Use Your Inside Voice: Intra-Party Social Pressure and the Avoidance of Political Speech

Patrick Liu and Daniel B. Markovits

September 4, 2025

Abstract

Americans increasingly report censoring their political beliefs for fear of social repercussions—not from adversaries, but from co-partisans. We assess if the conventional theory that social pressure stifles minority opinions extends to this lesser-studied context. Using a nationwide survey experiment fielded during the 2024 primaries ($N = 17,691$), we show that intra-party speech is more costly to dissenters than the party faithful. Social pressure, however, constrains both. We document widespread misperceptions about the likelihood of social sanctions, then randomly correct these exaggerated fears to manipulate social pressure's salience. This intervention reduces self-censorship for party minorities and majorities alike, suggesting that intra-party social pressure results less from injunctive norms dictating what right-thinking partisans should say than from a widely shared aversion to divisive debate. To explain dissenters' disproportionate self-censorship, we look beyond assumptions about differential exposure to social pressure and highlight other important motivations—namely, majorities' unusual eagerness to persuade co-partisan opponents.

Introduction

An enduring cause for concern among political observers is the idea that, even where legal protections for free speech persist, citizens themselves will deplete the quality of public debate by alienating those who hold unpopular views (e.g., Noelle-Neumann, 1974). Recent research on political speech in the U.S. has attended especially to salient culture war debates over political correctness (Chong, Citrin and Levy, 2024) and to the role of partisan animosity in limiting interaction and dialogue between ideological opponents (Iyengar et al., 2019; Settle and Carlson, 2019). Little attention has been paid, however, to the causes and consequences of self-censorship *within* co-partisan networks—a pattern that, at least in the public imagination, is now highly pervasive. During the 2024 presidential election, for instance, Republican voters reported that they risked ostracism or even outright violence if they expressed misgivings about Donald Trump to peers (Tabet, 2023), so much so that the Democratic campaign implored women in Republican households to secretly vote for Kamala Harris in the election's final weeks (McCammon, 2024).

These accounts of intra-party social pressure raise key questions about democratic competence, polarization, and the social nature of partisanship. In an era where sincere deliberation between ideological opponents is rare (Gentzkow and Shapiro, 2011; Settle and Carlson, 2019; Brown and Enos, 2021) and voters' receptiveness to factual information and opinion leadership is strongly shaped by partisan group cues (Barber and Pope, 2019; Coppock, 2023; Carey et al., 2024), co-partisans with counter-partisan views may be a crucial source of cross-cutting exposure in citizens' daily lives. Exposure to contrary perspectives is said to produce better-informed opinions and to reduce polarization by improving citizens' understanding of legitimate rationales for opposing viewpoints (Fishkin, 1991; Mutz, 2002). Further, cross-cutting exposure can accelerate party defection. How do parties keep the electoral effects of factional divisions in check? While social norms have proven vital in sustaining Democratic partisanship within Black communities (White and Laird, 2020), it is far from evident whether similar norms pervade partisan networks in general. Knowing whether, how, and why social pressure constrains intra-party debate thus offers important

insights into the persistence of mass polarization and partisan voting behavior. This is especially true given the politically homogeneous networks in which most citizens reside.

While a handful of recent studies affirm that hostilities between co-partisan factions are on the rise and that those who hold dissenting views from their party tend to perceive reputational costs for speaking their mind (Young and de Wit, 2025; Spelman et al., 2024), they do not establish whether fears of social sanction are widespread, if they affect dissenters alone, or how they compare to inter-party pressures. Because factional animosity extends both ways, members of the party faithful may be equally wary of the reputational costs for speaking candidly with friends who dissent.[1] The effect of intra-party social pressure may not be conformity to the majority opinion but rather a quelling of political conversation in its entirety, as both sides adhere to a social etiquette against raising contentious discussion topics (Eliasoph, 1998).

This paper investigates two central questions. We begin by measuring to what extent party dissenters and members of the party faithful self-censor in co-partisan social contexts. After verifying that dissenters—who constitute the minority in both parties—are indeed less willing than those in the majority to discuss their sincere views with co-partisan opponents, we ask if this pattern is explained by social pressure. We find that although fears of social sanction indeed fuel self-censorship in co-partisan contexts, these fears affect party dissenters and the party faithful alike. Were social pressure solely the result of injunctive norms dictating the preferences that right-thinking Democrats and Republicans should hold, then those in the majority would perceive no risk of social sanction in interactions with co-partisans. Instead, our findings suggest that intra-party social pressure stems in large part from forces that constrain party majorities and minorities alike. Therefore, social pressure alone is likely insufficient to explain the disproportionate self-censorship of party dissenters. We investigate what other psychological motivations could explain these patterns of political speech.

Our puzzle requires an empirical strategy that differs from existing research paradigms. Canonical interventions have made participants believe their behavior will be observed by

[1]Dissenting factions within both parties made their voices heard during the 2024 election, harnessing substantial resources and drawing significant media attention toward efforts to damage their parties' nominees (Aratani, 2024; Abdul-Hakim et al., 2024).

peers (Gerber, Green and Larimer, 2008), placed participants in artificial lab environments where they hold the minority position (Asch, 1951), or altered participants' perceptions about the popularity of their own preferences within a reference group (Tankard and Paluck, 2016). While studies utilizing these interventions have offered compelling evidence that political speech responds strongly to social observation, minority status, and descriptive norms (e.g., Braghieri, 2024; Ho and Huang, 2024), these designs assess how social pressure operates in group settings without addressing one-on-one conversations, and they alter a bundle of considerations at once rather than prime fears of social sanction alone. Randomizing the public revelation of people's private beliefs, for instance, can prime positive social motivations such as the desire to persuade.

We devise a measurement strategy that isolates the role of social pressure by assessing how people's willingness to engage in sensitive conversations with co-partisans responds to an exogenous change in the salience of social pressure. We show that respondents widely overestimate the frequency with which Americans at large face social sanction for political disagreement. Correcting these misperceptions via an "ask-tell" intervention (e.g., Braley et al., 2023; Mernyk et al., 2022) enables us to instrument for the reputational concerns brought to mind when respondents consider discussing their 2024 vote choice with co-partisans they know. This design therefore directly intervenes on respondents' perceptions of the likelihood of social sanctions for disagreeable speech. To examine if social pressure can explain an asymmetry in self-censorship, we then test whether the treatment exhibits heterogeneous effects between dissenters and the party faithful.

This experiment was embedded in a large, nationwide survey ($N = 21,400$) fielded during the summer before the 2024 U.S. presidential election. Capitalizing on tensions surrounding the presidential primary, we define dissenters as partisans who did not support their party's 2024 election nominee. Because dissenters constituted a small minority within each party, a large sample size is crucial for our assessments of treatment effect heterogeneity. To assess if the ask-tell correction improves willingness to speak, we used "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). Unique

among studies of self-censorship, we employ a common scale to measure respondents' willingness to speak in a diverse range of circumstances, enabling us to contrast the social fears engendered by disagreement within and between parties, as well as across group and one-on-one interactions.

We find that on the topic of 2024 vote choice, members of the party faithful were substantially more willing to speak with party defectors they knew (referred to as their "co-partisan opponents") than vice versa. Exposure to the ask-tell correction reduced the prices party defectors demanded to speak with co-partisan opponents, confirming that social pressure constrains intra-party speech for those in the minority. However, ask-tell exposure also reduced the prices demanded by members of the party faithful to speak with co-partisan opponents, with no significant difference in CATEs between defectors and the party faithful. We similarly find little evidence that being in the minority within one's private network is associated with a greater ask-tell effect. Further, ask-tell exposure reduced compensation demands by similar amounts for other peers in respondents' networks, such as out-partisans and people with whom respondents frequently discuss politics to begin with. Intra-party social pressure seems more often to resemble a general aversion to discussing contentious political issues than conformity to speech norms prescribed by one's partisanship.

Our conjecture that asymmetry in party majorities' and minorities' willingness to discuss politics cannot be fully explained by social pressure invites an alternative explanation. We demonstrate with a follow-up survey that when asked to consider the reasons why they may initiate discussions about the election with co-partisan opponents, party defectors are considerably less likely to see persuading their peers to change their mind as a possible benefit. Meanwhile, party defectors and the party faithful are equally likely to cite damaging relationships as reasons for avoiding discussions about the election. We offer this as suggestive evidence that party majorities' unusual zeal for discussing the election with possible defectors is better explained by the instrumental partisan benefits they derive from persuading wayward co-partisans.

In contrast to identities upon which strong community ties are built (see White and Laird, 2020), our findings suggest important limitations on the power of partisan identity

to constrain political behavior. Despite the durability of partisanship as social identity and the extreme favoritism voters show toward co- over out-partisans (Green, Palmquist and Schickler, 2002; Mason, 2018), partisanship ranks low on the list of identities Americans tend to name as personally important (Krupnikov and Ryan, 2022). Our evidence suggests that intra-party hostilities indeed mute political discussion, but that partisan attachments may not be sufficiently important to voters' self-concept that partisan identity *per se* constitutes an organic basis for injunctive speech norms. Instead, specific non-political social groups that overlap with partisan identity may be required for the development of norms that more effectively prohibit speech against the party line.

We begin by defining social pressure and placing it within the broader set of psychological motivations governing political speech. We then show that conventional theories widely conceive of social pressure as a force that principally stymies minority opinions. We define in contrast to this view a subtler conception of social pressure rooted in people's aversion to divisive political debate. The remainder of the paper elaborates methods and results. We conclude with theoretical implications as well as a discussion of how our empirical design advances the study of social influence more broadly.

Defining Social Pressure

Social pressure as employed in this essay refers to the psychological constraints that flow from "a basic human drive to win praise and avoid chastisement" (Green and Gerber, 2010, 331). Under classic models of social influence (e.g., Cialdini and Goldstein, 2004), social pressure is best understood as a function of people's "goals of affiliation" which deter actions that could endanger one's ability to create or maintain meaningful social relationships. It is by threatening these affiliative goals that social pressure becomes a powerful tool for enforcing injunctive norms, which circumscribe the types of behavior of which one's peers will approve. Making people aware that their actions will be observed by peers has proven an effective means of inducing electoral turnout (Gerber, Green and Larimer, 2008) and pro-social behavior (Schultz, 1999). Conformity to these injunctive norms is usually thought

to be driven by concerns about preserving one's image among peers.[2]

What distinguishes social pressure from other forms of interpersonal influence? Here, the difference between injunctive and descriptive norms is illustrative. Descriptive norms alter behavior by describing what others typically do, especially within some important reference group. "Norm change interventions" operate on this logic by drawing comparisons between a respondent's own behavior and the average behavior of their reference group (Tankard and Paluck, 2016); for instance, information about whether a majority of one's peers will turn out to vote can alter electoral participation (Gerber and Rogers, 2009). Descriptive norms need not be enforced through fears of social sanction. People may imitate their social surroundings because they infer that, if others unanimously or even widely act this way, the behavior must be somehow sensible or efficient (Cialdini, Kallgren and Reno, 1991, 203).[3]

Other Motivations for Political Speech

Choices about whether to speak and what to say are often governed by considerations having little to do with social scrutiny. These include the mere expressive value people can experience from defending sincere commitments and the instrumental value voters may see in persuading peers to support their preferred candidate.[4] In contrast to affiliative motivations, these considerations are sometimes characterized as rooted in people's goals of "affirmation" and "accuracy," respectively (Cialdini and Goldstein, 2004; Carlson and Settle, 2022).

The dichotomy between social pressure on the one hand and the remaining considerations that dictate speech is a central focus of this paper. Concerns about social pressure have featured prominently in recent media coverage of intra-party politics, with news outlets reporting that Republican voters secretly voted against Trump in the 2024 election without telling friends and family in order to avoid scorn (McCammon, 2024). Whether this sort of

[2]In centering *social* conformity, this paper mostly sets aside the equally important literature on conformity to demands from authority figures, beginning with Milgram (1974) and including studies of preference falsification under authoritarian regimes (e.g., Kuran, 1995).

[3]Alternatively, when the source of speech repression is not social but legal, descriptive norms can affect speech by altering the expected external costs of expressing sincere preferences. Kuran (1991) famously argued that in authoritarian contexts, as second-order beliefs about the popularity of mass dissent grow, the risks of speaking out decrease due to the increasing difficulty of arresting dissidents and increasing likelihood of a successful rebellion.

[4]That voters pay keen attention to whether their peers support their preferred candidate is evinced by strategic voting both in the U.S. (Abramson, Aldrich and Rohde, 1983) and abroad (e.g., Black, 1978; Kawai and Watanabe, 2013). Across varied settings, voters attend to candidate electability.

intra-party sanctioning is widespread, however, is far from self-evident. Social pressure may matter, but political discourse in the U.S. also frequently exaggerates its aggregate influence. The accusation that political debate is being stifled by citizens intolerant of opposing views has been a fixture of partisan and populist rhetoric in the U.S. since at least Nixon's "Silent Majority" address (Gillion, 2020), and the contemporary media environment features a constant drone of debates about "cancel culture" (Dias, Druckman and Levendusky, 2024).

Existing studies of intra-party social pressure do not offer clear evidence if it is widespread among all partisans. White and Laird (2020) argue that support for the Democratic Party is an injunctive social norm among Black Americans and that this norm preserves Democratic partisan identification, voting, and donations among otherwise ideologically conservative Black voters. This mechanism, however, depends on within-group dynamics among Black Democrats which may not be felt by American partisans more broadly. Spelman et al. (2024), in the paper perhaps most closely connected to our own, argue that partisan social identity more broadly serves to constrain speech that violates the majority view within parties. Spelman and colleagues illustrate not only that self-censorship is asymmetric between minority and majority opinion holders within each party, but also that misperceptions about co-partisans' enthusiasm to deliver social sanctions predict reluctance to speak. Yet, social pressure may not be the sole reason for this asymmetry. Dissidents within the party, such as anti-Trump Republicans, may have fewer positive motivations to persuade the party faithful to change their vote than vice versa.

Partisan minorities might also be less willing to speak not for reasons of social pressure but because their dueling partisan and electoral preferences reduce their positive motivations to speak. Dissidents within the party may be precisely the type of cross-pressured voters who, as Mutz (2002) argued, are less likely to share their beliefs, vote, or engage in a range of political activities. Weaker partisan attachments could also correlate with socioeconomic variables as well as attitudinal and informational traits associated with lower levels of participation (Brady, Verba and Schlozman, 1995). In this paper we assess to what extent social pressure versus motivations for persuasion are responsible for the asymmetry in co-partisans' self-censorship.

Two Tales of Social Pressure

Spanning the literatures on social influence and political communications, the standard view of social pressure is that it serves to promote conformity to majority opinion. In contrast to this view, however, is a subtler conception of social pressure as something that can stymie the speech of minorities and majorities alike by deterring divisive issues. While the latter is not often viewed within the literature as a form of social pressure, in this section we show how it flows from the same affiliative pressures that are often said to enforce injunctive norms of speech against opinion minorities.

Substantive Social Pressure Social scientific research on conformity traces its lineage to the seminal line-judgment experiments of Asch (1956), which in the paradigmatic case asked subjects to publicly state whether two plainly different lines were of equal length after a room of confederates unanimously pronounced the lines identical. One-third of subjects conformed, suggesting that majority influence in an immediate group context—while not ubiquitous—can powerfully shape how people express their beliefs. The idea that perceiving oneself to be in the minority constrains speech has since featured prominently in scholarship about public deliberation, particularly in Noelle-Neumann's (1974) "spiral of silence" theory. Noelle-Neumann surmised that social regulation reduces the quality of deliberation, as people maintain a "quasi-statistical" sense of the distribution of opinion in society and minority opinion-holders withhold their true preferences in public settings for fear of isolation.

Though empirical examinations have failed to substantiate the prediction that being of the minority view mechanically reduces people's willingness to express that opinion (Glynn, Hayes and Shanahan, 1997), there is weight to the notion that citizens will enforce injunctive norms against the expression of particular views, especially views held by a small fraction of the public. Decades of empirical study testify to a proclivity among Americans—seemingly on both sides of the political aisle—to condone speech restrictions on political opponents whose views they find intolerable or dangerous across a broad range of substantive areas (Sullivan, Piereson and Marcus, 1979; Chong, Citrin and Levy, 2024; Dias, Druckman and Levendusky, 2024). Drawing on the classic sociological work of Goffman (1963), Valentim

(2024*b*) analogizes the source of social pressure constraining support for fringe candidates to political *stigma*, a judgment of "undesired differentness" in one's beliefs and by extension one's character (1384). Valentim (2024*a*,*b*) argues that certain parties in European systems—notably the radical right—are subject to stigma, that other small parties do not face this concern, and that this mechanism can reduce vote share. This theory suggests that minority status is a necessary but insufficient condition for social pressure. Instead, stigma is a property of specific beliefs that violate injunctive norms, and citizens are willing punish others who violate these norms. We refer to this as a *substantive* kind of social pressure.

Demonstrating the causal effect of social pressure is a persistent empirical puzzle, and papers that study substantive cases of social pressure have sought to address this challenge with a variety of experimental designs. Take, for instance, a string of recent studies concerning "political correctness" on college campuses. Ho and Huang (2024) show that undergraduates make incorrect inferences about the meaning of silence, interpreting an unwillingness to share political beliefs as an endorsement of the majority view. Correcting this false interpretation of silence increases political minorities' willingness to speak without affecting the behavior of majorities. Although this design demonstrates asymmetries in the behavior of those who hold "socially appropriate" and "inappropriate" views, the mechanism at play is not an injunctive norm but rather a descriptive one, as the intervention alters second-order beliefs—beliefs about the popularity of the respondent's own views among their peers. Alternatively, Braghieri (2024) finds that on controversial social issues alone, randomly heightening the salience of social image concerns—namely, by hinting to undergraduates that their survey responses might be shared with peers—similarly affects the willingness to speak of those holding the minority opinion but not members of the majority. This design, like that of canonical social pressure experiments (e.g., Gerber, Green and Larimer, 2008), is effective for identifying how social scrutiny impacts public behavior but does not identify whether subjects are contemplating affiliative or alternative motivations.

