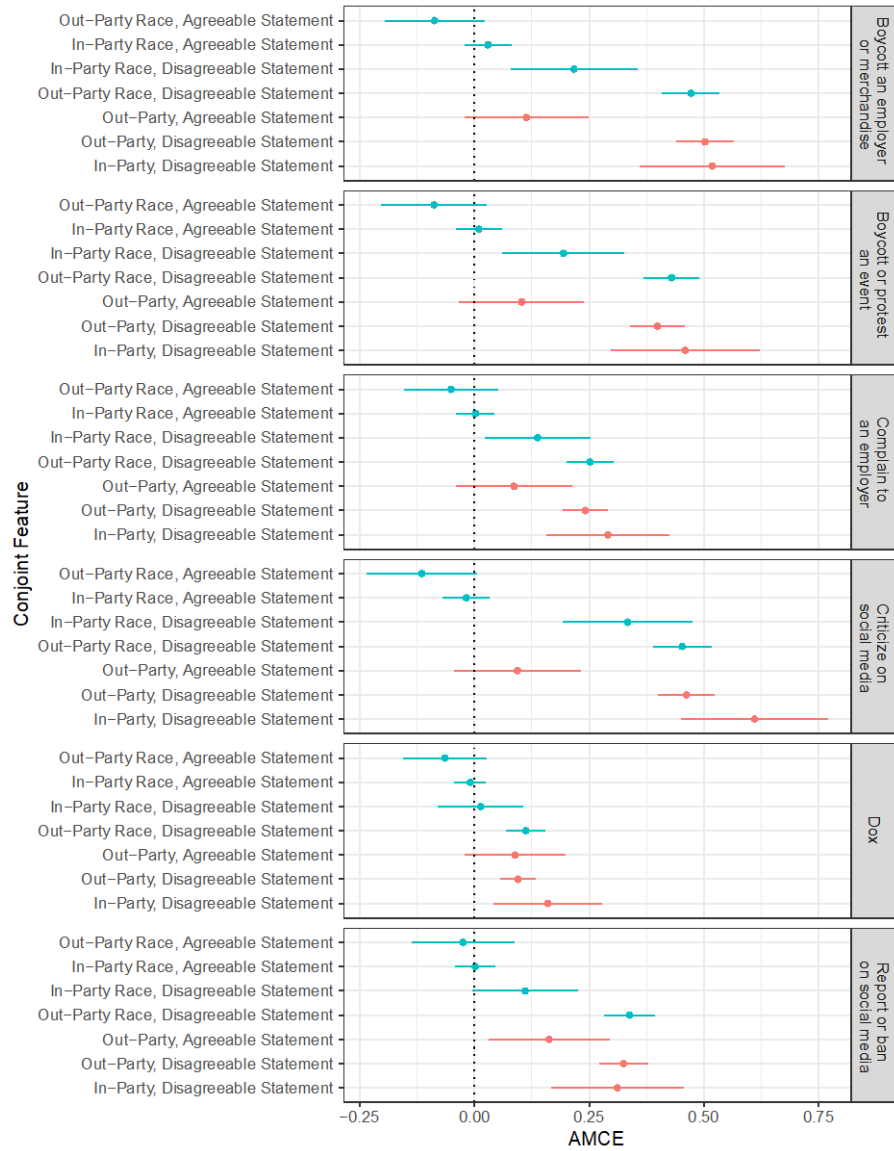
Here, we present several supplemental analyses we could not include in the body of the paper.

Analysis of First-Order Conjoint Experiments

Separating Canceling Behaviors

In the body of the paper, we analyzed our first-order conjoint results in terms of the sum of the likelihoods of engaging in our six canceling behaviors, as specified in our pre-analysis plan. Figure E1 replicates our analyses separately for each canceling behavior. These findings are generally consistent with our main results (Figure 1 in the body of the paper). However, conjoint features have weaker effects on more severe canceling behaviors (e.g., doxing).

Figure E1: Why Do People Engage in Canceling Behaviors?

