

The Ubiquity of Social Pressure: How Self-Censorship Constrains Speech Within and Across Parties

Patrick Liu* and Daniel B. Markovits†

October 31, 2026

Abstract

In conversations and on social media, citizens frequently censor their true political beliefs to avoid social backlash, not only from members of the opposing party but also from friends and co-partisans. We argue that self-censorship in disparate social contexts stems from a common source: exaggerated fears about the severity of social repercussions for honest political discussion. Using a nationwide survey experiment during the 2024 U.S. presidential election ($N=17,691$), we find that partisans overestimate the likelihood of social sanctions for expressing their views. Experimentally reducing these exaggerated fears significantly reduces aversion to expressing one's views across diverse social contexts, including inter- and intra-party environments. We use these results to evaluate prominent theories of political conversation, clarifying how asymmetries in partisan factions' aversion to debate can arise even if social pressure is ubiquitous. These results shed light on interpersonal speech and persuasion in an era of intensified partisan and factional conflict.

*PhD Candidate, Columbia University, Department of Political Science

†Postdoctoral Fellow, Northwestern University, Kellogg School of Management

Introduction

In everyday conversations and online exchanges with friends, family, and colleagues, citizens routinely dance around the subject of politics. Often this aversion is dispositional, reflecting plain disinterest or uncertainty of opinion (Carlson and Settle, 2022). Increasingly, however, Americans report avoiding politics for a more troubling reason: to avoid heated arguments or severe social repercussions from expressing their sincere views (Gibson and Sutherland, 2023). Self-censorship resulting from social pressure between ideological opponents is particularly concerning given how it directly curbs citizens’ cross-cutting exposure. Political scientists have long argued that exposure to contrary perspectives produces better-informed opinions and reduces polarization by improving citizens’ understanding of legitimate rationales for opposing viewpoints (Noelle-Neumann, 1974; Fishkin, 1991; Mutz, 2006).

Why, then, do Americans face social pressure for expressing their sincere opinions? Existing research has centered on partisan conflict and group conformity, drawing attention to particular types of interactions that make politics uniquely contentious. One such hypothesis attributes the social costs of speech to inter-party hostilities, linking rising self-censorship to the growth of affective polarization between Republicans and Democrats (Gibson and Sutherland, 2023; Carlson and Settle, 2022, 255). A second hypothesis highlights pressures to conform to majority opinion across a range of settings unrelated to partisanship. This view emphasizes “cancel culture” and the concern that making public remarks on sensitive topics could lead to ostracism for comments deemed offensive (Dias, Druckman and Levendusky, 2024). Studies have also documented how citizens self-censor or falsify their positions on political issues when placed in a room full of opponents, highlighting how the immediate group context shapes conformity (Levitan and Verhulst, 2016; Carlson and Settle, 2016; Braghieri, 2024). A third hypothesis addresses the social pressure faced by voters who dissent from a specific intra-party norm. Voters who hold beliefs at odds with their party’s orthodoxy often recognize their departure from the majority and fear ostracism from co-partisans for speaking their mind, both on policy issues and electoral choice (White and Laird, 2020; Groenendyk, Kimbrough and Pickup, 2023; Spelman et al., 2026).

Yet existing evidence also hints at a more universal dynamic that might underpin citizens’ self-censorship across inter-, intra-, and non-partisan settings: a widely held and exaggerated perception of contentious political issues as socially precarious. Americans are surrounded by and increasingly accustomed to media coverage and social discourse about political correctness, cancel culture, and the severity of political polarization—coverage that appears to warp Americans’ perceptions about polarization and the scope of permissible debate (Chong, Citrin and Levy, 2024; Heseltine et al., 2025; Lee and Ahn, 2025). Dissuaded by the contentiousness of contemporary politics, some citizens shun its mere mention and eschew peers too eager to discuss the subject (Eliasoph, 1998; Krupnikov and Ryan, 2022; Groenendyk et al., 2025). This aversion may be compounded by Americans’ well-documented tendency to overestimate the extent to which opponents hold harmful or hostile attitudes (Ahler and Sood, 2018; Mernyk et al., 2022; Braley et al., 2023). There is evidence in particular that Americans hold exaggerated beliefs about both the opposing party’s willingness to “cancel” and their own party’s propensity to police dissenting speech (Dias, Druckman and Levendusky, 2024; Spelman et al., 2026).

This paper argues that fears of social sanctions are more ubiquitous—and that self-censorship in disparate contexts has more in common—than existing accounts suggest. As we document, Americans systematically overestimate the prevalence of severe social sanctions for political expression, believing that other Americans have suffered in their personal relationships or been penalized at work due to their political beliefs at much higher rates than existing research indicates. These exaggerated fears (which need not be linked to any one partisan group or ideological faction) in turn shape people’s perceptions about the personal risk they face for speaking their mind, along with their discussion behaviors in wide-ranging contexts. We show that correcting misperceptions about social costs in personal and work contexts improves citizens’ willingness to express their sincere views on a contentious issue with opposing partisans they know, co-partisans with whom they agree *and* disagree, and even friends with whom they frequently discuss politics—as well as in group conversations with co- and out-partisans. These findings suggest that self-censorship is in part a product of a generalized fear of social sanctions that pervades numerous interactions and cannot be

principally attributed to any one source of conflict.

Identifying whether this generalized fear constrains political debate with both political allies and opponents in citizens’ existing social networks requires an empirical strategy that differs from existing paradigms in studies of social pressure (e.g., Asch, 1956; Gerber, Green and Larimer, 2008; Tankard and Paluck, 2016). We first show that respondents dramatically overestimate the frequency with which Americans at large face severe social sanctions for political disagreement. Altering these perceptions via an “ask-tell” intervention enables us to shape the reputational concerns brought to mind when respondents consider discussing their 2024 vote choice with co- and out-partisans they know. This design seeks to intervene directly on respondents’ perceptions of the social cost to disagreeable speech, isolating social pressure from other motivations primed by social observation. We then use the effects of this treatment on different kinds of speech to assess asymmetries of self-censorship between inter- and intra-party contexts, as well as between party dissenters and the party faithful, illustrating how marked differences in people’s baseline willingness to discuss politics can belie their uniformity in experiencing social pressure.

This experiment was embedded in a large, nationwide survey ($N = 21,400$) fielded during the summer before the 2024 U.S. presidential election in the immediate aftermath of the presidential primaries. Capitalizing on tensions surrounding the two parties’ presumptive nominees enabled us to investigate citizens’ willingness to speak about a high-salience topic contentious both within and between the parties. Further, we leverage this moment of heightened intra-party division to investigate if generalized fears of social sanctions may suppress majority and minority speech alike. Crucially, this mechanism may serve to counterbalance the effects of partisan norms, which should motivate self-censorship in intra-party discussion contexts only among those in the minority. Defining party minorities as respondents unwilling to support their party’s nominee, we assess if only these party dissenters face social pressure or if even the party faithful fear possible social costs from discussing the election with dissenters. To assess if the ask-tell correction improves willingness to speak, we use “compensation demand” or “willingness-to-accept” survey measures, which elicit the minimum price respondents would need to be paid to complete an assignment (Settle and

Carlson, 2019).

Our argument proceeds in four steps. First, we demonstrate that respondents widely overestimate the risk of severe social sanctions other Americans have faced for expressing their views. This does not mean Americans are wrong to perceive political discussions as uncomfortable or hostile—only that they exaggerate the likelihood of facing grave social or career repercussions. Second, we compare voters with varied partisan and factional allegiances in terms of their willingness to speak freely about the 2024 election and show that the patterns cohere with theories of both inter-party conflict and intra-party norms: most respondents were less willing to express their views to opposing partisans than to co-partisans, yet party minorities displayed precisely the opposite preference, favoring conversations with opposing partisans instead. Third, we demonstrate that correcting social pressure misperceptions reduced respondents’ own perceived risk of social costs within their network. Finally, we leverage this treatment to unpack how generalized social sanction fears shaped Americans’ willingness to discuss the 2024 election. In contrast to the differential patterns in willingness to speak, social pressure is ubiquitous: our corrections reduced partisans’ compensation demands for discussing the election with both co- and out-partisans in one-on-one and group settings alike. Moreover, both party minorities and majorities became more open to conversations with co-partisan opponents following the intervention.

In the remainder of the paper, we discuss additional experimental results that highlight how social media magnifies fears of severe social sanctions. While Americans vary meaningfully in the amount of personal risk they perceive for expressing political views within their social network, they widely agree that posting their opinions to social media would engender far more severe risks for their career or relationships. Here too we observe striking similarities across groups, with social media context increasing perceived risks irrespective of whether the respondent believed they held the minority or majority opinion within their social network. Further, we leverage an additional pre-election survey to help reconcile the asymmetry in party minorities and majorities’ willingness to speak with our finding that both groups self-censor due to social pressure. Compared to dissenters, party majorities were equally likely to cite possible damage to relationships as a reason to avoid discussion,

yet considerably more likely to see persuading their peers to change their mind as a possible benefit. Thus, *perceived instrumental benefits* can help give rise to asymmetries in speech between majorities and minorities even when both face social pressure.

As we elaborate in subsequent sections, these findings add important nuance to theories of political conversation, partisanship, and conflict in American democracy. We begin by clarifying how social pressure differs from other considerations shaping openness to political speech and by defining hypotheses about where social pressure is concentrated. We highlight these perspectives’ unique emphases on specific types of interactions as the primary loci of social pressure. The subsequent sections elaborate methods and results before discussing implications for self-censorship and political conversation across and within party lines.

Defining the Considerations Shaping Political Speech

From Nixon’s appeals to a “Silent Majority” (Gillion, 2020, chapter 1) to contemporary debates over “cancel culture” (Dias, Druckman and Levendusky, 2024), the idea that political discussion may be stifled by citizens’ intolerance is a recurring feature of American politics (Sullivan, Piereson and Marcus, 1979; Chong, Citrin and Levy, 2024). We describe the stifling of speech that results from fear of this intolerance as “social pressure”: the psychological influence on behavior that flows from “a basic human drive to win praise and avoid chastisement” (Green and Gerber, 2010, 331).

Crucially, decisions to self-censor or avoid politics are shaped not by social pressure alone but by a range of psychological inputs. Scholars of social influence have emphasized that social pressure bears principally on “affiliative” motivations related to cultivating and preserving key social ties (Cialdini and Goldstein, 2004; Carlson and Settle, 2022). Yet political speech is concurrently shaped by “affirmation” and “accuracy” considerations such as a fear of appearing uninformed or the desire to persuade. Social pressure is distinct from these considerations in its troubling normative implications and its external enforcement.

In intra-party contexts, for instance, dissidents may exhibit a greater aversion to political debate not only because they are socially pressured to abide by injunctive partisan norms

but also because they are less politically active, cross-pressured voters with bipartisan social networks (Mutz, 2002), whose social milieus furnish fewer motivations to acquire and share information (Connors, Pietryka and Ryan, 2022). A weak commitment to the party orthodoxy could also correlate with social and personality traits that deter political engagement (Gerber et al., 2012, 2013). Thus, even if we observe that voters who conform to partisan norms (hereafter, the party majority) are less prone to self-censorship than dissenters (party minority), party majorities may nonetheless perceive social costs to expressing their beliefs.

In short, understanding whether generalized fears of social sanction are truly ubiquitous—that is, felt by both party majorities and minorities, not only in group but in one-on-one conversations, and with co-partisans and out-partisans alike—requires an empirical strategy that does not rely solely on relative differences in these groups’ aversion to political debate. In the coming section, we review research investigating the myriad sources of partisans’ aversion to political debate, highlighting how our hypothesis takes a more holistic view of social pressure that does not hinge upon any one type of social dynamic.

Theorizing the Sources of Social Pressure

Inter-Party Conflict

In the summative chapter of their book characterizing the complex interplay of motivations shaping political conversation today, Carlson and Settle (2022) observe that political talk may be undergoing “a shift over time that coincides with increasing levels of polarization, though again the evidence to test this idea is largely circumstantial” (255). They note that while exploratory work suggests people may be discussing politics more than in the past, there is little evidence that people desire more political conversation, meaning these interactions could be increasingly undesired and contentious. Indeed, social networks have grown increasingly homophilous, disagreeable discussion is more frequently avoided, and rates of reported self-censorship are increasing with time (Butters and Hare, 2022). Gibson and Sutherland (2023) show that these trends coincide with the growth of inter-party ideological sorting and affective polarization. Partisans may be especially unwilling to speak to oppos-

ing partisans because, all else equal, conversations with opposing partisans are likely to be uniquely hostile and to bear heightened social costs.

The existing evidence for this hypothesis, however, is indeed largely circumstantial. Despite extensive research documenting Republicans and Democrats’ personal dislike of one another and the rarity of cross-partisan social interactions (Iyengar and Westwood, 2015; Klar, Krupnikov and Ryan, 2018; Brown and Enos, 2021), very little of the literature on affective polarization documents if partisans specifically are deterred from expressing their personal beliefs in their one-on-one engagements with partisan opponents. Clearly, voters consistently and explicitly favor discussing politics with co-partisans over out-partisans (Settle and Carlson, 2019; Carlson and Settle, 2023). Yet preference does not amount inherently to self-censorship in practice, and choices of discussion partners often appear better explained by incidental than purposive factors (Minozzi et al., 2020).

Further, scholars continue to disagree over whether hostile behaviors and attitudes between the parties reflect an affective disdain fueled by toxic stereotypes and misperceptions about the other side (Iyengar and Westwood, 2015; Mason, 2018) or simply liberals and conservatives’ ideological disagreements (Huckfeldt, Mendez and Osborn, 2004; Orr, Fowler and Huber, 2023). To the extent that Republicans and Democrats avoid political discussion with one another, the root cause may have little to do with partisan identity, attributable instead to a broad avoidance of ideological debate that scholars have long observed even prior to contemporary partisan sorting (Mutz, 2006; Gerber et al., 2012). In this paper, we investigate this broader conflict avoidance account by assessing if people perceive social costs to discussing contentious issues with opponents from all ends of the ideological spectrum, including both out-partisans and opponents within one’s own party.

Conformity in Non-Partisan Environments

While the partisan conflict hypothesis suggests that conflict between the parties creates a special risk of social pressure, another class of theories implies that majority influence constrains speech on individual issues that may be orthogonal to partisanship. The idea that perceiving oneself to be in the minority constrains speech on a given issue features

prominently in scholarship on public deliberation, particularly in Noelle-Neumann’s (1974) “spiral of silence” theory which posited that minority opinion-holders withhold their true preferences in public settings for fear of isolation.

As experimental studies have repeatedly demonstrated, people—often undergraduates in laboratory settings—conform to the expressed political opinions of the majority in their *immediate* context. Experiments modeled after Asch (1956) illustrate how group lab discussions shape students’ political expression (Carlson and Settle, 2016, 2022; Levitan and Verhulst, 2016), while other designs have shown that the possible public revelation of their beliefs shapes expression for opinion minorities far more than students in the majority (Ho and Huang, 2024; Braghieri, 2024). “Norm change interventions” outside the lab have also found that people conform when their perceptions of majority behavior change (Tankard and Paluck, 2016).

These designs may struggle to isolate social pressure as the key mechanism, since social learning is a competing factor: subjects could infer that a behavior must be somehow sensible or efficient because it is popular with others (Cialdini, Kallgren and Reno, 1991, 203). In addition, people’s tendency to conform to the expressed political views in their immediate surroundings does not guarantee their conformity to the majority position in other conversational settings, such as one-on-one conversations with an opponent in the majority where mechanisms for norm enforcement may be less apparent (Glynn, Hayes and Shannahan, 1997). Since the goal in this paper is to assess if a generalized fear of social sanctions deters speech both in one-on-one and group conversations, our empirical strategy will seek to manipulate social pressure without relying on group discussion interventions or descriptive norm changes.

Crucially, some minority beliefs—such as support for radical right parties in Europe (Valentim, 2024; Alvarez-Benjumea and Valentim, 2024) or comments that people across the aisle deem potentially harmful (Dias, Druckman and Levendusky, 2024)—may be constrained by social pressure even in one-on-one settings because they are uniquely stigmatized. One setting in which scholars are increasingly concerned about this dynamic taking shape is within parties. Thus, in the next section, we take up the question of whether intra-party

networks are shaped by injunctive norms that stigmatize dissenting speech.

The Partisan Norms Hypothesis

An emerging literature within political science has sought to apply the logic of conformity to political parties, treating parties as salient social groups that impose a distinctive set of social pressures on dissenters. The theoretical building blocks of this literature are theories that frame partisanship as an enduring social identity (Campbell et al., 1960; Green, Palmquist and Schickler, 2002). Some scholars have increasingly emphasized the importance of partisanship to one’s self-concept, arguing that this internalization shapes how partisans behave socially, making them hostile toward opponents on social media and resistant to information or arguments that threaten their partisan identity (Mason, 2018; Van Bavel and Pereira, 2018). These ardent partisans may in turn frown upon co-partisans who vocalize dissent or express opinions that resemble those held by the other party (Suhay, 2015; Groenendyk, Kimbrough and Pickup, 2023; Spelman et al., 2026). With policy and ideological disagreements growing among intra-party factions (Bloch Rubin, 2017; Henderson et al., 2022), so too are animosities between the voters who affiliate with these factions (Young and de Wit, 2025).

However, other perspectives call into question the power of partisan identity alone to enforce social norms against dissenting speech, without the aid of external enforcement mechanisms. As White and Laird (2020) show, partisan social norms are so strong among Black voters that dissenters are unwilling to share dissenting beliefs with other Black Democrats. Yet the same authors do not extend this finding to other Democrats, perhaps due to the decline of rooted party organizations, to which the Black church remains an exception (Schlozman and Rosenfeld, 2019, 2024). Moreover, partisans can dissent in a given election without undermining their partisan self-identity (Green, Palmquist and Schickler, 2002), suggesting that partisan social identity alone may generate little psychological pressure to conform.

Despite these caveats, a growing empirical literature has sought to document how partisan ties may deter dissent from party orthodoxy. Priming descriptive norms about co-partisans’ issue beliefs and political values consistently increases adherence to party or-

thodoxy (Groenendyk, Kimbrough and Pickup, 2023; Connors, 2020). Gerber, Huber and Washington (2010) show that encouraging voters to formally register with a party increases stereotypical partisan attitudes and improves evaluations of party leaders. Finally, in the paper most closely related to our own, Spelman et al. (2026) find that party minorities overestimate the social costs of deviating from majority beliefs, an effect that diminishes when partisans are reminded of other issues on which they hold party-conforming views.

Though this paper does not test directly whether injunctive partisan norms generate social pressures that uniquely affect intra-party minorities, we do assess if party minorities and majorities are both deterred by generalized fears of social sanction from discussing a contentious issue involving a clearly defined majority-minority dynamic (support for the party’s presidential nominee). If, as we hypothesize, these concerns constrain speech from both partisan factions, it suggests that the contentious nature of contemporary political conversation could partially weaken the asymmetric effects of injunctive partisan norms, creating reasons to self-censor even among those who conform to party orthodoxy.

The Ubiquity of Social Pressure

Despite its dominant focus on partisan conflict and group conformity, the existing political conversations literature also offers hints that social pressure may be more ubiquitous than these accounts reflect. Being forced to debate conflict-ridden issues is thoroughly unpleasant for many people, particularly the majority of Americans whose daily lives are not deeply saturated with politics to begin with (Berelson, Lazarsfeld and McPhee, 1954; Prior, 2005; Guess, 2021). Studies in political communication have long observed that citizens across the aisle avoid political discussion with likely opponents (Huckfeldt, Mendez and Osborn, 2004; Gerber et al., 2012), an aversion that appears to stem in no small part from a widespread association of the mere *idea* of politics with discomfort. Conflict-averse individuals may be willing to discuss substantive policy questions but become reluctant once those same questions are explicitly framed as political (Groenendyk et al., 2025; Spelman et al., 2026). Eliasoph (1998) similarly argues that political avoidance represents an effort to maintain the

boundaries of appropriate social manners or “etiquette” (21). Political discussion appears especially fraught to the less politically involved, who may perceive highly invested partisans as eager to engage in bad faith debate (Krupnikov and Ryan, 2022).

Crucially, these perceptions of political conversation as socially precarious need not depend on the popularity of one’s beliefs, making them likely to span the partisan divide and unlikely to be held exclusively by either party majorities or party minorities. These perceptions may very well have become more exaggerated in recent years, helping to explain rising rates of reported self-censorship over time (Gibson and Sutherland, 2023): Americans have been bombarded with news coverage and social discourse about societal polarization as well as cancel culture (Robison and Mullinix, 2016; Ploger, 2024; Dias, Druckman and Levendusky, 2024), and there is substantial evidence that this coverage warps perceptions of the true extent of societal divisions (Levendusky and Malhotra, 2016*a,b*; Heseltine et al., 2025). Recent work further suggests that increasing citizens’ perceptions of societal polarization causally increases their likelihood of self-censorship (Lee and Ahn, 2025).

What remains unclear is whether Americans specifically hold exaggerated misperceptions about the severity of social costs they are likely to bear for expressing their sincere opinions with intra- and inter-party opponents. But there is substantial reason to expect these inferences to be overly pessimistic. Americans systematically exaggerate the homogeneity, anti-democratic values, and partisan hostility of both opposing partisans and co-partisans (Ahler and Sood, 2018; Mernyk et al., 2022; Braley et al., 2023; Moore-Berg and Hameiri, 2024; Dias, Lelkes and Pearl, 2025). Even more pertinent is the evidence that Americans overestimate the hostility associated with dissent from their own party (Spelman et al., 2026) and the willingness of political opponents to “cancel” (Dias, Druckman and Levendusky, 2024). Our hypothesis, however, is that Americans hold these misperceptions about political discussion not just in specific social environments but in general, and that these generalized fears of social pressure motivate self-censorship in wide-ranging contexts.

Empirical Strategy

The theoretical accounts we described offer distinct though complementary predictions about aversion to speech: the inter-party conflict view emphasizes voters’ greater receptivity to political conversations with co-partisans than with out-partisans, while the partisan norms view suggests that this is evident only for party majorities. Party minorities should be particularly averse to political conversations with co-partisans in the majority, and may be equally or *more* willing to discuss politics with opposing partisans. Our hypothesis about the ubiquity of social pressure does not diverge from these predictions. Instead, our argument is that these asymmetries in Americans’ willingness to speak may belie a surprising unanimity in their generalized fears of social sanctions. Our empirical strategy therefore seeks to disentangle *willingness to speak* from *social pressure*, the latter being but one of many considerations that collectively determine the former.

To measure willingness to speak, we adopt the “compensation demand” approach of Settle and Carlson (2019), asking respondents the price they would need to be paid to express their real opinions among peers.¹ For each respondent, we elicit their compensation demands (hereafter, CDs) for one-on-one conversations with individuals of varying identities—including intra-party allies, intra-party opponents, and out-partisans—as well as for group conversations.

We operationalize social pressure as the extent to which this measure of willingness to speak *responds* to an exogenous change in the perceived risk of social sanction for speech, treating it as an essentially causal quantity: the amount by which constraints on speech are alleviated when generalized social sanction risks are reduced. Specifically, our intervention offers information about the true frequency with which other Americans have faced severe damage to their career or relationships. Examining whether this intervention subsequently reduces CD prices for hypothetical political conversations with discussants of various identities enables us to measure whether social pressure is prevalent in each given interaction.

¹This approach somewhat resembles “willingness-to-accept” (WTA) measures developed in contingent valuation studies (Venkatachalam, 2004). While some scholars are concerned that such measures are poor barometers of real world behavior (Hanemann, 1991; Shogren et al., 1994), we are interested only in assessing whether different groups perceive differential costs, and not in an accurate price estimate for its own sake.