This paper examines to what extent intra-party social pressure is substantive in nature. Crucially, this view has clear and testable implications for the relationship between individual positions on intra-party conflict (e.g., 2024 party nominee) and self-censorship. Specifically,

social pressure should affect partisan minorities yet be all but absent in interactions where the speaker is either of the majority opinion or agrees with their conversation partner.

Substance-Neutral Social Pressure In contrast to the substantive view, which predicts asymmetry between minorities and majorities, other studies suggest fear of social sanction might be considerably more uniform—in large part because most of the population is unenthusiastic about political speech, regardless of its content. For as long as scholars have recognized the role of majority influence in constraining speech, they have understood in equal measure that daily life is, for most people, not deeply steeped in politics (Berelson, Lazarsfeld and McPhee, 1954; Prior, 2005). In her canonical examination of why even privately political volunteers and activists project an apathetic attitude around their peers, Eliasoph (1998) argued that rather than reflecting true disinterest, avoiding politics often reflects a considered effort to assess and maintain the boundaries of "political manners" or "etiquette" (21). The injunctive norms governing this etiquette are not substantive but rather apply to all, irrespective of the content of their beliefs. People recognize that, in addition to the inevitable discomfort raised with certain political topics, appearing excessively invested in politics may turn off those less politically involved (Krupnikov and Ryan, 2022). Social pressure to self-censor may derive from the simple desire to avoid experiencing discomfort or inflicting it upon others.

Understanding social pressure as substance-neutral generates alternative predictions about the effects of social influence in same-party networks. If the concern that talking politics will damaging existing affiliations hinges not upon the stigmatization of one's specific beliefs but upon a general wariness of political tension, majority and minority opinion-holders should differ little in the extent to which they experience social pressure. The contribution of this essay is to help assess not only the prevalence of social pressure in interactions among copartisans or the asymmetry in self-censorship between party defectors and the party faithful, but also the degree of asymmetry in affiliative concerns. To do so, we intervene directly on respondents' beliefs about the willingness of peers to socially sanction for political speech and examine how these treatment effects vary between party defectors and the party faithful.

Measuring Willingness to Speak in Intra-Party Contexts

Identifying the social pressure latent in respondents' real networks presents a unique set of measurement challenges. A common approach to studying self-censorship in everyday interactions is to examine whether, in surveys, self-reported perceptions of a climate of intolerance predict a lower willingness to speak (Gibson and Sutherland, 2023; Gibson, 1992). While perceptions of the overall opinion climate are useful, they may be unreliable given the politicized nature of accusations that free speech is stifled (Menzner and Traunmüller, 2023). Data of this variety is also blunt, offering little opportunity to examine whether or how social pressure varies across the many sets of possible conversants.

A vast literature spanning the sociology, political science, and communication disciplines has standardized the use of "name generators," a category of survey methods for mapping core social networks that invites respondents to name three or more "alters" with whom they most frequently discuss "important" or political matters (Burt, 1984; Huckfeldt and Sprague, 1987). While this method estimates an upper bound on the effects of exposure to within-party disagreement, it inherently precludes analysis of the dyad-level predictors of discussion frequency, since it provides no data on the attitudes, partisanship, or perceived tolerance of alters not mentioned.[5]

In light of these methodological concerns, Settle and Carlson (2019) approach their study of political discussion avoidance from the angle of willingness-to-accept measures, and in doing so make it feasible to study communication preferences with substantially greater granularity. Arguing that variance in willingness to hold discussion in different group compositions can be elicited through respondents' "compensation demands" for participating in hypothetical scenarios, they find that people consistently exhibit greater willingness to discuss politics with co-partisans than out-partisans across various issues. We use this style of outcome as our primary measure.

[5]Some surveys circumvent this concern by employing role-based generators—which modify the selection criteria for alters by asking respondents to name peers who fill specific roles, such as neighbors or friends. We are not aware of any recently collected surveys that employ role-based name generator items that also elicit respondents' affiliative concerns about the alters mentioned. Name generators also may result in long, time-intensive survey instruments, possibly limiting their use.

The challenge of this method for our research questions is that compensation demands (hereafter, CDs) capture the full set of considerations bearing on one's choice to talk politics—including, as Carlson and Settle (2022, chapter 3) highlight, perceived benefits such as persuading peers to share one's view or verbally expressing one's deeply held commitments. The average difference in CD prices when varying the identity of the discussant therefore cannot be interpreted solely as quantifying an asymmetry in social pressure. For instance, respondents may report higher CD prices for out-partisans than for co-partisans not only due to fears of social pressure, but also because they perceive a lower likelihood of productive debate and an increased risk of having to defend one's uncertain opinions—a cost tied to goal of maintaining a positive self-concept (Cialdini and Goldstein, 2004) but that is both normatively and substantively distinct from social pressure. Quantifying intra-party social pressure requires a novel measurement strategy that can *decompose* the CD price for co-partisan conversation into its constitutive elements.

The measurement strategy employed in this paper is to quantify social pressure as the extent to which willingness to speak *responds* to an exogenous change in the salience of affiliative concerns. In doing so we define social pressure as an essentially causal quantity. Examining whether an intervention that mitigates fears of social sanction subsequently reduces CD prices for hypothetical political conversations with discussants of various identities enables us to measure (a) whether social pressure is prevalent in co-partisan interactions and (b) whether heterogeneous treatment effects reveal symmetry or asymmetry in the degree to which party defectors and the party faithful experience social pressure in these contexts.

Methods

We embedded an experiment in a large, nationwide survey (N = 21,400) fielded through the Civic Health and Institutions Project (CHIP50).[6] The data were part of the broader wave 32 collection and were collected between June 18th and July 28th, 2024 by PureSpectrum,

[6]We thank the Civic Health and Institutions Project, a 50 States Survey (CHIP50), NSF Grants SES-2241884, SES-2241885, and SES-2241886, led by Matthew Baum, James Druckman, David Lazer, and Katherine Ognyanova, Principal Investigators.

an online panel management platform. Respondents were collected from all 50 states and D.C. Following our pre-registration, we subset our main analyses to the 17,691 partisans, including self-identified partisans and Democrat- or Republican-leaning independents.[7]

The survey occurred at a crucial moment for intra-party minorities, as both parties' nominees appeared certain but influential factions—including Never-Trump Republicans and Democrats opposed to the Biden administration's handling of Israel-Palestine—spoke against their respective nominees. Within each group, respondents who said they would vote for their party's respective nominee (Biden for Democrats, Trump for Republicans) if the "election were held today" are considered part of the *party majority*, whereas respondents who gave any other response (including "Not sure" or "I would not vote") are categorized as the *party minority*. Due to the scarcity of party minorities (21.25% in our final data), 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.[8]

Experimental Design

Our survey asked respondents to consider a series of hypothetical scenarios involving political discussion. We randomized several features of the survey in sequence, measuring different outcomes after each randomization. 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.[9]

[7]The 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 per the pre-registration.

[8]All 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.

[9]The 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 large variance in how respondents

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.[10] 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 and analogize our results to other areas where respondents hold exaggerated negative beliefs, including about the anti-democratic attitudes (Mernyk et al., 2022; Braley et al., 2023) and homogeneity (Ahler and Sood, 2018) of the opposing party.

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 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 democratic meta-perceptions (e.g., Braley et al., 2023; Druckman et al., 2023), where respondents are asked priors about second-order beliefs about other Americans' views, then randomly corrected to investigate how false beliefs affect democratic attitudes. This paper uses the ask-tell to exogenously

approach survey questions involving numerical scales. To improve precision, we include this baseline CD price in our pre-registered vector of covariates.

[10]The wording and true percentage for the first statement are adapted from the following 2022 Knight Foundation report: https://knightfoundation.org/wp-content/uploads/2022/01/KF_Free_Expression_2022.pdf. For the second and third statements, refer to the following October 2022 Times/Siena Poll: https://www.nytimes.com/interactive/2022/10/18/upshot/times-siena-poll-registered-voters-crosstabs.html.

reduce perceived affiliative costs, which we verify using a series of mechanism checks.

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. Two outcomes were subsequently measured:

- *Private Compensation Demands*: Respondents stated the lowest payments they would accept to send their one-minute video to the frequent discussant and two randomly selected targets. CDs were measured on a left-right integer scale from $0 to $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 and outcomes described above. An anonymous version of the pre-analysis plan is provided in Appendix 1.11. We hypothesized that CDs 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), diminish the perceived likelihood of social sanctions (H2B), and reduce all CDs on average (H3A). As our main test of intra-party social pressure, we hypothesized that the ask-tell correction would reduce CDs foor co-partisan opponents (H3C) and a co-partisan group discussion (H3D).

The remainder of the experiment was designed to assess the mechanisms and scope conditions governing intra-party social pressure. First, if social pressure is in fact rooted in reasoned judgments about the risk of social sanction, then the ask-tell effect should be larger for respondents with more pessimistic priors. Second, we address concerns that dissenters could show greater concern than the party faithful about social sanctions either (i) for speech posted to social media or (ii) in group conversations, even if social pressure is equally salient for these groups in a one-on-one environment. Finally, we test if our findings are limited to specific discussion topics or to a specific political party. For instance, we hypothesized that the ask-tell CATEs on prices demanded for co-partisan conversation may be larger for party minorities than for party majorities among Republicans alone (H5B), as Republicans have been the subject of most reports about intra-party censorship. 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 varying whether the topic would be the 2024 presidential election (prob = 0.5), "economic issues like taxation and social security'' (prob = 0.25), or "social issues like abortion and immigration'' (prob = 0.25).

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

Estimation

A full discussion of the model specifics for each outcome can be found 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 $\mathbf{\Phi}_i$ 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.

$$\text{Outcome}_i = \beta_0 + \beta_1 \tau_1 + \eta \mathbf{\Phi}_i + \epsilon_i \tag{1}$$

Model 2 accounts for the within-subjects nature of the private 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.

$$\text{Outcome}_{ic} = \beta_0 + \beta_1 \text{CD1}_c + \beta_2 \text{CD2}_c + \beta_3 \text{CD3}_c + \eta \mathbf{\Phi}_i + \epsilon_i \tag{2}$$

Results

In Figure 1, we illustrate the stark differences between party majorities and party minorities in the compensations they demand for the private video-sharing assignment. For party minorities, Hypothesis 1 is partially affirmed: 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 striking, 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.

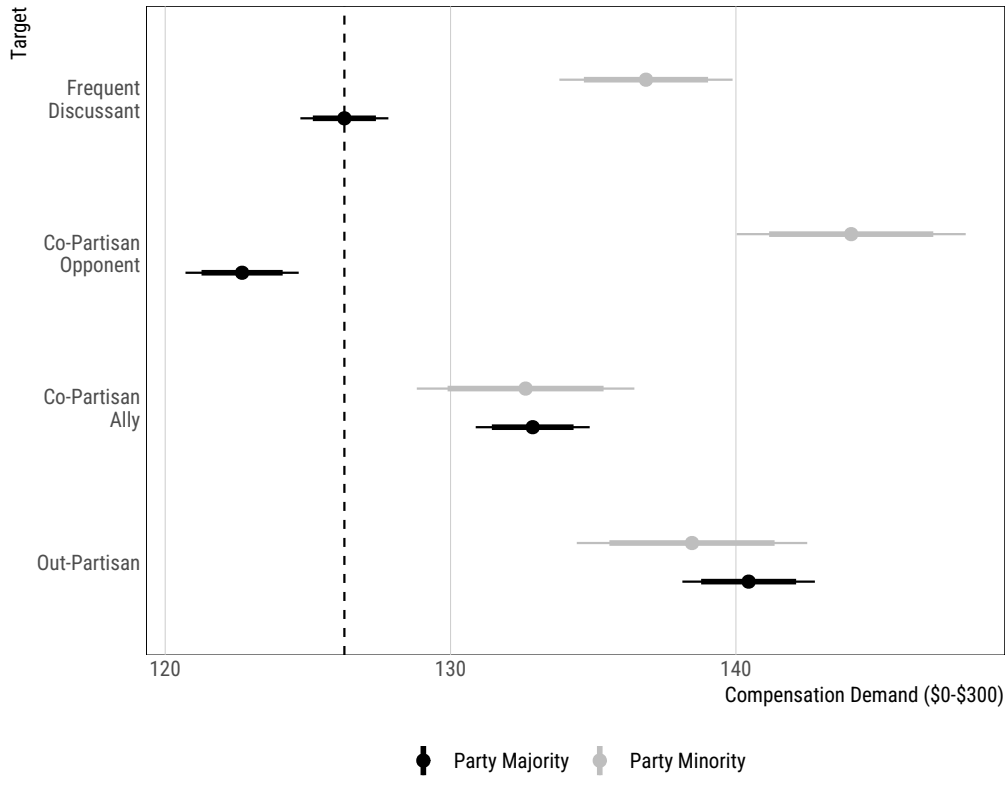

Figure 1: Private CDs by Party Majority/Minority, Full Model in A6.

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 private 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.

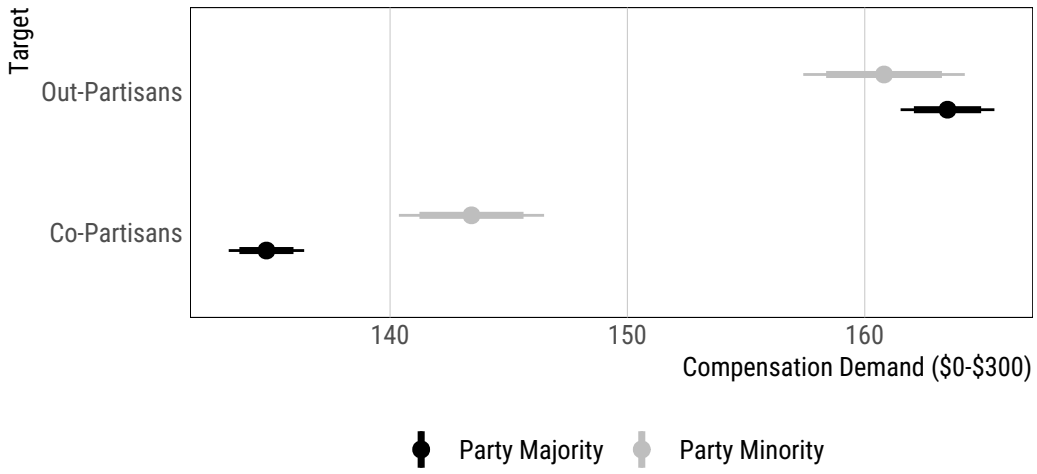

Figure 2: Group Conversation CDs by Party Majority/Minority, Full Model in Table A7.

At first glance, the CD results offer strong support for the theory that self-censorship in the 2024 election cycle resulted from partisan injunctive norms favoring majority opinion. Respondents were least willing to participate when they held reservations about their party's nominee and were asked to share these opinions with co-partisans in strong support of the nominee. However, as discussed in our theoretical section, price demands alone cannot isolate social pressure, instead representing a bundle of many considerations at once.

To determine whether this arises from an asymmetry in social pressure, we turn to the ask-tell correction which was designed to reduce fears of social sanction while leaving other motivations to speak unaffected. First, in Figure A11 in the Appendix, we plot the distributions of respondents' estimates for each of three survey items as well as their ask-tell

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***	−2.368**	−7.643***	−1.490	−0.337***
	(0.383)	(0.912)	(0.399)	(0.909)	(0.021)
Prior	0.436***	0.536***	0.454***	0.522***	0.015***
	(0.011)	(0.014)	(0.011)	(0.015)	(0.001)
Ask-Tell:Prior		−0.197***		−0.134***	
		(0.019)		(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 A1.

prior. 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, on average, under-estimated the true level of social sanctions. Like other misperceptions that Americans hold regarding the opinions of their peers (Ahler and Sood, 2023; Braley et al., 2023), these beliefs may contribute to baleful outcomes—in our case, the suppression of political speech. Misperceptions were relatively uniform across lines of party (and minority status within a party), race, education and gender, suggesting that Americans almost ubiquitously overestimate the extent to which their fellow citizens may retaliate against political speech.