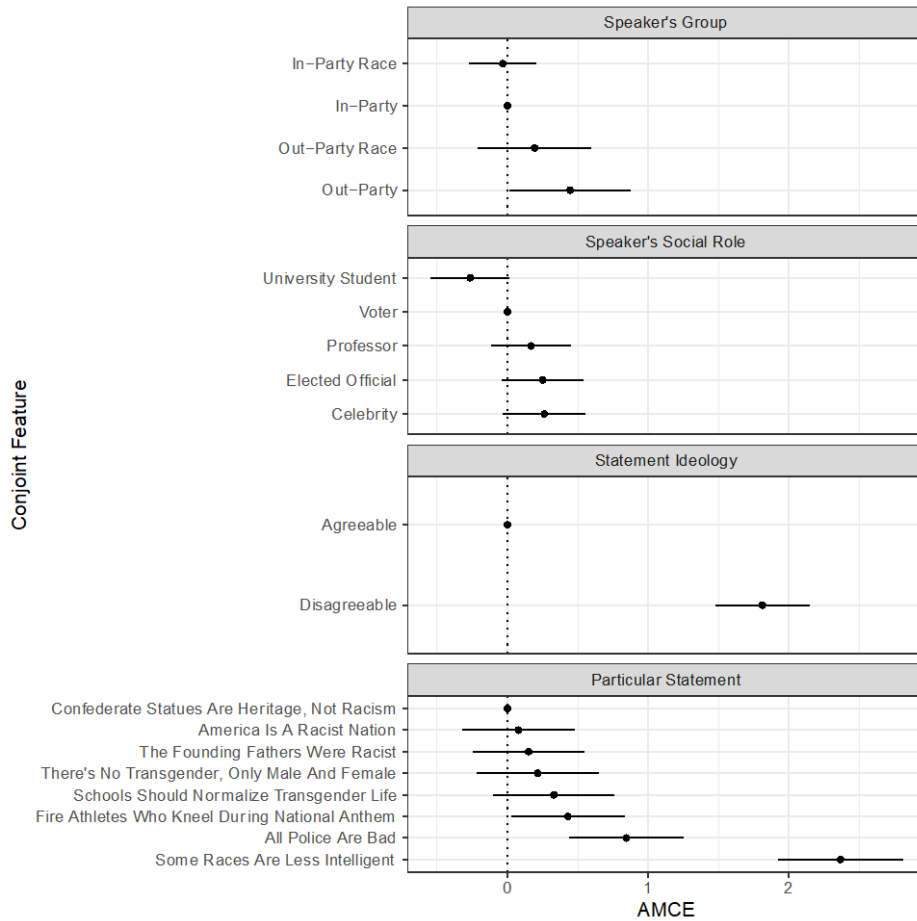


Note: Points are the AMCEs with bars representing 95 percent confidence intervals. The reference category for the Speaker's Party and Speaker's Race levels is the In-Party, Agreeable Statement condition. Blue points represent race source attributes, whereas red points represent partisan source attributes.

Separating Speaker Identity and Statement Conjoint Features

For the sake of presentation, in the body of the paper, we crossed our speaker identity and statement ideology factors. However, we randomized these features separately. To show that crossing these factors did not bias our results, we present below the “raw” analysis of our data, as well as the effects of particular statements, in Figure E2. These results are consistent with those in the body of the paper: the ideological thrust of a statement, not the speaker's identity, drives canceling behavior. Some statements—such as the assertion that some races are less intelligent than others—are particularly cancelable.

Figure E2: Why Do People Cancel?