To assess our hypothesized mechanism—that this information alters willingness to speak by reducing the perceived risk of sanctions within one’s *own* network—we utilize a series of manipulation checks along with tests of alternative mechanisms (see Appendix A.6).

Methods

We embedded an experiment in a large, nationwide survey ($N = 21,400$) fielded through the Civic Health and Institutions Project (CHIP50). The data were part of the broader wave 32 collection and were collected between June 18th and July 28th, 2024 by PureSpectrum, an online panel management platform. Following our pre-registration, we subset our main analyses to the 17,691 partisans, including self-identified partisans and Democrat- or Republican-leaning independents.²

The survey was fielded at a crucial moment of heightened inter- and intra-party tensions. In the final months of the 2024 presidential primaries, just as the general election was kicking off, influential factions—including Never-Trump Republicans and Democrats opposed to the Biden administration’s handling of Israel–Palestine—continued to speak publicly against their parties’ presumptive nominees. We define respondents who said they would vote for their party’s respective nominee (Biden for Democrats, Trump for Republicans) if the “election were held today” as part of the *party majority*, while respondents who gave any other response (including “Not sure” or “I would not vote”) are categorized as the *party minority*.³ Importantly, in an era of increased partisan loyalty and issue sorting (Levendusky, 2009; Smidt, 2017) partisan minorities constitute small shares of their parties (21.25% in our final data), suggesting that a sample of CHIP50’s size is crucial for testing whether supporters of and defectors from the party leaders alike experience pressures to withhold their true preferences from peers.⁴

²The remaining 3,685 “pure” independents were assigned to the survey wording for Democrats if they self-identified as ideologically liberal, assigned to the survey wording for Republicans if they self-identified as conservative, or randomized to one of the two conditions otherwise. These pure independents are dropped from the main analysis.

³In Appendix A.4, we show that our results are robust to alternative definitions of party minorities.

⁴All but 27 responses (20 from partisans) were recorded before Biden announced his withdrawal from the race via his personal X account at 1:46pm ET on July 21: <https://x.com/JoeBiden/status/1815080881981190320>.

Experimental Design

Our survey asked respondents to consider their willingness to share their political beliefs across a range of scenarios with randomly assigned attributes. The survey began with standard demographic questions, and our module asked respondents to estimate the shares of their personal networks that are co-partisans, out-partisans, and co-partisans who “are hesitant” to support their party’s nominee before asking their beliefs about pressures to avoid sharing their political beliefs. In order to familiarize respondents with the compensation demand outcomes later in the survey, we first presented a pre-treatment CD question based on a more standard assignment in an online survey context. Respondents were asked for their CD price to “write a few paragraphs explaining why” they would or would not vote for their party’s nominee.⁵

From here, respondents were asked to estimate the percentages of Americans who (i) “say they have been penalized at work for something they said on social media,” (ii) “say they have recently had political disagreements with family or friends that hurt their relationship,” and (iii) “say a person’s political views tell you a lot about whether they are a good or bad person.” These statements are based on real survey questions asked of representative samples of Americans in 2022, with the true percentages being respectively 3, 19, and 14 percent.⁶ A clear majority of our respondents overestimated the correct percentage for each, indicating widespread misperceptions about the prevalence of social sanctions for disagreeable speech in everyday life. We refer to these beliefs as respondents’ “priors” regarding the likelihood of social sanction.

Ask-Tell Treatment (τ_1) Our first randomized treatment corrected these misperceptions in the form of an “ask-tell” intervention (e.g., Mernyk et al., 2022), so-called because after all

⁵The aim of the baseline CD is both to reduce the likelihood that our later CD outcomes present hypothetical assignments that respondents have difficulty imagining and to mitigate concerns about respondents’ unfamiliarity with the question format. To improve precision, we include this baseline CD price in our pre-registered vector of covariates. Despite being a source of bias in contingent valuation studies, anchoring is not a cause for concern here as we do not interpret the CD prices as inherently meaningful.

⁶The wording and true percentage for the statements come respectively from a 2022 Knight Foundation report: <https://perma.cc/HN2A-SDRK> and a Times/Siena Poll: <https://www.nytimes.com/interactive/2022/10/18/upshot/times-siena-poll-registered-voters-crosstabs.html>.

respondents were “asked” for their priors, respondents were randomized with 0.5 probability to be “told” the true percentages. In the tell condition, respondents were shown a table that contained their own answers, the true percentages based on real data, the average percent across their three guesses, and the true average (12 percent). Finally, they read a summary statement: “Overall, Americans rarely punish others for their speech choices, and very few have ever lost a job or even a friend over their political opinions.” In the control condition, respondents were shown a table only reminding them of their own estimates and average guess. This design builds on one often employed in studies of pluralistic ignorance and meta-perceptions (e.g., Bursztyn and Yang, 2022; Braley et al., 2023).

Target Randomization (τ_2) We presented three CD outcomes following the ask-tell. The CHIP50 core survey items elicited respondents’ 2024 vote intention. Accounting for this answer, we asked respondents to consider an assignment in which they would make a one-minute video explaining why they did or did not support their party’s 2024 presidential nominee. They were then asked to think of three people they know (hereafter, “targets”) and to write down their initials. The first target listed, which serves as a benchmark, was always “A person with whom you frequently discuss politics.” The second and third were randomly selected from three possible targets: a co-partisan “opponent” who, unlike the respondent, did (did not) intend to vote for the party leader; a co-partisan “ally” who, like the respondent, did not (did) intend to vote for the party leader; or an out-partisan. Each of these three possible targets was mentioned to two-thirds of the sample. This design allows for both between- and within-subject comparisons. We then measured these items:

- *One-on-One Compensation Demands:* Respondents stated the lowest payments they would accept to send their one-minute video to a frequent discussant and two randomly selected targets. CDs were measured on a left-right integer scale (\$0–\$300 USD).
- *Mechanism Check:* To assess if the ask-tell affects the CDs by altering beliefs about the likelihood of social sanction, we asked respondents to estimate what percentages of Americans and of people they know agree with the statement, “When people I know express political views with which I disagree, I lose some respect for them.”

Our main hypotheses center around the treatments, outcomes, and mechanism checks described above. An anonymous version of the pre-analysis plan is provided in supplementary materials. Consistent with theories of inter-party conflict and intra-party norms, we hypothesized that CDs on average would be higher for co-partisan opponents than for frequent discussants (H1A) or co-partisan allies (H1B) but lower for co-partisan opponents than for out-partisans (H1C); party minorities, however, would report higher CDs for co-partisan opponents than would party majorities (H5A). We then predicted that the ask-tell correction would move our mechanism check outcomes by reducing the estimated shares of Americans and peers who would judge political opponents (H2A) and by diminishing the perceived likelihood of social sanctions (H2B).

To assess the *ubiquity* of generalized social pressure concerns, we hypothesized that the ask-tell treatment would reduce all one-on-one and group CDs: reducing generalized social sanction fears should improve willingness to discuss politics not just with out-partisans but with co-partisan allies and opponents (H3A,C–D). We also assess if the correction reduces co-partisan conversation CDs not only among party minorities but party *majorities*, as well.

The remainder of the experiment was designed to assess the mechanisms and scope conditions governing our social pressure hypothesis. First, if social pressure is in fact rooted in reasoned judgments about generalized risks of social sanction, then the ask-tell effect should be larger for respondents with more pessimistic priors. Second, we assess if our findings regarding the effects of generalized fears of social sanction extend to social media environments and group conversation settings. For concision, our remaining hypotheses are described in the pre-registration. The experiment proceeded as follows:

Social Media Prime (τ_3) Respondents were asked to reflect on the concerns that came to mind as they contemplated sharing a one-minute video about the party nominee with peers. We randomly assigned two question wordings with equal probability. In the control, respondents considered these questions in the context of the individual CD targets previously described. Respondents assigned to the prime instead considered sharing the video to a social media platform of their choice, “where many people you know could see it.”

- *Sanction Likelihood*: On a 7-point Likert scale, respondents stated the perceived likelihood of: “suffering consequences for your career,” “damaging relationships with people you care about,” and “experiencing hostility from strangers or people you barely know.”

Group Conversation Randomization (τ_4) The final grid in our module measured respondents’ willingness to have five-minute discussions with a group of Republicans and a group of Democrats. We randomized question wording by randomizing with equal probability whether the topic would be the 2024 presidential election, “economic issues like taxation and social security,” or “social issues like abortion and immigration.”

- *Group Conversation CDs*: We measured respondents’ compensation demand for each of the two hypothetical discussion assignments on \$0 to \$300 scales.

Estimation

Model specifications for each outcome are described in our pre-analysis plan. Here, we outline in broad strokes our two approaches to estimation. All models include an error term ϵ and control for a pre-registered vector Φ_i containing party affiliation, the average of respondents’ three “ask-tell” estimates, two measures of self-reported network diversity, baseline CD price, and the “core items” shared across CHIP50 modules. All models employ robust standard errors at the individual level, which serves to cluster standard errors for the estimates that rely on multiple observations per individual. For some iterations of these models, other prior treatment conditions are included in the set of control variables.

Model 1 applies to respondent-level outcomes. The treatment τ_1 refers alternatively to the ask-tell correction, the social media prime, or the content of a group discussion.

$$\text{Outcome}_i = \beta_0 + \beta_1\tau_1 + \eta\Phi_i + \epsilon_i \quad (1)$$

Model 2 accounts for the within-subjects nature of the one-on-one CD outcomes, measured at the respondent-target level. The co-partisan opponent, co-partisan ally, and out-partisan are represented by dummy variables CD1, CD2, and CD3 respectively, while the

frequent discussant CD serves as a baseline. Respondents are indexed by i and compensation demand observations by c .

$$\text{Outcome}_{ic} = \beta_0 + \beta_1 \text{CD1}_c + \beta_2 \text{CD2}_c + \beta_3 \text{CD3}_c + \eta \Phi_i + \epsilon_i \quad (2)$$

Results

Willingness to Speak

In Figure 1, we illustrate stark asymmetries in different partisan groups' willingness to discuss the election in one-on-one settings. The results partially affirm Hypothesis 1: party minorities demanded higher prices for the co-partisan opponent they know than for any other target, including an out-partisan they know, though party majorities show a different pattern. More strikingly, Figure 1 confirms Hypothesis 5a. Not only did party minorities report higher demands for co-partisan opponents than did party majorities, but whereas party minorities demanded the highest CD for this target, party majorities demanded the *lowest* compensation for co-partisan opponents, even lower than people with whom they frequently discuss politics. The asymmetry between minorities and majorities in willingness to discuss the election is specific to co-partisan opponents and frequent discussants, not extending to co-partisan allies or out-partisans.

CDs for the group assignment exhibited asymmetry for co- and out-partisans alike. Figure 2 indicates that relative to majorities, party minorities demanded higher prices for the co-partisan conversation and lower prices for the out-partisan conversation. Unlike in the one-on-one CD results, minorities did not prefer speech with the opposing party to their co-partisans. A plausible explanation is that these survey items did not specify the distribution of preferences, meaning we left to respondents' imagination what proportion of the group would share their beliefs. Nonetheless, party minorities exhibited greater reluctance toward intra-party speech than party majorities in both one-on-one and group environments.

In Appendix A.9, we validate these patterns using a simple set of descriptive pre-treatment questions. Consistent with concerns about rising intra-party tensions, 41.5% of

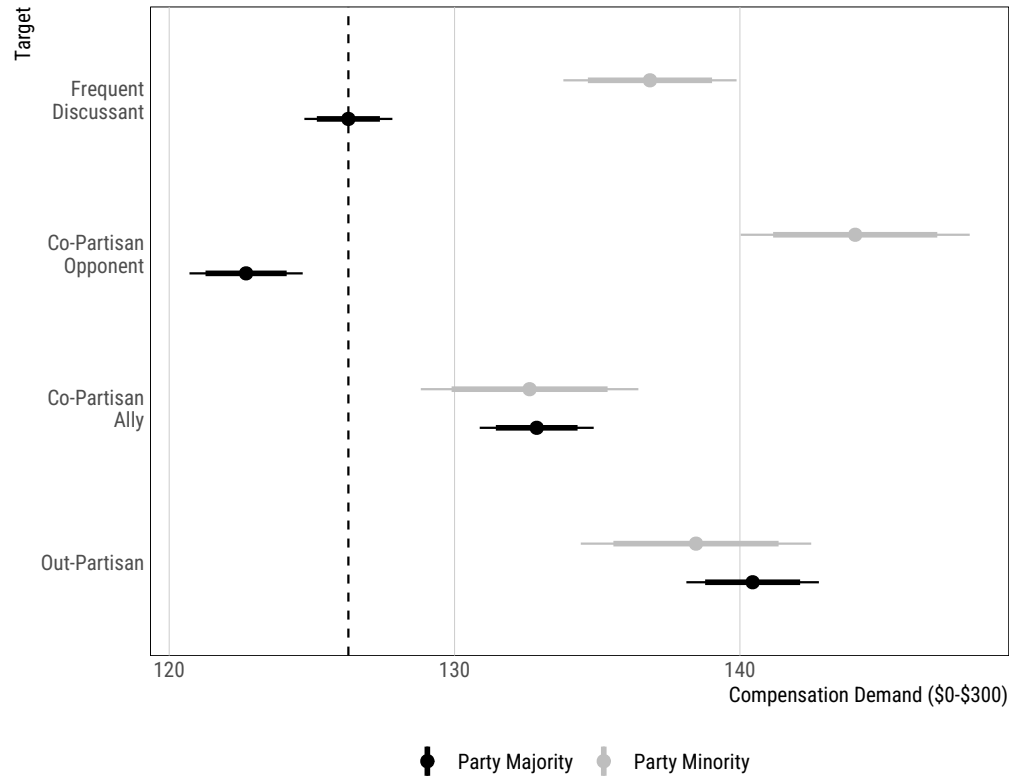


Figure 1: One-on-One CDs by Party Status, Full Model in B6.

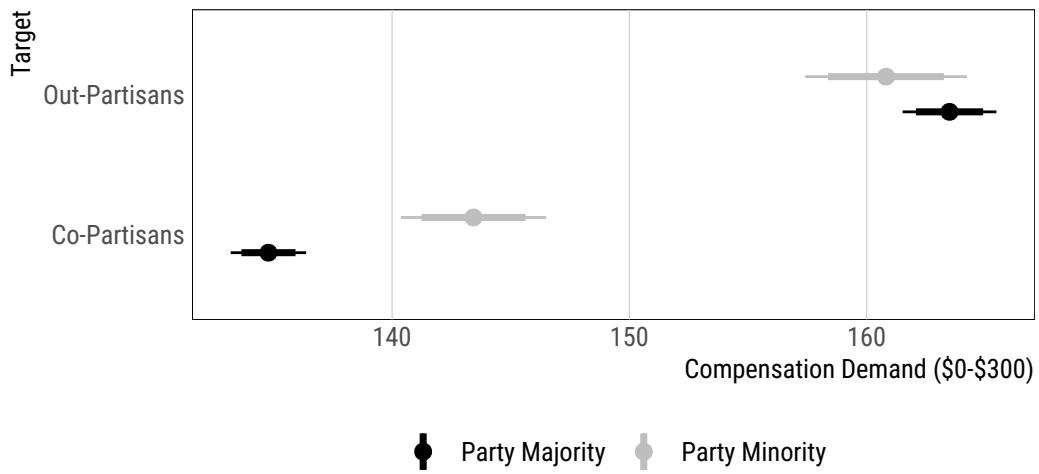


Figure 2: Group Conversation CDs by Party Status, Full Models in Table B7.

respondents in total agreed that they avoid discussing politics with some co-partisans they know. Moreover, 54.5% of party minorities agreed compared to only 37.9% of party majorities, affirming that party minorities are substantially more averse to intra-party debate.

On the flip side, 63.9% of party minorities said they avoid politics with some out-partisans compared to 72.1% of party majorities, further corroborating the patterns in Figures 1 and 2. We test these differences formally in Appendix A.3.

Perceptions of Social Sanctions

To determine whether these differential patterns in willingness to speak belie more ubiquitous fears of social sanction, we turn to the ask-tell correction, which was designed to reduce fears of severe sanctions for speech while leaving other motivations unaffected. First, in Figure 3, we plot response distributions for all three survey items as well as their mean. The dotted line in each faceted histogram corresponds to the true answer. The vast majority of our respondents overestimate the true percentage for each question, indicating widespread and consistently pessimistic misperceptions about the likelihood of social sanction for political speech in the US; the average answer was 46% compared to a true value of 12%. Only 5.9% of respondents underestimated the true average level of social risk.

Perceptions about social sanction are broadly uniform between party minorities and majorities (see Appendix A.3), suggesting that our later tests of heterogeneous effects do not bundle treatments of different intensities. Priors were also uniform across lines of race and education, though they vary somewhat with gender, political interest, partisan identity, and respondents’ stated propensity to self-censor with co- and out-partisans. These predictors explain only a small share of variation in priors: Americans almost ubiquitously overestimate the extent to which their fellow citizens may retaliate against political speech.⁷

Crucially, these misperceptions are about the prevalence of severe political division at a *societal* level. We cannot know the true extent of sanctioning within respondents’ own networks and thus cannot establish if Americans over- or underestimate their personal risk of social sanctions. This being said, respondents themselves may have only imprecise guesses about the true likelihood of sanctioning within their networks. We test if (a) correcting mis-

⁷Dias et al. (2024) offer a rejoinder to past work on misperceptions about out-partisans’ democratic commitments, arguing that these misperceptions are unstable over time and thus unlikely to be deeply rooted beliefs that meaningfully shape people’s own attitudes or behavior. We believe the misperceptions measured in this paper are more likely to be meaningful given their closer connection to day-to-day concerns, as discussed further in Appendix A.5.

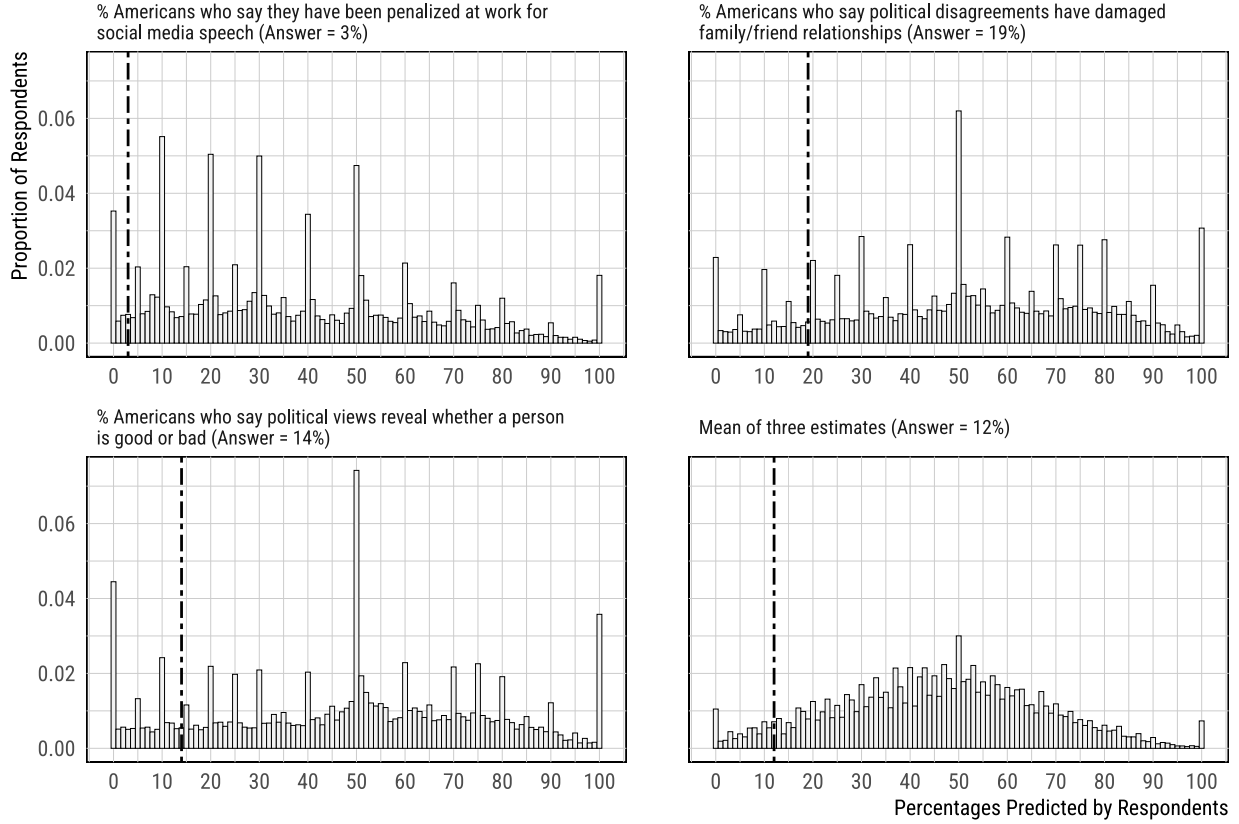


Figure 3: Respondents’ “Ask-Tell” Prior Estimates

perceptions about social sanctions across American society (b) shifts perceptions about social pressure from one’s own co- and out-partisan peers, which in turn (c) improves willingness to speak.

Ask-Tell Intervention

Do these misperceptions help explain why Americans widely avoid political discussion? Table 1 (column 1) and Figure 4 (top row) confirm that the ask-tell treatment substantially reduced respondents’ estimates of the percentage of “All Americans” who would say they lose respect for people who express political views with which they disagree (-11.46 percentage points, SE = 0.38)—full models are presented in Appendix Table B1. Consistent with our intuition that people hold less diffuse but nonetheless malleable priors about familiars, exposure to the ask-tell also substantially reduced the estimated percentage for “People you know” (-7.6 pp, SE = 0.40). Though the magnitude of this effect is smaller, this manipulation check is crucial,

confirming that our treatment successfully alters perceptions not only about Americans writ large but about social sanctions within respondents’ own networks. With regard to the risks of specific sanctions for the one-on-one CD task, ask-tell exposure significantly reduced the perceived likelihood of all three outcomes as well as their mean (-0.34 , $SE = 0.02$).⁸

Though scholars have argued that social pressure is best understood as not merely a gut response but the product of a concrete if subconscious cost–benefit assessment (Cialdini and Goldstein, 2004), this intuition has yet to be empirically demonstrated to our knowledge. We offer compelling evidence that people indeed reason about the risks of social sanction in a concrete and considered manner. In columns 2 and 4 in Table 1, the negatively signed interaction effect between ask-tell assignment and ask-tell prior indicates that the ask-tell ATE substantially increased in magnitude for larger priors—that is, for respondents who more drastically overestimated the likelihood of social sanctions for political speech.

Table 1: ATE on Mechanism Check Results with Interactions by Prior (Continuous)

	“All Americans” Manip Check		“People you know” M.C.		Sanction Mean
	(1) H2a	(2) H2c	(3) H2a	(4) H2c	(5) H2b
Ask-Tell	-11.455^{***} (0.383)	-2.368^{**} (0.912)	-7.643^{***} (0.399)	-1.490 (0.909)	-0.337^{***} (0.021)
Prior	0.436^{***} (0.011)	0.536^{***} (0.014)	0.454^{***} (0.011)	0.522^{***} (0.015)	0.015^{***} (0.001)
Ask-Tell:Prior		-0.197^{***} (0.019)		-0.134^{***} (0.020)	
Num.Obs.	16 650	16 650	16 477	16 477	16 901
R2	0.210	0.215	0.196	0.199	0.119

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Full Models in Table B1.

Turning to the main effects of our misperception correction, Figure 5 shows that ask-tell exposure reduced compensation demands for discussing the election with co-partisan opponents and opposing partisans, evincing the prevalence of social pressure both in intra-party networks and between partisan opponents (full models in B9). Importantly, the ask-tell correction substantially reduced reluctance toward both one-on-one and group conversations. It must be noted, however, that ask-tell exposure reduced compensation demands not just for

⁸In Appendix A.6, we leverage additional outcomes measured after our module in the CHIP50 Wave 32 survey to assess alternative mechanisms.