Do these misperceptions help explain why Americans widely avoid political discussion? Table 1 (column 1) and Figure 3 (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 (-11.50 percentage points, SE = 0.38) - full models are available in Appendix Table A1. 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" (-7.71 pp, SE = 0.40), though the magnitude of this effect is smaller, as respondents

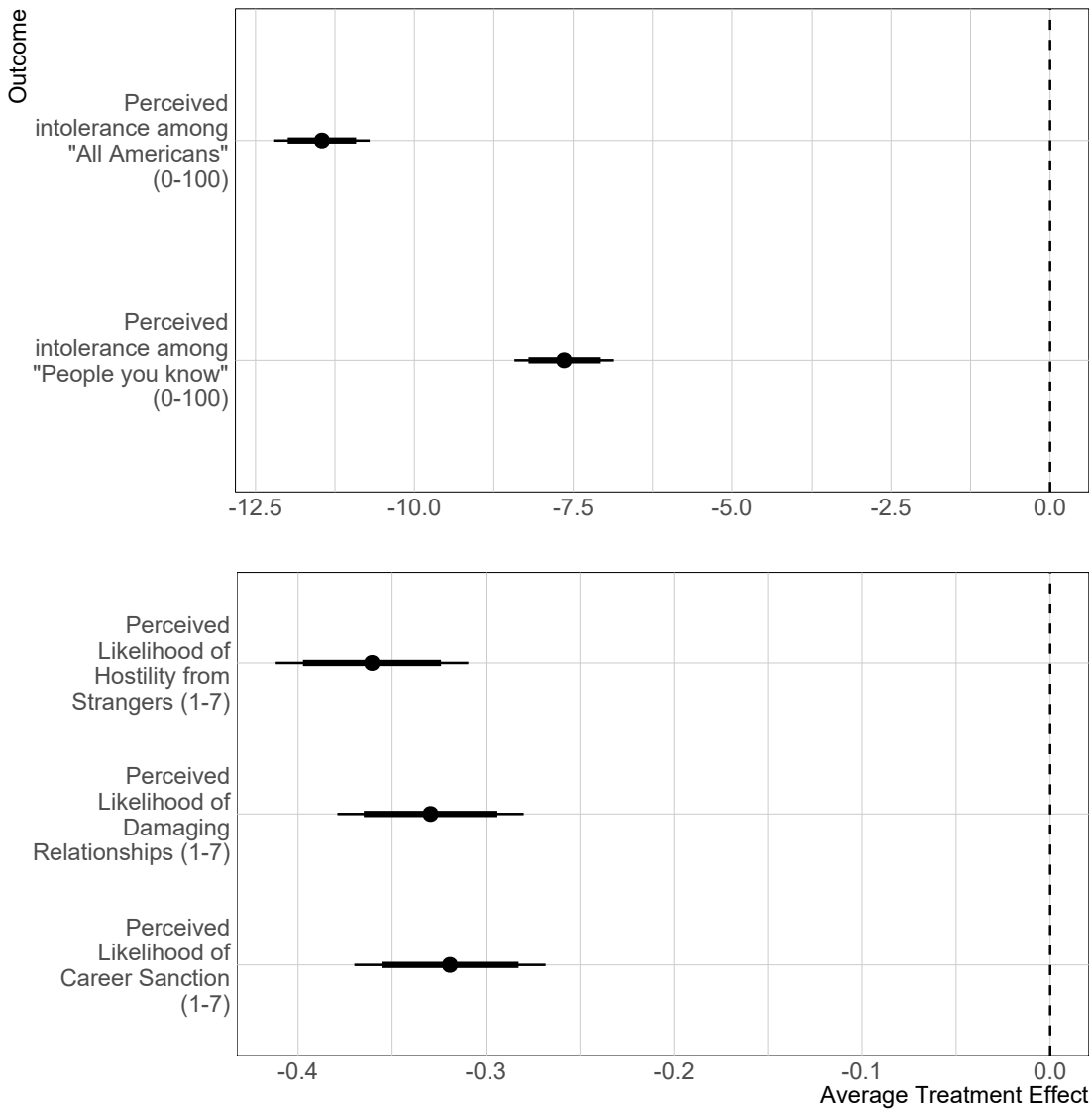

Figure 3: Ask-Tell ATEs on Mechanism Checks (Separate Regressions), Full Model in Table A8.

likely have stronger priors about their social circles than "All Americans" in aggregate. With regard to the risks of specific sanctions for the private CD task, ask-tell exposure significantly reduced the perceived likelihood of all three outcomes as well as their mean (-0.34, SE = 0.02).

Though social pressure is well-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 consid-

ered 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.

Turning to the main effects of the ask-tell, Figure 4 shows that ask-tell exposure reduced compensation demands for discussing the election with co-partisan opponents, evincing the prevalence of social pressure in intra-party networks (full models in A9). The lower the perceived risk of social sanction, the more people were willing to speak about their party's nominee with co-partisan peers. Importantly, the ask-tell correction substantially reduced reluctance toward both one-on-one and group conversations (confirming H3a and H3b).

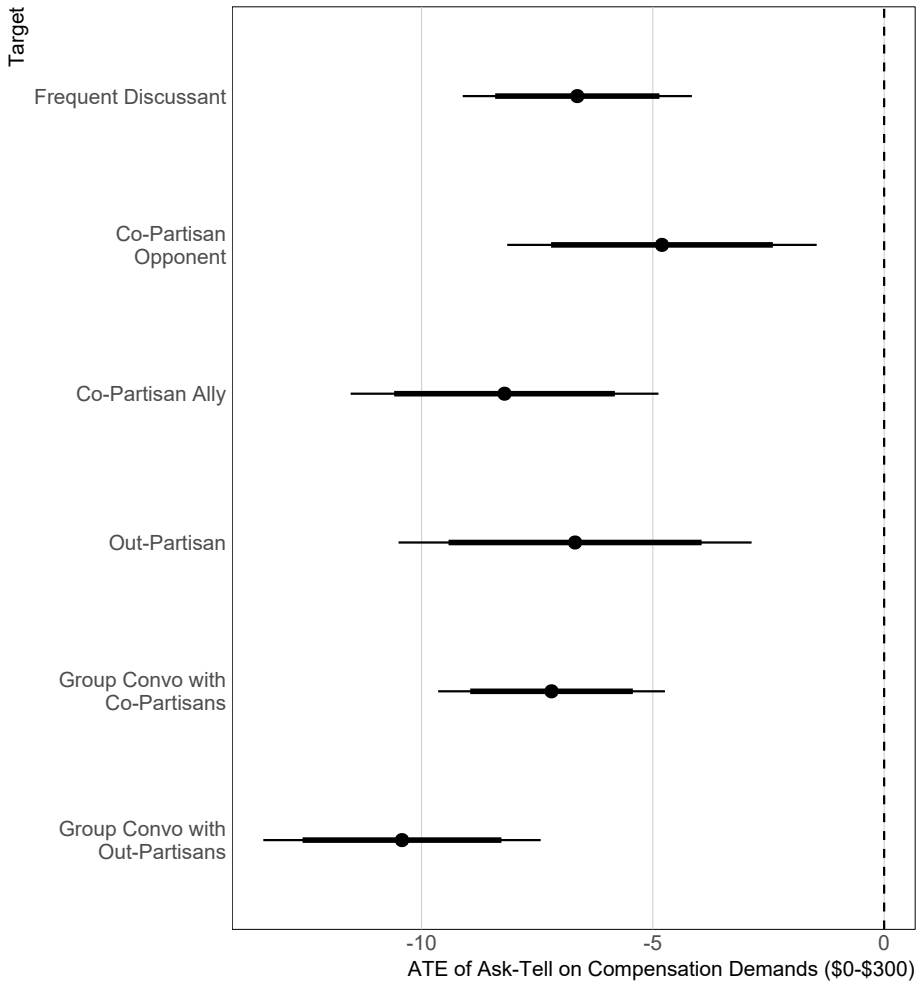

Figure 4: Ask-Tell ATE on Private and Group CDs, Full Models in Table A9.

It must be noted, however, that ask-tell exposure reduced compensation demands not just for co-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 we might expect ideological or partisan differences should be smaller or at least less salient. These results accord with an understanding of social pressure as less about the specific substance of one's beliefs than about a broader aversion to political discussion irrespective of one's own minority or majority status. Intra-party social pressure is real, but may not be rooted in a unique set of norms specific to one's party network.

Consistent with this interpretation, we find that reduction in social pressure from the ask-tell is not driven solely by members of the party minority. Interacting the ask-tell treatment indicator with another for party minority status, the main effects listed in Table 2 indicate that for party majorities alone, our correction reduced compensation demands for discussing the election with co-partisan opponents privately (-\$3.79, SE = 1.89) and with a group of co-partisans (-\$6.43, SE = 2.39). Though the average effect sizes for party minorities more than double those of party majorities in magnitude, suggesting minorities may face increased social pressure, we cannot reject the null of no difference between the sub-group effects. Concerning the hypothesis that asymmetry in social pressure may be more pertinent for Republicans than Democrats, the effect is in the right direction: Table 3 shows that, for both outcomes, the difference in CATEs between party majorities and minorities is smaller for Democrats than for Republicans. But we again fail to reject the null of no difference across parties. While social pressure is not uniform, fear of social sanction serves to constrain intra-party speech across the board, irrespective of party, social context, or the normative status of one's beliefs.

Alternative Routes to Majority Influence

One concern with the analysis above may be that it fails to uncover evidence of asymmetry between majorities and minorities because minority status is defined only with respect to a global distribution of opinion within the Democratic and Republican parties. Majority influence is said to occur through varied channels according to the literature on social pressure.

Asch's (1956) line-judgment experiments, for instance, leveraged majority influence within an immediate group context, as do recent laboratory experiments (Ho and Huang, 2024). The hypothetical assignment described in the group conversation CD was intended to bring such a context to mind, implying to party minorities that they would likely be in the minority if forced into conversation with a group of strangers. But another way majority influence may operate is through network composition. Perhaps perceived stigmatization is salient primarily for people who find themselves in the minority within their personal networks, not just the party writ large. Mutz (2002) famously argued that people who encounter interpersonal political disagreements the most often tend to subsequently withdraw from political activity as a result of cross-pressures. Or perhaps asymmetry is conditional upon the type of speech, such that party defectors disproportionately self-censor in online speech acts that would broadcast their opinions publicly to their social media network (Schulz, 2024), even if they act similarly to the party faithful in private exchanges or group conversations with strangers.

We leverage two components of our design to examine whether considerations about position within one's personal network may be a source of meaningful asymmetry in social pressure. First, we explore the effects of our social media treatment. As Table 4 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

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

	Co-Partisan Opp. CD	Co-Partisan Group CD
Ask-Tell	−3.804*	−6.467**
	(1.889)	(2.384)
Party Min	21.609***	11.183**
	(3.184)	(4.161)
Ask-Tell:Party Min	−4.985	−7.348
	(4.364)	(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 A2.

Table 3: Does Majority/Minority Asymmetry in Ask-Tell Effects Vary by Party?

	Co-Partisan Opp. CD	Co-Partisan Group CD
Ask-Tell Treatment	−4.549	−4.861*
	(2.786)	(1.946)
Party Minority	28.131***	14.830***
	(5.021)	(3.542)
Democrat	−9.135**	6.097*
	(3.197)	(2.374)
Ask-Tell:Party Min	−7.159	−6.260
	(6.999)	(5.044)
Ask-Tell:Dem	1.459	−4.158
	(3.787)	(2.773)
Party Min:Dem	−11.354+	−10.160*
	(6.453)	(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 A3.

Table 4: Social Media Prime Main Effects

	(1) Sanction Mean (H6a)	(2) Career	(3) Affiliative	(4) Strangers
Social Media Prime	0.178***	0.131***	0.122***	0.283***
	(0.021)	(0.026)	(0.025)	(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 A4.

to the social media prime with party minority status, we find no evidence that the CATE was stronger for party minorities than majorities (Figure 5).

Second, we leverage pre-treatment questions which 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.

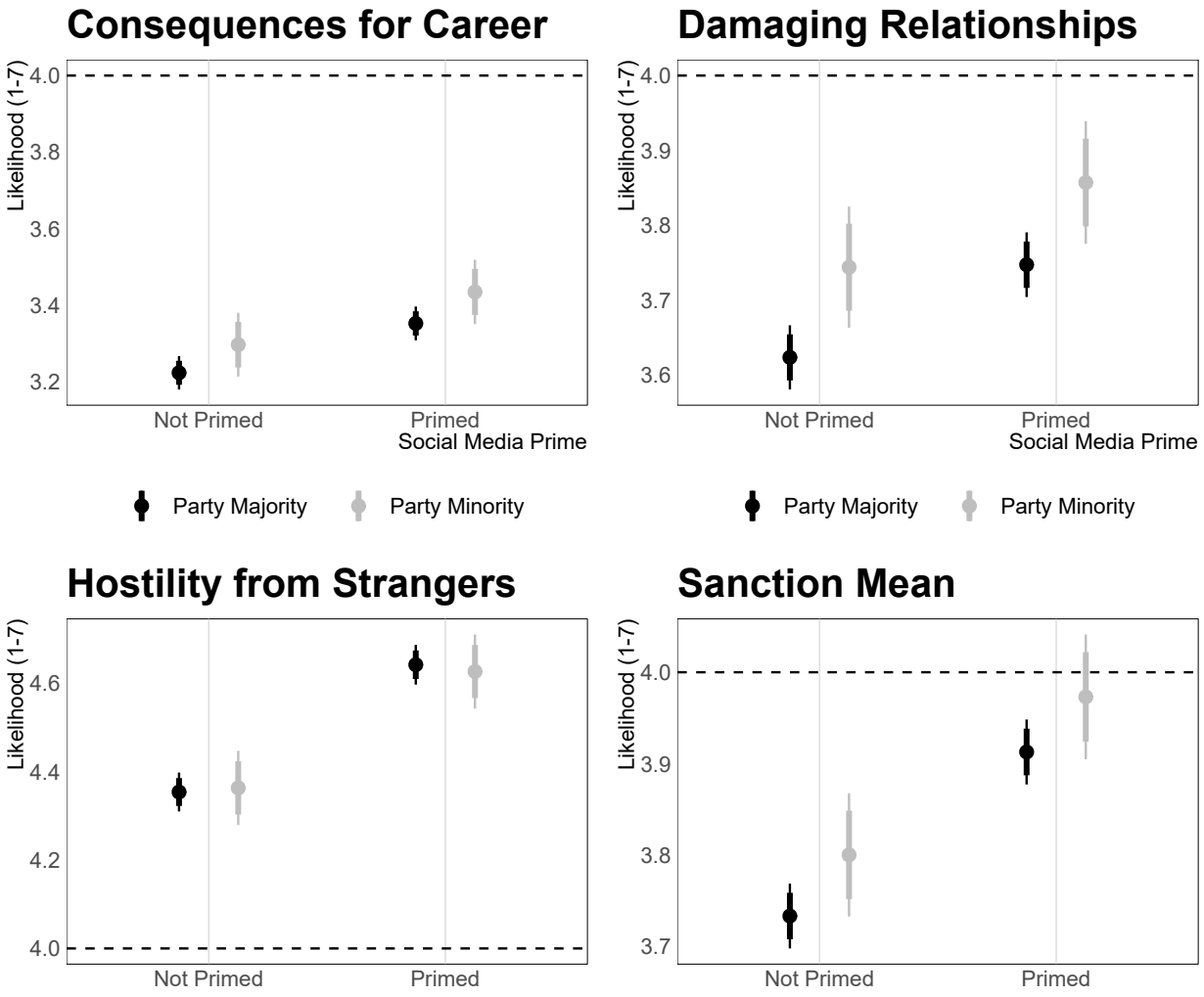

Figure 5: Estimated Marginal Means by Social Media Prime and Party Minority Status, Full Models in Table A10.

Table 5 shows that both proxies for self-reported network share strongly predict the 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. Publicly sharing views appears no more costly for those with more adverse 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 5: 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 A5.

Why do Party Minorities Speak Less if Social Pressure is Roughly Uniform?

Given these findings, we conducted a follow-up survey on a different sample to investigate alternate explanations to the gap in willingness to speak to co-partisans between partisan minorities and majorities. In light of our observed treatment effects, we hypothesized that differential social pressure cannot explain the full scope of party minorities' disproportionate self-censorship. We fielded a short module in a survey with $N = 3,400$ partisans or partisan-leaners. Respondents were recruited through CloudResearch Connect between October 31 and November 4, just before Election Day on November 5. Respondents were asked whether they personally knew a co-partisan who disagreed with them about whom to vote for in the 2024 presidential election, to which 1,001 responded affirmatively. Respondents were then shown a list of four reasons why they might start a discussion about the election with this person (or a co-partisan opponent they envision) and asked to select all that apply or "None of the above." A similar question was asked containing four reasons why they might

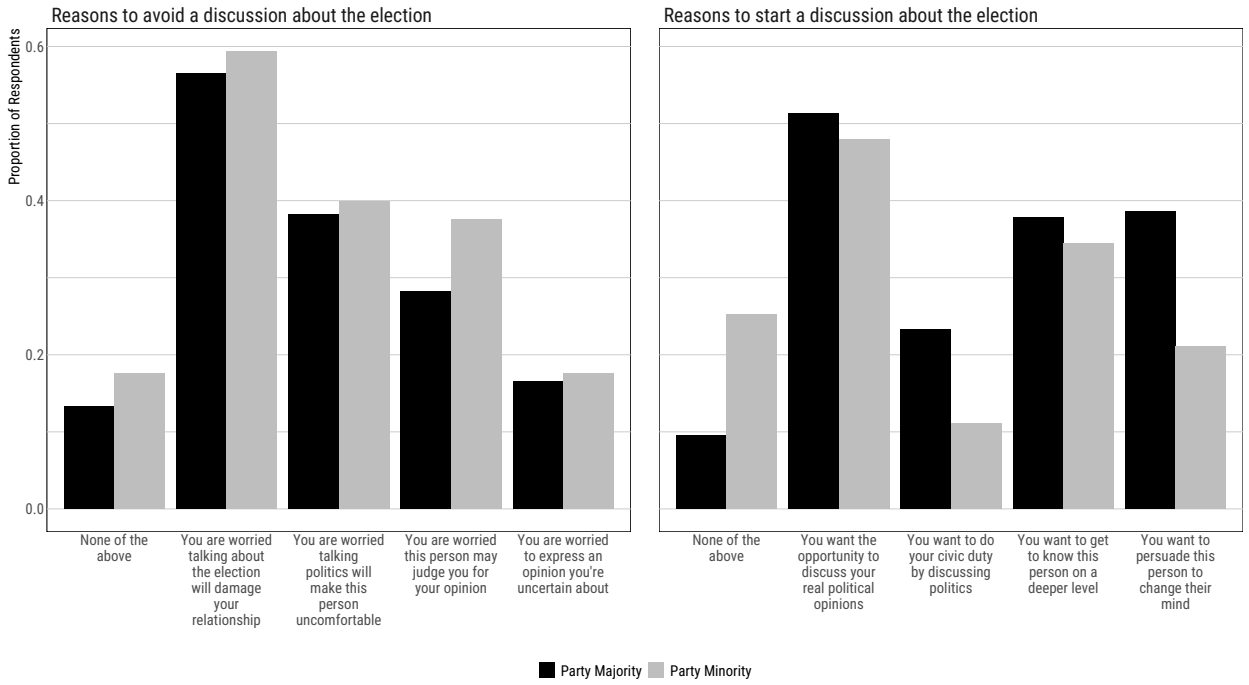

Figure 6: Considerations for Election Discussions with Co-Partisan Opponents (Know a Co-Opp), Full Models in Figures A11 and A12.

avoid a discussion about the election. Figure 6 visualizes the proportion of respondents who selected each reason, disaggregated for party majorities and minorities. This analysis includes only those respondents who said they know a co-partisan opponent, 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. Regressing whether each reason is cited only on an indicator for party minority status, we find that party majorities were about 17.6 percentage points more likely to cite persuasion as a perceived benefit to starting a conversation on the election (SE = 0.03). Party majorities were also 12.1 percentage points more likely to cite a civic duty of discussing politics (SE = 0.02), potentially illuminating dispositional differences in engagement between the party faithful and party defectors (e.g., Krupnikov and Ryan, 2022). Party minorities were 15.7 percentage points more likely to say they see none of the listed considerations as potential benefits for speaking with co-partisan opponents (SE = 0.03).