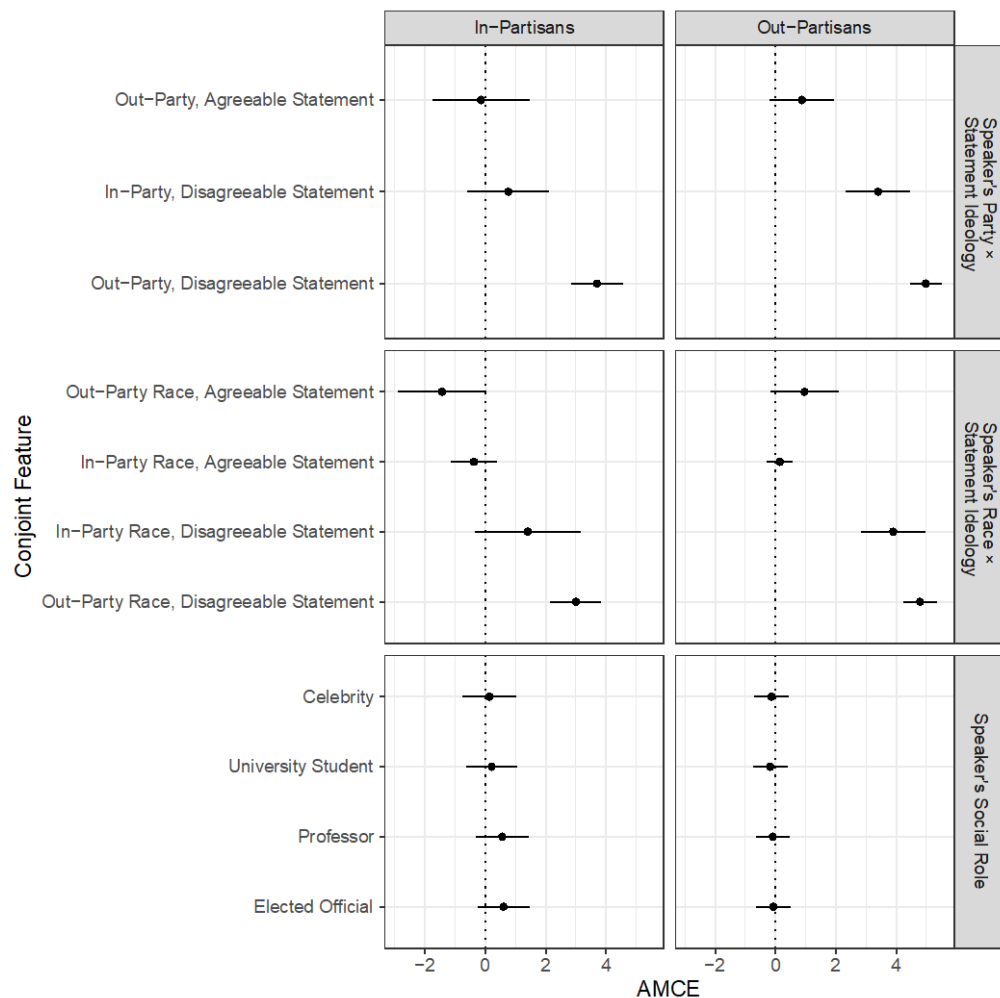


Note: Points are the AMCEs with bars representing 95 percent confidence intervals. The reference category for the Speaker's Party and Speaker's Race levels is the In-Party, Agreeable Statement condition.

Analysis of Second-Order Conjoint Experiments

In Figure E3, we visualize the effects of each conjoint feature on the perceived likelihood that in-partisans and out-partisans would engage in canceling behavior. Here, conjoint attributes such as “in-party” are defined relative to the supposed canceler. That is, when out-partisans are the cancelers in question, attributes referencing the “in-party” refer to the supposed canceler’s in-party and the respondent’s out-party.

Figure E3: Why Do Others’ Perceive In-Partisans and Out-Partisans Cancel?



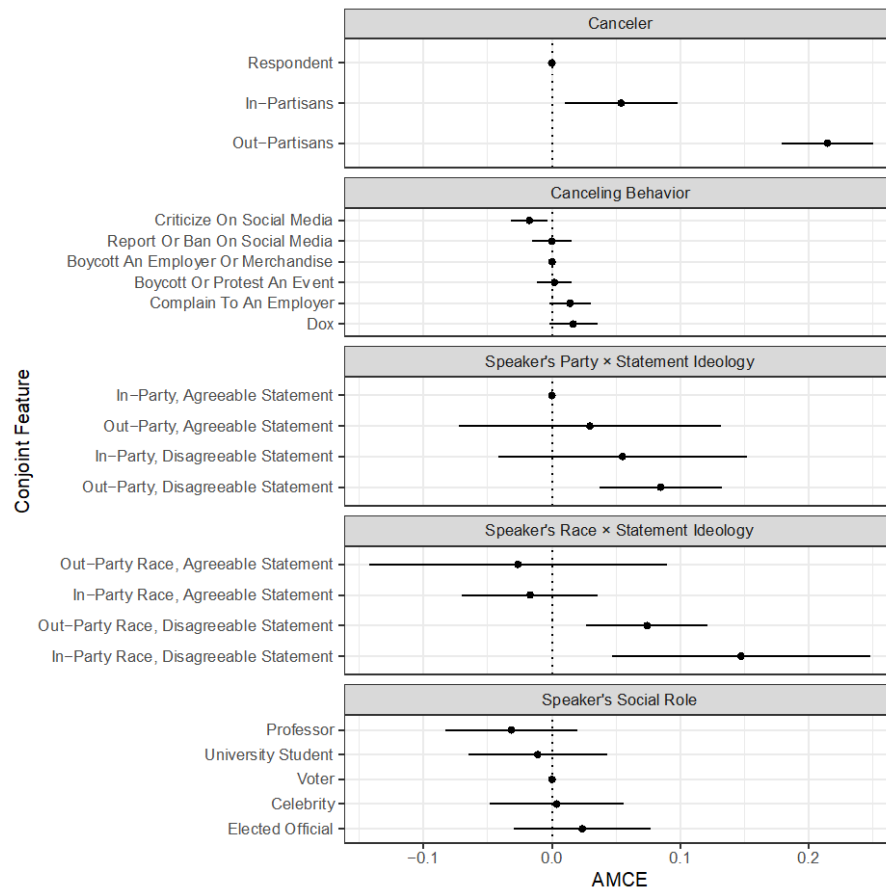
Note: Points are the AMCEs with bars representing 95 percent confidence intervals. The reference category for the Speaker’s Party and Speaker’s Race levels is the In-Party, Agreeable Statement condition. Conjoint attributes are defined relative to the supposed canceler.