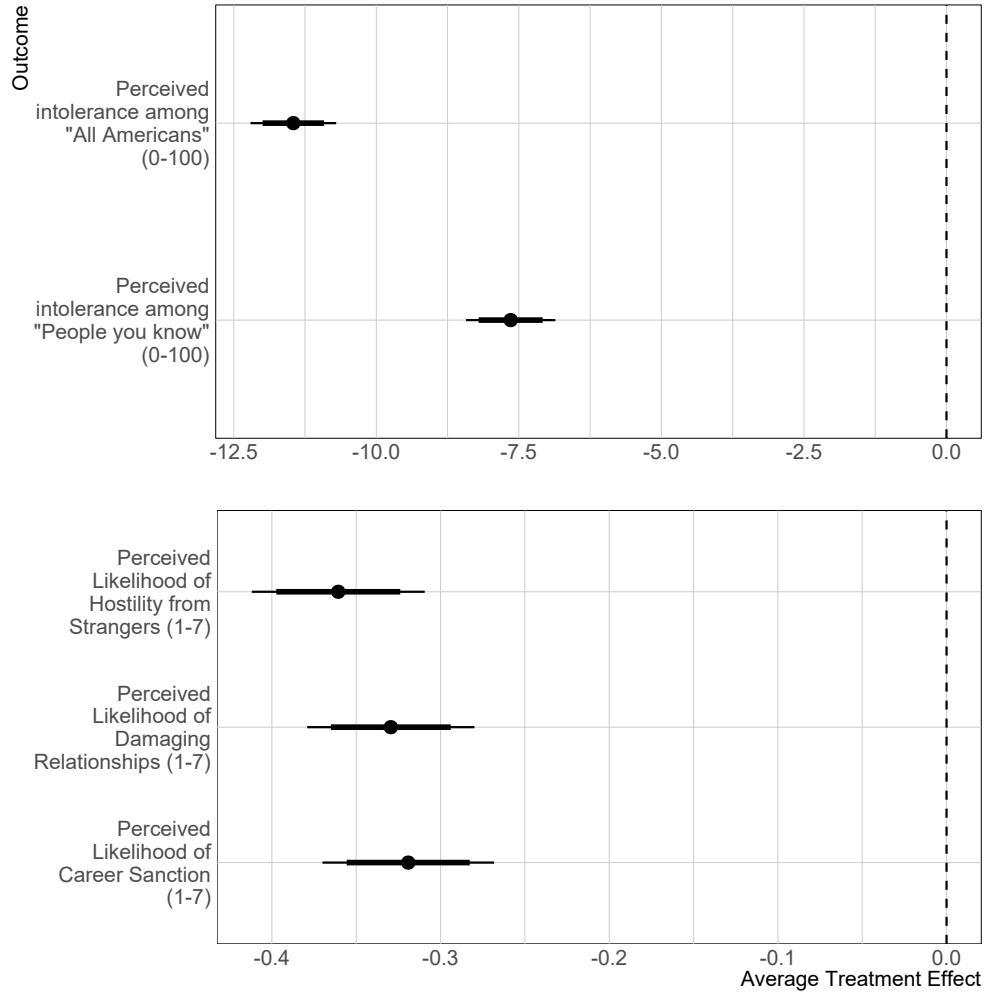


Figure 4: Ask-Tell ATEs on Mechanism Checks (Separate Regressions), Full Models in Table B8.

co-partisan opponent and out-partisan CDs but across the board. Especially striking is that the ask-tell reduced CD prices for the frequent discussant and co-partisan ally, among whom we might expect ideological or partisan differences to be smaller or at least less salient. These results accord with an understanding of social pressure as deriving in part from a general exaggeration of social risk, such that corrections affect all categories of conversation.

Consistent with this interpretation, we find that the ask-tell effects on co-partisan conversation CDs are not driven solely by party minorities. Interacting the treatment indicator with another for party minority status, the main effects listed in Table 2 illustrate that even for party majorities, our correction reduced compensation demands for discussing the election with co-partisan opponents one-on-one ($-\$3.79$, $SE = 1.89$) and with a group of co-

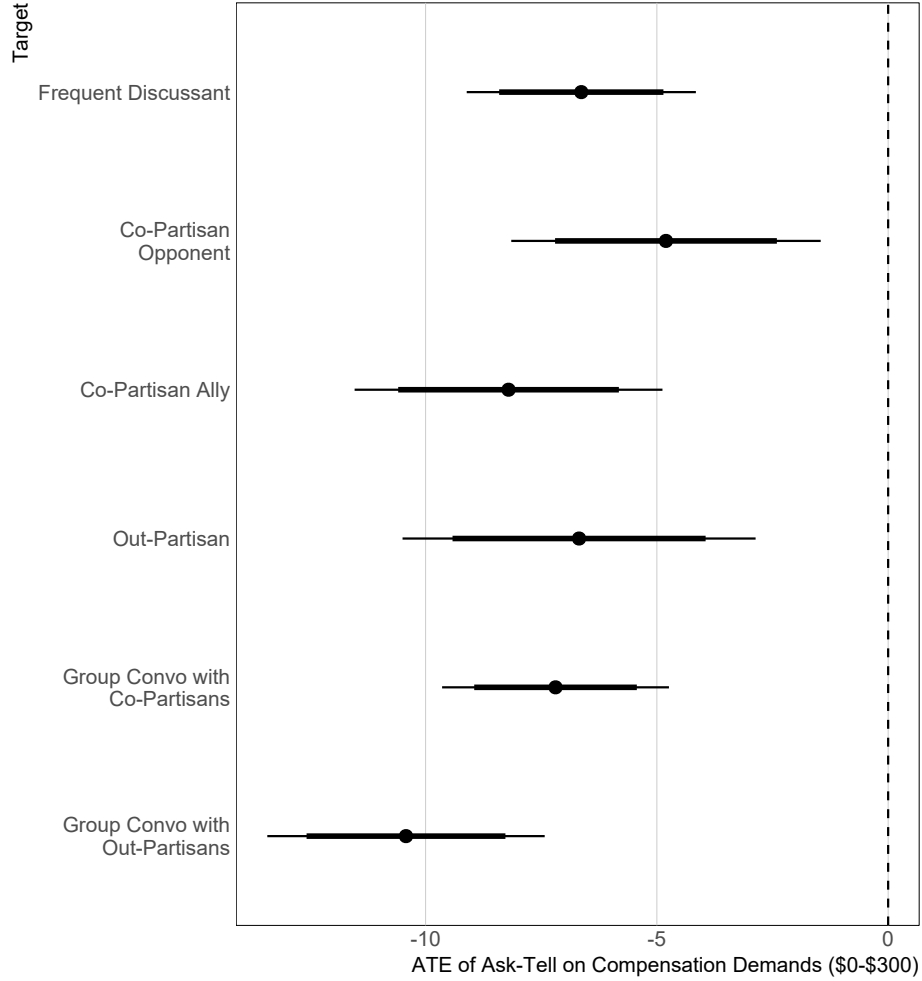


Figure 5: Ask-Tell ATE on One-on-One and Group CDs, Full Models in Table B9.

partisans ($-\$6.43$, $SE = 2.39$). These results corroborate the hypothesis that misperceptions about severe social sanctions motivate self-censorship in a ubiquitous fashion, irrespective of whether one holds the minority or majority view within their network.

The efficacy of our ask-tell correction among party majorities does not rule out meaningful heterogeneity between party minorities and majorities. Though we cannot reject the null of no difference between the sub-group effects, the average effect sizes for party minorities approximately double those of party majorities in magnitude, suggesting minorities may face increased social pressure. It is striking given our uniquely large sample size that these interaction tests still have limited power to detect moderate heterogeneity—an unavoidable challenge given the relatively small share of partisan dissenters. A minimum detectable effect

Table 2: Do Ask-Tell Effects Vary Between Party Majority and Minority?

	Co-Partisan Opp. CD	Co-Partisan Group CD
Ask-Tell	−3.804* (1.889)	−6.467** (2.384)
Party Min	21.609*** (3.184)	11.183** (4.161)
Ask-Tell:Party Min	−4.985 (4.364)	−7.348 (5.646)
Num.Obs.	11 043	5490
R2	0.284	0.402

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Full Models in Table B2.

(MDE) analysis (Appendix A.1) indicates that the design could reliably detect a differential ask-tell effect between party majorities and minorities of approximately 0.12–0.15 standard deviations at 80% power ($\alpha = 0.05$). The observed interaction estimates fall below this threshold. Thus, we cannot rule out meaningful but moderate asymmetry in how social pressure corrections affect majority versus minority partisans. In the Appendix, we assess the robustness of this conclusion to a wide range of alternative specifications: using a tobit model to account for truncated responses at the \$300 ceiling, recoding party minority status to only include voters who actively support an opposing- or third-party candidate, and assessing heterogeneous effects across pre-treatment measures of network homogeneity.

Social Media Treatment

A limitation of the prior analyses is that they examine speech only in the enclosed contexts of one-on-one or group conversations, rather than in public forums like online social media platforms. Our finding that generalized fears of social sanction constrain speech across the board—irrespective of party, social context, or the normative status of one’s beliefs—might fail to extrapolate to a social media setting in which self-censorship could be biased in a more overt way toward majority attitudes.

We leverage an additional randomization embedded in our survey design. As described in the Experimental Design section, we measured respondents’ perceptions about the likelihood they would personally face three types of severe social sanctions for expressing their

election preferences with peers. We randomly assigned respondents to answer these questions either with respect to (i) the one-on-one CD interactions they already considered or (ii) posting their election views to social media. This social media prime lets us examine whether social media environments exacerbate asymmetries in social pressure between opinion majorities and minorities.

Table 3: Social Media Prime Main Effects

	(1) Sanction Mean (H6a)	(2) Career	(3) Affiliative	(4) Strangers
Social Media Prime	0.178*** (0.021)	0.131*** (0.026)	0.122*** (0.025)	0.283*** (0.026)
Num.Obs.	16 901	16 925	16 931	16 930
R2	0.107	0.110	0.086	0.066

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001, Full Models in Table B4.

As Table 3 illustrates, the social media prime heightened the perceived risk for all three sanctions, consistent with the intuition that public speech acts bring to mind a wide range of one’s existing affiliations and the associated costs and benefits for honest speech. However, upon interacting exposure to the social media prime with party minority status, we find no evidence that the CATE was stronger for party minorities than majorities (Figure 6). Party majorities and party minorities alike perceive social media as a setting that engenders uniquely large social costs.

We also assess a second form of possible heterogeneity, leveraging pre-treatment questions that elicited self-reported network composition. Respondents provided estimates for the shares of their network who are co-partisans (N1), out-partisans (N2), and co-partisans hesitant to support the party nominee (N3). Responses were recorded on a 7-point scale (options: None, Almost none, A few, About half, A lot, Nearly all, All) then rescaled to be 0-1. Rescaled responses to N2 are employed as a rough proxy for the proportion of the respondent’s network who are out-partisans. We construct a proxy for co-partisan opponent network share by calculating $N1 * N3$ for party majorities and $N1 * (1 - N3)$ for party minorities.

Table 4 shows that both proxies for self-reported network share strongly predict the

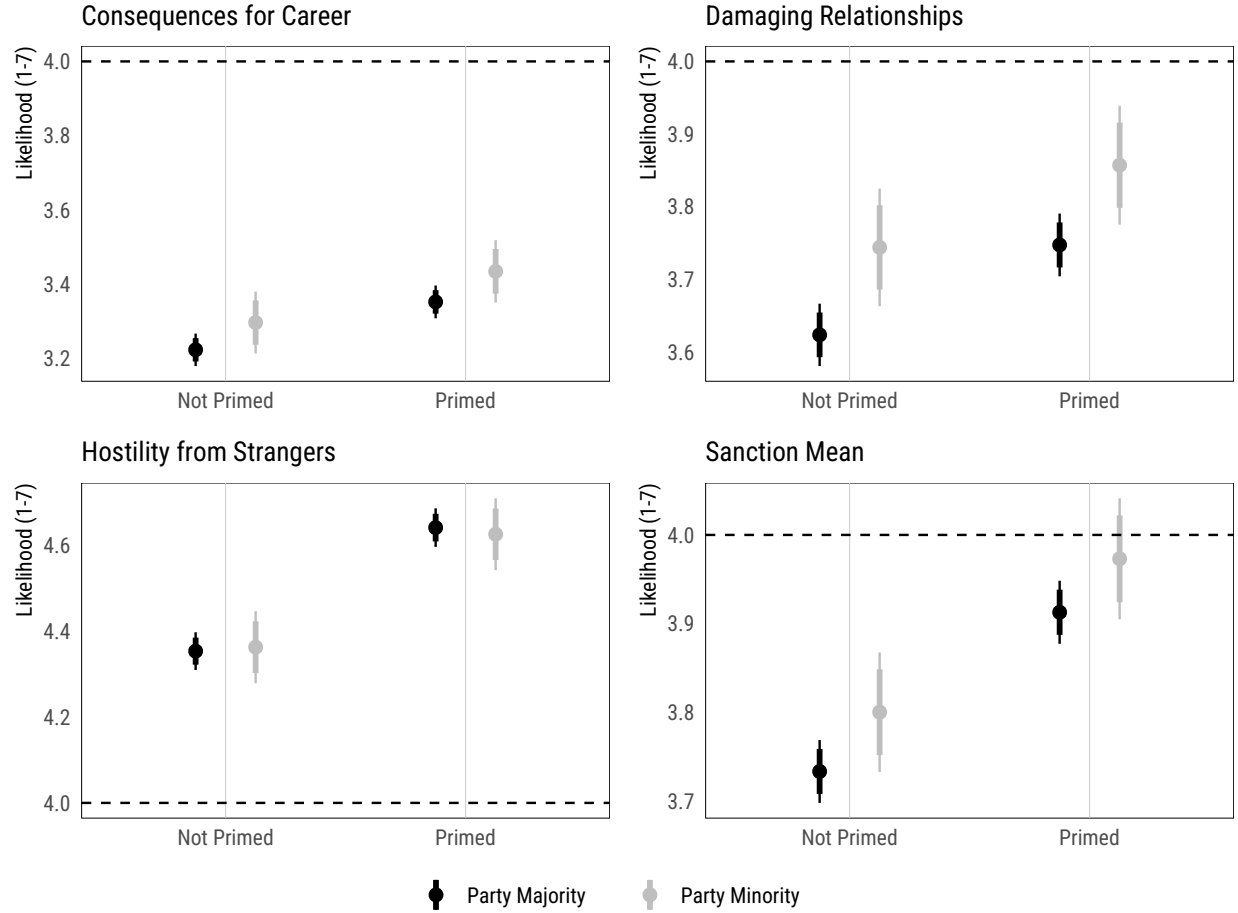


Figure 6: Estimated Marginal Means by Social Media Prime and Party Minority Status, Full Models in Table B10.

perceived risk of sanctions for the video-sharing assignment. That is, the perception of risk increases with the share of out-partisans and share of co-partisan opponents in the respondent's network. Again, however, we find no evidence of interaction effects between the social media prime and either proxy of network share. If anything, the signs on the coefficients suggest respondents with more opponents in their network respond somewhat less strongly to the social media prime. Respondents with adverse networks do not experience greater additional costs for public sharing than those with agreeable networks. These results run contrary to our expectation that individuals with more hostile networks would confront more aggressive concerns about publicly sharing a video, since public as opposed to private conduct implicates an individual's entire network.

Table 4: Interaction Effects for Social Media Prime by Network Proxies

	Sanction Mean
Social Media Prime	0.246*** (0.052)
Party Minority	0.057 (0.038)
Out-Partisan Network Share	0.975*** (0.080)
Co-Opp Network Share	0.410*** (0.087)
SM Prime X Party Minority	0.007 (0.053)
SM Prime X Opponent Network Share	-0.095 (0.112)
SM Prime X Co-Opp Network Share	-0.163 (0.123)
Num.Obs.	16 901
R2	0.107
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001, Full Model in Table B5.	

Explaining Party Minorities’ Disproportionate Self-Censorship

Our finding that even party majorities self-censor in intra-party settings due to generalized fears of social sanctions does not preclude the possibility that party minorities face *greater* social pressure, as theories of intra-party norms may predict. Even so, self-censorship by party majorities raises a crucial question: what explains the stark asymmetry in party majorities’ and minorities’ willingness to speak? While partisan norms are one plausible explanation, our follow-up survey offers compelling evidence these norms are unlikely to explain the full scope of party minorities’ disproportionate self-censorship.

We fielded a short module in a survey with $N = 3,400$ partisans recruited through CloudResearch Connect between October 31 and November 4, 2024. Respondents were shown a list of four reasons why they might initiate a discussion about the election with a co-partisan opponent and were asked to select all that apply or “None of the above.” A similar question was asked containing four reasons why they might *avoid* a discussion about the election. Figure 7 visualizes the proportion of respondents who selected each reason,

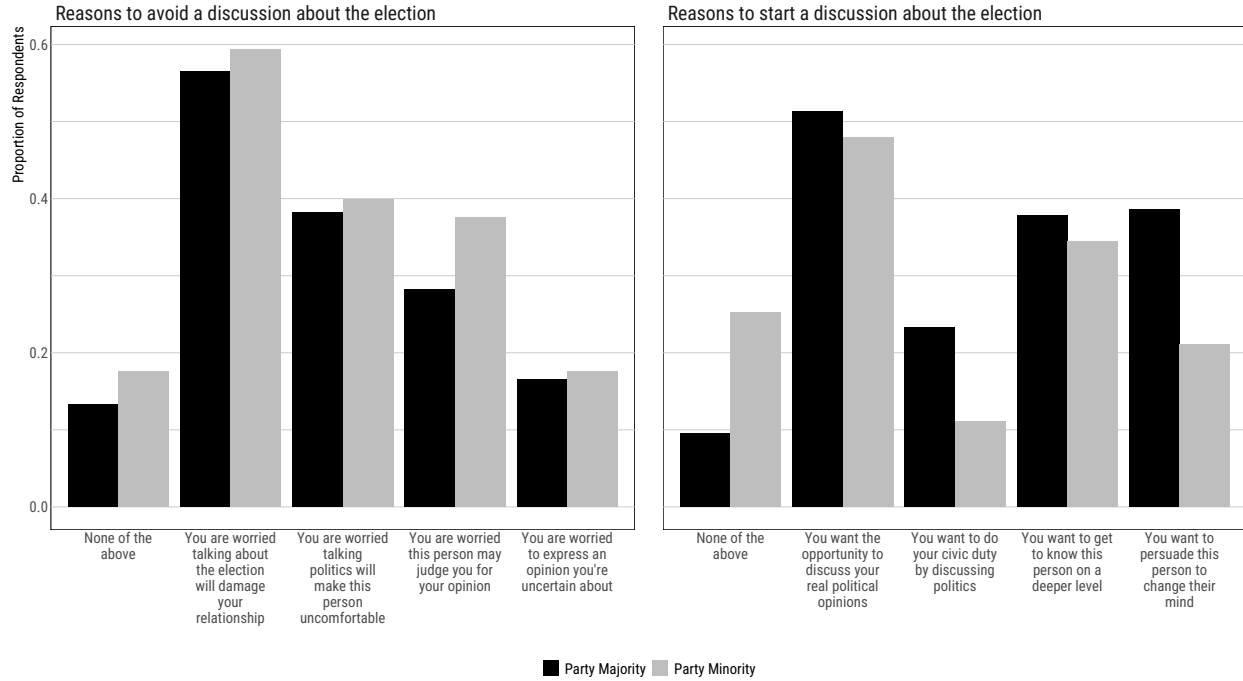


Figure 7: Considerations for Election Discussions with Co-Partisan Opponents (Know a Co-Opp), Full Models in Tables B11 and B12.

disaggregated for party majorities and minorities. This analysis includes only the 1,001 respondents who said they personally knew a co-partisan who disagreed with them about whom to vote for in the 2024 presidential election, though the substantive conclusions of the analysis remain when all partisans are included, as shown in the Appendix.

The most salient difference between party majorities and minorities in terms of the considerations bearing on co-partisan speech concerned the opportunity to persuade the discussant to change their mind. We find that party majorities were about 17.6 percentage points ($SE = 3$) more likely to cite persuasion as a perceived benefit to starting a conversation on the election. Party majorities were also 12.1 percentage points ($SE = 2$) more likely to cite a civic duty of discussing politics, potentially illuminating dispositional differences in engagement between the party faithful and party defectors. Party minorities were 15.7 percentage points ($SE = 3$) more likely to say they see none of the listed considerations as potential benefits for speaking with co-partisan opponents.

By contrast, we find no statistically significant difference in the frequency with which the two groups cited concerns about damaging relationships or making the other discussant

uncomfortable as a reason to avoid discussing the election. Among majorities and minorities alike, nearly 60 percent cite affiliative costs as a reason to avoid intra-party speech—by far the most salient consideration. There was only one concern that party minorities were meaningfully more likely to mention: fearing judgment for their opinion (9.3 pp, $SE = 3$). This squares with our experimental findings, suggesting that although dissenters perceive their minority opinion as stigmatized, dyadic conversation settings render affiliative costs salient to party minorities and majorities alike. Differing strategic motivations play a role in explaining the large asymmetry between party majorities and minorities in the compensation they demand for speaking to co-partisan opponents.

Discussion

In this paper, we have presented evidence that Americans dramatically overestimate the severity of social risks for expressing political disagreement in America today, and that correcting these misperceptions reduces self-censorship. We highlight the sheer breadth of social contexts in which these generalized fears of social sanction constrain political speech, ranging from one-on-one conversations with intra-party opponents to group conversations with out-partisans to social media platforms. Moreover, these misperceptions about social sanctions seem to affect all Americans, irrespective of whether they hold the minority or majority status in either (a) their political party or (b) their own social network.

We contrast the ubiquity of social pressure with meaningful asymmetries between different partisan groups in their willingness to share their preferences about the 2024 election. Party majorities far prefer discussing politics with co-partisans over out-partisans, yet party minorities are profoundly unwilling to speak with co-partisans who toe the party line. We suggest that while these patterns could be driven in part by partisan norms, asymmetries in perceived positive benefits (i.e., persuasion) also play a crucial role. These results offer conditional support for multiple existing theories of social pressure—namely, the partisan norms and inter-party conflict accounts—while advancing a new, more holistic account of social pressure rooted in Americans’ widespread misperceptions about the severity of societal

divisions.

These results shed light on recent trends in public opinion and public discourse regarding constraints on speech. Though the rise in reported self-censorship among Americans has often been attributed to partisan polarization and intra-party conflict, our results suggest that these claims may be self-fulfilling insofar as they fuel generalized fear of social sanction. News coverage and discourse about the severity of polarization and cancel culture may send repeated signals to citizens about the social risks of political debate, seeding these exaggerated fears (Levendusky and Malhotra, 2016*a*; Heseltine et al., 2025). In turn, those who continue to debate politics may be citizens who derive particular utility from politics, such as the instrumental benefits of persuasion or personal enjoyment (Hersh, 2020; Krupnikov and Ryan, 2022). These voters may engage in political debate not for lack of social pressure, but rather *despite* believing the act to be socially precarious.

Political scientists have long decried people’s aversion to discussing politics with ideological opponents, arguing that cross-cutting debate makes better citizens by improving their understanding of the legitimacy of opposing viewpoints (Fishkin, 1991; Mutz, 2002). In theory, intra-party ideological disagreements should offer a silver lining: even if voters avoid political debates with out-partisans, they may be exposed to cross-cutting views through co-partisans who hold counter-stereotypical beliefs. Indeed, co-partisans are often the most effective messengers for counter-stereotypical information (Clayton and Willer, 2023; Larsen et al., 2023; Dabruzzi, 2024). Yet recent evidence suggests rising hostilities among intra-party factions deter political conversation. During the 2024 presidential election campaign, Republican voters described fearing ostracism from co-partisans if they expressed misgivings about Donald Trump being the party nominee (Tabet, 2023; McCammon, 2024).