By contrast, we find no statistically significant difference in the frequency with which the two groups cite concerns about damaging relationships or making the other discussant comfortable as a reason to avoid discussing the election. Among majorities and minorities alike, nearly 60 percent affiliative costs as a reason to avoid intra-party speech—by far the most salient consideration. Party minorities were more likely to mention only one concern: fearing judgment for their opinion (9.3 pp, SE = 0.03).

Differing strategic motivations, not a significant gap in social pressure, may be the principal explanation for the large asymmetry between party majorities and minorities in the compensations they demand for co-partisan opponents. This interpretation coheres with the striking finding that party majorities demanded lower prices for co-partisan opponents than even their frequent discussants.

Discussion

In this paper, we have presented evidence that minorities within their own party are profoundly unwilling to speak to co-partisans who toe the party line and that this disparity in self-censorship extends across parties and topics. Strikingly, individuals who dissent from their party's majority view in some cases prefer sharing their views with opposing partisans even over members of their party's majority.

In the ask-tell portion of our experiment, we documented how Americans across party lines overestimate the share of their fellow citizens who have either experienced or sought to inflict social sanctions in response to political speech. This misperception is substantially large, and future studies should examine how these misperceptions find their origin—for instance, in mass media or elite rhetoric. In addition, our design advances the study of social influence by demonstrating how scholars can elucidate the particular role played by social pressure in political behavior, decomposing its causal effect from that of other social considerations, with which it is often bundled. To date, studies of social pressure have almost ubiquitously employed group-level interventions (whether in the form of descriptive norm change or social observation primes), leading scholars to equate social pressure with majority

influence. By manipulating the salience of social pressure directly, our intervention is able to demonstrate concretely that social pressure exists in people's one-on-one interactions, in line with the intuition that social pressure represents "a basic human drive to win praise and avoid chastisement" (Green and Gerber, 2010). This affiliative goal governs all interpersonal interactions, not just those in group environments. While we cannot claim that our treatment eliminates social pressure, we show that it created substantially large shifts in perceptions of others' beliefs and in fears of social sanction. In contrast to similar research designs, this treatment served to vary social pressures concerns alone.

Using this empirical strategy, we find—contrary to some theoretical expectations—that fears of social sanction for political speech are not confined to any one group. Social pressure is prevalent across factional lines, with members of the party majority and minority alike exhibiting significant concern about affiliative costs to honest speech. Both groups were substantially more willing to speak when we reduced those concerns with an informational correction, even as effects differed modestly between sub-groups. This effect persisted over all possible conversation pairs, contexts, and topics, indicating that social pressure is ubiquitous in American political life, although there are some contexts in which it may be more severe. Even presumably congenial discussions between agreeable conversation partners are not free of social pressure concerns.

While there is partial evidence that minorities feel more social pressure (including in descriptive analyses of pre-treatment questions), this gap does not come close to explaining different groups' willingness to speak. Instead—as we demonstrated in our descriptive follow-up survey—partisan majorities have greater positive motivations to speak to co-partisan dissenters than vice-versa. That party majorities' willingness to discuss the election responds to ask-tell exposure and the CATEs for majorities and minorities differ minimally suggest intra-party social pressure is borne more from a substance-neutral aversion to a highly contentious political climate than from the perceived stigmatization of one's personal preferences.

While prior work finds that party identification can survive electoral defection in a general election (Green, Palmquist and Schickler, 2002), our findings suggest those who do not support their party's nominee find interactions with co-partisans costly, which may affect

their long-term partisan identities through their selection into discussions and social circles. This mechanism can transform electoral defection in a single election into longer-term shifts in partisan social identity. Our treatment effects serve only to measure a snapshot, holding fixed individuals' networks at a given moment in time, but future studies should examine how social isolation may produce cascading effects on partisan commitments over successive periods.

Our findings call attention to an important distinction in research on self-censorship and free speech. Both public media coverage, especially in partisan outlets, and some scholarship (e.g., Noelle-Neumann, 1974; Ho and Huang, 2024) have drawn attention to a form of social sanctioning that is carried out by one political faction to suppress the beliefs espoused by another, often making analogies to authoritarian repression more explicit. Through a large, nationwide sample, we show that this is not how most Americans experience pressure to self-censor. Not only are social fears not concentrated in any one faction, but they are also relevant to diverse inter- and intra-party contexts and extend to hypothetical discussions likely to be avoid or minimize substantive disagreement. In today's political climate, Americans across all party and factional lines fear social repercussions from political speech, even as party minorities may feel more pressure to avoid sharing beliefs that deviate from their group norms.

References

Abdul-Hakim, Gabriella, Fritz Farrow, Libby Cathey, Oren Oppenheim and Adam Carlson. 2024. "Anti-Biden 'uncommitted' voters set to win 2 delegates in Michigan primary, boosting protest." *ABC News* .

Abramson, Paul R., John H. Aldrich and David W. Rohde. 1983. *Change and Continuity in the 1980 Elections*. CQ Press.

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. Publisher: The University of Chicago Press.

Ahler, Douglas J. and Gaurav Sood. 2023. "Typecast: A Routine Mental Shortcut Causes Party Stereotyping." *Political Behavior* 45(4):1581–1607.

Aratani, Lauren. 2024. "Vote against Trump, former supporters urge in $50m video campaign." *The Guardian* .

Asch, S. E. 1951. Effects of group pressure upon the modification and distortion of judgments. In *Groups, leadership and men; research in human relations*. Oxford, England: Carnegie Press pp. 177–190.

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.

Bail, Chris. 2022. Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing. In *Breaking the Social Media Prism*. Princeton University Press.

Barber, Michael and Jeremy C. Pope. 2019. "Does Party Trump Ideology? Disentangling Party and Ideology in America." *American Political Science Review* 113(1):38–54.

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.

Black, Jerome H. 1978. "The Multicandidate Calculus of Voting: Application to Canadian Federal Elections." *American Journal of Political Science* 22(3):609–638.

Brady, Henry E., Sidney Verba and Kay Lehman Schlozman. 1995. "Beyond SES: A Resource Model of Political Participation." *The American Political Science Review* 89(2):271–294.

Braghieri, Luca. 2024. "Political Correctness, Social Image, and Information Transmission." *American Economic Review* 114(12):3877–3904.

Braley, Alia, Gabriel S. Lenz, Dhaval Adjodah, Hossein Rahnama 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.

Burt, Ronald S. 1984. "Network Items and the General Social Survey." *Social Networks* 6(4):293–339.

Carey, John, Brian Fogarty, Marília Gehrke, Brendan Nyhan and Jason Reifler. 2024. "Prebunking and Credible Source Corrections Increase Election Credibility: Evidence from the U.S. and Brazil." *Working Paper* .

Carlson, Taylor N. and Jaime E. Settle. 2022. *What Goes Without Saying: Navigating Political Discussion in America*. Cambridge: Cambridge University Press.

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.

Coppock, Alexander. 2023. *Persuasion in Parallel: How Information Changes Minds about Politics*. University of Chicago Press.

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* .

Druckman, James N., Suji Kang, James Chu, Michael N. Stagnaro, Jan G. Voelkel, Joseph S. Mernyk, Sophia L. Pink, Chrystal Redekopp, David G. Rand and Robb Willer. 2023. "Correcting misperceptions of out-partisans decreases American legislators' support for undemocratic practices." *Proceedings of the National Academy of Sciences* 120(23):e2301836120.

Dunn, Amina. 2021. "Two-thirds of Republicans want Trump to retain major political role; 44% want him to run again in 2024.".

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.

Gentzkow, Matthew and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *The Quarterly Journal of Economics* 126(4):1799–1839.

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. and Todd Rogers. 2009. "Descriptive Social Norms and Motivation to Vote: Everybody's Voting and so Should You." *The Journal of Politics* 71(1):178–191.

Gibson, James L. 1992. "The Political Consequences of Intolerance: Cultural Conformity and Political Freedom." *American Political Science Review* 86(2):338–356.

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.

Goffman, Erving. 1963. *Stigma: Notes on the Management of Spoiled Identity*. Touchstone.

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.

Ho, Yuen and Yihong Huang. 2024. "Breaking the Spiral of Silence.".

Huckfeldt, Robert and John Sprague. 1987. "Networks in Context: The Social Flow of Political Information." *American Political Science Review* 81(4):1197–1216.

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.

Kalla, Joshua L. and David E. Broockman. 2020. "Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments." *American Political Science Review* 114(2):410–425.

Kawai, Kei and Yasutora Watanabe. 2013. "Inferring Strategic Voting." *The American Economic Review* 103(2):624–662.

Krupnikov, Yanna and John Barry Ryan. 2022. *The Other Divide*. Cambridge: Cambridge University Press.

Kuran, Timur. 1991. "Now out of Never: The Element of Surprise in the East European Revolution of 1989." *World Politics* 44(1):7–48.

Kuran, Timur. 1995. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge: Harvard University Press.

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* .

Menzner, Jan and Richard Traunmüller. 2023. "Subjective Freedom of Speech: Why Do Citizens Think They Cannot Speak Freely?" *Politische Vierteljahresschrift* 64(1):155–181.

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.

Milgram, Stanley. 1974. *Obedience to Authority: An Experimental View*. Harper & Row.

Mutz, Diana C. 2002. "Cross-Cutting Social Networks: Testing Democratic Theory in Practice." *The American Political Science Review* 96(1):111–126.

Noelle-Neumann, Elisabeth. 1974. "The Spiral of Silence A Theory of Public Opinion." *Journal of Communication* 24(2):43–51.

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.

Schultz, P. Wesley. 1999. "Changing Behavior With Normative Feedback Interventions: A Field Experiment on Curbside Recycling." *Basic and Applied Social Psychology* 21(1):25–36.

Schulz, William Small. 2024. "What Would You Say on the Internet? How Moderate Speakers' Self-Censorship Polarizes Online Discourse." *Draft Manuscript* pp. 1–23.

Settle, Jaime E. and Taylor N. Carlson. 2019. "Opting Out of Political Discussions." *Political Communication* 36(3):476–496.

Sheagley, Geoffrey, Logan Dancey and John Henderson. 2023. "A Good Partisan? Ideology, Loyalty, and Public Evaluations of Members of Congress." *Legislative Studies Quarterly* 48(3):657–679.

Spelman, Trevor, Abdo Elnakouri, Nour Kteily and Eli Finkel. 2024. "Overestimating the Social Costs of Political Belief Change.".

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.

Sunstein, Cass R. 2018. *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press.

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*a*. "Norms of Democracy, Staged Democrats, and Supply of Exclusionary Ideology." *Comparative Political Studies* p. 00104140241283009.

Valentim, Vicente. 2024*b*. "Political Stigma and Preference Falsification: Theory and Observational Evidence." *The Journal of Politics* 86(4):1382–1402.

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

1.1 Full Models

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

Table A1: 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***	16.625***	13.747***	10.622***	2.879***
	(1.323)	(1.380)	(1.359)	(1.417)	(0.072)
Ask-Tell	−11.455***	−2.368**	−7.643***	−1.490	−0.337***
	(0.383)	(0.912)	(0.399)	(0.909)	(0.021)
Dem	2.603***	2.640***	3.275***	3.305***	−0.033
	(0.522)	(0.521)	(0.533)	(0.532)	(0.029)
Demand Base	0.021***	0.021***	0.024***	0.024***	0.001***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.000)
Party Minority	−3.140***	−3.108***	−3.632***	−3.609***	0.060*
	(0.486)	(0.484)	(0.506)	(0.505)	(0.027)
Ideology	−0.532*	−0.529*	−1.002***	−0.996***	0.047***
	(0.251)	(0.250)	(0.261)	(0.260)	(0.014)
Pol. Int. (5-point)	1.355***	1.370***	1.466***	1.480***	0.009
	(0.204)	(0.203)	(0.210)	(0.210)	(0.011)
Education	−0.106	−0.107	−0.249	−0.252	0.066***
	(0.216)	(0.215)	(0.223)	(0.223)	(0.012)
Prior	0.436***	0.536***	0.454***	0.522***	0.015***
	(0.011)	(0.014)	(0.011)	(0.015)	(0.001)
Network Co-Opp	8.172***	8.148***	9.046***	9.024***	0.326***
	(1.132)	(1.134)	(1.188)	(1.188)	(0.062)
Network Opposing	1.370	1.398	1.730	1.736	0.927***
	(1.045)	(1.042)	(1.098)	(1.097)	(0.056)
Female	−2.791***	−2.830***	−2.307***	−2.334***	0.099***
	(0.397)	(0.396)	(0.415)	(0.415)	(0.022)
White	0.419	0.418	−0.346	−0.340	0.052*
	(0.475)	(0.473)	(0.496)	(0.495)	(0.025)
Married	0.568	0.574	0.084	0.092	0.054*
	(0.425)	(0.424)	(0.440)	(0.439)	(0.023)
Fully Employed	−0.162	−0.198	0.575	0.555	0.059*
	(0.457)	(0.455)	(0.479)	(0.479)	(0.025)
SM Prime					0.175***
					(0.021)
Ask-Tell:Prior		−0.197***		−0.134***	
		(0.019)		(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 A2: Full Models, Ask-Tell Effects by Minority Status (Table 2)

	Co-Partisan Opponent CD	Co-Partisan Group CD
Intercept	38.581***	47.233***
	(5.691)	(7.346)
Ask-Tell	−3.706*	−6.418**
	(1.888)	(2.380)
Dem	−11.059***	3.618
	(2.316)	(3.072)
Demand Base	0.510***	0.612***
	(0.010)	(0.012)
Party Minority	21.652***	10.938**
	(3.181)	(4.168)
Ideology	−3.044**	0.689
	(1.080)	(1.448)
Pol. Int. (5-point)	−2.812**	−4.448***
	(0.879)	(1.134)
Education	5.276***	−1.935
	(0.940)	(1.229)
Prior	0.329***	0.309***
	(0.046)	(0.056)
Network Co-Opp	31.755***	17.908**
	(4.845)	(6.115)
Network Opposing	19.974***	10.654+
	(4.439)	(5.642)
Female	4.447*	8.159***
	(1.764)	(2.246)
White	5.248**	−2.416
	(2.004)	(2.612)
Married	1.347	2.926
	(1.909)	(2.441)
Fully Employed	4.268*	7.636**
	(2.008)	(2.595)
Ask-Tell:Party Min	−5.203	−7.213
	(4.363)	(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 A3: 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***	58.807***
	(5.771)	(4.361)
Ask-Tell	−4.464	−4.896*
	(2.786)	(1.945)
Dem	−9.234**	6.001*
	(3.193)	(2.371)
Demand Base	0.509***	0.619***
	(0.010)	(0.007)
Party Minority	27.884***	14.800***
	(5.032)	(3.542)
Ideology	−2.565*	0.539
	(1.100)	(0.833)
Pol. Int. (5-point)	−2.820**	−4.930***
	(0.878)	(0.649)
Education	5.159***	−0.391
	(0.942)	(0.702)
Prior	0.332***	0.292***
	(0.046)	(0.033)
Network Co-Opp	30.623***	14.458***
	(4.875)	(3.533)
Network Opposing	19.793***	6.014+
	(4.439)	(3.301)
Female	4.484*	2.926*
	(1.763)	(1.291)
White	5.332**	−2.406
	(2.005)	(1.506)
Married	1.278	3.544*
	(1.909)	(1.400)
Fully Employed	4.286*	3.294*
	(2.007)	(1.483)
Ask-Tell:Party Min	−7.197	−6.236
	(7.004)	(5.042)
Ask-Tell:Dem	1.486	−4.065
	(3.787)	(2.772)
Party Min:Dem	−10.835+	−10.250*
	(6.457)	(4.637)
Ask-Tell:Party Min:Dem	3.197	9.527
	(8.946)	(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 A4: Full Models, Social Media Prime Main Effects (Table 4)