What Gets Called “Cancel Culture”?

In our study, we were careful not to describe any activity as “cancel culture” given concerns about social desirability biases and the term’s negative connotations, especially among Republicans. A fortunate byproduct of this decision is that we can analyze what, precisely, respondents label as cancel culture. Recall that, in our conjoint experiments, respondents could use any of 11 descriptors, including cancel culture, to describe any given canceling behavior performed by any given canceler (themselves, in-partisans, or out-partisans) in response to any given speaker making any given statement. As such, we can examine how use of the term cancel culture varies with (1) what a canceler did, (2) who the canceler is, (3) who the speaker being canceled is, and (4) what the speaker said.

We model whether something is labeled as cancel culture as a function of these four factors and present the results in Figure E4. These results suggest that behavior by others, and especially out-partisans, is more likely to be labeled cancel culture. Moreover, when a canceler is responding to a statement presumably perceived to be disagreeable to the canceler, their actions are more likely to be considered cancel culture. Remarkably, the particular action taken by the canceler has almost no effect of whether their actions are labeled as cancel culture.

Figure E4: What’s Labeled Cancel Culture? Out-Partisans Responding to Disagreeable Statements



Note: Points are the AMCEs with bars representing 95 percent confidence intervals. The reference category for the Speaker's Party and Speaker's Race levels is the In-Party, Agreeable Statement condition.

Appendix F: Deviations from the Pre-Analysis Plan

Redesignating Political Correctness, Wokeness, and Cancel Culture as Ambiguous

As noted in the text, our pre-registration designated the terms “political correctness,” “wokeness,” and “cancel culture” as negative. However, given that the significance of these terms varies across the political spectrum, we re-classified them as ambiguous. As a robustness check, we re-compute our tests of the personal attribution, out-party attribution, and Republican out-party attribution hypotheses using our original designations.

Consistent with our personal attribution hypothesis, respondents described their own canceling behavior with at least one positive term 73 percent of the time and at least one negative term 49 percent of the time (this difference is statistically significant, $p < 0.01$). We also still find support for the out-party attribution hypothesis: respondents described in-partisans’ canceling behavior negatively only 64 percent of the time, but described identical behavior by out-partisans negatively 80 percent of the time ($p < .001$).