Despite recent research on partisan norms, few studies have directly assessed whether social pressure is the mechanism of injunctive norm enforcement against intra-party dissent. The institutions that White and Laird (2020) show have historically sustained norms of Democratic support among Black conservatives do not generalize in an obvious fashion to other partisan contexts in the current era of relatively weak party institutions (Schlozman and Rosenfeld, 2024). Our results offer some of the first direct empirical evidence that partisan

dissenters indeed perceive substantial social costs to discussing points of disagreement with members of the party faithful. Yet we also show that party majorities face social pressure, and that party minorities’ aversion to intra-party speech may be explained not solely by a stigmatization of their beliefs but also by majorities’ propensity to perceive instrumental benefits in persuading dissenters to change their mind. Future work could unpack to what extent this association is dispositional or a causal effect stemming from electoral incentives.

A crucial question going forward concerns the external validity of our findings. We show that during the 2024 election, the topic of parties’ presidential nominees was sufficiently contentious within and between parties to motivate self-censorship. While factional conflict is on the rise, historically these factional coalitions have been notoriously short-lived (Key, 1949; Hopkins and Sigler, 2026). Further, presidential elections are times of heightened partisan conflict (Iyengar et al., 2019). These factors mean that our surveys’ timing may have supplied an easy test for our generalized social pressure hypothesis. Still, the political environment our paper studies is likely to resurface in future election cycles as parties strive to maintain loyalty to their nominees and new factional divisions emerge.

Future research should investigate questions that our design poses but cannot fully answer. For instance, do misperceptions about social sanction stem from media highlighting exceptional cases of “cancel culture,” similar to an exaggerated media focus on partisan hostility (Levendusky and Malhotra, 2016a)? Are these beliefs stable over time? Future work should also benchmark self-censorship concerns against positive motivations for political speech, including social norms in favor of participation (Gerber, Green and Larimer, 2008; Bond et al., 2012).

Media coverage, especially in partisan outlets—along with academic papers on censorship (e.g., Noelle-Neumann, 1974; Ho and Huang, 2024)—has drawn sustained attention to a form of social sanctioning in which one political faction specifically suppresses the beliefs espoused by another, often making analogies to authoritarian repression. Through a large, nationwide sample, we show that this is not the only or principal mechanism through which Americans experience pressure to self-censor. Social pressure is not singularly concentrated in any one faction, but is relevant to diverse inter- and intra-party contexts.

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):964–981. The University of Chicago Press.
- Alvarez-Benjumea, Amalia and Vicente Valentim. 2024. “The Enforcement of Political Norms.” *British Journal of Political Science* 54(4):1043–1066.
- Asch, Solomon E. 1956. “Studies of independence and conformity: I. A minority of one against a unanimous majority.” *Psychological Monographs: General and Applied* 70(9):1–70.
- Berelson, Bernard R., Paul F. Lazarsfeld and William N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago, IL: University of Chicago Press.
- Bloch Rubin, Ruth. 2017. *Building the Bloc: Intraparty Organization in the US Congress*. 1 ed. Cambridge University Press.
- Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle and James H. Fowler. 2012. “A 61-million-person experiment in social influence and political mobilization.” *Nature* 489(7415):295–298.
- Braghieri, Luca. 2024. “Political Correctness, Social Image, and Information Transmission.” *American Economic Review* 114(12):3877–3904.
- Braley, Alia, Gabriel S. Lenz, Dhaval Adjudah, Hossein Rahnema and Alex Pentland. 2023. “Why voters who value democracy participate in democratic backsliding.” *Nature Human Behaviour* 7(8):1282–1293.
- Brown, Jacob R. and Ryan D. Enos. 2021. “The measurement of partisan sorting for 180 million voters.” *Nature Human Behaviour* 5(8):998–1008.
- Bursztyn, Leonardo and David Y. Yang. 2022. “Misperceptions About Others.” *Annual Review of Economics* 14(1):425–452.
- Butters, Ross and Christopher Hare. 2022. “Polarized Networks? New Evidence on American Voters’ Political Discussion Networks.” *Political Behavior* 44(3):1079–1103.
- Campbell, Angus, Philip E. Converse, Warren E. Miller and Donald E. Stokes. 1960. *The American Voter*. New York: John Wiley & Sons, Inc.
- Carlson, Taylor N. and Jaime E. Settle. 2016. “Political Chameleons: An Exploration of Conformity in Political Discussions.” *Political Behavior* 38(4):817–859.
- Carlson, Taylor N. and Jaime E. Settle. 2022. *What Goes Without Saying: Navigating Political Discussion in America*. Cambridge: Cambridge University Press.
- Carlson, Taylor N. and Jaime E. Settle. 2023. “Freedom of Expression in Interpersonal Interactions.” *PS: Political Science & Politics* 56(2):245–249.
- Chong, Dennis, Jack Citrin and Morris Levy. 2024. “The Realignment of Political Tolerance in the United States.” *Perspectives on Politics* 22(1):131–152.

- Cialdini, Robert B., Carl A. Kallgren and Raymond R. Reno. 1991. A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. In *Advances in Experimental Social Psychology*, ed. Mark P. Zanna. Vol. 24 Academic Press pp. 201–234.
- Cialdini, Robert B. and Noah J. Goldstein. 2004. “Social Influence: Compliance and Conformity.” *Annual Review of Psychology* 55(1):591–621.
- Clayton, Katherine and Robb Willer. 2023. “Endorsements from Republican politicians can increase confidence in U.S. elections.” *Research & Politics* 10(1):20531680221148967.
- Connors, Elizabeth C. 2020. “The Social Dimension of Political Values.” *Political Behavior* 42(3):961–982.
- Connors, Elizabeth C., Matthew T. Pietryka and John Barry Ryan. 2022. “Examining Motivations in Interpersonal Communication Experiments.” *Elements in Experimental Political Science* .
- Dabruzzi, Anthony. 2024. “How groups opposed to Trump are using the RNC as their platform.” *Spectrum News* 1 .
- Dias, Nicholas C, Laurits F Aarslew, Kristian Vrede Skaaning Frederiksen, Yphtach Lelkes, Lea Pradella and Sean J Westwood. 2024. “Correcting misperceptions of partisan opponents is not effective at treating democratic ills.” *PNAS Nexus* 3(8):pgae304.
- Dias, Nicholas C., Yphtach Lelkes and Jacob Pearl. 2025. “American partisans vastly underestimate the diversity of other partisans’ policy attitudes.” *Political Science Research and Methods* 13(3):725–735.
- Dias, Nicholas, James N. Druckman and Matthew S. Levendusky. 2024. “Unraveling a “Cancel Culture” Dynamic: When, Why, and Which Americans Sanction Offensive Speech.” *The Journal of Politics* .
- Eliasoph, Nina. 1998. *Avoiding Politics: How Americans Produce Apathy in Everyday Life*. Cambridge Cultural Social Studies Cambridge: Cambridge University Press.
- Fishkin, James S. 1991. *Democracy and Deliberation: New Directions for Democratic Reform*. Yale University Press.
- Gerber, Alan S., Donald P. Green and Christopher W. Larimer. 2008. “Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment.” *American Political Science Review* 102(1):33–48.
- Gerber, Alan S., Gregory A. Huber, David Doherty and Conor M. Dowling. 2012. “Disagreement and the Avoidance of Political Discussion: Aggregate Relationships and Differences across Personality Traits.” *American Journal of Political Science* 56(4):849–874.
- Gerber, Alan S., Gregory A. Huber, David Doherty and Conor M. Dowling. 2013. “Is There a Secret Ballot? Ballot Secrecy Perceptions and Their Implications for Voting Behaviour.” *British Journal of Political Science* 43(1):77–102.

- Gerber, Alan S., Gregory A. Huber and Ebonya Washington. 2010. "Party Affiliation, Partisanship, and Political Beliefs: A Field Experiment." *American Political Science Review* 104(4):720–744.
- Gibson, James L and Joseph L Sutherland. 2023. "Keeping Your Mouth Shut: Spiraling Self-Censorship in the United States." *Political Science Quarterly* 138(3):361–376.
- Gillion, Daniel Q. 2020. *The Loud Minority: Why Protests Matter in American Democracy*. In *The Loud Minority*. Princeton University Press.
- Glynn, Carroll J., Andrew F. Hayes and James Shanahan. 1997. "Perceived Support for One's Opinions and Willingness to Speak Out: A Meta-Analysis of Survey Studies on the "Spiral of Silence"." *The Public Opinion Quarterly* 61(3):452–463.
- Green, Donald, Bradley Palmquist and Eric Schickler. 2002. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.
- Green, Donald P. and Alan S. Gerber. 2010. "Introduction to Social Pressure and Voting: New Experimental Evidence." *Political Behavior* 32(3):331–336.
- Groenendyk, Eric, Erik O. Kimbrough and Mark Pickup. 2023. "How Norms Shape the Nature of Belief Systems in Mass Publics." *American Journal of Political Science* 67(3).
- Groenendyk, Eric, Yanna Krupnikov, John Barry Ryan and Elizabeth C. Connors. 2025. "Selecting Out of "Politics": The Self-Fulfilling Role of Conflict Expectation." *American Political Science Review* 119(1):40–55.
- Guess, Andrew M. 2021. "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets." *American Journal of Political Science* 65(4):1007–1022. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12589>.
- Hanemann, W. Michael. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *The American Economic Review* 81(3):635–647. Publisher: American Economic Association.
- Henderson, John A., Geoffrey Sheagley, Stephen N. Goggin, Logan Dancey and Alexander G. Theodoridis. 2022. "Primary Divisions: How Voters Evaluate Policy and Group Differences in Intraparty Contests." *The Journal of Politics* 84(3):1760–1776.
- Hersh, Eitan. 2020. *Politics Is for Power: How to Move Beyond Political Hobbyism, Take Action, and Make Real Change*. Simon and Schuster.
- Heseltine, Michael, Bernhard Clemm von Hohenberg, Ericka Menchen-Trevino, Tomasz Gackowski and Magdalena Wojcieszak. 2025. "Effects of Over-Time Exposure to Partisan Media and Coverage of Polarization on Perceived Polarization." *Political Communication* 42(3):432–453.
- Ho, Yuen and Yihong Huang. 2024. "Breaking the Spiral of Silence."
- Hopkins, Daniel J. and Gall Sigler. 2026. "Not of Primary Concern: Assessing Ideological Voting Over Time in U.S. Primaries, 2008–2024." *Political Behavior* 48(2):695–719.

- Huckfeldt, Robert, Jeanette Morehouse Mendez and Tracy Osborn. 2004. "Disagreement, Ambivalence, and Engagement: The Political Consequences of Heterogeneous Networks." *Political Psychology* 25(1):65–95.
- Iyengar, Shanto and Sean J. Westwood. 2015. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59(3):690–707.
- Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra and Sean J. Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22(Volume 22, 2019):129–146.
- Key, V. O. 1949. *Southern Politics in State and Nation*. New York: A.A. Knopf.
- Klar, Samara, Yanna Krupnikov and John Barry Ryan. 2018. "Affective Polarization or Partisan Disdain? Untangling a Dislike for the Opposing Party from a Dislike of Partisanship." *Public Opinion Quarterly* 82(2):379–390.
- Krupnikov, Yanna and John Barry Ryan. 2022. *The Other Divide*. Cambridge: Cambridge University Press.
- Larsen, Bradley J., Timothy J. Ryan, Steven Greene, Marc J. Hetherington, Rahsaan Maxwell and Steven Tadelis. 2023. "Counter-stereotypical messaging and partisan cues: Moving the needle on vaccines in a polarized United States." *Science Advances* 9(29):eadg9434.
- Lee, Amber Hye-Yon and Chloe Ahn. 2025. "Talking Politics in a Polarized America: How Perceived Polarization Shapes Political Self-Censorship." *Political Behavior* .
- Levendusky, Matthew. 2009. *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. University of Chicago Press. Google-Books-ID: O0mTXyhQnD8C.
- Levendusky, Matthew and Neil Malhotra. 2016a. "Does Media Coverage of Partisan Polarization Affect Political Attitudes?" *Political Communication* 33(2):283–301.
- Levendusky, Matthew S. and Neil Malhotra. 2016b. "(Mis)perceptions of Partisan Polarization in the American Public." *Public Opinion Quarterly* 80(S1):378–391.
- Leviton, Lindsey C. and Brad Verhulst. 2016. "Conformity in Groups: The Effects of Others' Views on Expressed Attitudes and Attitude Change." *Political Behavior* 38(2):277–315.
- Mason, Lilliana. 2018. *Uncivil Agreement: How Politics Became Our Identity*. Chicago, IL: University of Chicago Press.
- McCammon, Sarah. 2024. "Meet the conservative women who are keeping their votes for Kamala Harris a secret." *NPR* .
- Mernyk, Joseph S., Sophia L. Pink, James N. Druckman and Robb Willer. 2022. "Correcting inaccurate metaperceptions reduces Americans' support for partisan violence." *Proceedings of the National Academy of Sciences* 119(16):e2116851119.

- Minozzi, William, Hyunjin Song, David M. J. Lazer, Michael A. Neblo and Katherine Ognyanova. 2020. "The Incidental Pundit: Who Talks Politics with Whom, and Why?" *American Journal of Political Science* 64(1):135–151.
- Moore-Berg, Samantha L. and Boaz Hameiri. 2024. "Improving intergroup relations with meta-perception correction interventions." *Trends in Cognitive Sciences* 28(3):190–192.
- Mutz, Diana C. 2002. "Cross-Cutting Social Networks: Testing Democratic Theory in Practice." *The American Political Science Review* 96(1):111–126.
- Mutz, Diana C. 2006. *Hearing the Other Side: Deliberative versus Participatory Democracy*. Cambridge: Cambridge University Press.
- Noelle-Neumann, Elisabeth. 1974. "The Spiral of Silence: A Theory of Public Opinion." *Journal of Communication* 24(2):43–51.
- Orr, Lilla V., Anthony Fowler and Gregory A. Huber. 2023. "Is Affective Polarization Driven by Identity, Loyalty, or Substance?" *American Journal of Political Science* 67(4):948–962.
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12796>.
- Ploger, Gavin. 2024. "Polarization All the Way Down: How Coverage of Elite and Partisan Polarization Spills Over to Perceptions of the U.S. Mass Public." *Political Communication* 41(3):393–412.
- Prior, Markus. 2005. "News vs. Entertainment: How Increasing Media Choice Widens Gaps in Political Knowledge and Turnout." *American Journal of Political Science* 49(3):577–592.
- Robison, Joshua and Kevin J. Mullinix. 2016. "Elite Polarization and Public Opinion: How Polarization Is Communicated and Its Effects." *Political Communication* 33(2):261–282.
- Schlozman, Daniel and Sam Rosenfeld. 2019. The Hollow Parties. In *Can America Govern Itself?*, ed. Frances E. Lee and Nolan McCarty. Cambridge University Press pp. 120–153.
- Schlozman, Daniel and Sam Rosenfeld. 2024. *The Hollow Parties: The Many Pasts and Disordered Present of American Party Politics*. Princeton University Press.
- Settle, Jaime E. and Taylor N. Carlson. 2019. "Opting Out of Political Discussions." *Political Communication* 36(3):476–496.
- Shogren, Jason F., Seung Y. Shin, Dermot J. Hayes and James B. Kliebenstein. 1994. "Resolving Differences in Willingness to Pay and Willingness to Accept." *The American Economic Review* 84(1):255–270.
- Smidt, Corwin D. 2017. "Polarization and the Decline of the American Floating Voter." *American Journal of Political Science* 61(2):365–381.
- Spelman, Trevor, Abdo Elnakouri, Nour Kteily and Eli J. Finkel. 2026. "Overestimating the social costs of political belief change." *Journal of Personality and Social Psychology* 130(4):742–762.
- Suhay, Elizabeth. 2015. "Explaining Group Influence: The Role of Identity and Emotion in Political Conformity and Polarization." *Political Behavior* 37(1):221–251.

- Sullivan, John L., James Piereson and George E. Marcus. 1979. "An Alternative Conceptualization of Political Tolerance: Illusory Increases 1950s-1970s." *The American Political Science Review* 73(3):781–794.
- Tabet, Alex. 2023. "Trump's secret weapon consolidating the GOP: Fear." *NBC News* .
- Tankard, Margaret E. and Elizabeth Levy Paluck. 2016. "Norm Perception as a Vehicle for Social Change." *Social Issues and Policy Review* 10(1):181–211.
- Valentim, Vicente. 2024. "Political Stigma and Preference Falsification: Theory and Observational Evidence." *The Journal of Politics* 86(4):1382–1402.
- Van Bavel, Jay J. and Andrea Pereira. 2018. "The Partisan Brain: An Identity-Based Model of Political Belief." *Trends in Cognitive Sciences* 22(3):213–224.
- Venkatachalam, L. 2004. "The contingent valuation method: a review." *Environmental Impact Assessment Review* 24(1):89–124.
- White, Ismail K. and Chryl N. Laird. 2020. *Steadfast Democrats: How Social Forces Shape Black Political Behavior*. Princeton University Press.
- Young, David J. and Lee H. de Wit. 2025. "Affective polarization within parties." *Political Psychology* 46(1):49–69.

Appendix

A	Exploratory Analyses	1
A.1	Minimum Detectable Effects for Heterogeneity by Party Minority Status . .	1
A.2	Compensation Demand Scale Design and Ceiling Effects	1
A.3	Demographic Predictors of Prior Beliefs	4
A.4	Robustness to Alternate Specifications of Party Minority	7
A.5	Do Meta-Perceptions of Social Pressure Reflect Genuine Misperceptions? . .	9
A.6	Alternative Mechanism Tests	12
A.7	Robustness to Network Missingness	15
A.8	Sample Composition and Balance Tables	16
A.9	Reexamining Surveys about Self-Censorship	18
A.10	Placebo and Spillover Tests	20
A.11	Differences in Ask-Tell Asymmetry by Party	21
A.12	Additional Exploratory Analyses	22
B	Full Models	24
C	Additional Pre-Registered Analyses	35
C.1	Hypothesis 3	35
C.2	Hypothesis 4	36
C.3	Hypothesis 5	37
D	Ethical Statement	38

A Exploratory Analyses

A.1 Minimum Detectable Effects for Heterogeneity by Party Minority Status

Table 2 in the main text tests whether the ask-tell treatment effect on compensation demands varies by party minority status. The interaction term (Ask-Tell $\times$ Party Minority) is negative in both models but does not reach conventional levels of statistical significance. To clarify the inferential limits of these tests, we report minimum detectable effects (MDEs)—the smallest true interaction effect each model could detect with 80% power at $\alpha = 0.05$ (two-sided).

Following standard practice, we compute the MDE as $(z_{0.95} + z_{0.80}) \times \text{SE} \approx 2.49 \times \text{SE}$ in the case of a one-sided test and $(z_{0.975} + z_{0.80}) \times \text{SE} \approx 2.8 \times \text{SE}$ for a two-sided test, where SE is the HC2 standard error of the interaction coefficient from the fitted model. We also express MDEs in standard deviation units of the relevant outcome to aid interpretation.

Table A1: Minimum Detectable Effects for Table 2 Interaction Terms

	N	$\hat{\beta}_{\text{int}}$	SE	p	MDE (one-sided)		MDE (two-sided)	
					\$	SD	\$	SD
Co-Partisan Opp. CD	11,043	-5.20	4.36	0.233	-10.85	-0.10	-12.22	-0.12
Co-Partisan Group CD	5,490	-7.21	5.66	0.203	-14.08	-0.14	-15.86	-0.15

Note: MDEs computed at 80% power, $\alpha = 0.05$. One-sided MDEs are in the direction of the point estimate; two-sided MDEs are signed to match. $\hat{\beta}_{\text{int}}$ is the coefficient on Ask-Tell $\times$ Party Minority.

For compensation demands directed at a co-partisan opponent, the one-sided MDE is -\$10.85 (-0.10 SD); for group conversation demands with co-partisans on the election topic, the MDE is -\$14.08 (-0.14 SD). The more conservative two-sided MDEs are -\$12.22 (-0.12 SD) and -\$15.86 (-0.15 SD), respectively. The observed interaction term coefficients (-\$5.20 and -\$7.21, respectively) fall below these thresholds. Thus, while we cannot reject equal treatment effects across majority and minority partisans, we also cannot rule out meaningful heterogeneity.

A.2 Compensation Demand Scale Design and Ceiling Effects

In this section we describe how we pre-tested the slider scale used for the compensation demand (CD) outcomes and then assess the robustness of our conclusions to its ceiling. The choice of ceiling for a CD item is consequential: ceilings can both “cut off” responses that would otherwise exceed the maximum and “anchor” the broader distribution of responses (e.g., some people will respond at the midpoint of the scale even as the value of the midpoint changes). These are two distinct mechanisms with different empirical implications. Pinning at the ceiling can also mechanically attenuate OLS estimates of treatment effects, and disproportionately so within subgroups (such as party minorities) whose baseline CDs are higher and who are therefore more likely to bunch at the ceiling.

Pilot Testing Prior to designing this study, we conducted a pilot experiment to evaluate how scale ceilings shape compensation-demand responses. Results are summarized publicly in an online blog post hosted by the survey provider YouGov. The pilot randomized the slider ceiling (\$50, \$100, or \$200) for a CD item similar to the one in our final survey—the price

respondents would accept to record a one-minute video expressing their political views—in a 3×2 factorial design crossed with a separate experimental treatment ($N \approx 2,300$). Raising the ceiling increased the mean reported price as well as lowered the share of respondents pinned at the ceiling, from 61% at \$50 to 49% at \$100 and 42% at \$200, confirming that both anchor and cutoff mechanisms operate in these CD measures. The pilot did not find evidence that ceilings systematically interact with treatment effects, though we attributed this null in part to the pilot treatment itself having no effect on the outcome.

These pilot results directly motivated two design choices in the present study. First, we set the slider ceiling for the primary CD outcomes at \$300—substantially above the upper end of the pilot’s tested range—to reduce ceiling effects and thus preserve room for downward or upward movement near the ceiling. Second, because the pilot showed that anchoring meaningfully shifts the distribution, we do not interpret the absolute magnitude of CD prices as substantively meaningful. We are interested only in relative prices and how these prices shift under treatment. Figure A1 shows the realized distribution of CDs for the co-partisan opponent and co-partisan group outcomes by treatment condition. Even at the \$300 ceiling, the share of respondents at the maximum is substantial—about 18% for the co-partisan opponent CD and about 22% for the co-partisan group CD among control respondents—but it is markedly lower than what we observed at the lower ceilings tested in the pilot, indicating that our scale choice meaningfully reduced ceiling effects.