	(1) Sanction Mean (H6a)	(3) Career	(4) Affiliative	(5) Strangers
Intercept	2.696***	2.857***	2.213***	3.021***
	(0.072)	(0.089)	(0.087)	(0.089)
Dem	−0.035	−0.084*	0.008	−0.029
	(0.029)	(0.035)	(0.035)	(0.035)
Demand Base	0.001***	0.001***	0.001***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Party Minority	0.064*	0.078*	0.115***	−0.003
	(0.028)	(0.034)	(0.033)	(0.034)
Ideology	0.045**	0.114***	0.048**	−0.028+
	(0.014)	(0.017)	(0.017)	(0.017)
Pol. Int. (5-point)	0.011	−0.032*	−0.016	0.082***
	(0.011)	(0.014)	(0.013)	(0.014)
Education	0.065***	0.035*	0.056***	0.102***
	(0.012)	(0.015)	(0.014)	(0.015)
Prior	0.015***	0.015***	0.016***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)
Network Co-Opp	0.329***	0.364***	0.448***	0.176*
	(0.062)	(0.077)	(0.074)	(0.076)
Network Opposing	0.924***	0.668***	1.301***	0.792***
	(0.057)	(0.070)	(0.070)	(0.070)
Female	0.099***	0.005	0.107***	0.184***
	(0.022)	(0.027)	(0.026)	(0.027)
White	0.054*	−0.199***	0.145***	0.215***
	(0.026)	(0.032)	(0.031)	(0.032)
Married	0.049*	0.013	0.094***	0.041
	(0.023)	(0.029)	(0.028)	(0.029)
Fully Employed	0.060*	0.271***	0.000	−0.096**
	(0.025)	(0.031)	(0.030)	(0.031)
SM Prime	0.178***	0.131***	0.121***	0.283***
	(0.021)	(0.026)	(0.025)	(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 A5: 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 A6: 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 A7: Full Models, Group Convos (Figure 2)

	Co-Partisans	Out-Partisans
Intercept	56.209***	78.297***
	(4.255)	(5.109)
Dem	2.384	4.768*
	(1.738)	(1.992)
Demand Base	0.620***	0.481***
	(0.007)	(0.008)
Party Minority	8.640***	−2.675
	(1.681)	(1.885)
Ideology	0.233	−4.593***
	(0.819)	(0.961)
Pol. Int. (5-point)	−4.878***	−2.844***
	(0.650)	(0.788)
Education	−0.382	6.662***
	(0.702)	(0.861)
Prior	0.290***	0.367***
	(0.033)	(0.041)
Network Co-Opp	15.268***	−26.868***
	(3.505)	(4.392)
Network Opposing	6.045+	3.276
	(3.305)	(4.056)
Female	2.935*	8.985***
	(1.293)	(1.585)
White	−2.383	11.643***
	(1.508)	(1.808)
Married	3.498*	3.318+
	(1.401)	(1.705)
Fully Employed	3.348*	−1.763
	(1.485)	(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 A8: Full Models, Ask-Tell ATEs on Mechanism Checks (Figure 3)

	Career Sanction	Relation Sanction	Stranger Sanction	All Americans	People You Know
Intercept	3.031***	2.393***	3.217***	21.205***	13.747***
	(0.090)	(0.087)	(0.090)	(1.323)	(1.359)
Ask-Tell	−0.319***	−0.329***	−0.361***	−11.455***	−7.643***
	(0.026)	(0.025)	(0.026)	(0.383)	(0.399)
Dem	−0.081*	0.011	−0.026	2.603***	3.275***
	(0.035)	(0.035)	(0.035)	(0.522)	(0.533)
Demand Base	0.001***	0.001***	0.000	0.021***	0.024***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)
Party Minority	0.075*	0.112***	−0.007	−3.140***	−3.632***
	(0.034)	(0.033)	(0.034)	(0.486)	(0.506)
Ideology	0.116***	0.050**	−0.026	−0.532*	−1.002***
	(0.017)	(0.017)	(0.017)	(0.251)	(0.261)
Pol. Int. (5-point)	−0.034*	−0.018	0.080***	1.355***	1.466***
	(0.014)	(0.013)	(0.014)	(0.204)	(0.210)
Education	0.036*	0.057***	0.104***	−0.106	−0.249
	(0.014)	(0.014)	(0.015)	(0.216)	(0.223)
Prior	0.015***	0.016***	0.015***	0.436***	0.454***
	(0.001)	(0.001)	(0.001)	(0.011)	(0.011)
Network Co-Opp	0.361***	0.444***	0.172*	8.172***	9.046***
	(0.077)	(0.074)	(0.076)	(1.132)	(1.188)
Network Opposing	0.671***	1.305***	0.796***	1.370	1.730
	(0.070)	(0.070)	(0.069)	(1.045)	(1.098)
Female	0.006	0.108***	0.185***	−2.791***	−2.307***
	(0.027)	(0.026)	(0.027)	(0.397)	(0.415)
White	−0.202***	0.142***	0.212***	0.419	−0.346
	(0.031)	(0.031)	(0.032)	(0.475)	(0.496)
Married	0.018	0.098***	0.046	0.568	0.084
	(0.029)	(0.028)	(0.029)	(0.425)	(0.440)
Fully Employed	0.270***	−0.001	−0.097**	−0.162	0.575
	(0.031)	(0.030)	(0.031)	(0.457)	(0.479)
SM Prime	0.128***	0.119***	0.280***		
	(0.026)	(0.025)	(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 A9: Full Models, Ask-Tell Effects (Figure 4)

	Group Out	Group In	Out-Partisan	Co-Ally	Co-Opponent	Discussant
Intercept	83.959***	60.115***	47.083***	48.896***	39.031***	25.641***
	(5.181)	(4.317)	(6.534)	(5.710)	(5.676)	(4.263)
Ask-Tell	−10.423***	−7.191***	−6.682***	−8.207***	−4.803**	−6.634***
	(1.530)	(1.251)	(1.947)	(1.698)	(1.706)	(1.264)
Dem	4.824*	2.440	2.248	−2.326	−11.047***	−3.720*
	(1.989)	(1.736)	(2.549)	(2.305)	(2.317)	(1.700)
Demand Base	0.480***	0.619***	0.452***	0.575***	0.510***	0.648***
	(0.008)	(0.007)	(0.010)	(0.009)	(0.010)	(0.007)
Party Minority	−2.800	8.565***	0.287	−0.192	19.056***	10.423***
	(1.883)	(1.679)	(2.371)	(2.225)	(2.293)	(1.696)
Ideology	−4.552***	0.269	−3.580**	0.945	−3.031**	−0.794
	(0.960)	(0.818)	(1.212)	(1.093)	(1.081)	(0.779)
Pol. Int. (5-point)	−2.904***	−4.919***	−1.236	−0.784	−2.808**	−1.434*
	(0.787)	(0.649)	(0.986)	(0.880)	(0.879)	(0.650)
Education	6.716***	−0.345	8.597***	−0.434	5.259***	3.537***
	(0.860)	(0.702)	(1.100)	(0.963)	(0.940)	(0.704)
Prior	0.367***	0.290***	0.407***	0.380***	0.329***	0.235***
	(0.041)	(0.033)	(0.052)	(0.046)	(0.046)	(0.034)
Network Co-Opp	−27.026***	15.149***	−15.846**	10.109*	31.802***	8.210*
	(4.389)	(3.503)	(5.345)	(4.764)	(4.844)	(3.391)
Network Opposing	3.423	6.122+	29.797***	−16.448***	19.992***	7.640*
	(4.050)	(3.300)	(5.154)	(4.420)	(4.437)	(3.244)
Female	9.011***	2.950*	6.668***	4.165*	4.448*	1.319
	(1.583)	(1.292)	(2.021)	(1.762)	(1.764)	(1.303)
White	11.558***	−2.448	20.463***	−0.063	5.234**	3.718*
	(1.806)	(1.507)	(2.303)	(2.025)	(2.004)	(1.483)
Married	3.463*	3.590*	5.847**	3.422+	1.398	3.664**
	(1.703)	(1.400)	(2.186)	(1.893)	(1.909)	(1.402)
Fully Employed	−1.805	3.324*	−2.508	4.212*	4.304*	2.548+
	(1.786)	(1.483)	(2.282)	(1.975)	(2.007)	(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 A10: Full Models, Social Media Prime and Minority Status (Figure 5)

	Career	Damage Relationships	Hostility Strangers	Sanction Mean
Intercept	2.858***	2.212***	3.018***	2.695***
	(0.089)	(0.087)	(0.089)	(0.072)
Dem	−0.084*	0.008	−0.029	−0.035
	(0.035)	(0.035)	(0.035)	(0.029)
Demand Base	0.001***	0.001***	0.000	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Party Minority	0.073	0.120**	0.009	0.067+
	(0.047)	(0.046)	(0.047)	(0.038)
Ideology	0.114***	0.048**	−0.028+	0.045**
	(0.017)	(0.017)	(0.017)	(0.014)
Pol. Int. (5-point)	−0.032*	−0.016	0.082***	0.011
	(0.014)	(0.013)	(0.014)	(0.011)
Education	0.035*	0.056***	0.102***	0.065***
	(0.015)	(0.014)	(0.015)	(0.012)
Prior	0.015***	0.016***	0.015***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)
Network Co-Opp	0.364***	0.448***	0.176*	0.329***
	(0.077)	(0.074)	(0.076)	(0.062)
Network Opposing	0.668***	1.301***	0.792***	0.923***
	(0.070)	(0.070)	(0.070)	(0.057)
Female	0.005	0.107***	0.184***	0.098***
	(0.027)	(0.026)	(0.027)	(0.022)
White	−0.199***	0.145***	0.214***	0.054*
	(0.032)	(0.031)	(0.032)	(0.026)
Married	0.013	0.094***	0.041	0.049*
	(0.029)	(0.028)	(0.029)	(0.023)
Fully Employed	0.271***	0.000	−0.096**	0.060*
	(0.031)	(0.030)	(0.031)	(0.025)
SM Prime	0.129***	0.124***	0.288***	0.179***
	(0.029)	(0.028)	(0.029)	(0.023)
SM Prime:Party Min	0.009	−0.011	−0.025	−0.007
	(0.065)	(0.063)	(0.065)	(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 A11: Full Models, Perceived Benefits to Conversations with Co-Partisan Opponents

		Discuss Real		Get to Know	
	None	Opinions	Civic Duty	Discussant	Persuade
Intercept	0.096***	0.514***	0.232***	0.378***	0.386***
	(0.011)	(0.018)	(0.016)	(0.018)	(0.018)
Party Minority	0.157***	−0.035	−0.121***	−0.034	−0.176***
	(0.029)	(0.036)	(0.025)	(0.034)	(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 A12: 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***	0.565***	0.382***	0.282***	0.165***
	(0.012)	(0.018)	(0.018)	(0.017)	(0.014)
Party Minority	0.044	0.029	0.016	0.093**	0.011
	(0.027)	(0.035)	(0.035)	(0.034)	(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

1.2 Additional Pre-Registered Analyses

1.2.1 Hypothesis 3

Table A13: Full Models, Hypothesis 3

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

1.2.2 Hypothesis 4

Table A14: Full Models, Hypotheses 4

	Group Demands	Private Demands
Intercept	56.184	35.621
	(4.055)	(3.866)
Co-partisan Ally		4.868
		(1.108)
Opposing Partisan		11.274
		(1.386)
Ask-Tell	−6.920	−6.800
	(1.548)	(1.272)
Dem	3.573	−3.783
	(1.572)	(1.528)
Party Minority	2.846	7.675
	(1.557)	(1.528)
Pol. Int. (5-point)	−3.884	−1.581
	(0.591)	(0.587)
Prior	0.329	0.328
	(0.031)	(0.031)
Network Co-Opp	−5.812	8.194
	(3.101)	(3.036)
Network Opposing	4.677	9.714
	(2.986)	(2.928)
Female	6.017	3.848
	(1.195)	(1.182)
White	4.561	6.861
	(1.371)	(1.325)
Married	3.538	3.709
	(1.288)	(1.273)
Fully Employed	0.752	2.228
	(1.362)	(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

1.2.3 Hypothesis 5

Table A15: Full Models, Hypothesis 5

	Group Demands	Private Demands	Group Demands
Intercept	38.581	31.465	56.457
	(5.771)	(3.805)	(4.290)
Co-partisan Ally		6.599	
		(0.876)	
Opposing Partisan		14.166	
		(1.116)	
Ask-Tell	−4.464		
	(2.786)		
Dem	−9.234	−3.807	2.388
	(3.193)	(1.530)	(1.739)
Party Minority	27.884	10.569	9.067
	(5.032)	(1.685)	(2.000)
Pol. Int. (5-point)	−2.820	−1.556	−4.886
	(0.878)	(0.587)	(0.650)
Prior	0.332	0.328	0.290
	(0.046)	(0.031)	(0.033)
Network Co-Opp	30.623	8.326	15.243
	(4.875)	(3.038)	(3.506)
Network Opposing	19.793	9.512	6.058
	(4.439)	(2.930)	(3.305)
Female	4.484	3.798	2.934
	(1.763)	(1.183)	(1.293)
White	5.332	6.918	−2.384
	(2.005)	(1.327)	(1.508)
Married	1.278	3.600	3.491
	(1.909)	(1.274)	(1.401)
Fully Employed	4.286	2.218	3.352
	(2.007)	(1.341)	(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	