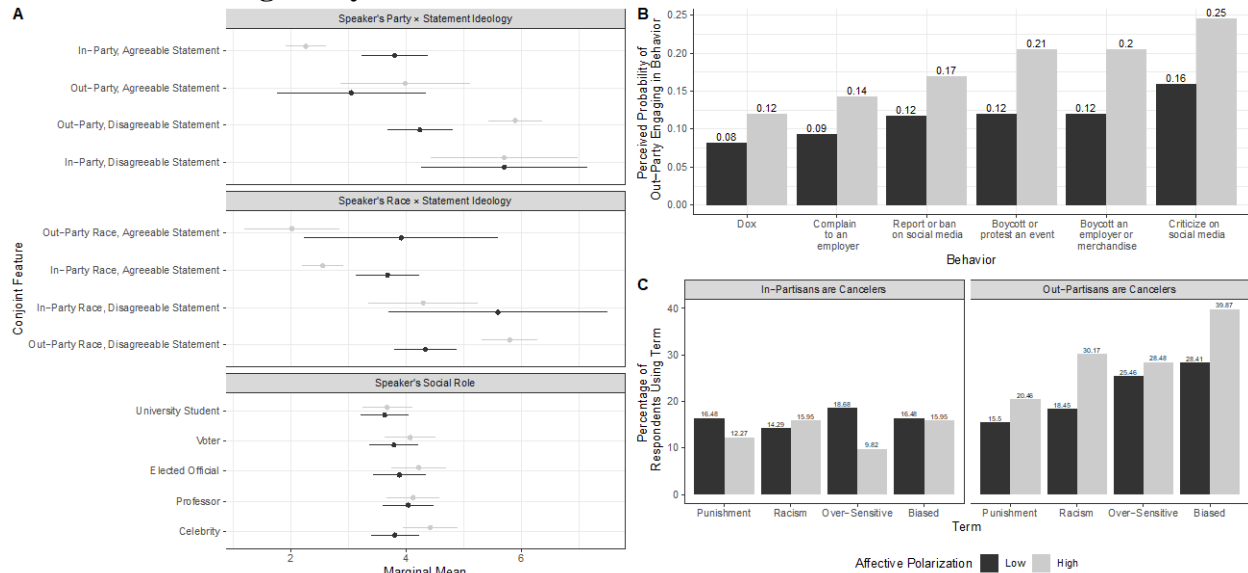
Finally, both Democrats and Republicans are more likely describe the behavior of the other party, as opposed to their own party, in negative terms ($p < .001$). When political correctness, wokeness, and cancel culture count as negative terms, however, *Republicans* are somewhat more likely to describe the out-party in more negative terms ($p < .01$). This finding contrasts with our in-paper results, where Democrats were somewhat more likely to describe out-partisans’ canceling behaviors using other negative terms. However, it is consistent with our Republican attribution hypothesis, which suggests that Republicans would view out-partisans’ canceling behavior more negatively than in-partisans’ canceling behavior. This finding also confirms our expectation that Republicans would be especially likely to describe out-partisan canceling behavior with politically charged terms like “cancel culture,” “political correctness,” and “wokeness.”

Moderation Analyses

Due to space constraints, we excluded from the body of the paper two analyses from section six of our pre-analysis plan, on the moderating effects of affective polarization and Fox News consumption (among Republicans). We expected the affectively polarized to be especially likely to confirm the partisan source, canceling perception, and out-party attribution hypotheses. Moreover, given the extensive coverage of cancel culture on conservative media outlets—and especially Fox News—we expected that Republican Fox News consumers would be especially likely to confirm these hypotheses.

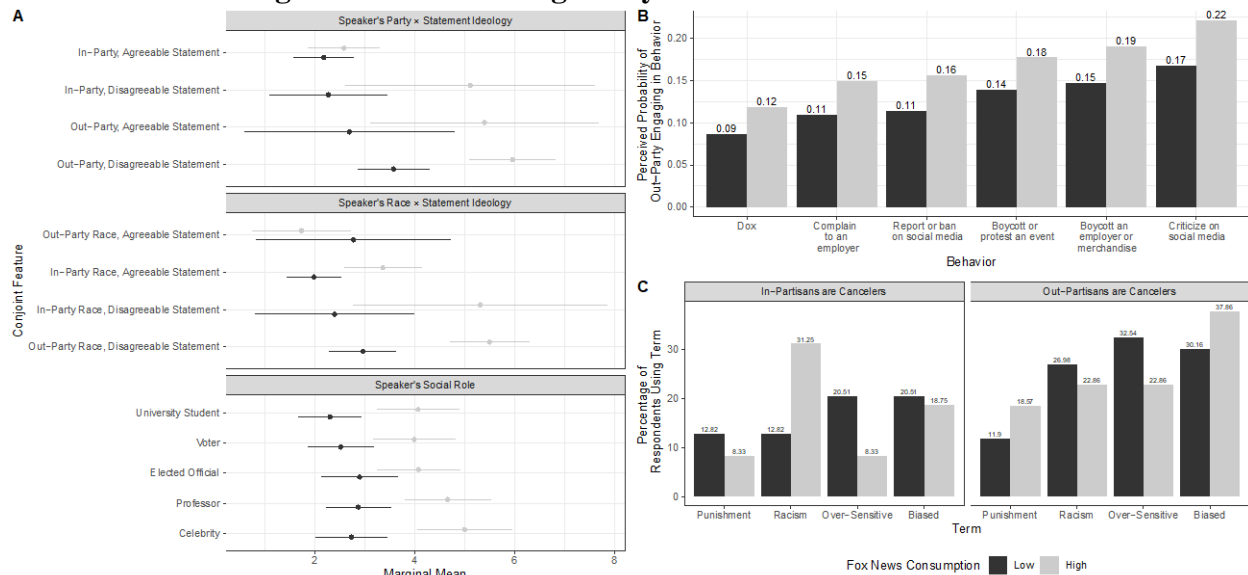
As shown in in Figures F1 and F2, our analyses generally confirm these expectations. The affectively polarized are more likely to cancel those from the out-party, think out-partisans engage in canceling behavior, and describe out-partisans’ canceling behavior in negative terms. Likewise, among Republicans, Fox News consumers are more likely to cancel Democrats, think Democrats engage in canceling behavior, and describe Democrats’ canceling behavior as “punishment” and “biased.” However, Republican Fox News consumers are less likely to describe Democrats’ canceling behavior as “racism” or “over-sensitivity.”