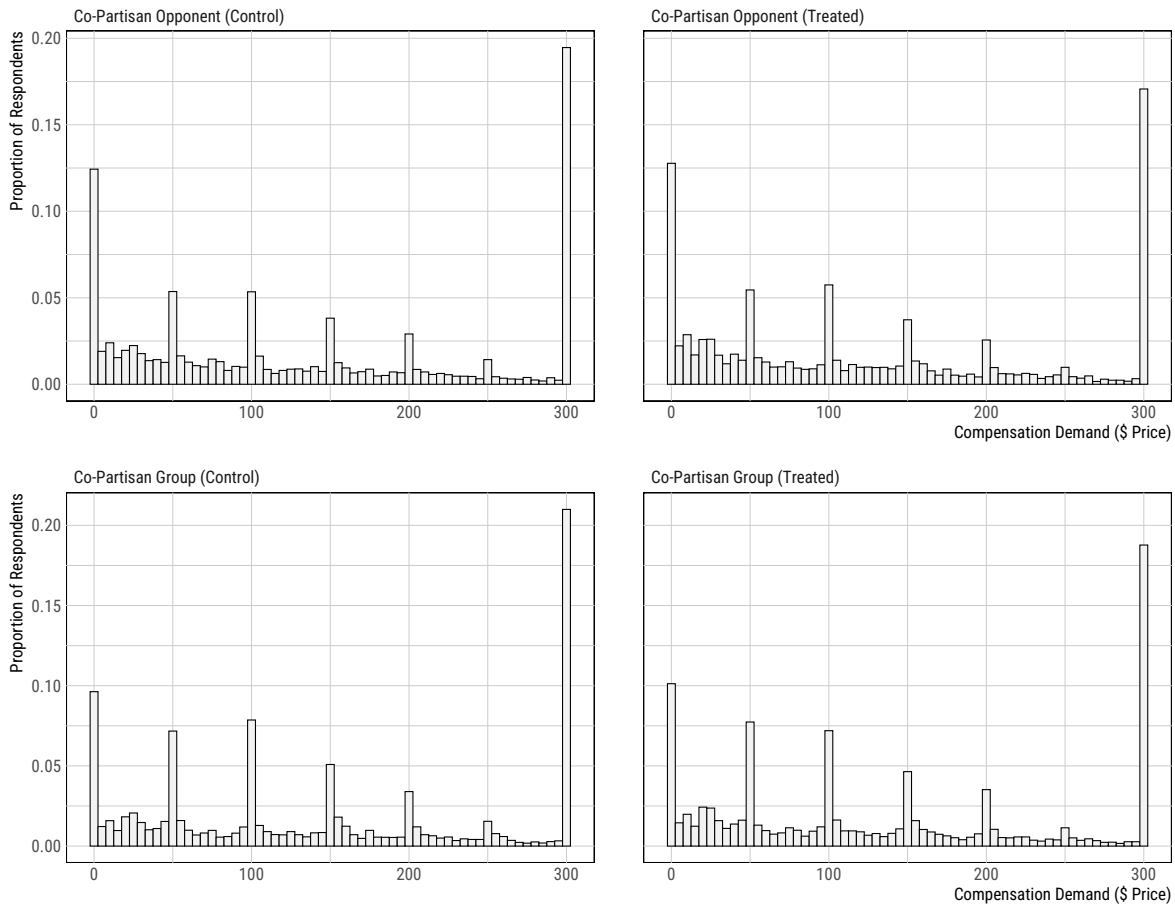


Figure A1: Distribution of Co-Partisan Compensation Demands

Robustness of Main Analyses Although the \$300 ceiling reduces ceiling effects relative to lower ceilings, residual bunching at \$0 and \$300 remains visible in Figure A1, motivating two robustness checks designed to assess whether ceiling effects bias our main estimates. Ceiling effects could attenuate the estimated treatment effect, and disproportionately so within subgroups for whom ceiling effects are more severe at baseline. In particular, since party minorities demand higher prices at baseline and could experience higher rates of cutoff at \$300, this attenuation could mask additional heterogeneity in ask-tell treatment effects. In this section we explore this possibility using two alternative specifications of our main model.

A close look at Figure A1 suggests that the ask-tell intervention does reduce the share of respondents at the \$300 ceiling: in both the opponent and group panels, the treated histograms have visibly shorter ceiling bars than the control histograms. Our first alternative specification, a binary outcome model, tests this pattern directly by recoding the outcome as one if the respondent demands the maximum (\$300) and zero otherwise, then re-estimating the same linear model. This model assesses whether the intervention reduced the share of respondents hitting the ceiling and if this effect is larger or smaller for party minorities. If the interaction effect between the ask-tell and party minority status is positive—in particular, despite party minorities being more likely to hit the ceiling in the control condition—this indicates that ceiling effects might artificially attenuate the ask-tell effect to a greater degree for party minorities. If the interaction effect is negative, however, then the attenuating effect is less evident.

Our second specification is a robustness check using a tobit model (left-censored at \$0, right-censored at \$300). This approach treats observed CDs as truncated realizations of a latent willingness-to-accept that may extend beyond the slider’s bounds and estimates treatment effects on the latent (uncensored) scale. Under the model, the tobit coefficient recovers what OLS would have estimated had the slider no bounds and respondents could enter any non-negative amount. If ceiling effects are attenuating OLS, the tobit coefficients should be larger in magnitude than the OLS coefficients on the same models and the interaction effect with party minority would be statistically significant.

Table A2: Robustness of Table 2 Estimates to Ceiling Effects

	OLS		Binary		Tobit	
	Co-Opp. (1)	Group (2)	Co-Opp. (3)	Group (4)	Co-Opp. (5)	Group (6)
Ask-Tell	−3.706* (1.888)	−6.418** (2.380)	−0.008 (0.007)	−0.014 (0.010)	−4.479+ (2.572)	−7.276* (3.231)
Party Minority	21.652*** (3.181)	10.938** (4.168)	0.095*** (0.013)	0.074*** (0.018)	30.002*** (4.087)	16.618** (5.137)
Ask-Tell × Party Min	−5.203 (4.363)	−7.213 (5.662)	−0.025 (0.018)	−0.045+ (0.024)	−7.749 (5.635)	−11.994+ (7.001)
Num.Obs.	11 043	5490	11 043	5490	11 043	5490
R2	0.283	0.401	0.163	0.241		

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2 reports the results. The first two columns contain the OLS results reported in

the main text for the co-partisan opponent CD and co-partisan group CD. The main effects from the tobit model corroborate our OLS finding that the ask-tell treatment reduces prices demanded to discuss the election with co-partisans ($-\$4.48$, $p = 0.082$ for the co-partisan opponent CD; $-\$7.28$, $p = 0.024$ for the group CD). The binary estimates, however, are not significant for either outcome (-0.008 , $p = 0.233$ for the one-on-one CD; -0.014 , $p = 0.151$ for the group CD): we do not find evidence that the intervention measurably pulls party majorities off the ceiling. The combined picture is that the intervention reduces majorities' CDs in dollar terms (as OLS and tobit indicate) but does not detectably shift the share of majorities who hit the ceiling (meaning either their price is above \$300 or they would refuse the task at any price).

The interaction terms tell a more nuanced story. First, columns 3–4 show that in the control condition, party minorities were substantially more likely to be at the ceiling than party majorities (9.5 percentage points more likely for the co-partisan CD and 7.4 percentage points for the group CD). How does this affect the CATEs? Under OLS, the interaction is $-\$5.20$ ($p = 0.233$) for demands to a co-partisan opponent and $-\$7.21$ ($p = 0.203$) for group conversation demands. Under tobit, the latent-scale interactions are slightly larger in magnitude at $-\$7.75$ ($p = 0.169$) and $-\$11.99$ ($p = 0.087$). The latter is marginally significant at the $p < 0.1$ level. At the same time, under our binary specification, the interaction effect is -0.025 ($p = 0.154$) for the co-partisan opponent CD and -0.045 ($p = 0.057$) for the group CD: the ask-tell intervention reduces the probability of demanding the maximum by an additional 2.5–4.5 percentage points for party minorities relative to majorities, though the effect is only marginally significant in one case.

Taken together, these results broadly corroborate the central conclusions of our paper. First, the ask-tell intervention reduces compensation demands for co-partisan conversation among even party majorities. This finding is robust even when accounting for ceiling effects in dollar terms: the Ask-Tell main effect is negative and significant under OLS and tobit alike. Second, we cannot reject modest heterogeneity between party majorities and minorities in the ask-tell effect. The difference becomes more stark but remains imprecisely estimated even when accounting for the scale maxima under a tobit model. Moreover, we find some suggestive evidence that the ask-tell effect was *more*, not less, likely to move party minorities off the ceiling than party majorities. This difference in CATEs might not overcome the larger share of party minorities hitting the ceiling at baseline but in tandem with the tobit results helps mitigate concern that the absence of starker heterogeneous effects is merely a measurement artifact.

A.3 Demographic Predictors of Prior Beliefs

Whether party majorities and party minorities hold systematically different prior beliefs about the prevalence of social sanctions for political speech affects the interpretation of our results. Substantive differences in priors across these groups would imply that the ask-tell intervention induces corrections of different magnitudes for party minorities versus party majorities, complicating interpretation of our CATEs. To assess this, we regress the prior (the mean of the three ask-tell items, scale 0–100) on party minority status and a vector of demographic covariates. To offer face validity regarding whether priors meaningfully capture respondents' experiences with or intuitions about political speech and censorship, we also assess how pre-treatment items measuring experiences with avoiding political conversation predict these priors. These items include "Censorship: Climate"—which measures

the extent to which the respondent agrees that the current political climate prevents them from saying things they believe—along with “Censorship: Co-Partisans” and “Censorship: Out-Partisans”, which measure the extent to which the respondent avoids conversation with certain co-partisans and out-partisans they know. All items were measured on 1–7 Likert scales. Question wording along with response distributions are available in Appendix A.9.

Notably, party minority status correlates meaningfully with these items. Table A3 shows that party minorities were more likely than party majorities to say they avoid discussing politics with some co-partisans they know (col. 3). They were correspondingly less likely to say they avoid discussing politics with some out-partisans they know (col. 4), and both of these results are robust to model specification (to the inclusion of party ID and measures capturing the share of respondents’ networks that are co-partisan opponents and out-partisans). The first column also suggests that party minorities were more likely to say the political climate these days censors them, but this difference is not robust to model specification. These results suggest that party minorities and majorities do have different experiences with avoiding political conversation among co-partisans. Does this then shape their ask-tell priors?

Table A3: Predictors of Pre-Treatment Items about Speech and Censorship

	Climate		Co-partisans		Out-partisans	
	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	3.974*** (0.015)	3.587*** (0.036)	3.758*** (0.016)	3.466*** (0.035)	5.276*** (0.014)	4.766*** (0.036)
Party Minority	0.108*** (0.031)	0.021 (0.032)	0.807*** (0.031)	0.714*** (0.032)	−0.286*** (0.030)	−0.369*** (0.031)
Democrat		−0.384*** (0.027)		−0.434*** (0.027)		0.182*** (0.025)
Network Co-Partisans		0.613*** (0.076)		1.336*** (0.079)		0.051 (0.075)
Network Out-Partisans		1.353*** (0.070)		0.721*** (0.070)		1.190*** (0.068)
Num.Obs.	17 672	17 595	17 647	17 574	17 655	17 580
R2	0.001	0.039	0.033	0.069	0.005	0.027

The first column of table A4 suggests that party minorities’ ask-tell priors were on average one point higher than those of party majorities on a 0–100 scale, a statistically significant but substantively small difference. The difference, however, is highly sensitive to model specification. Including demographic covariates as in columns 2–3 cuts the magnitude of the difference and renders it nonsignificant. We interpret this as suggesting little systematic difference in priors between party majorities and party minorities. They appear to enter the experiment with comparable perceptions about the prevalence of social sanctions.

We do detect statistically significant differences in priors along several other lines. Gender produces the largest effect: depending on specification, women report priors 2.60–3.25 points higher than men on the 0–100 scale ($p < 0.001$), indicating that women on average overestimate the prevalence of sanctions by a meaningful margin more than men. Political interest is the second largest source of variation: each one-point increase on the 5-point political interest scale is associated with a 2.54- to 2.57-point increase in priors ($p < 0.001$), so respondents at the top of the political-interest scale overestimate sanction prevalence by

Table A4: Predictors of Prior Beliefs about Social Sanctions

	(1)	(2)	(3)
	Basic Model	Demographics	All Covariates
(Intercept)	45.799*** (0.179)	54.640*** (0.806)	55.208*** (0.817)
Party Minority	0.948* (0.389)	−0.058 (0.383)	−0.440 (0.382)
Democrat		−1.241** (0.410)	−0.834* (0.405)
Female		3.254*** (0.311)	2.595*** (0.309)
White		−0.013 (0.367)	−0.897* (0.366)
Age		−0.354*** (0.009)	−0.345*** (0.009)
Education (std.)		0.126 (0.170)	−0.167 (0.168)
Ideology (std.)		0.276 (0.196)	0.128 (0.193)
Income		0.000 (0.000)	0.000 (0.000)
Political Interest		2.565*** (0.160)	2.540*** (0.160)
Censorship: Climate (std.)			2.816*** (0.165)
Censorship: Co-Partisans (std.)			0.839*** (0.167)
Censorship: Out-Partisans (std.)			1.502*** (0.177)
Num.Obs.	17 358	17 335	17 265
R2	0.000	0.091	0.124

roughly 10 points more than those at the bottom—an unsurprising pattern given that politically attentive respondents are most likely to encounter highly publicized cases of political backlash. Party identity also matters at the margin: Democrats report priors 0.83–1.24 points lower than Republicans on average. The difference between White and non-White respondents is less than one point across columns 2–3 and sensitive to specification. Older respondents hold somewhat more accurate priors ($b = -0.35$ per year, $p < 0.001$). Education, ideology, and income are not significant predictors.

The final three rows in column 3 of Table A4 use our pre-treatment measures of respondents’ self-perceptions regarding avoidance of political conversation. All items are presented here using standardized z-scores. Each of these items strongly predicts prior beliefs about social sanctions. A one SD increase in beliefs that the current political climate stymies their speech, for instance, corresponds in column 3 to a 2.82-point increase in the respondent’s prior. These results lend some face validity to our measures of social sanction perceptions.

A.4 Robustness to Alternate Specifications of Party Minority

Our primary definition of party minority status classifies respondents as minorities if they do not intend to vote for their party’s presidential nominee. This broad definition—which includes respondents who actively support an alternative candidate (the opposing party’s nominee, a third-party candidate, or another candidate) as well as those who report they would not vote or are unsure whom to support—was pre-registered because party minorities are scarce (21.25% of partisans in our sample) and a narrower definition would have limited the statistical power needed for the heterogeneity tests at the heart of the paper.

However, respondents categorized as minorities for different reasons may experience substantively different constraints on speech: a respondent who actively supports the opposing party’s nominee, for example, may face a different kind of dissent from peers than a respondent who simply prefers to remain politically uninvolved (Krupnikov and Ryan, 2022). To assess whether our results are sensitive to how minority status is defined, this section reports robustness checks against an alternate operationalization of minority status. We re-estimate our Table 2 CATE models after dropping respondents who reported they would not vote or were unsure. This approach excludes the disengaged but retains partisans who actively supported the opposing party’s nominee or a third-party candidate. We also, as an extension, examine how the ask-tell effect varies with measures of political interest.

Excluding the Disengaged We first re-estimate the Table 2 CATE models. The “Recode” columns in Table A5 recode those that reported they would not vote or were unsure whom to support as party majorities instead of party minorities, leaving a minority sample of $N = 2,316$ whose preferences differ definitively from their party’s nominee. The “Exclude” columns entirely drop these 1,443 respondents (from both the majority and minority groups).

Table A5: Table 2 CATEs Excluding Disengaged Minorities

	Co-Partisan Opp. CD			Co-Partisan Group CD		
	Original	Recode	Exclude	Original	Recode	Exclude
Ask-Tell	−3.706* (1.888)	−5.054** (1.819)	−3.695+ (1.888)	−6.418** (2.380)	−7.173** (2.308)	−6.266** (2.377)
Party Minority	21.652*** (3.181)	11.702** (3.739)	15.227*** (3.771)	10.938** (4.168)	10.919* (4.750)	11.440* (4.795)
Ask-Tell × Party Min.	−5.203 (4.363)	1.832 (5.301)	0.473 (5.326)	−7.213 (5.662)	−5.508 (6.724)	−6.499 (6.770)
Num.Obs.	11 043	11 043	10 164	5490	5490	5044

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Models include the full covariate vector and use HC2 robust SEs.

Results are substantively unchanged. The ask-tell treatment effects are nearly identical in magnitude and significance to the original. The baseline minority premium in compensation demands decreases somewhat for the one-on-one co-partisan opponent CD (from \$21.65 to \$11.70–\$15.23), suggesting that disengaged minorities demand more compensation at baseline, but active dissenters still demand significantly more than majority members. The interaction terms remain firmly null. Indeed, in the co-partisan opponent CD models, the

interaction effects attenuate toward zero, suggesting politically uninvolved respondents may be a primary cause of minority-majority asymmetry in responses to the ask-tell. In the group CD models they are similar to the original, though also attenuated. The inclusion of disengaged respondents does not drive our failure to detect statistically significant differences in CATEs; however, it could be that the modest interaction effects we do observe are partly driven by group composition and personality (i.e., minorities are simply more politically timid than majorities) rather than by normative standing.

Heterogeneity by Political Interest Next, we extend our assessment of how political interest, a strong demographic correlate of prior beliefs about social sanctions, shapes the ask-tell treatment effect itself. Our political interest measure is a 5-point scale ($SD = 1.1$). Theoretically, we might expect more politically interested respondents to be less responsive to the ask-tell treatment because they are more accustomed to disagreement.

Table A6: Ask-Tell Effects by Target and Political Interest

	All Targets	Discussants	Out-Partisans	Co-Ally	Co-Opponent
Ask-Tell Treatment	-10.435** (3.983)	-14.272** (5.392)	-5.949 (8.140)	-11.323 (7.163)	-16.627* (7.216)
Political Interest	-2.127** (0.803)	-2.615* (1.018)	-1.489 (1.568)	-1.401 (1.404)	-4.496** (1.419)
Ask-Tell:Interest	1.111 (1.088)	2.086 (1.416)	-0.194 (2.160)	1.041 (1.901)	3.189+ (1.900)
Num.Obs.	49 868	13 787	9135	9157	9113
R2	0.298	0.421	0.193	0.339	0.287
Std.Errors	by: id	by: id	by: id	by: id	by: id

The direction of effects is consistent with this intuition, though the evidence is weak. Across all targets, each one-point increase on the political interest scale attenuates the ask-tell effect by approximately \$1, but the interaction is not statistically significant when pooling across targets ($b = 1.11$, $p > 0.10$). The clearest evidence appears for compensation demands directed at co-partisan opponents, where the interaction reaches marginal significance ($b = 3.19$, $p < 0.10$): politically interested respondents are slightly less moved by the correction when imagining conversations with in-party opponents. With the interaction between treatment condition and party minority status, we cannot easily disentangle the roles of political views and of compositional differences; lower political interest could in part drive larger ask-tell effects among minorities.

A.5 Do Meta-Perceptions of Social Pressure Reflect Genuine Misperceptions?

A common concern in studies of meta-perceptions (perceptions about the opinions or experiences of others) is whether these are genuine beliefs or unstable, “doorstep” survey responses (Dias et al., 2024). The extent to which respondents’ priors consist of genuine misperceptions versus mere survey responses is important, both intrinsically and to better understand the mechanism behind our intervention. We should distinguish upfront between two distinct concerns about measures of meta-perceptions. The first is a concern about *non-attitudes*: because respondents are unlikely to have given consideration to the exact percentage of Americans who experienced a specific social sanction for speech, they may respond randomly or expressively. A second concern is about granularity: perhaps our measures meaningfully capture variation in respondents’ general intuitions about the current political climate, but a 0–100 scale is simply too granular.

We posit that the beliefs investigated in our ask-tell are likely closer to genuine beliefs for two reasons. First, beliefs about social sanctions are of far more proximate day-to-day concern to individuals than concerns about opposing partisans’ positions on democratic norms and political violence (e.g., Mernyk et al., 2022; Braley et al., 2023).

Second, we offer evidence from the beliefs themselves. While we cannot measure the stability of respondents’ meta-perceptions without access to panel data, we can examine these concerns using other methods. In Appendix A.3, we show that pre-treatment survey questions measuring respondents’ stated likelihood of avoiding political conversation with certain subsets of their personal networks strongly predict respondents’ meta-perception priors, offering suggestive evidence that these are not merely non-attitudes. This, however, can be true even if there remains a subgroup in our sample who responded randomly or expressively. To better disentangle the non-attitudes and granularity critiques, we explore the volume and nature of anchoring—pooling at specific points—in the distribution.

If respondents’ priors are largely non-attitudes, then we should expect severe pooling at the most ambivalent (namely, 50%) and/or most extreme (0% and 100%) response options. On the other hand, if respondents are mostly giving considered responses but their intuitions are not granular to the degree implied by our 0–100 scale, then we might observe anchoring also at less granular intervals, e.g., in increments of 5% or 10%. We assess the degree of anchoring both in absolute terms and by benchmarking our results against three prominent studies that examine perceptions about the opinions or traits of opposing party voters using 0–100 scales. Results are summarized in Table A7. We do not show the full set of histograms due to space constraints but visualize four example items from each paper in Figures A2–A4 as a comparison for Figure 3 in the main text.

As Figure 3 and Table A7 suggest, pooling at the 50% midpoint was more common across all three items in our study than at 0% or 100%. However, the share of respondents answering 50% in our study is comparable to the share in our three benchmark studies. This share is highest for our question about the percentage of Americans saying that political views reveal one’s character, but this remains meaningfully smaller than the share answering 50% for several items studied by Ahler and Sood (2018). The share of respondents in our study who responded with 0% and 100% is also meaningfully smaller than in either Mernyk et al. (2022) or Dias et al. (2024), suggesting a relatively low amount of expressive responding in our study. Turning to the “% Intervals of 5” column, our respondents were far more likely to anchor at other multiples of 5 than at 0%, 50%, or 100%, consistent with the “granularity” view. The share of responses at these intervals is higher than for Mernyk et al. (2022) and

comparable to Dias et al. (2024) but substantially lower than Ahler and Sood (2018).

These results offer speculative evidence that the perceptions measured by our ask-tell design capture coarse but meaningful intuitions about the current political climate. We emphasize, however, that much remains ambiguous in this exploratory analysis. At a glance, the distributions in Figure 3 appear slightly more uniform than the distributions in the three benchmark studies, perhaps suggesting ambiguity in responses. Yet there is meaningful variation between our items (with a notable right skew in the workplace penalties distribution) and it is difficult to know *ex ante* how uniformity and instability correlate. We encourage future work to assess the stability of (mis)perceptions about social sanctions for speech.