1.3 Additional Exploratory Analyses

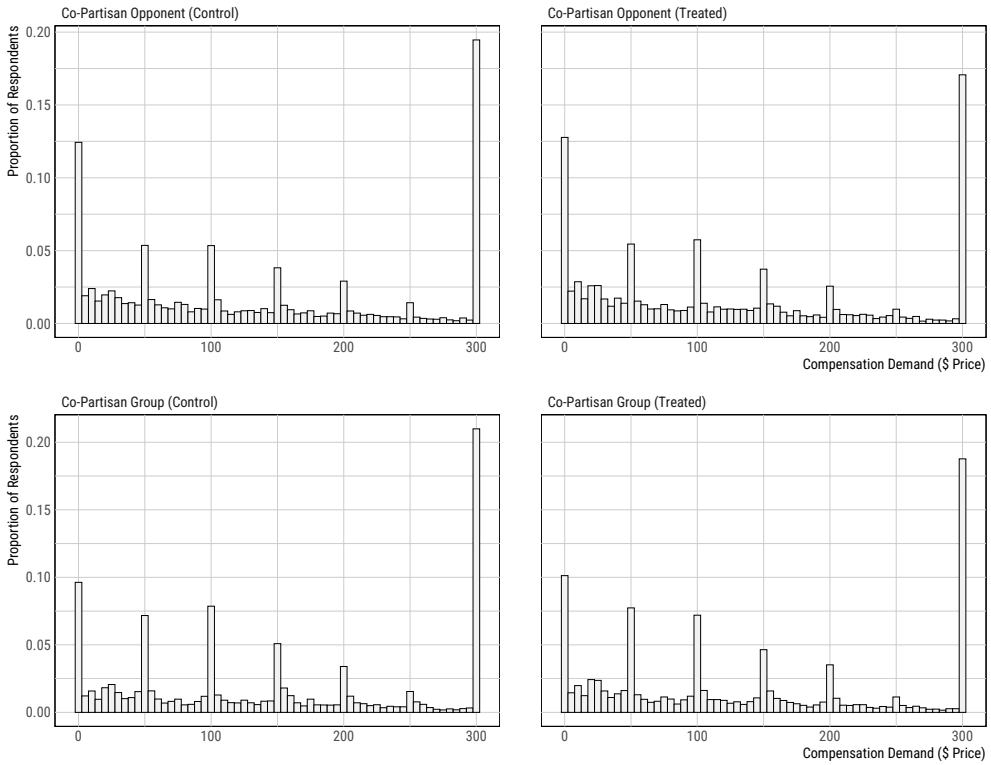

Figure A1: Distribution of Co-Partisan Compensation Demands

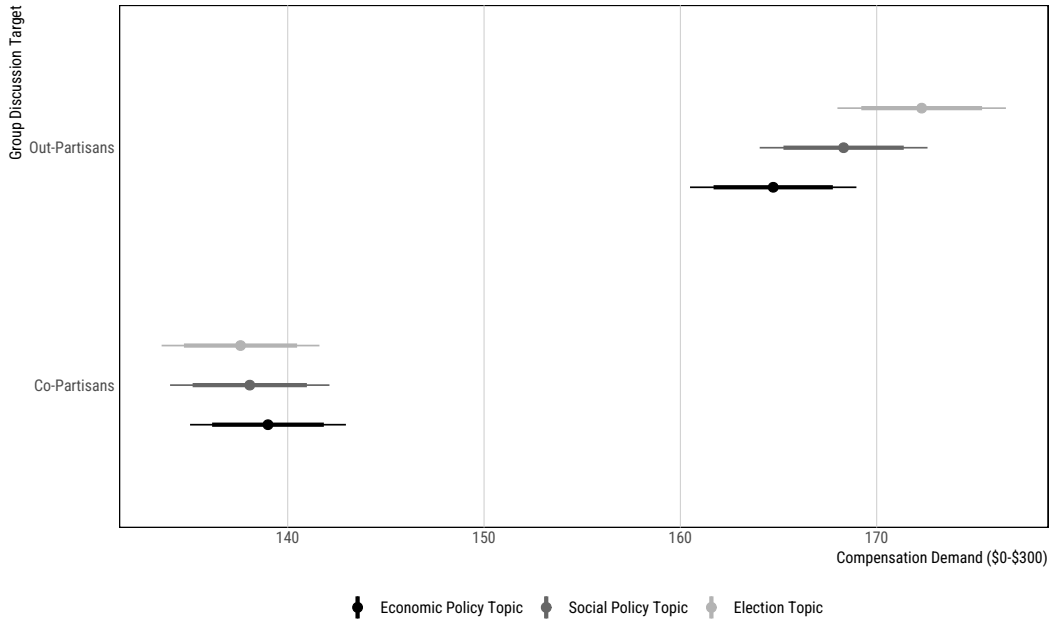

Figure A2: Group Conversation CDs by Topic

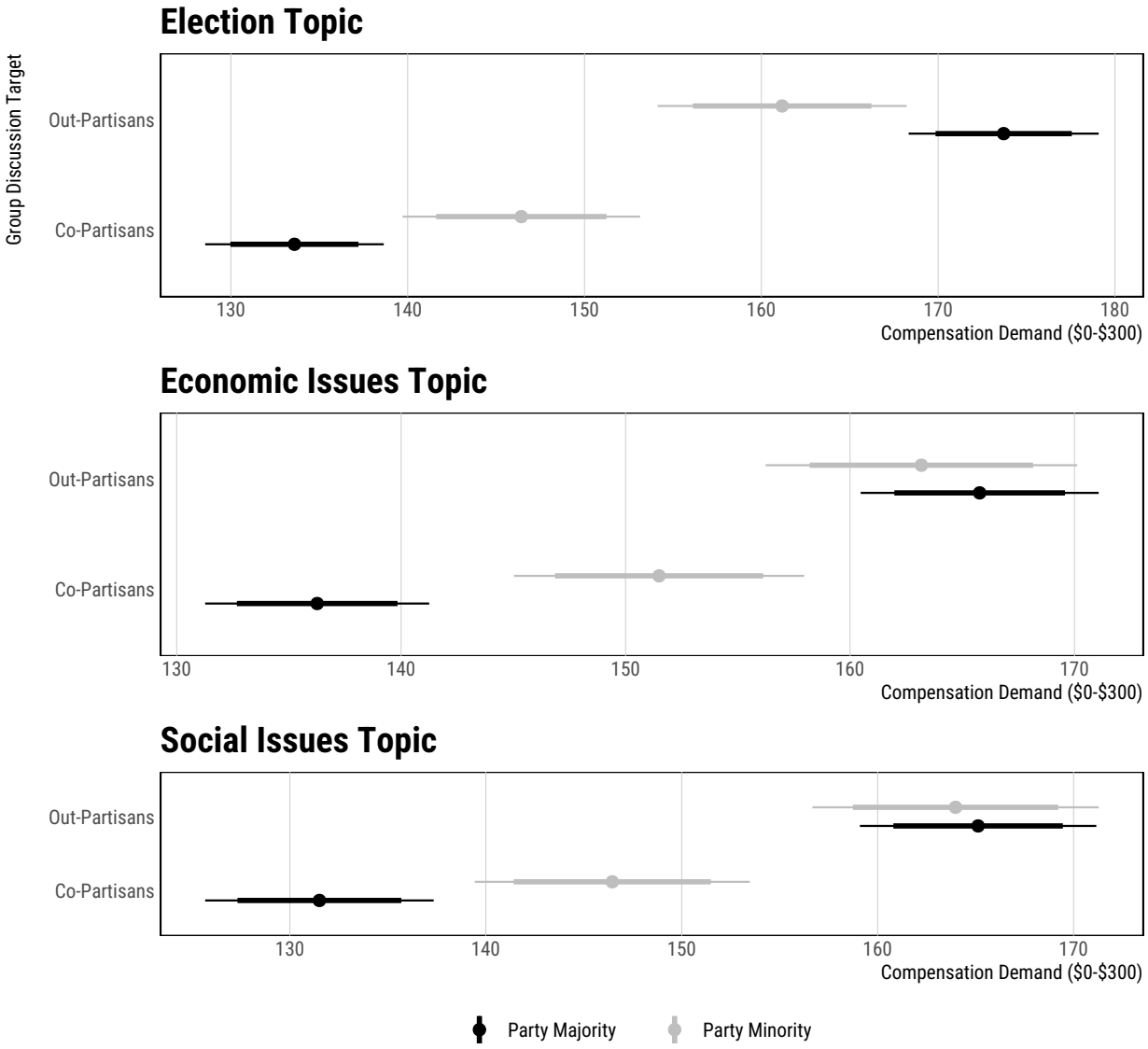

Figure A3: Group Conversation CDs by Topic and Party Majority/Minority

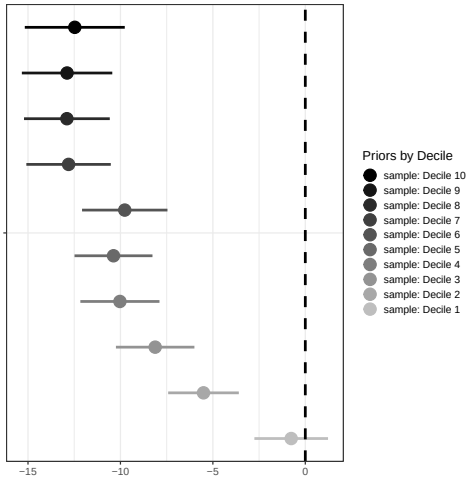

Figure A4: CATEs by Priors: Manipulation Index

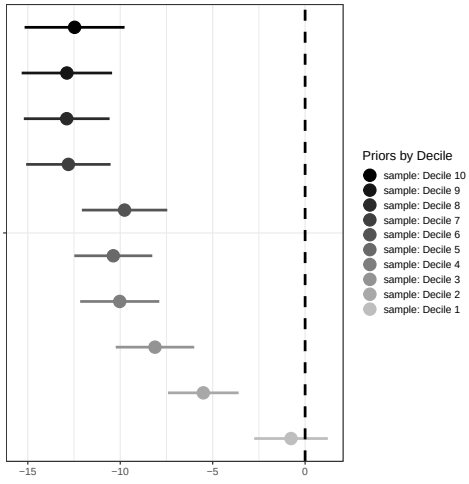

Figure A5: CATEs by Priors: Main Outcome Measure

1.4 Sample Composition and Balance Tables

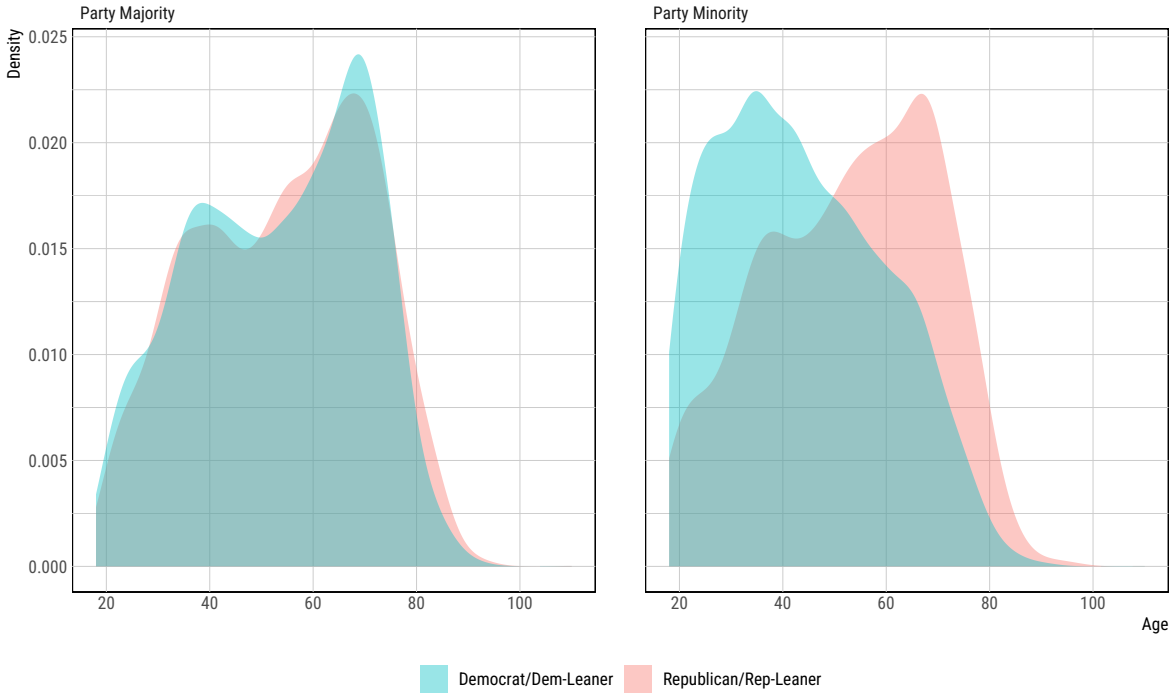

Figure A6: Density of Age by Party Minority Status

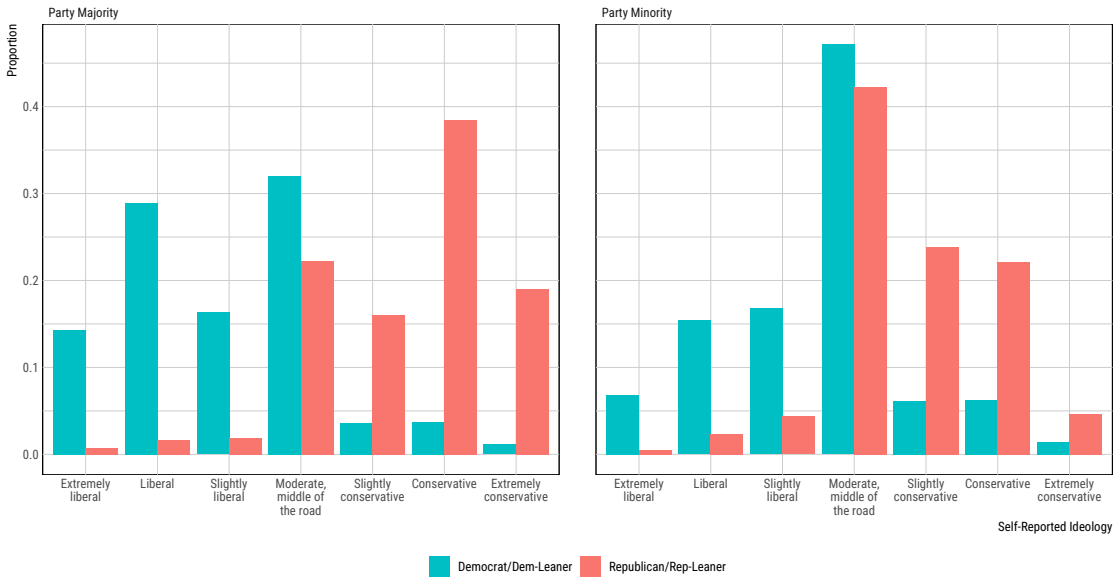

Figure A7: Self-Reported Ideological Identification by Party Minority Status

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 A16: Balance Tables Across Targets

	discussant (N=21400)		coopp (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 A17: 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 A18: 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

1.5 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, while the last statement mirrors survey language used by the Cato Institute.

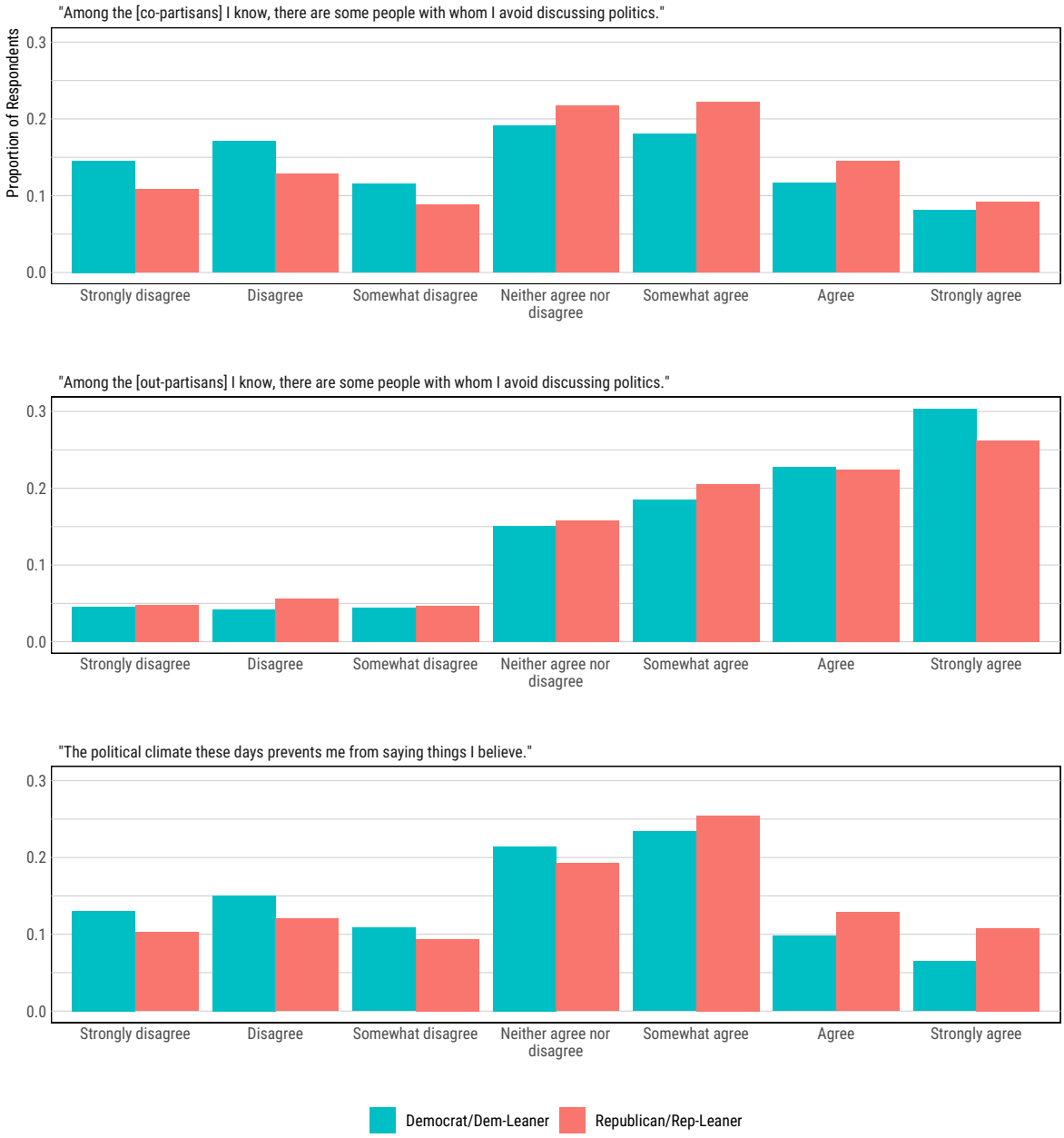

Figure A8: APM and Cato Survey Questions by Party

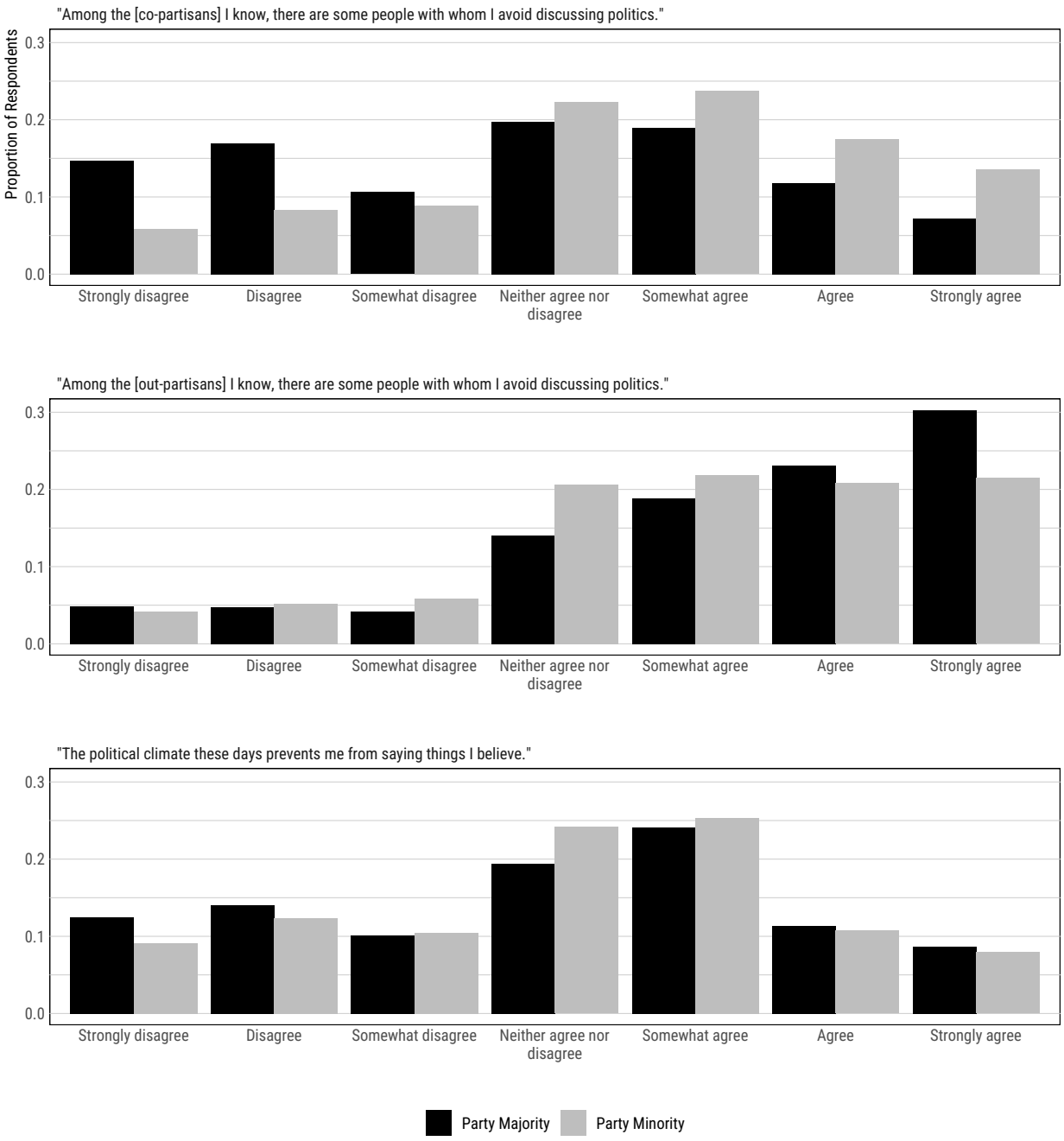

Figure A9: APM and Cato Survey Questions by Party Minority Status

1.6 Alternate Specifications

We first explore heterogeneity by a more stringent definition of intra-party conflict - defining minorities as respondents who say they will vote for the opposing party's nominee (as opposed to the broader pre-registered definition). Using this alternate specification we replicate two results: the baseline gap in willingness to speak (Figure 1) and the response to the correction treatment (Figure 6). We note, consistent with expectations from public polling and academic literature, that far smaller fractions of self-identified partisans affirmatively supported the opposing party's nominee as opposed to merely declining to support their own party's candidate.