Figure F1: The Affectively Polarized Cancel More and Perceive Out-Partisans' Canceling Behavior More Negatively



Note: The affectively polarized are more likely to cancel those from the out-party (Panel A), think out-partisans engage in canceling behavior (Panel B), and describe out-partisans' canceling behavior negatively (Panel C). Data points for low and high affective polarization represent respondents in the bottom and top terciles for affective polarization, respectively.

Figure F2: Among Republicans, Fox News Consumers Cancel More and Perceive Out-Partisans' Canceling Behavior More Negatively



Note: Among Republicans, Fox News consumers are more likely to cancel Democrats (Panel A), think Democrats engage in canceling behavior (Panel B), and describe Democrats' canceling behavior as "punishment" and "biased" (Panel C). However, Republican Fox News consumers are less likely to describe Democrats' canceling behavior as "racism" or "over-sensitivity" (Panel C). Data points for low and high Fox News consumption represent Republicans in the bottom and top terciles for Fox News Consumption, respectively. We measured Fox News consumption by having respondents select which Fox News programs they regularly watch, on television or online, from a list of Fox News' ten most popular programs (Dilliplane et al. 2013).

Reference

Dilliplane, Susanna, Seth Goldman, and Diana Mutz. 2013. "Televised Exposure to Politics: New Measures for a Fragmented Media Environment." *American Journal of Political Science* 57(1): 236-48.

Appendix G: Pre-Test Analyses

Offensiveness Drives cancellation

We performed an unregistered pre-test with a sample of 662 respondents collected on Amazon's Mechanical Turk platform on July 15, 2021. In this pre-test, we ran conjoint experiments similar to those in the body of the paper and asked respondents (1) how offensive each statement was and (2) what canceling behaviors, if any, they would engage in. The design of this study was nearly identical to that reported in this paper, excepting slight differences in the phrasing of the canceling behaviors.

This pre-test allows us to test our implicit assumption that that perceived offensiveness drives respondents' willingness to cancel statements. We concatenate participants' responses to all four first-order vignettes and estimate participant random effects. To test our assumption, we regress the sum of the likelihoods of engaging in our six canceling behaviors on statement offensiveness, while controlling for the speaker's job or social role, the speaker's race or party, and the particular statement made by the speaker. We measured statement offensiveness on a 101-point semantic differential scale ranging from "Not at all offensive" to "Extremely offensive".

Confirming our expectations, offensive statements were more likely to be canceled ($\beta = 0.05$, $p < .001$). For perspective, this means moving from the bottom to the top of the offensiveness scale is predicted to increase the sum of the likelihoods of engaging in our six canceling behaviors by 4.75—or roughly 1.2 standard deviations—even after controlling for all presented features of the speaker's identity and the particular statement made by the speaker.

Furthermore, using a similar random-effects model, we regress statement offensiveness statement ideology, while controlling for the speaker's job or social role, the speaker's race or party, and the particular statement made by the speaker. This lets us further confirm that statement ideology is determinative of statement offensiveness. Indeed, ideologically agreeable statements are far less offensive ($\beta = -36.53$, $p < .001$).