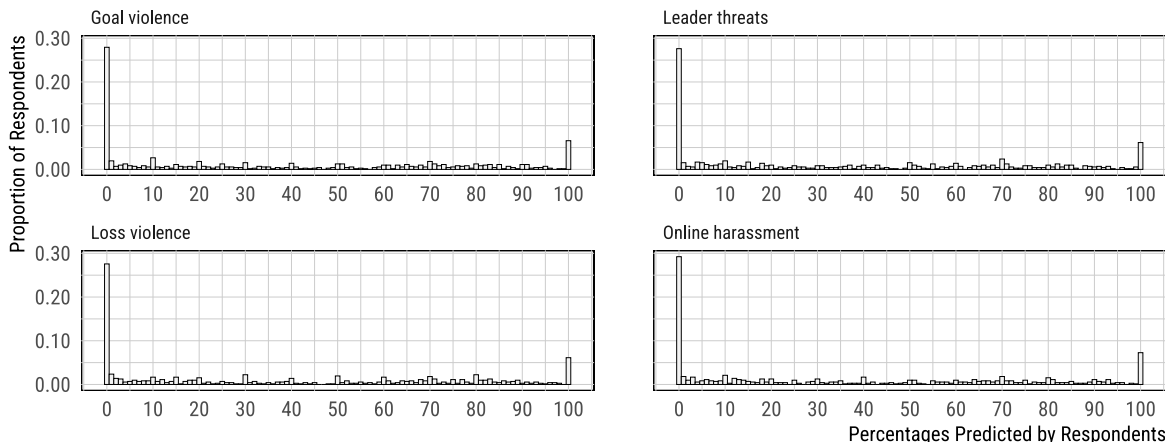


Figure A2: Mernyk et al. (2022), Study 1 Meta-Perceptions

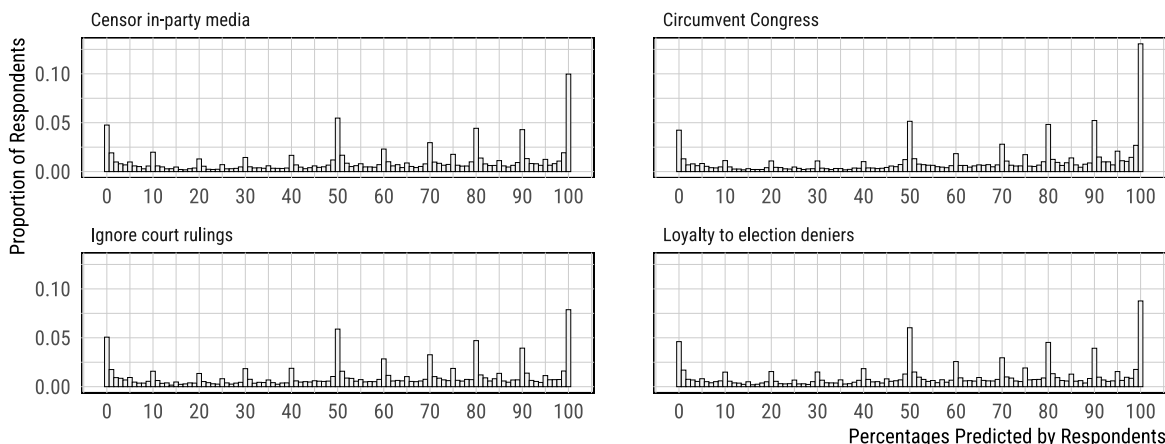


Figure A3: Dias et al. (2024), Sample of Panel Wave 1 Meta-Perceptions

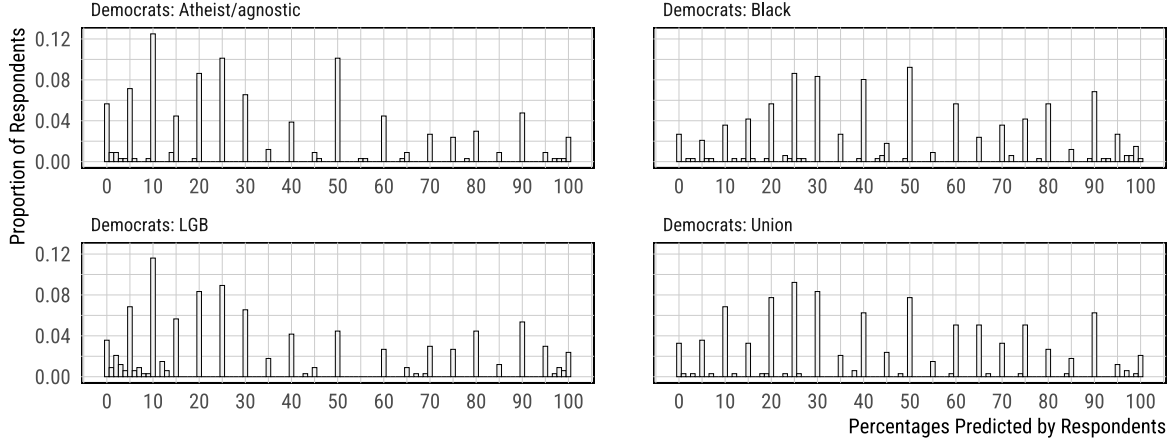


Figure A4: Ahler and Sood (2018), Republican Perceptions of Opposing Democratic Party Voters

Table A7: Benchmarking Anchoring Against Other Studies

Item	% at 0	% at 50	% at 100	% at {0,50,100}	% Intervals of 5
Workplace penalties	3.8	5.1	2.0	10.9	35.2
Damaged relationships	2.6	6.4	3.3	12.4	32.2
Politics reveals character	4.9	7.8	3.6	16.4	28.2
<i>Mernyk et al. (2022; Averaging Across Four Studies)</i>					
Online harassment	21.2	4.5	7.1	32.9	23.6
Leader threats	22.2	5.1	5.8	33.0	23.9
Goal violence	23.4	4.6	6.3	34.3	22.9
Loss violence	23.6	4.9	6.2	34.7	22.2
<i>Dias et al. (2024; Panel Survey, Wave 1)</i>					
Throw rocks at in-partisans	3.1	5.5	7.1	15.7	35.7
Stab prominent in-partisan	6.6	4.9	5.3	16.8	31.4
Ignore court rulings	5.1	5.9	7.9	18.8	30.3
Loyalty to election deniers	4.6	6.0	8.8	19.4	29.5
Reduce polling stations	6.1	5.6	8.4	20.2	29.6
Censor in-party media	4.8	5.5	10.0	20.2	28.8
Circumvent Congress	4.2	5.1	13.1	22.4	27.3
Punish in-party companies	14.2	4.6	3.6	22.6	22.5
<i>Ahler & Sood (2018)</i>					
Democrats: LGB	3.6	4.5	2.4	10.4	78.0
Democrats: Black	2.7	9.2	0.3	12.2	78.0
Democrats: Union	3.3	7.7	2.1	13.1	81.5
Republicans: Evangelical	6.2	6.6	2.5	15.3	78.3
Republicans: Southern	4.8	10.0	2.5	17.4	76.3
Republicans: 65+	3.2	12.6	2.3	18.0	75.3
Democrats: Atheist/agnostic	5.7	10.1	2.4	18.2	75.6
Republicans: \$250K+	3.7	8.0	7.8	19.4	69.4

Note: Percentages computed on non-missing responses to each individual slider item. “% Intervals of 5” is the share of responses landing on $\{5, 10, 15, \dots, 45, 55, 60, \dots, 95\}$, i.e. multiples of 5 other than 0, 50, and 100.

A.6 Alternative Mechanism Tests

While our mechanism checks in the main text indicate that the ask-tell intervention successfully reduced fears about social sanctions from one’s peers, in this section we assess plausible alternative mechanisms. We consider two such accounts. The first is that the ask-tell could move our main CD outcomes not by reducing preexisting fears about social sanction but instead by making the case for a civic norm for respondents to participate in political conversations as a good citizen. That control respondents were shown all the same stimuli besides the true information suggests any priming should be mitigated by the design, though we cannot entirely rule out that the “tell” component of the design makes a stronger case for this type of norm. A second alternative mechanism would be if, in the process of altering others’ perceived tolerance toward political disagreement, the ask-tell treatment not only made potential discussants seem less likely to impose social sanctions, but also made them appear as more desirable conversation partners (i.e., by making them seem more similar in disposition or ideology). This would mean the treatment not only lowers perceived affiliative costs but concurrently raises perceived affiliative or instrumental benefits.

To assess alternative mechanisms, we take advantage of the fact that our module was embedded in a larger CHIP50 Wave 32 survey that included other modules. We use a series of separate regressions (with our pre-registered covariate vector) to assess how the ask-tell affected relevant items asked after our module. Of the available items, there are two batteries for self-rated morality and out-group morality (specifically, the morality of those in some opposing *religious* group), which each include four positive traits (Trustworthy, Fair, Respectful, Principled) and four negative traits (Insincere, Prejudiced, Manipulative, Deceptive). The self-rated morality battery may capture change in respondents’ beliefs about their own adherence to pro-social norms, allowing us to test the first hypothesized mechanism. The out-group battery can be divided into positive and negative items. A reduction in perceived negative traits would be consistent with our intended mechanism of reducing the perceived censoriousness of peers and therefore the risks associated with political conversation. However, changes in both the perceived positive and negative traits could also be consistent with the alternative mechanism, suggesting the ask-tell supplies newfound positive benefits for speech. The out-group battery thus supplies only an imperfect test, owing both to ambiguous implications of each item and to the focus on opposing religionists. The final post-module items we test for include social support and indicators of depression, stress, and irritability. Changes in these mental health indicators could corroborate our theorized mechanism that the treatment mitigates the anxiety respondents face when contemplating stressful political conversations.

Table A8 displays the ask-tell treatment effect on each item along with the standardized z-score in the rightmost column. We find little evidence of change in self-rated morality with the sole exception of self-perceived manipulative behavior. Generally, perceptions about one’s own character or adherence to social norms seem unfazed by our treatment. In addition, we find small and nonsignificant effects for all four positive out-group morality items (with a marginally significant effect at the $p < 0.1$ level for Respectful). By contrast, the treatment exhibited substantial effects on all four out-group morality items, in all cases reducing negative attitudes toward opposing religionists’ character. At the same time, we observe meaningful effects on various depression, stress, and irritability indicators, suggesting the treatment plausibly delivered temporary relief toward generalized anxiety.

To put these effects in perspective, Figure A5 displays the ask-tell treatment effect on

our own mechanism checks alongside the statistically significant effects on these post-module outcomes (along with a partial sample of the null items). All effects are standardized using the control group SD. While the effects on out-group morality items are larger than for any of the mental health indicators or self-rated manipulative behavior, the out-group morality effects are substantially smaller than the effects on our own mechanism checks. The ask-tell produced twice as large of an effect on the perceived intolerance toward speech of “people you know” (our main mechanism check) than on perceived out-group prejudice (the out-group morality item with the largest effect)—and three times as large of an effect on the perceived intolerance of “all Americans.” These results suggest that although our treatment could plausibly bundle more than one mechanism, our main effects are primarily driven by reducing perceived social pressure risks from one’s personal network (in part by reducing the perceived censoriousness of peers).

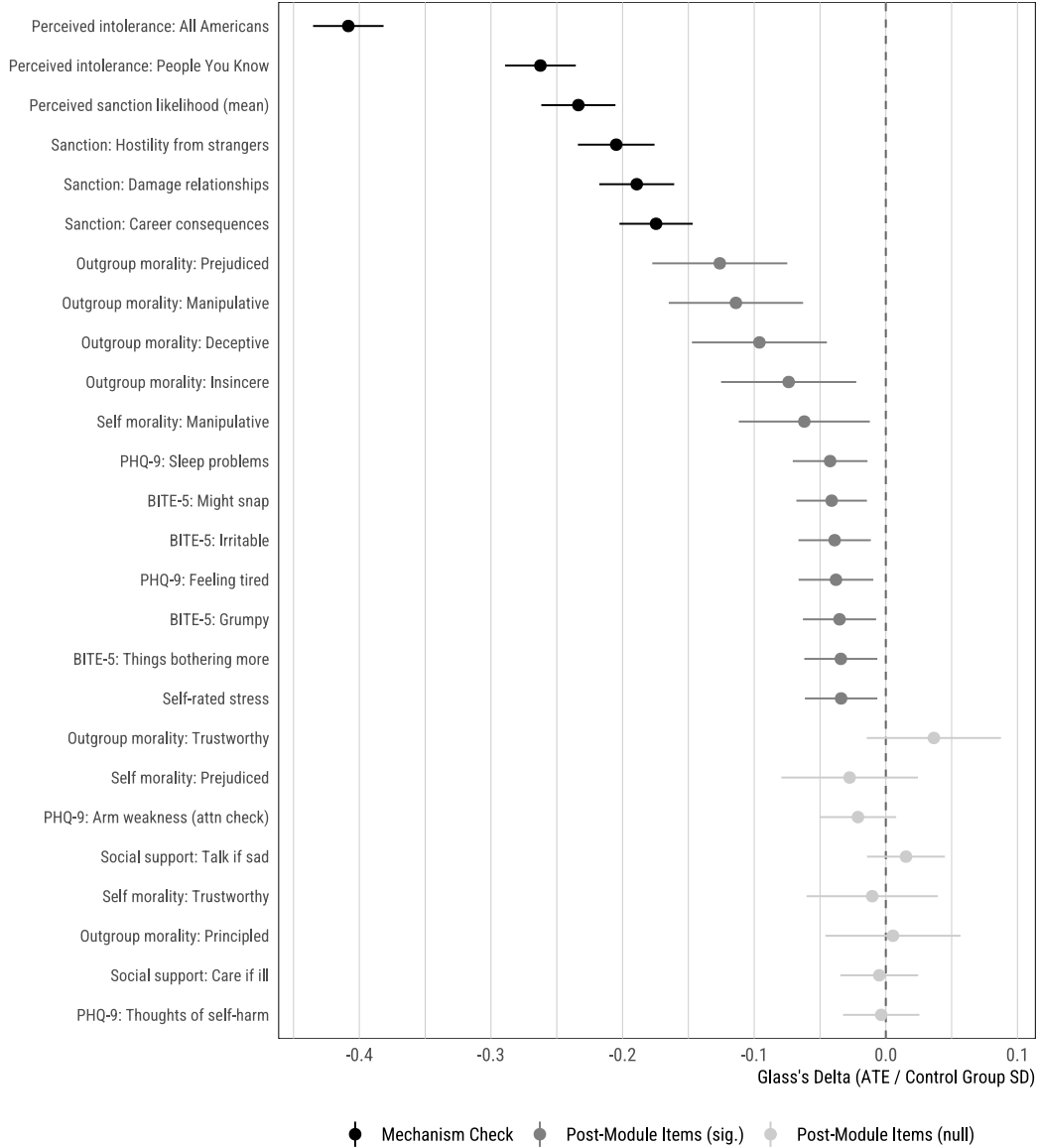


Figure A5: Benchmarking Our Mechanism Checks Against Additional Post-Module Outcomes

Table A8: Ask-Tell Treatment Effects on Additional Post-Module CHIP50 Outcomes

Category	Outcome	Est.	SE	<i>p</i>	<i>N</i>	Std. Est.
Self-Rated	Self: Trustworthy	-0.189	(0.465)	0.685	5,848	-0.010
Morality	Self: Fair	-0.246	(0.508)	0.628	5,848	-0.012
	Self: Respectful	0.211	(0.460)	0.646	5,852	0.012
	Self: Principled	0.011	(0.620)	0.986	5,804	0.000
	Self: Insincere	-1.094	(0.601)	0.069	5,324	-0.047
	Self: Prejudiced	-0.673	(0.647)	0.298	5,384	-0.028
	Self: Manipulative	-1.433*	(0.588)	0.015	5,364	-0.063
	Self: Deceptive	-0.867	(0.579)	0.135	5,329	-0.039
Outgroup	Outgroup: Trustworthy	0.994	(0.708)	0.161	5,707	0.037
Morality	Outgroup: Fair	0.967	(0.693)	0.163	5,716	0.037
	Outgroup: Respectful	1.251	(0.705)	0.076	5,720	0.047
	Outgroup: Principled	0.146	(0.703)	0.836	5,699	0.005
	Outgroup: Insincere	-2.121**	(0.753)	0.005	5,623	-0.074
	Outgroup: Prejudiced	-3.759***	(0.780)	<0.001	5,641	-0.127
	Outgroup: Manipulative	-3.375***	(0.771)	<0.001	5,638	-0.114
	Outgroup: Deceptive	-2.789***	(0.760)	<0.001	5,627	-0.096
Depression (PHQ-9)	Little interest/pleasure	-0.023	(0.013)	0.077	17,052	-0.026
	Nervous/anxious	-0.015	(0.013)	0.263	17,036	-0.016
	Uncontrolled worry	-0.017	(0.013)	0.198	17,017	-0.019
	Weakness in arm (attn. check)	-0.013	(0.009)	0.153	16,985	-0.022
	Feeling down/depressed	-0.023	(0.013)	0.078	17,036	-0.026
	Sleep problems	-0.043**	(0.015)	0.003	17,046	-0.043
	Feeling tired	-0.037**	(0.014)	0.009	17,038	-0.038
	Appetite problems	-0.020	(0.013)	0.127	17,001	-0.022
	Feeling bad about self	-0.016	(0.013)	0.233	17,031	-0.017
	Trouble concentrating	-0.009	(0.012)	0.461	17,037	-0.011
	Moving slowly/fidgety	-0.019	(0.010)	0.056	17,020	-0.028
	Thoughts of self-harm	-0.002	(0.010)	0.814	17,028	-0.003
Stress	Self-rated stress	-0.089*	(0.037)	0.016	16,875	-0.034
Irritability (BITE-5)	Grumpy	-0.040*	(0.016)	0.013	17,065	-0.035
	Feeling like might snap	-0.053**	(0.018)	0.003	17,055	-0.042
	Others getting on nerves	-0.036	(0.018)	0.052	17,051	-0.028
	Bothered more than normal	-0.046*	(0.019)	0.016	17,055	-0.035
	Feeling irritable	-0.051**	(0.018)	0.006	17,032	-0.039
Social Support	Care if ill	-0.014	(0.043)	0.742	17,015	-0.005
	Lend money	-0.003	(0.037)	0.931	16,992	-0.001
	Talk to if sad	0.048	(0.047)	0.309	16,972	0.015
	Help find job	0.066	(0.048)	0.168	16,975	0.021

Note: Separate regressions. Std. Est. = coefficient divided by outcome SD. ****p* < 0.001; ***p* < 0.01; **p* < 0.05.

A.7 Robustness to Network Missingness

Some respondents may not know or may not be able to think of peers who are co-partisan opponents. We assess the consequences of this missingness using our pre-treatment network measures, which allow us to measure the self-reported share of respondents' networks comprising co-partisan opponents and out-partisans. 18.23% of partisan respondents in our sample reported not knowing any co-partisan opponents, with 20.99% of party majorities saying they did not know a co-partisan hesitant to vote for the party nominee and 7.99% of party minorities saying they did not know a co-partisan set on voting for the party nominee. By contrast, 9.19% of partisan respondents overall said they did not know a single out-partisan.

The first row of Table A9 suggests that excluding respondents who claim to know zero co-partisan opponents increases, not decreases, the estimated main effect of the ask-tell among party majorities. Meanwhile, excluding these respondents does not alter our conclusions about the heterogeneity between party majorities and minorities. While the magnitude of the interaction effect increases for the co-partisan opponent CD, the effect in column 3 does not reach statistical significance. In addition, the interaction effect for the group CD meaningfully shrinks toward zero (column 4). On the whole, not only are our results robust to this alternative specification, but excluding those respondents who fail to call to mind co-partisan opponents somewhat strengthens our initial conclusions.

Table A9: Ask-Tell Effects Excluding Those Without Co-Partisan Opponents in Network

	Full Sample		Restricted	
	Co-Partisan Opp. (1)	Group (2)	Co-Partisan Opp. (3)	Group (4)
Ask-Tell	-3.706* (1.888)	-6.418** (2.380)	-4.924* (2.067)	-8.103** (2.607)
Party Minority	21.652*** (3.181)	10.938** (4.168)	22.124*** (3.289)	8.016+ (4.326)
Ask-Tell $\times$ Party Minority	-5.203 (4.363)	-7.213 (5.662)	-6.999 (4.551)	-2.864 (5.897)
Num.Obs.	11 043	5490	9092	4497
R2	0.283	0.401	0.284	0.400

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.8 Sample Composition and Balance Tables

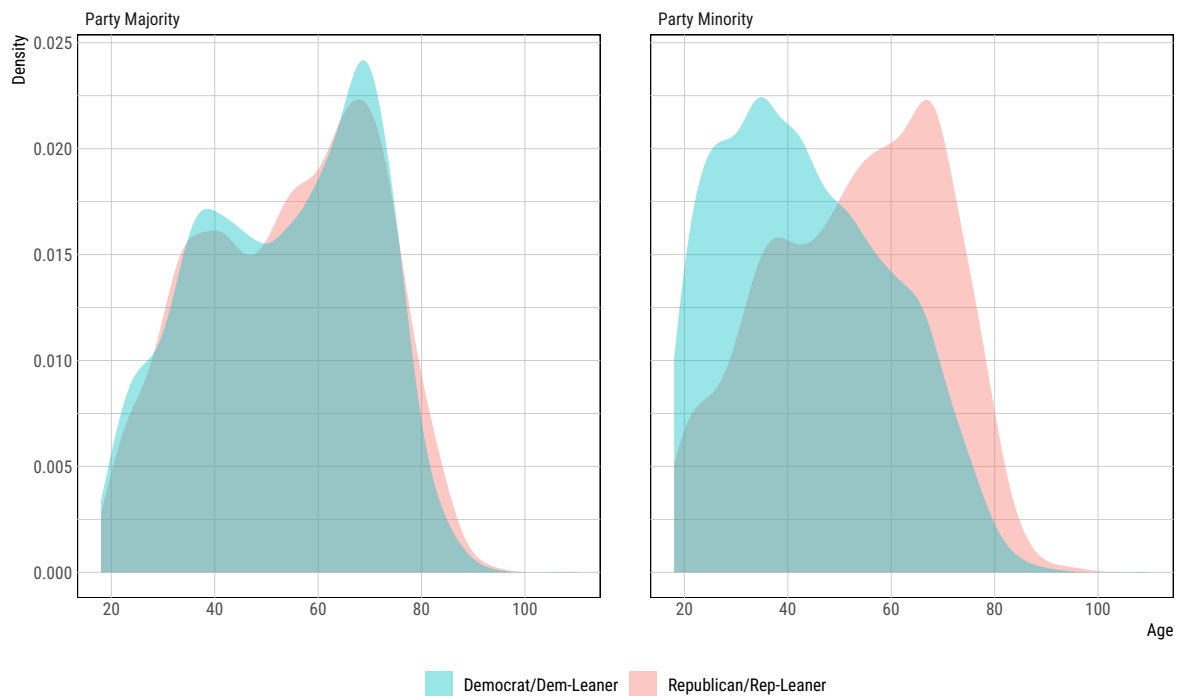


Figure A6: Density of Age by Party Minority Status

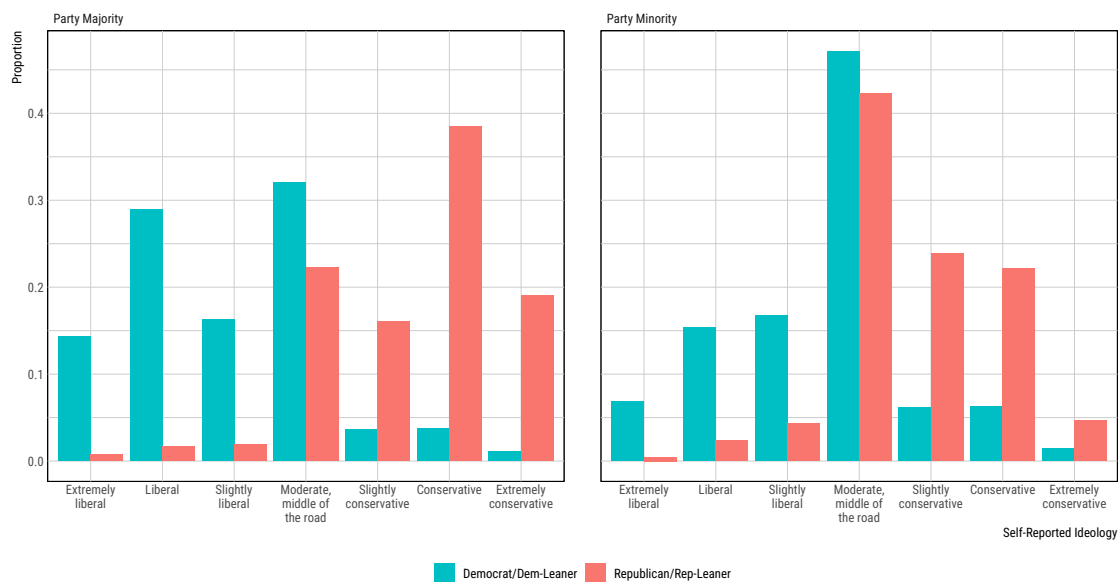


Figure A7: Self-Reported Ideological Identification by Party Minority Status

Here, we explore balance across pre-registered covariates for all four randomizations. Of note, the “target” randomization includes 3 observations per individual so there is overlap across the assignments.