Next, we explore heterogeneity by expressed turnout intention. While these self-reported answers are likely over-estimates of realized voter behavior (Jackman and Spahn 2019), we might expect that individuals who do not intend to vote have less strong feelings about their preferred candidate. Consistent with this interpretation, 76% of the sample reported a certainty to vote, compared to 64% of the American population that voted in 2024. Below, we compare the certain voters to the rest of the sample for out main outcomes. Further, most of the self-reported non-voters in our sample were pure independents leaving us with only 3064 partisans who expressed anything other than certainty to vote. We present results for certain-to-turnout voters and possible abstainers. The ask-tell correction operates similarly for both groups but abstainers have lower compensation demands to speak to opposing partisans and co-partisan allies.

Similarly, we explore heterogeneity by political interest. Our political interest variable

Table A19: Target and Ask-Tell Effects by Turnout Intention

	Voters	Abstainers
Co-partisan Ally	4.833***	1.850
	(0.849)	(1.943)
Opposing Partisan	12.876***	4.761*
	(1.078)	(2.167)
Ask-Tell Treatment	−6.466***	−6.915*
	(1.254)	(2.805)
Num.Obs.	41 192	8594
R2	0.307	0.276
Std.Errors	by: id	by: id

is on a 5 point scale. Theoretically, we might expect that more interested subjects are less responsive to our treatment because they are less effected by the etiquette mechanism we described. While exploratory, this analysis confirms our intuition, albeit directionally. Across all targets, each 1 point increase in political interest (SD = 1.1) is associated with a $2 decline in the effectiveness of the correction.

Table A20: Ask-Tell Effects by Target and Political Interest

	All Targets	Discussants	Out-Partisans	Co-Ally	Co-Opponent
Ask-Tell Treatment	−10.435**	−14.272**	−5.949	−11.323	−16.627*
	(3.983)	(5.392)	(8.140)	(7.163)	(7.216)
Political Interest	−2.127**	−2.615*	−1.489	−1.401	−4.496**
	(0.803)	(1.018)	(1.568)	(1.404)	(1.419)
Ask-Tell:Interest	1.111	2.086	−0.194	1.041	3.189+
	(1.088)	(1.416)	(2.160)	(1.901)	(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

1.7 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 A21, the baseline category for the omit 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.

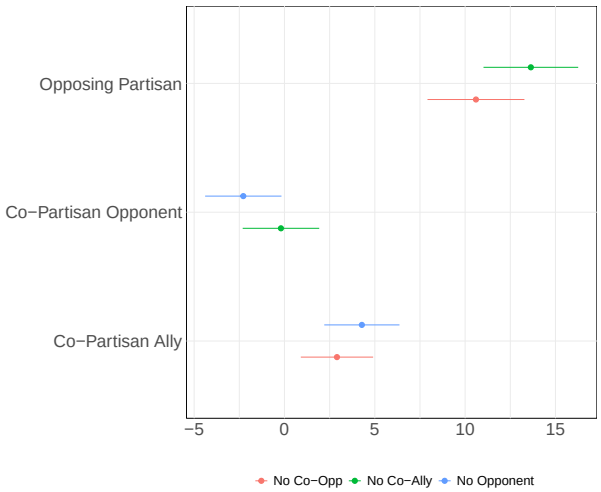

Figure A10: Comparison of Target Effects by Omitted Category

Table A21: 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

1.8 Descriptive Statistics and Self-Reported Censorship Fears

We next explore how partisanship and party minority status correlate with pre-treatment questions about self-censorship. While we believe these questions are less likely to effectively reveal self-censorship than our randomized ask-tell exposure (in part because of partisan cheerleading surrounding the opposing party and party faction's supposed censorship), they nonetheless offer an important self-reported measure of this type of consideration and provide circumstantial evidence surrounding our explanations.

We asked three questions in this vein, one about general self-censorship, one about fears

of Republicans and one about fears of Democrats; notably the wording of all three questions could implicate differential positive motivations. Using these questions, we explore which groups are more likely to express self-censorship concerns.

We show that while party minorities express more self-censorship fears overall, these effects are fairly small and are inconsistent across parties, less than 1/10th of a standard deviation for Democrats and a small correlation of minority status with less self-censorship among Republicans. We do find that minority partisans describe a substantially larger concern about co-partisans than partisan majorities. Despite frequent allegations of partisan blame regarding self-censorship, we find no substantively or statistically significant gaps how self-reported censorship changes by minority status across partisan identity, as shown in Tables A22 and A23 below.

Table A22: Descriptive Statistics for Censorship Fears among Dems

	General Censorship	Fear of Republicans	Fear of Democrats
Democrat	2.481***	3.346***	3.993***
	(0.157)	(0.157)	(0.165)
Partisan Minority	0.042	−0.164**	0.345***
	(0.052)	(0.050)	(0.055)
Network Copartisan Opponent	0.553***	−0.099	0.982***
	(0.128)	(0.129)	(0.139)
Network Opposing	1.375***	1.014***	0.663***
	(0.117)	(0.113)	(0.119)
Num.Obs.	8998	8993	8987
R2	0.087	0.136	0.095

Table A23: Descriptive Statistics for Censorship Fears among Reps

	General Censorship	Fear of Republicans	Fear of Democrats
Partisan Minority	0.028	0.464***	−0.076
	(0.061)	(0.062)	(0.062)
Network Copartisan Opponent	0.645***	1.444***	0.152
	(0.130)	(0.131)	(0.137)
Network Opposing	0.731***	0.760***	0.825***
	(0.130)	(0.126)	(0.128)
Num.Obs.	7936	7928	7929
R2	0.057	0.102	0.055

1.9 Descriptives of Mediators and Outcomes

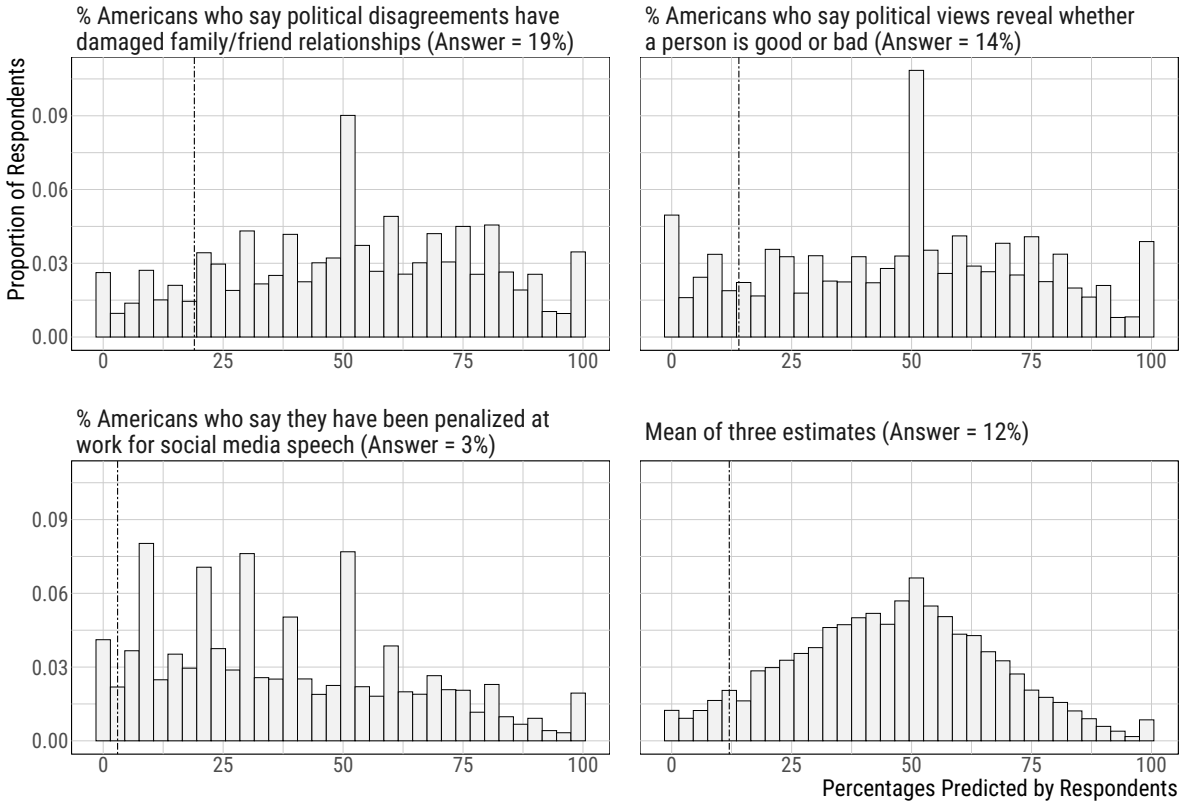

Figure A11: Respondents' "Ask-Tell" Prior Estimates

1.10 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 reper-

cussions 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.

1.11 Full Anonymized Pre-Analysis Plan

1.11.1 Motivation

Intra-party tensions are running high in American politics. Although Joe Biden and Donald Trump clearly won their respective presidential primaries, minorities of each party continue to express deep dissatisfaction with partisan nominees in public polls. Donald Trump's consistent under-performance compared to the polls and a wave of "uncommitted" protest votes in Democratic primaries (Abdul-Hakim et al., 2024)—not to mention protests and unrest on collegiate campuses—corroborate claims that discontent with party leadership is brewing in the wings of both parties.

The bottom-up strategies of these intra-party challenges hinge to a substantial degree on *interpersonal influence among voters*. Partisans are not only more willing to engage in conversation with co-partisans than out-partisans (Settle and Carlson, 2019), but also more receptive to persuasive messages that are inconvenient to their party when shared by a co-partisan messenger (Carey et al., 2024). In keeping with this intuition, prominent anti-Trump campaigns organized by conservative activists have poured millions into disseminating homemade video testimonials by former-Trump voters about the private concerns that have led them to abandon Trump in 2024 (Aratani, 2024). Moreover, peer conversations may be uniquely effective at bridging deep social divides over issues that divide co-partisans, such as the conflict in Gaza. In-depth exchanges with strangers can change deep-seated prejudices, on the key condition that such exchanges remain non-judgmental (Kalla and Broockman, 2020). Beyond resource intensive campaigns, intimate social networks are the most likely sites for such exchanges to take place.

However, anecdotal and survey evidence suggests that discussions about election issues among co-partisans are expected to be hostile and thus frequently constrained by heightened concerns about social sanction. Among Republicans, never-Trump voters frequently describe avoiding criticism of the former president for fear of criticism or ostracism by their peers

(Tabet, 2023). Surveys indicate large fractions of both parties are indeed hostile toward co-partisan elites who dissent from party leaders (Dunn, 2021; Sheagley, Dancey and Henderson, 2023). But pressures to conform with peers may not be limited to those within each party's dissenting minority. Irrespective of majority status, moderate factions may be less inclined to engage in public-facing online speech than those with more polarized attitudes (Krupnikov and Ryan, 2022), which resonates with the theory that the distribution of political attitudes expressed online has polarized through a process of self-selection (Bail, 2022).

Limited empirical evidence exists to substantiate whether experiences of *intra*-party social pressure are widespread, much less whether they disproportionately impact certain voters—such as party minorities or members of one party in particular. Prominent research on self-censorship in the U.S. has examined whether social pressure biases against minority opinion-holders in society at large (e.g., Noelle-Neumann, 1974; Glynn, Hayes and Shanahan, 1997), how perceptions of political intolerance vary across parties (Gibson and Sutherland, 2023), and how distaste for encounters with members of the out-party contributes to sorting into partisan echo chambers (Settle and Carlson, 2019; Sunstein, 2018). Implicit to the conclusion that people far prefer discussing politics in homogeneous networks is an intuition that voters mostly feel free to express their true preferences when surrounded by co-partisans, one we aim to complicate.

How universal is intra-party social pressure, and does it serve party unity or disunity in the aggregate? In this project, we seek to document the prevalence of intra-party social pressure in contemporary discourse about the 2024 presidential election. This paper serves as a pre-analysis plan for a large survey experiment ($N = 15,000 - 25,000$) being fielded starting in late June 2024 through the CHIP50 nationwide survey[11]. We examine to what extent *affiliative* concerns undergird self-censorship and how the prevalence of these concerns varies with the target audience (co-partisan allies, co-partisan opponents, out-partisans, or some bundle) and with respondents' perceptions of their networks' preferences and tolerance for dissent.

[11]Sample size is uncertain within this range due to the organization of the survey. Authors will have no control of the sample size and will not have access to the data before the sample size is finalized.

1.11.2 Survey Design

Covariates

We begin by defining the covariates that we will use, both from our module and the common content across the different elements of CHIP50 (“core items’’). These covariates will be used for survey piping, to increase precision, and to estimate heterogeneous effects.

All respondents are first defined as identifying with or leaning toward one of the two major political parties. Based on CHIP50’s core items, we partition respondents into two party labels (“Republican/Lean Republican’’ or “Democrat/Lean Democrat’’) and use these categories to define “co-partisans’’ and “out-partisans’’ for each respondent. The remaining respondents (pure Independents or members of third parties who do not lean toward a major party) will be filtered out of the main analysis but will be included as part of our exploratory analyses. These respondents will be assigned to “Dem/Lean Dem” if they identify their ideology as liberal, assigned to “Rep/Lean Rep” if they identify their ideology as conservative, or randomly assigned to one of the two if they do not identify as liberal or conservative.

CHIP50 core items include the following question about respondents’ 2024 vote intention: “If the 2024 U.S. presidential election were held today, which candidate would you vote for?” {Response options: Joe Biden (Democrat); Donald Trump (Republican); Another candidate:_; I would not vote; Not sure}. We code respondents as belonging to a party minority if they do not respond that they would vote for their party’s presumptive nominee (**party_minority** = 1).

We then ask the following questions about network heterogeneity {Response options: None; Almost none; A few; About half; A lot; Nearly all; All}:

- N1: How many of the people you know are [co-partisans] or lean toward the [party name] party?
- N2: How many of the people you know are [out-partisans] or lean toward the [out-party name] party?
- N3: Of the [co-partisans] you know, how many are hesitant to vote for [presumptive nominee for respondent’s party]?

We use these questions to construct two measures of exposure to disagreement. We

rescale responses to the first and third questions (N1 and N3) from a 1-7 scale to a 0-1 scale (0 = None, 1 = All), then multiply these two values to construct a **network_co_opponent** measure (0-1) which approximates the proportion of a respondent's network who are "co-partisan opponents''—same-party voters who differ from the respondent in terms of support for the party leader. This measure is calculated as N1*N3 for respondents who intend to vote for their party's leader and as N1*(1-N3) for respondents who do not intend to. Finally, we rescale responses to N2 from a 1-7 scale to a 0-1 scale and define **network_out** to capture exposure to out-partisans in the respondent's network.

Respondents are then asked to write their agreement using a 7-point Likert scale toward three statements: "The political climate these days prevents me from saying things I believe"; "Among the Republicans I know, there are some people with whom I avoid discussing politics"; "Among the Democrats I know, there are some people with whom I avoid discussing politics." These 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.[12] We will generate three measures from respondents' answers—**censor_climate**, **censor_co**, and **censor_out**—and use these questions to examine the degree of inter-party asymmetry in self-censorship.

To increase precision in our analyses and anchor respondents' later compensation demand responses, we ask respondents to first consider a simpler assignment: to write a few paragraphs explaining why they would or would not vote for their party's leader if the 2024 presidential election happened today. Respondents provide the price they would need to be paid to participate on a $0 to $300 slider, which we call **demand_base**.

All respondents are then asked to provide three estimates of the percentage of Americans who have reported various experiences with political speech. We define a dummy variable, **above_prior**, which is coded as 1 if the average of these three guesses is less than 12%. We further define a numeric variable, **prior**, which equals the numeric average of the three answers.

Our main vector of covariates includes the following variables: party affiliation, prior, party_minority, network_co_opponent, network_out, demand_base, political ideology, political interest, household income, employment status, marital status, religion, age, gender,

[12]The first statement mirrors survey language used by the Cato Institute, while the second and third statement resemble survey language used by APM Research Lab.

race/ethnicity, and education.

Randomizations

We define four randomizations. Because these occur sequentially in the survey, not all randomizations affect all outcome variables.