Table A10: Balance Tables Across Targets

	discussant (N=21400)		coop (N=14237)		coally (N=14300)		out (N=14263)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
age	51.6	17.1	51.5	17.0	51.6	17.0	51.7	17.1
ask_1	38.2	25.5	38.2	25.6	38.3	25.5	38.0	25.4
ask_2	51.3	26.2	51.3	26.3	51.3	26.2	51.3	26.0
ask_3	47.6	27.4	47.5	27.5	47.6	27.4	47.7	27.3
dem	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
education	3.3	1.1	3.3	1.1	3.3	1.1	3.3	1.1
race_white	0.7	0.5	0.7	0.5	0.7	0.5	0.7	0.5

Table A11: Balance Tables for Ask-Tell Randomization

	control (N=10717)		treat (N=10683)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	Std. Error
age	51.8	17.1	51.4	17.0	-0.4	0.2
ask_1	38.0	25.4	38.3	25.6	0.3	0.4
ask_2	51.4	26.1	51.2	26.2	-0.1	0.4
ask_3	47.4	27.3	47.8	27.5	0.4	0.4
dem	0.5	0.5	0.5	0.5	0.0	0.0
education	3.3	1.1	3.3	1.1	0.0	0.0
race_white	0.7	0.5	0.7	0.5	0.0	0.0

Table A12: Balance Tables for Conversation Topic Randomization

	Econ issues (N=5774)		Social issues (N=5833)		2024 Election (N=5751)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
age	52.4	17.2	52.3	17.1	52.4	17.1
ask_1	38.1	25.3	38.0	25.4	37.9	25.6
ask_2	51.2	25.8	52.2	25.9	51.9	26.2
ask_3	47.9	27.4	48.6	27.4	48.2	27.3
dem	0.5	0.5	0.5	0.5	0.5	0.5
education	3.4	1.1	3.4	1.0	3.4	1.0
race_white	0.7	0.5	0.7	0.5	0.7	0.5
party_minority	0.2	0.4	0.2	0.4	0.2	0.4
turnout	0.0	0.2	0.0	0.2	0.0	0.1
interest	3.5	1.1	3.5	1.1	3.4	1.1

A.9 Reexamining Surveys about Self-Censorship

Respondents were asked to rate their agreement toward three statements. The wordings come from two public survey reports which provide conflicting accounts about the degree to which Democrats and Republicans asymmetrically engage in self-censorship and avoid political discussion. The first pair resembles survey language used by APM Research Lab ([link](#)), while the last statement mirrors survey language used by the Cato Institute ([link](#)).

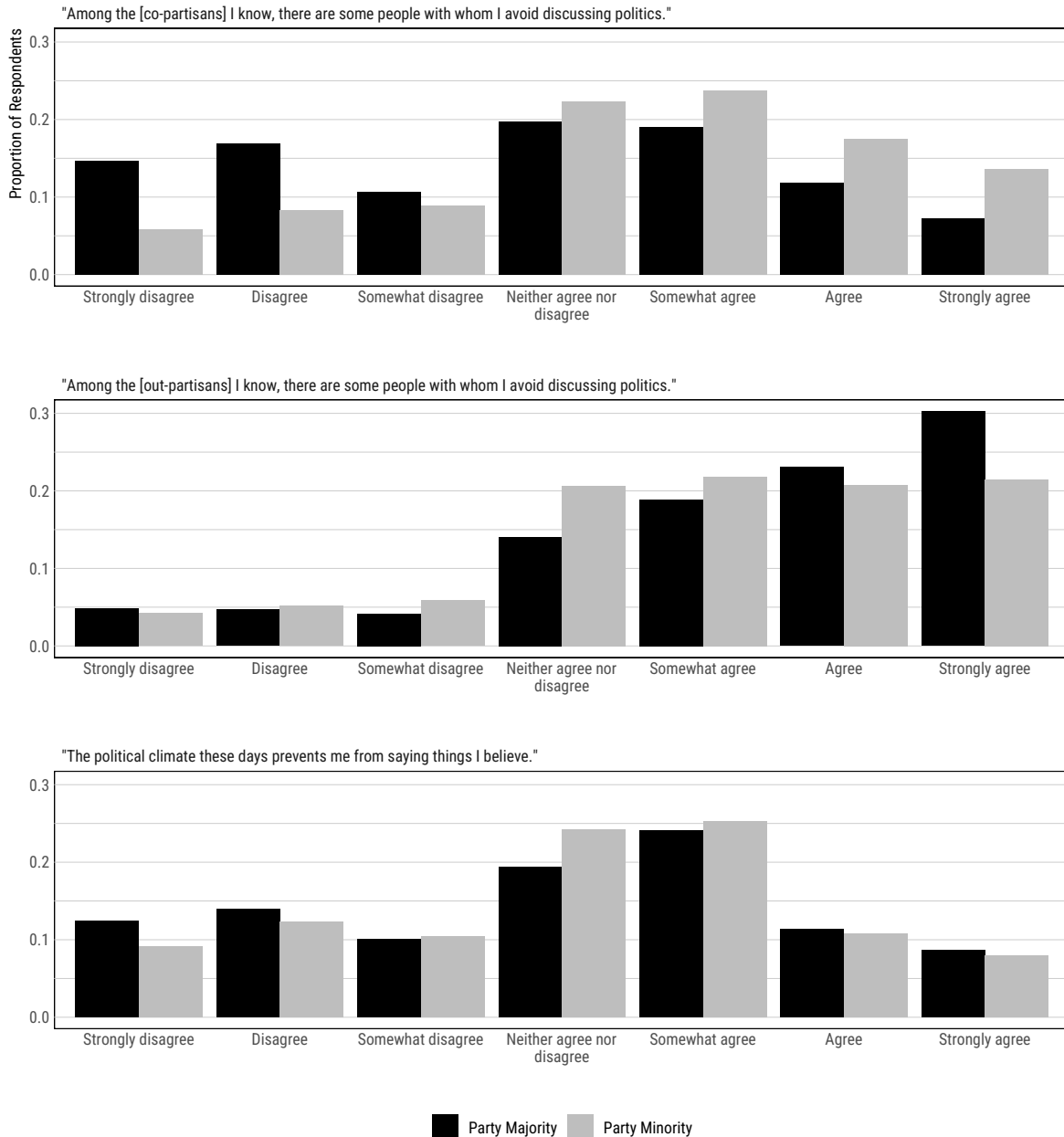


Figure A8: APM and Cato Survey Questions by Party Minority Status

Figure A8 shows 41.5% of respondents agreed (i.e., at least somewhat agreed) that they avoid discussing politics with some co-partisans they know. 54.5% of party minorities agreed but only 37.9% of party majorities. By contrast, 70.3% in total agreed they avoid discussing politics with some out-partisans, including 63.9% of minorities and 72.1% of majorities. Figure A9 shows that 45.3% of Republicans agreed they avoid discussing politics with some co-partisans, compared to 36.3% of Democrats. Meanwhile, 68.5% of Republicans said they avoid discussing politics with some out-partisans, similar to the 70.6% of Democrats who agreed.

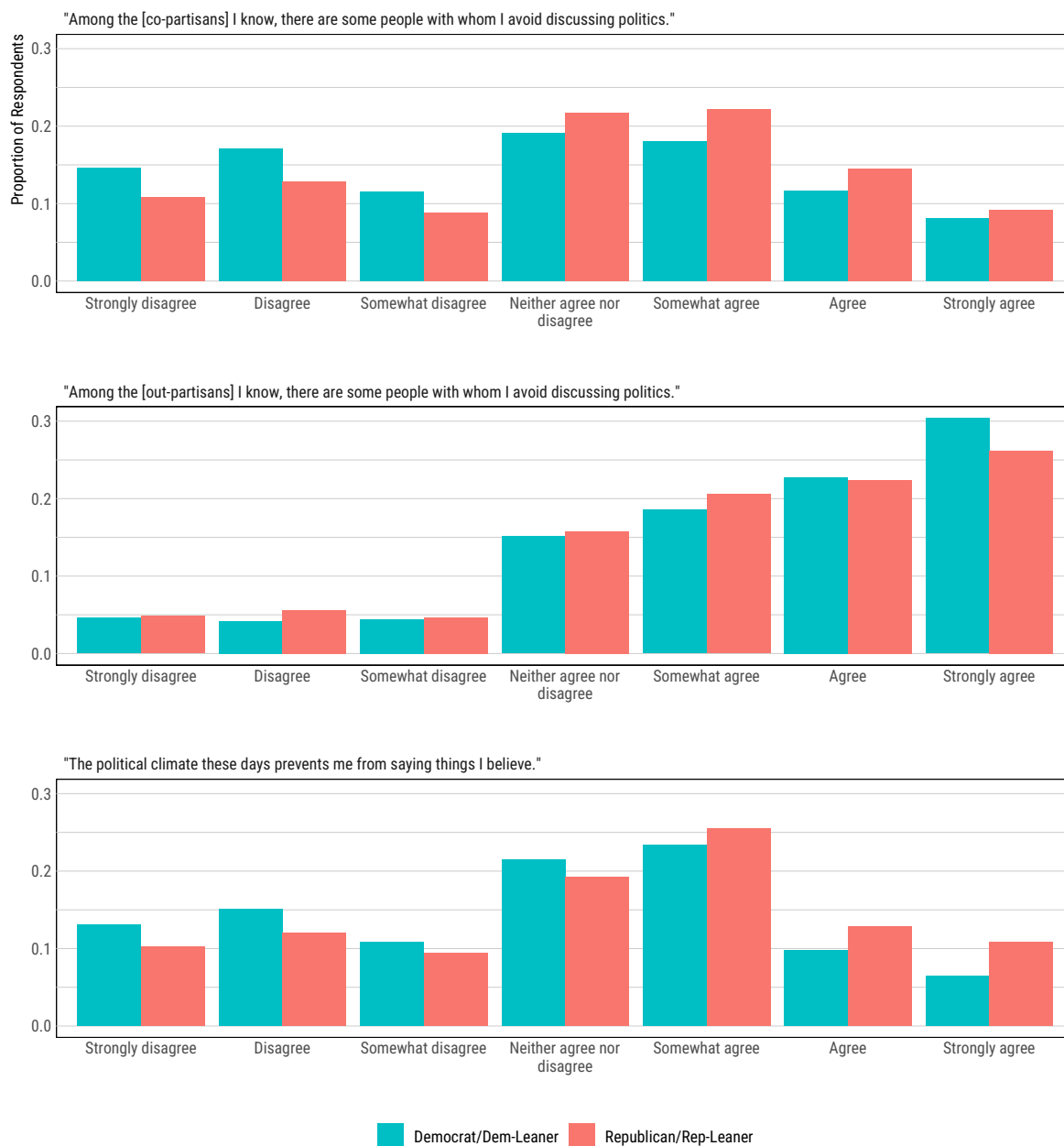


Figure A9: APM and Cato Survey Questions by Party

A.10 Placebo and Spillover Tests

Does the presence of a target affect the CD for the other target? To test this within-subject stable-unit-treatment value assumption (which is that the potential outcomes of observation c for individual i are unchanged for different assignments to the same individual for observations $c + 1$ and $c + 2$) we compared CDs for each of the three categories as follows: does the CD of Target 1 differ if the other randomly assigned Target is Target 2 compared to Target 3? We present these results for each of the three randomly assigned targets. We can see that there is no statistically or substantively significant violation of our assumptions. In Table A13, the baseline category for the omitted variables is “co-partisan ally” and the baseline target is the frequent discussant. The first two coefficients show the change in the “frequent discussant” CD when the co-opp and out-partisan targets are omitted while the next three interaction terms show how the treatment effect of those targets is modified depending on the omitted category.

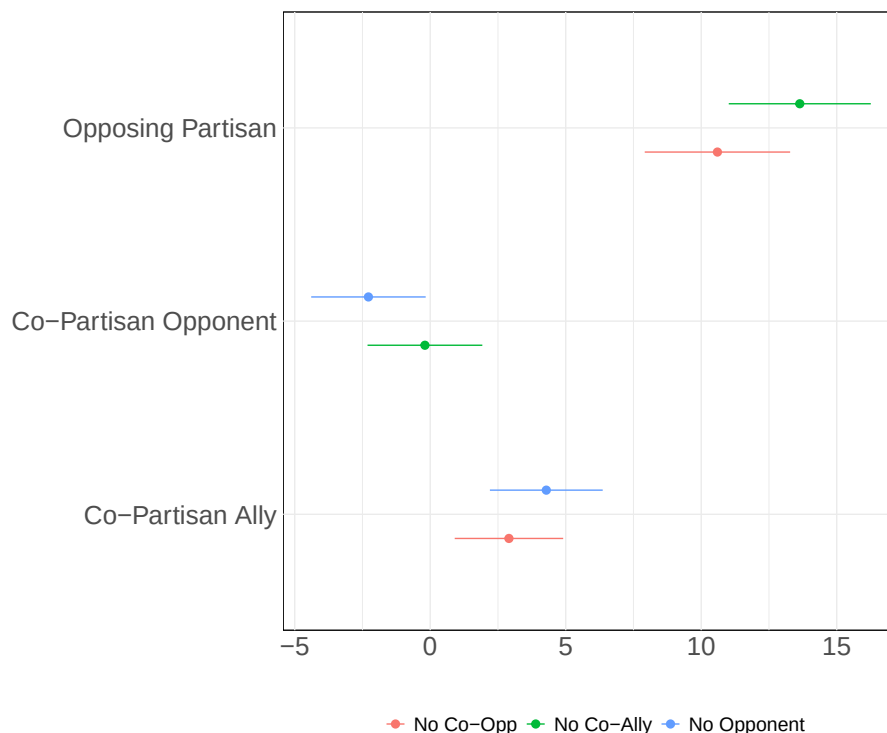


Figure A10: Comparison of Target Effects by Omitted Category

Table A13: Spillover Effects Model

	(1)
Omit Co-opp	1.587 (1.557)
Omit Out	2.644 ⁺ (1.562)
Coally:Omit Co-Opp (baseline = Omit Out)	-1.379 (1.473)
Opponent:Co-Opp (baseline = Omit Co-Ally)	-3.078 (1.913)
Coally:Co-Opp (baseline = Omit Co-Ally)	-2.094 (1.526)
Num. Obs.	49 868
R^2	0.300
Std. Errors	by: id

A.11 Differences in Ask-Tell Asymmetry by Party

Per our pre-registered H5B, we tested if the ask-tell CATEs on prices demanded for co-partisan conversation are larger for party minorities than for party majorities among Republicans specifically, as Republicans were the subject of most reports about intra-party censorship during the 2024 campaign. The effect is in the right direction: Table A14 shows that, for both outcomes, the difference in CATEs between party majorities and minorities is smaller for Democrats than for Republicans. But here, we again fail to reject the null of no difference.

Table A14: Does Majority/Minority Asymmetry in Ask-Tell Effects Vary by Party?

	Co-Partisan Opp. CD	Co-Partisan Group CD
Ask-Tell Treatment	-4.549 (2.786)	-4.861* (1.946)
Party Minority	28.131*** (5.021)	14.830*** (3.542)
Democrat	-9.135** (3.197)	6.097* (2.374)
Ask-Tell:Party Min	-7.159 (6.999)	-6.260 (5.044)
Ask-Tell:Dem	1.459 (3.787)	-4.158 (2.773)
Party Min:Dem	-11.354+ (6.453)	-10.160* (4.640)
Num.Obs.	11 043	16 661
R2	0.285	0.402

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Full Models in Table B3.

A.12 Additional Exploratory Analyses

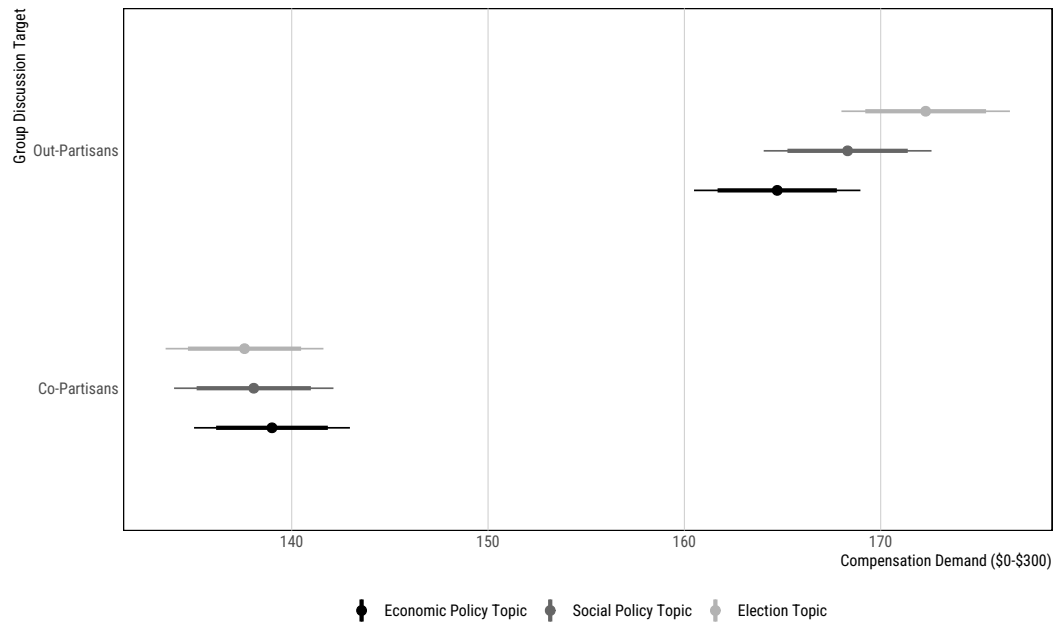


Figure A11: Group Conversation CDs by Topic

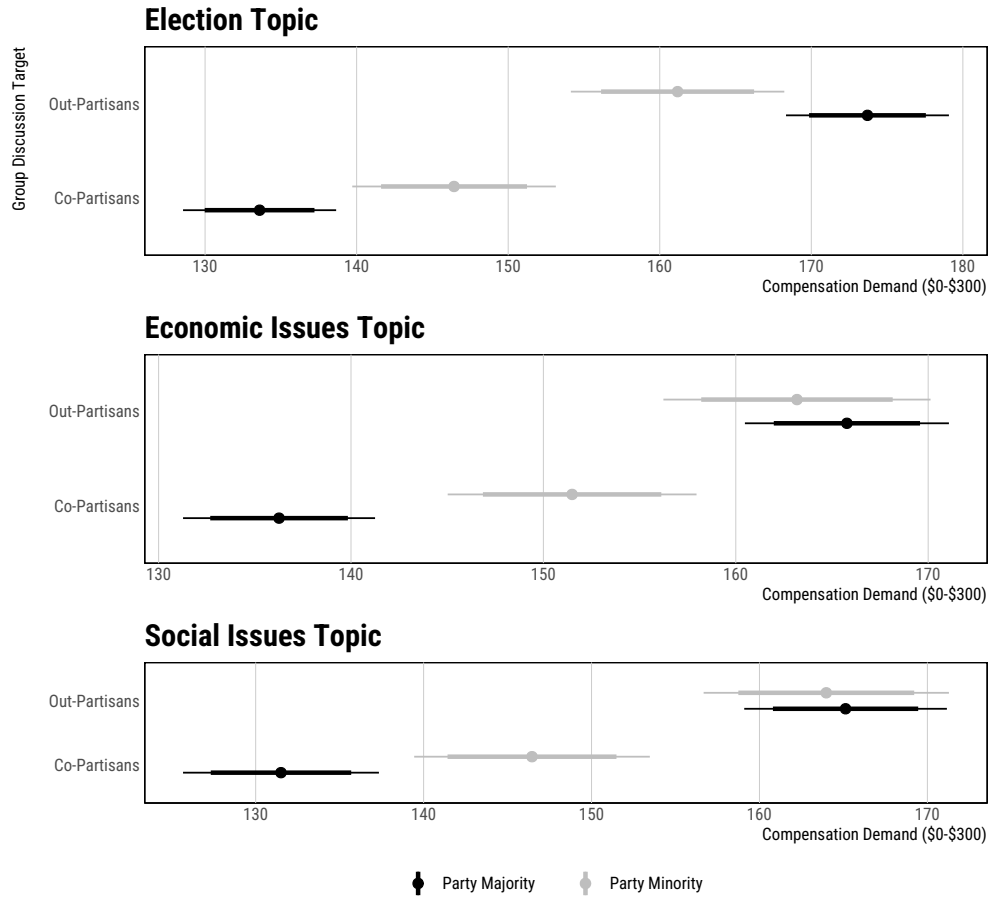


Figure A12: Group Conversation CDs by Topic and Party Majority/Minority

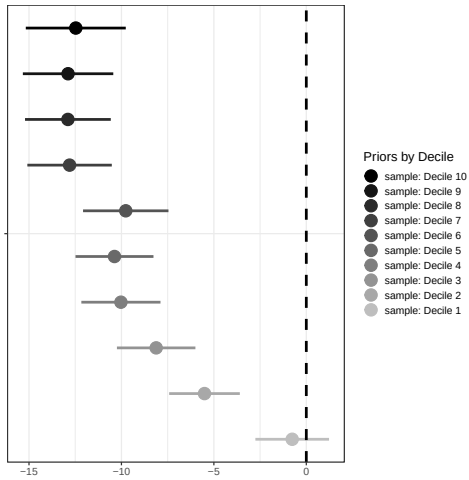


Figure A13: CATEs by Priors: Manipulation Index

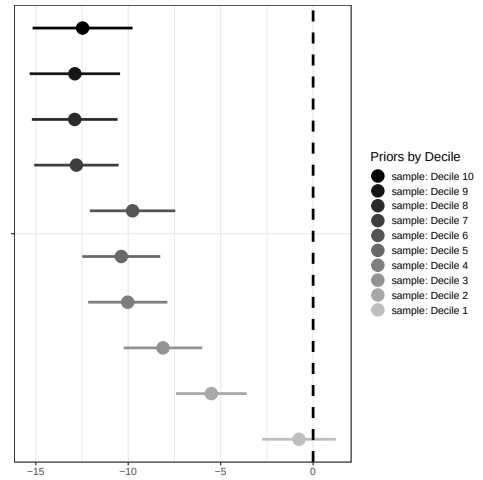


Figure A14: CATEs by Priors: Main Outcome Measure

B Full Models

We begin by showing full models, with all covariates, for all models used in the main analysis.

Table B1: Full Models, ATE on Mechanism Check Results with Interactions by Prior (Continuous) (Table 1)

	“All Americans”		“People you know”		Sanction Mean
	(1) H2a	(2) H2c	(3) H2a	(4) H2c	(5) H2b
Intercept	21.205*** (1.323)	16.625*** (1.380)	13.747*** (1.359)	10.622*** (1.417)	2.879*** (0.072)
Ask-Tell	-11.455*** (0.383)	-2.368** (0.912)	-7.643*** (0.399)	-1.490 (0.909)	-0.337*** (0.021)
Dem	2.603*** (0.522)	2.640*** (0.521)	3.275*** (0.533)	3.305*** (0.532)	-0.033 (0.029)
Demand Base	0.021*** (0.002)	0.021*** (0.002)	0.024*** (0.002)	0.024*** (0.002)	0.001*** (0.000)
Party Minority	-3.140*** (0.486)	-3.108*** (0.484)	-3.632*** (0.506)	-3.609*** (0.505)	0.060* (0.027)
Ideology	-0.532* (0.251)	-0.529* (0.250)	-1.002*** (0.261)	-0.996*** (0.260)	0.047*** (0.014)
Pol. Int. (5-point)	1.355*** (0.204)	1.370*** (0.203)	1.466*** (0.210)	1.480*** (0.210)	0.009 (0.011)
Education	-0.106 (0.216)	-0.107 (0.215)	-0.249 (0.223)	-0.252 (0.223)	0.066*** (0.012)
Prior	0.436*** (0.011)	0.536*** (0.014)	0.454*** (0.011)	0.522*** (0.015)	0.015*** (0.001)
Network Co-Opp	8.172*** (1.132)	8.148*** (1.134)	9.046*** (1.188)	9.024*** (1.188)	0.326*** (0.062)
Network Opposing	1.370 (1.045)	1.398 (1.042)	1.730 (1.098)	1.736 (1.097)	0.927*** (0.056)
Female	-2.791*** (0.397)	-2.830*** (0.396)	-2.307*** (0.415)	-2.334*** (0.415)	0.099*** (0.022)
White	0.419 (0.475)	0.418 (0.473)	-0.346 (0.496)	-0.340 (0.495)	0.052* (0.025)
Married	0.568 (0.425)	0.574 (0.424)	0.084 (0.440)	0.092 (0.439)	0.054* (0.023)
Fully Employed	-0.162 (0.457)	-0.198 (0.455)	0.575 (0.479)	0.555 (0.479)	0.059* (0.025)
SM Prime					0.175*** (0.021)
Ask-Tell:Prior		-0.197*** (0.019)		-0.134*** (0.020)	
Num.Obs.	16 650	16 650	16 477	16 477	16 901
R2	0.210	0.215	0.196	0.199	0.119

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B2: Full Models, Ask-Tell Effects by Minority Status (Table 2)

	Co-Partisan Opponent CD	Co-Partisan Group CD
Intercept	38.581*** (5.691)	47.233*** (7.346)
Ask-Tell	-3.706* (1.888)	-6.418** (2.380)
Dem	-11.059*** (2.316)	3.618 (3.072)
Demand Base	0.510*** (0.010)	0.612*** (0.012)
Party Minority	21.652*** (3.181)	10.938** (4.168)
Ideology	-3.044** (1.080)	0.689 (1.448)
Pol. Int. (5-point)	-2.812** (0.879)	-4.448*** (1.134)
Education	5.276*** (0.940)	-1.935 (1.229)
Prior	0.329*** (0.046)	0.309*** (0.056)
Network Co-Opp	31.755*** (4.845)	17.908** (6.115)
Network Opposing	19.974*** (4.439)	10.654+ (5.642)
Female	4.447* (1.764)	8.159*** (2.246)
White	5.248** (2.004)	-2.416 (2.612)
Married	1.347 (1.909)	2.926 (2.441)
Fully Employed	4.268* (2.008)	7.636** (2.595)
Ask-Tell:Party Min	-5.203 (4.363)	-7.213 (5.662)
Num.Obs.	11 043	5490
R2	0.283	0.401

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B3: Full Models, Does Majority/Minority Asymmetry in Ask-Tell Effects Vary by Party?
(Table 3)

	Co-Partisan Opp. CD	Co-Partisan Group CD
Intercept	38.581*** (5.771)	58.807*** (4.361)
Ask-Tell	-4.464 (2.786)	-4.896* (1.945)
Dem	-9.234** (3.193)	6.001* (2.371)
Demand Base	0.509*** (0.010)	0.619*** (0.007)
Party Minority	27.884*** (5.032)	14.800*** (3.542)
Ideology	-2.565* (1.100)	0.539 (0.833)
Pol. Int. (5-point)	-2.820** (0.878)	-4.930*** (0.649)
Education	5.159*** (0.942)	-0.391 (0.702)
Prior	0.332*** (0.046)	0.292*** (0.033)
Network Co-Opp	30.623*** (4.875)	14.458*** (3.533)
Network Opposing	19.793*** (4.439)	6.014+ (3.301)
Female	4.484* (1.763)	2.926* (1.291)
White	5.332** (2.005)	-2.406 (1.506)
Married	1.278 (1.909)	3.544* (1.400)
Fully Employed	4.286* (2.007)	3.294* (1.483)
Ask-Tell:Party Min	-7.197 (7.004)	-6.236 (5.042)
Ask-Tell:Dem	1.486 (3.787)	-4.065 (2.772)
Party Min:Dem	-10.835+ (6.457)	-10.250* (4.637)
Ask-Tell:Party Min:Dem	3.197 (8.946)	9.527 (6.524)
Num.Obs.	11 043	16 661
R2	0.283	0.401

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B4: Full Models, Social Media Prime Main Effects (Table 4)

	(1) Sanction Mean (H6a)	(3) Career	(4) Affiliative	(5) Strangers
Intercept	2.696*** (0.072)	2.857*** (0.089)	2.213*** (0.087)	3.021*** (0.089)
Dem	-0.035 (0.029)	-0.084* (0.035)	0.008 (0.035)	-0.029 (0.035)
Demand Base	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.000 (0.000)
Party Minority	0.064* (0.028)	0.078* (0.034)	0.115*** (0.033)	-0.003 (0.034)
Ideology	0.045** (0.014)	0.114*** (0.017)	0.048** (0.017)	-0.028+ (0.017)
Pol. Int. (5-point)	0.011 (0.011)	-0.032* (0.014)	-0.016 (0.013)	0.082*** (0.014)
Education	0.065*** (0.012)	0.035* (0.015)	0.056*** (0.014)	0.102*** (0.015)
Prior	0.015*** (0.001)	0.015*** (0.001)	0.016*** (0.001)	0.015*** (0.001)
Network Co-Opp	0.329*** (0.062)	0.364*** (0.077)	0.448*** (0.074)	0.176* (0.076)
Network Opposing	0.924*** (0.057)	0.668*** (0.070)	1.301*** (0.070)	0.792*** (0.070)
Female	0.099*** (0.022)	0.005 (0.027)	0.107*** (0.026)	0.184*** (0.027)
White	0.054* (0.026)	-0.199*** (0.032)	0.145*** (0.031)	0.215*** (0.032)
Married	0.049* (0.023)	0.013 (0.029)	0.094*** (0.028)	0.041 (0.029)
Fully Employed	0.060* (0.025)	0.271*** (0.031)	0.000 (0.030)	-0.096** (0.031)
SM Prime	0.178*** (0.021)	0.131*** (0.026)	0.121*** (0.025)	0.283*** (0.026)
Num.Obs.	16 901	16 925	16 931	16 930
R2	0.105	0.106	0.085	0.065

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B5: Full Model, Interaction Effects for Social Media Prime by Network Proxies (Table 5)

	Sanction Mean
Intercept	2.664*** (0.075)
Dem	−0.035 (0.029)
Demand Base	0.001*** (0.000)
Party Minority	0.059 (0.038)
Ideology	0.045** (0.014)
Pol. Int. (5-point)	0.011 (0.011)
Education	0.065*** (0.012)
Prior	0.015*** (0.001)
Network Co-Opp	0.412*** (0.087)
Network Opposing	0.967*** (0.080)
Female	0.098*** (0.022)
White	0.054* (0.026)
Married	0.049* (0.023)
Fully Employed	0.060* (0.025)
SM Prime	0.242*** (0.052)
SM Prime:Party Min	0.009 (0.053)
SM Prime:Network Out	−0.085 (0.112)
SM Prime:Network Co-Opp	−0.165 (0.123)
Num.Obs.	16 901
R2	0.105
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Table B6: Full Models, Private CDs by Majority/Minority Status (Figure 1)

	CD Price (\$0-\$300)
Intercept	31.465*** (3.805)
Co-partisan Opponent	-3.584*** (0.874)
Co-partisan Ally	6.599*** (0.876)
Opposing Partisan	14.166*** (1.116)
Dem	-3.807* (1.530)
Demand Base	0.558*** (0.006)
Party Minority	10.569*** (1.685)
Ideology	-1.620* (0.708)
Pol. Int. (5-point)	-1.556** (0.587)
Education	4.141*** (0.635)
Prior	0.328*** (0.031)
Network Co-Opp	8.326** (3.038)
Network Opposing	9.512** (2.930)
Female	3.798** (1.183)
White	6.918*** (1.327)
Married	3.600** (1.274)
Fully Employed	2.218+ (1.341)
Target Co-Opp: Party Min	10.776*** (2.060)
Target Co-Ally: Party Min	-10.818*** (1.898)
Target Out: Party Min	-12.553*** (2.192)
Num.Obs.	49 868
R2	0.299
Std.Errors	by: id

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B7: Full Models, Group Convos (Figure 2)

	Co- Partisans	Out-Partisans
Intercept	56.209*** (4.255)	78.297*** (5.109)
Dem	2.384 (1.738)	4.768* (1.992)
Demand Base	0.620*** (0.007)	0.481*** (0.008)
Party Minority	8.640*** (1.681)	−2.675 (1.885)
Ideology	0.233 (0.819)	−4.593*** (0.961)
Pol. Int. (5-point)	−4.878*** (0.650)	−2.844*** (0.788)
Education	−0.382 (0.702)	6.662*** (0.861)
Prior	0.290*** (0.033)	0.367*** (0.041)
Network Co-Opp	15.268*** (3.505)	−26.868*** (4.392)
Network Opposing	6.045+ (3.305)	3.276 (4.056)
Female	2.935* (1.293)	8.985*** (1.585)
White	−2.383 (1.508)	11.643*** (1.808)
Married	3.498* (1.401)	3.318+ (1.705)
Fully Employed	3.348* (1.485)	−1.763 (1.789)
Num.Obs.	16 661	16 644
R2	0.400	0.215
Std.Errors	by: id	by: id

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B8: Full Models, Ask-Tell ATEs on Mechanism Checks (Figure 4)

	Career Sanction	Relation Sanction	Stranger Sanction	All Americans	People You Know
Intercept	3.031*** (0.090)	2.393*** (0.087)	3.217*** (0.090)	21.205*** (1.323)	13.747*** (1.359)
Ask-Tell	-0.319*** (0.026)	-0.329*** (0.025)	-0.361*** (0.026)	-11.455*** (0.383)	-7.643*** (0.399)
Dem	-0.081* (0.035)	0.011 (0.035)	-0.026 (0.035)	2.603*** (0.522)	3.275*** (0.533)
Demand Base	0.001*** (0.000)	0.001*** (0.000)	0.000 (0.000)	0.021*** (0.002)	0.024*** (0.002)
Party Minority	0.075* (0.034)	0.112*** (0.033)	-0.007 (0.034)	-3.140*** (0.486)	-3.632*** (0.506)
Ideology	0.116*** (0.017)	0.050** (0.017)	-0.026 (0.017)	-0.532* (0.251)	-1.002*** (0.261)
Pol. Int. (5-point)	-0.034* (0.014)	-0.018 (0.013)	0.080*** (0.014)	1.355*** (0.204)	1.466*** (0.210)
Education	0.036* (0.014)	0.057*** (0.014)	0.104*** (0.015)	-0.106 (0.216)	-0.249 (0.223)
Prior	0.015*** (0.001)	0.016*** (0.001)	0.015*** (0.001)	0.436*** (0.011)	0.454*** (0.011)
Network Co-Opp	0.361*** (0.077)	0.444*** (0.074)	0.172* (0.076)	8.172*** (1.132)	9.046*** (1.188)
Network Opposing	0.671*** (0.070)	1.305*** (0.070)	0.796*** (0.069)	1.370 (1.045)	1.730 (1.098)
Female	0.006 (0.027)	0.108*** (0.026)	0.185*** (0.027)	-2.791*** (0.397)	-2.307*** (0.415)
White	-0.202*** (0.031)	0.142*** (0.031)	0.212*** (0.032)	0.419 (0.475)	-0.346 (0.496)
Married	0.018 (0.029)	0.098*** (0.028)	0.046 (0.029)	0.568 (0.425)	0.084 (0.440)
Fully Employed	0.270*** (0.031)	-0.001 (0.030)	-0.097** (0.031)	-0.162 (0.457)	0.575 (0.479)
SM Prime	0.128*** (0.026)	0.119*** (0.025)	0.280*** (0.026)		
Num.Obs.	16 925	16 931	16 930	16 650	16 477
R2	0.113	0.094	0.075	0.210	0.196

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B9: Full Models, Ask-Tell Effects (Figure 5)

	Group Out	Group In	Out-Partisan	Co-Ally	Co-Opponent	Discussant
Intercept	83.959*** (5.181)	60.115*** (4.317)	47.083*** (6.534)	48.896*** (5.710)	39.031*** (5.676)	25.641*** (4.263)
Ask-Tell	-10.423*** (1.530)	-7.191*** (1.251)	-6.682*** (1.947)	-8.207*** (1.698)	-4.803** (1.706)	-6.634*** (1.264)
Dem	4.824* (1.989)	2.440 (1.736)	2.248 (2.549)	-2.326 (2.305)	-11.047*** (2.317)	-3.720* (1.700)
Demand Base	0.480*** (0.008)	0.619*** (0.007)	0.452*** (0.010)	0.575*** (0.009)	0.510*** (0.010)	0.648*** (0.007)
Party Minority	-2.800 (1.883)	8.565*** (1.679)	0.287 (2.371)	-0.192 (2.225)	19.056*** (2.293)	10.423*** (1.696)
Ideology	-4.552*** (0.960)	0.269 (0.818)	-3.580** (1.212)	0.945 (1.093)	-3.031** (1.081)	-0.794 (0.779)
Pol. Int. (5-point)	-2.904*** (0.787)	-4.919*** (0.649)	-1.236 (0.986)	-0.784 (0.880)	-2.808** (0.879)	-1.434* (0.650)
Education	6.716*** (0.860)	-0.345 (0.702)	8.597*** (1.100)	-0.434 (0.963)	5.259*** (0.940)	3.537*** (0.704)
Prior	0.367*** (0.041)	0.290*** (0.033)	0.407*** (0.052)	0.380*** (0.046)	0.329*** (0.046)	0.235*** (0.034)
Network Co-Opp	-27.026*** (4.389)	15.149*** (3.503)	-15.846** (5.345)	10.109* (4.764)	31.802*** (4.844)	8.210* (3.391)
Network Opposing	3.423 (4.050)	6.122+ (3.300)	29.797*** (5.154)	-16.448*** (4.420)	19.992*** (4.437)	7.640* (3.244)
Female	9.011*** (1.583)	2.950* (1.292)	6.668*** (2.021)	4.165* (1.762)	4.448* (1.764)	1.319 (1.303)
White	11.558*** (1.806)	-2.448 (1.507)	20.463*** (2.303)	-0.063 (2.025)	5.234** (2.004)	3.718* (1.483)
Married	3.463* (1.703)	3.590* (1.400)	5.847** (2.186)	3.422+ (1.893)	1.398 (1.909)	3.664** (1.402)
Fully Employed	-1.805 (1.786)	3.324* (1.483)	-2.508 (2.282)	4.212* (1.975)	4.304* (2.007)	2.548+ (1.486)
Num.Obs.	16 644	16 661	11 037	11 091	11 043	16 697
R2	0.217	0.401	0.194	0.320	0.283	0.408

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B10: Full Models, Social Media Prime and Minority Status (Figure 6)

	Career	Relationships	Strangers	Sanction Mean
Intercept	2.858*** (0.089)	2.212*** (0.087)	3.018*** (0.089)	2.695*** (0.072)
Dem	-0.084* (0.035)	0.008 (0.035)	-0.029 (0.035)	-0.035 (0.029)
Demand Base	0.001*** (0.000)	0.001*** (0.000)	0.000 (0.000)	0.001*** (0.000)
Party Minority	0.073 (0.047)	0.120** (0.046)	0.009 (0.047)	0.067+ (0.038)
Ideology	0.114*** (0.017)	0.048** (0.017)	-0.028+ (0.017)	0.045** (0.014)
Pol. Int. (5-point)	-0.032* (0.014)	-0.016 (0.013)	0.082*** (0.014)	0.011 (0.011)
Education	0.035* (0.015)	0.056*** (0.014)	0.102*** (0.015)	0.065*** (0.012)
Prior	0.015*** (0.001)	0.016*** (0.001)	0.015*** (0.001)	0.015*** (0.001)
Network Co-Opp	0.364*** (0.077)	0.448*** (0.074)	0.176* (0.076)	0.329*** (0.062)
Network Opposing	0.668*** (0.070)	1.301*** (0.070)	0.792*** (0.070)	0.923*** (0.057)
Female	0.005 (0.027)	0.107*** (0.026)	0.184*** (0.027)	0.098*** (0.022)
White	-0.199*** (0.032)	0.145*** (0.031)	0.214*** (0.032)	0.054* (0.026)
Married	0.013 (0.029)	0.094*** (0.028)	0.041 (0.029)	0.049* (0.023)
Fully Employed	0.271*** (0.031)	0.000 (0.030)	-0.096** (0.031)	0.060* (0.025)
SM Prime	0.129*** (0.029)	0.124*** (0.028)	0.288*** (0.029)	0.179*** (0.023)
SM Prime:Party Min	0.009 (0.065)	-0.011 (0.063)	-0.025 (0.065)	-0.007 (0.052)
Num.Obs.	16 925	16 931	16 930	16 901
R2	0.106	0.085	0.065	0.105

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B11: Full Models, Perceived Benefits to Conversations with Co-Partisan Opponents

	Discuss Real			Get to Know	
	None	Opinions	Civic Duty	Discussant	Persuade
Intercept	0.096*** (0.011)	0.514*** (0.018)	0.232*** (0.016)	0.378*** (0.018)	0.386*** (0.018)
Party Minority	0.157*** (0.029)	-0.035 (0.036)	-0.121*** (0.025)	-0.034 (0.034)	-0.176*** (0.031)
Num.Obs.	1001	1001	1001	1001	1001
R2	0.040	0.001	0.018	0.001	0.027

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table B12: Full Models, Perceived Costs to Conversations with Co-Partisan Opponents

	Damage		Make Discussant	Be Judged	Uncertainty
	None	Relationship	Uncomfortable	for Opinion	about Opinions
Intercept	0.132*** (0.012)	0.565*** (0.018)	0.382*** (0.018)	0.282*** (0.017)	0.165*** (0.014)
Party Minority	0.044 (0.027)	0.029 (0.035)	0.016 (0.035)	0.093** (0.034)	0.011 (0.027)
Num.Obs.	1001	1001	1001	1001	1001
R2	0.003	0.001	0.000	0.008	0.000

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

C Additional Pre-Registered Analyses

C.1 Hypothesis 3

Table C1: Full Models, Hypothesis 3

	(1)	(2)	(3)	(4)
Intercept	51.888*** (3.525)	49.214*** (3.708)	39.031*** (5.676)	60.115*** (4.317)
Ask-Tell	-7.175*** (1.037)	-1.899 (2.624)	-4.803** (1.706)	-7.191*** (1.251)
Dem	-0.780 (1.398)	-0.757 (1.397)	-11.047*** (2.317)	2.440 (1.736)
Party Minority	6.049*** (1.380)	6.060*** (1.379)	19.056*** (2.293)	8.565*** (1.679)
Pol. Int. (5-point)	-2.691*** (0.531)	-2.682*** (0.531)	-2.808** (0.879)	-4.919*** (0.649)
Prior	0.324*** (0.028)	0.382*** (0.038)	0.329*** (0.046)	0.290*** (0.033)
Network Co-Opp	2.162 (2.752)	2.161 (2.751)	31.802*** (4.844)	15.149*** (3.503)
Network Opposing	7.399** (2.670)	7.420** (2.671)	19.992*** (4.437)	6.122+ (3.300)
Female	4.911*** (1.072)	4.897*** (1.072)	4.448* (1.764)	2.950* (1.292)
White	5.944*** (1.214)	5.948*** (1.214)	5.234** (2.004)	-2.448 (1.507)
Married	3.611** (1.154)	3.614** (1.154)	1.398 (1.909)	3.590* (1.400)
Fully Employed	1.893 (1.222)	1.876 (1.222)	4.304* (2.007)	3.324* (1.483)
Num.Obs.	16 115	16 115	11 043	16 661
R2	0.444	0.445	0.283	0.401

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Models include covariate vector

C.2 Hypothesis 4

Table C2: Full Models, Hypothesis 4

	Group Demands	Private Demands
Intercept	56.184 (4.055)	35.621 (3.866)
Co-partisan Ally		4.868 (1.108)
Opposing Partisan		11.274 (1.386)
Ask-Tell	-6.920 (1.548)	-6.800 (1.272)
Dem	3.573 (1.572)	-3.783 (1.528)
Party Minority	2.846 (1.557)	7.675 (1.528)
Pol. Int. (5-point)	-3.884 (0.591)	-1.581 (0.587)
Prior	0.329 (0.031)	0.328 (0.031)
Network Co-Opp	-5.812 (3.101)	8.194 (3.036)
Network Opposing	4.677 (2.986)	9.714 (2.928)
Female	6.017 (1.195)	3.848 (1.182)
White	4.561 (1.371)	6.861 (1.325)
Married	3.538 (1.288)	3.709 (1.273)
Fully Employed	0.752 (1.362)	2.228 (1.340)
Ask-Tell:Topic Election	-1.084 (2.683)	
Ask-Tell:Target Out Partisans	-3.412 (1.938)	
Topic Election:Target Out Partisans	6.071 (2.492)	
Ask-Tell:Topic Election: Target Out Partisans	1.125 (3.450)	
Num.Obs.	33 305	49 868
R2	0.305	0.299
Std.Errors	by: id	by: id

C.3 Hypothesis 5

Table C3: Full Models, Hypothesis 5

	Group Demands	Private Demands	Group Demands
Intercept	38.581 (5.771)	31.465 (3.805)	56.457 (4.290)
Co-partisan Ally		6.599 (0.876)	
Opposing Partisan		14.166 (1.116)	
Ask-Tell	-4.464 (2.786)		
Dem	-9.234 (3.193)	-3.807 (1.530)	2.388 (1.739)
Party Minority	27.884 (5.032)	10.569 (1.685)	9.067 (2.000)
Pol. Int. (5-point)	-2.820 (0.878)	-1.556 (0.587)	-4.886 (0.650)
Prior	0.332 (0.046)	0.328 (0.031)	0.290 (0.033)
Network Co-Opp	30.623 (4.875)	8.326 (3.038)	15.243 (3.506)
Network Opposing	19.793 (4.439)	9.512 (2.930)	6.058 (3.305)
Female	4.484 (1.763)	3.798 (1.183)	2.934 (1.293)
White	5.332 (2.005)	6.918 (1.327)	-2.384 (1.508)
Married	1.278 (1.909)	3.600 (1.274)	3.491 (1.401)
Fully Employed	4.286 (2.007)	2.218 (1.341)	3.352 (1.485)
Ask-Tell:Party Min	-7.197 (7.004)		
Ask-Tell:Dem	1.486 (3.787)		
Party Min:Dem	-10.835 (6.457)		
Ask-Tell:Party Min:Dem	3.197 (8.946)		
Target Co-Opp: Party Min		10.776 (2.060)	
Target Co-Ally: Party Min		-10.818 (1.898)	
Target Out: Party Min		-12.553 (2.192)	
Num.Obs.	11 043	49 868	16 661
R2	0.283	0.299	0.400
Std.Errors		by: id	

D Ethical Statement

The authors declare that the human subjects research in this article was reviewed by the pertinent universities' Human Subjects Committees. All aspects of this project adhered to the ethical principles outlined by the American Political Science Association.

Participants were recruited through two widely used online platforms—PureSpectrum and CloudResearch—which provide panels for academic research. The PureSpectrum component was part of a larger study through the Civic Health and Institutions Project. All participants were required to provide informed consent prior to participation. The consent forms explicitly disclosed that participation was voluntary and that respondents had the right to withdraw at any point without penalty.

To the extent our study affected participants' beliefs about the real world, we provided correct information about the share of Americans reporting or seeking to inflict social repercussions for political speech, and this correction reduced harmful beliefs for the vast majority of participants. While the study explored preferences surrounding the 2024 presidential election, at no point were participants asked to share their political beliefs in their real social circles or to take any political action that affected real-world political processes.