- **(Ask-Tell Intervention) T1**: (0, 1) Earlier in the study, we asked respondents to provide estimates about the percentage of Americans who have reported various experiences with political speech. Respondents are then randomized to a tell treatment (probability = 0.5) or control (probability = 0.5). In the tell condition, respondents are told the true percentage of Americans based on real data and are reminded of their own answers. In the control condition, respondents are only reminded of their own estimates.

- **(Targets) T2**: (0, 1, 2) Respondents report their compensation demand for sending a one-minute video about their 2024 presidential election vote choice to one of three "targets'' (a co-partisan "opponent" who unlike the respondent does (does not) intend to vote for the party leader; a co-partisan "ally'' who also does not (does) intend to vote for the party leader; or an out-partisan). We assign respondents to two of these targets and in a random order, such that they report two compensation demands (on a $0 to 300$ scale). See the section "Outcome Measures" for more on how we define our variables for this randomization.

- **(Social Media Prime) T3**: (0, 1) When shown the "Sanction" question (see Outcome Measures below), respondents are randomly assigned to two possible question wordings. Respondents are either reminded of the previous questions and asked to think about sharing a one-minute video to a social media platform of their choice (T3 = 1) or only reminded of the previous questions, which asked them to consider privately sharing the video (T3 = 0).

- **(Discussion Topic) T4**: (0, 1, 2) The final grid in our module measures respondents' willingness (via a compensation demand scale) to have a five-minute discussion with a group of Republicans and a group of Democrats. We randomize the question wording

by varying whether the proposed discussion topic is the 2024 presidential election (prob = 0.5) or one of two issues: "economic issues like taxation and social security'' (prob = 0.25) or "social issues like abortion and immigration'' (prob = 0.25). We define two dummy variables which indicate whether the randomized topic is the 2024 election (**topic_election**) or social issues (**topic_social**).

Outcomes

Below we define key outcomes and the randomizations which affect each (in parentheses).

- Target Demands (T1, T2): All respondents are asked to give compensation demands for sending a video to a person with whom they "frequently discuss politics'' and to two randomly selected targets. We employ a within-subjects design that defines a single outcome, **demand_target**, and four indicators corresponding to our one universal and three randomized targets: CD0 (frequent political discussant, the baseline target), CD1 (co-partisan opponent), CD2 (co-partisan ally), and CD3 (out-partisan). Compensation demands are measured in U.S. dollars on a horizontal slider ranging from $0 to $300.

- Manipulation Check (T1, T2): All respondents are asked to estimate the percentage of two groups who would agree with the following statement: "When people I know express political views with which I disagree, I lose some respect for them.'' These outcomes are referred to as **manip_check1** (for "All Americans") and **manip_check2** (for "People you know"), with responses measured on a horizontal slider ranging from 0 to 100.

- Sanction (T1, T2, T3): Three item grid eliciting respondents' perceptions about the likelihood of experiencing various sanctions for sharing their views on their party's leader. The question wording is randomized to relate to either social media or not. The grid items are referred to as **sanction1**, **sanction2**, and **sanction3**, with responses measured on a 7-point Likert scale ranging from "Extremely unlikely" to "Extremely likely." We also define an average of the three sanction measures as **sanction_mean**.

- Conversation Demands (T1, T2, T3, T4): Two further compensation demands to speak to a group of co-partisans and a group of out-partisans. We employ a within-subjects

design that defines a single outcome, **demand_convo**, and dummy variables indicating whether the discussion group comprises co-partisans (CD4) or out-partisans (CD5). We define the average of these two prices as **convo_mean**.

- Aggregated outcomes: We calculate the average of all five compensation demands reported by a respondent as their **demand_mean**.

Hypotheses

In all hypotheses that specify directional effects or differences in CATEs, we hypothesize that the difference will be statistically significant to the $p = 0.05$ level.

Hypothesis 1: Perceived Costs of Speech

Our first aim is to establish how costly respondents perceive the act of revealing their 2024 vote intention to a co-partisan opponent they know. We do so by benchmarking respondents' compensation demands against three other targets: a person with whom they ordinarily discuss politics, a co-partisan ally they know, and an out-partisan they know. We adopt the following model specification,

$$\text{DemandTarget}_c = \beta_0 + \beta_1 \,\text{CD1}_c + \beta_2 \,\text{CD2}_c + \beta_3 \,\text{CD3}_c + \eta \mathbf{\Phi}_i + \epsilon_c \,, \tag{3}$$

where c indexes compensation demand observations, i indexes respondents, and $\mathbf{\Phi}_i$ is a vector of individual-level covariates. CD1-CD3 are dummy variables indicating whether the target is a co-partisan opponent, co-partisan ally, or out-partisan, respectively (with each respondent being randomly assigned to view two of these). For this within-subjects design, we will use OLS with cluster robust standard errors, clustering at the respondent level. We define the following hypotheses:

- H1a: Respondents will report higher compensation demands for sharing their 2024 vote choice to a co-partisan opponent they know than to a person with whom they frequently discuss politics ($\beta_1 > 0$).
- H1b: Respondents will report higher demands for sharing their 2024 vote choice to a co-partisan opponent they know than to a co-partisan ally they know ($\beta_1 > \beta_2$).

- H1c: Respondents will report lower demands for sharing their 2024 vote choice to a co-partisan opponent they know than to an out-partisan they know ($\beta_1 < \beta_3$).

To provide perspective on the relative sizes of perceived costs associated with these four targets, we will examine the *ratios* of co-partisan opponent, co-partisan ally, and baseline demands to the out-partisan demand. We also define the following research question to examine whether the co-partisan opponent demand more closely resembles that of co-partisan allies or out-partisans:

- RQ1: Will the difference between compensation demands for co-partisan opponents and opponents be smaller in magnitude than the difference between co-partisan opponent and co-partisan ally demands?

Hypothesis 2: Ask-Tell Mechanisms

Next, we examine to what extent these perceived costs are alleviated by an ask-tell intervention that corrects misperceptions of other Americans' intolerance for political disagreement and dissenting speech.[13] We begin with a set of mechanism checks using three outcomes—manip_check1, manip_check2, and sanction_mean—and adopt the following model specification for each,

$$\text{Outcome}_i = \beta_0 + \beta_1 \, \text{T1}_i + \eta \mathbf{\Phi}_i + \epsilon_i \,, \tag{4}$$

where T1 is a dummy variable indicating if the respondent was shown the tell treatment. For our hypothesis test of H2b (described below), our vector of covariates $\mathbf{\Phi}_i$ additionally includes T3, a dummy variable indicating whether the respondent was shown a social media prime before answering the sanction questions. We will use OLS with HC2 robust standard errors. We define the following hypotheses:

- H2a: Exposure to the tell treatment will decrease respondents' estimates of the percentage of "All Americans'' (Outcome = manip_check1) and "People [they] know''

[13]Though our main analysis will include respondents for whom above_prior = 1 (for whom the tell is an upward rather than a downward shock), in our exploratory analyses we will examine whether our result for the hypothesis tests defined below are sensitive to their exclusion.

(Outcome = manip_check2) who would agree with a statement expressing intolerance for political disagreement ($\beta_1 < 0$).[14]

- H2b: Exposure to the tell treatment will decrease respondents' perceived likelihood of experiencing social sanctions for sharing their 2024 vote intention with people they know (Outcome = sanction_mean; $\beta_1 < 0$).[15]

We verify with H2c whether respondents with priors about Americans' intolerance for political disagreement that are further from the true statistic are affected to a greater extent by the ask-tell treatment. For this hypothesis, we modify model 2 to include the interaction term $\beta_2(\text{T1}_i \times \text{Prior}_i)$ (note that Prior is already included in $\mathbf{\Phi}_i$).

- H2c: The ATEs of the ask-tell treatment on manip_check1 and manip_check2 increase in magnitude as respondents' average prior estimates about Americans' intolerance for political disagreement increase ($\beta_2 < 0$).

Hypothesis 3: Ask-Tell Treatment Effects

For H3 and H4, we shift attention to our compensation demand outcomes, examining whether and in what contexts perceived costs to speech are reduced by exposure to the ask-tell treatment. For H3a and H3b, we employ model 2 using demand_mean as the outcome (the average of the five compensation demand measures viewed by each subject) and including in $\mathbf{\Phi}_i$ a set of dummy variables indicating the T2 and T4 conditions to which the respondent was assigned.

- H3a: Compared to control, respondents shown the ask-tell treatment will report lower compensation demands on average ($\beta_1 < 0$).

- H3b: The ATE of the ask-tell treatment on demand mean increases in magnitude as respondents' average prior estimates about Americans' intolerance for political disagreement increase (Model 2 with $\beta_2(\text{T1}_i \times \text{Prior}_i)$: $\beta_2 < 0$).

[14] We anticipate that the (covariate-adjusted) average treatment effect will be larger in magnitude for "All Americans'' than for "People [they] know'' and will examine this as part of our exploratory analyses.

[15] To characterize in greater detail the mechanisms of our ask-tell intervention, our exploratory analyses will examine the constituent measures of sanction_mean (sanction1, sanction2, and sanction3) as distinct outcomes using separate regressions. That is, we will examine which of these perceived risks are affected most or least by the ask-tell treatment: "Suffering consequences for your career," "Damaging relationships with people you care about," or "Experiencing hostility from strangers or people you barely know."

In order to examine how the ask-tell ATE varies across six CD outcomes, we use model 2 with CD0 through CD5 as our outcomes. We are particularly interested in the ATEs for sharing the 2024 election video to a co-partisan opponent and for participating in a five-minute discussion with co-partisans. We define two hypotheses and a broader research question which compares effects for CD0-CD3 and CD4-CD5:

- H3c: Compared to control, respondents shown the ask-tell treatment will report lower compensation demands for sharing their views on their party's 2024 presidential nominee to a co-partisan opponent they know (Outcome = DemandTarget; subset = CD1==1; $\beta_1 < 0$).
- H3d: Compared to control, respondents shown the ask-tell treatment will report lower compensation demands for participating in a five-minute political discussion with a group of co-partisans (Outcome = DemandConvo; subset = CD4==1; $\beta_1 < 0$).
- RQ2: Are the effect sizes of the ask-tell treatment on compensation demands larger for group discussion assignments (CD4-CD5) than for private video-sharing assignments (CD0-CD3)?

Hypothesis 4: Variation in Ask-Tell Effects Across Settings

Building on RQ2, Hypothesis 4 examines whether the ask-tell ATE on CD price varies across targets and discussion topics. For the video-sharing assignments, we hypothesize that concerns about preserving important interpersonal relationships pertain more to co-partisan peers than out-partisan peers on average. To test this hypothesis, we define the following interaction model.

$$\begin{aligned} \text{DemandTarget}_c = \beta_0 &+ \beta_1\,\text{T1}_i + \beta_2\,\text{CD1}_c + \beta_3\,\text{CD2}_c + \beta_4\,\text{CD3}_c \\ &+ \beta_5(\text{T1}_i \times \text{CD1}_c) + \beta_6(\text{T1}_i \times \text{CD2}_c) \\ &+ \beta_7(\text{T1}_i \times \text{CD3}_c) + \eta\mathbf{\Phi}_i + \epsilon_c \end{aligned} \tag{5}$$

- H4a: The ask-tell treatment decreases demand price more for the co-partisan opponent CD than for the out-partisan CD (Model 3: $\beta_5 < \beta_7$).

To understand how the topic of discussion affects the relevance of intra- versus inter-party self-censorship, model 4 turns to the group discussion CDs alone. Again, outcomes are measured at the CD level and standard errors are clustered at the respondent level.

$$\begin{aligned}\text{DemandConvo}_c = \beta_0 &+ \beta_1\,\text{T1}_i + \beta_2\,\text{TopicElection}_i + \beta_3\,\text{CoPartisans}_c \\ &+ \beta_4(\text{T1}_i \times \text{TopicElection}_i) + \beta_5(\text{TopicElection}_i \times \text{CoPartisans}_c) \\ &+ \beta_6(\text{T1}_i \times \text{CoPartisans}_c) + \beta_7(\text{T1}_i \times \text{TopicElection}_i \times \text{CoPartisans}_c) \\ &+ \eta\mathbf{\Phi}_i + \epsilon_c\end{aligned} \tag{6}$$

- H4b: The ATE of the ask-tell treatment is larger (in magnitude) for the topic of the 2024 election than for salient policy issues when the discussion group comprises co-partisans rather than out-partisans (Model 4: $\beta_7 < 0$).[16]

Hypothesis 5: Heterogeneous Treatment Effects

H5 uses treatment-by-covariate interactions to investigate how the effects examined in H1-H4 vary with respondents' characteristics. First, we examine how party minorities and party majorities differ in their perceived costs for speech with peers. For H5a, PartyMinority_i is substituted for Covariate_i:

$$\begin{aligned}\text{DemandTarget}_c = \beta_0 &+ \beta_1\,\text{CD1}_c + \beta_2\,\text{CD2}_c + \beta_3\,\text{CD3}_c \\ &+ \beta_4(\text{CD1}_c \times \text{Covariate}_i) + \beta_5(\text{CD2}_c \times \text{Covariate}_i) \\ &+ \beta_6(\text{CD3}_c \times \text{Covariate}_i) + \eta\mathbf{\Phi}_i + \epsilon_c\end{aligned} \tag{7}$$

- H5a: Co-partisan opponent CD prices will be larger for respondents who hold the minority view in their party regarding their party's 2024 presumptive nominee than for party majority members ($\beta_4 > 0$).

Next, we hone in on a single CD outcome to examine how party minorities' experiences with intra-party pressure may differ across parties. While Trump defectors among Republicans are almost exclusively centrists, Biden defectors among Democrats may be centrists

[16]The policy issue discussion topic condition is further randomized across economic issues and social issues. In our exploratory analyses, we will examine how the ask-tell ATE varies across these topics.

or progressives. The model below interacts T1 with party minority status and an indicator for identifying with or leaning toward the Democratic party. For RQ3 and H5b, we will use Bonferroni corrections to account for the two outcomes with which we perform each test.

$$\begin{aligned}\text{Outcome}_i = \beta_0 &+ \beta_1 \text{T1}_i + \beta_2 \text{PartyMinority}_i + \beta_3 \text{Democrat}_i + \beta_4(\text{T1}_i \times \text{PartyMinority}_i) \\ &+ \beta_5(\text{T1}_i \times \text{Democrat}_i) + \beta_6(\text{PartyMinority}_i \times \text{Democrat}_i) \\ &+ \beta_7(\text{T1}_i \times \text{PartyMinority}_i \times \text{Democrat}_i) + \eta \mathbf{\Phi}_i + \epsilon_i\end{aligned} \tag{8}$$

- RQ3 (Observational): Will the co-partisan opponent video sharing CD (Outcome = Demand; subset = CD1==1) and the co-partisan group discussion CD (Outcome = Demand; subset = CD4==1) be higher for party minorities than party majorities for all subjects (i.e., $\beta_2 > 0$ and $\beta_6 > 0$), or will this pattern differ between Republicans and Democrats?

- H5b: The CATEs of the ask-tell treatment on CD1 and CD4 will be larger in magnitude for members of the Republican minority than for members of the Republican majority ($\beta_4 < 0$ for Outcome = Demand; subset = CD1==1 *and* subset = CD4==1). The same will *not* be true for Democrats ($\beta_7 \not< 0$).

Alternative accounts of self-censorship argue that it is driven by local networks in specific rather than the distribution of public opinion at large. We define model 7 to examine this:

$$\begin{aligned}\text{Outcome}_i = \beta_0 &+ \beta_1 \text{T1}_i + \beta_2 \text{PartyMinority}_i + \beta_3 \text{NetworkCoOpponent}_i \\ &+ \beta_4(\text{T1}_i \times \text{PartyMinority}_i) + \beta_5(\text{T1}_i \times \text{NetworkCoOpponent}_i) + \eta \mathbf{\Phi}_i + \epsilon_i\end{aligned} \tag{9}$$

- RQ4: Will the CATE of the ask-tell treatment on CD1 both decrease in magnitude as the share of a respondent's network comprising co-partisan opponents increases *and* be larger in magnitude for party minorities than party majorities ($\beta_4 < 0$ *and* $\beta_5 > 0$)?

We use model 8 to regress the CD4 outcome (co-partisan discussion demand) on T4 condition and an interaction with party minority status.

$$\begin{aligned}\text{Outcome}_i = \beta_0 + \beta_1 \text{TopicElection}_i + \beta_2 \text{PartyMinority}_i \\ + \beta_3(\text{TopicElection}_i \times \text{PartyMinority}_i) + \eta \mathbf{\Phi}_i + \epsilon_i\end{aligned} \tag{10}$$

- H5c: The CATE of the election topic on the co-partisan group discussion CD is larger for respondents who hold the minority view in their party regarding their party's 2024 presumptive nominee ($\beta_3 > 0$).

Finally, we investigate claims about the current political climate. Using model 5 and substituting censor_climate for Covariate, we examine to what extent variation in self-reports of self-censorship predicts perceived costs to co-partisan versus out-partisan speech. We hypothesize that perceptions of a hostile climate may capture the perceived costs to speech with out-partisans but not co-partisans despite the prevalence of intra-party pressures.

- H5d: Out-partisan CD prices will be higher for respondents who agree more strongly that the current political climate causes them to self-censor, (Model 5: $\beta_6 > 0$).
- H5e: Perceptions of the political climate will be more predictive of out-partisan CD prices than of co-partisan opponent CD prices (Model 5: $\beta_6 > \beta_4$).

Hypothesis 6: Private Versus Social Media Speech

H6 examines the effects of our social media prime on respondents' perceived likelihood of suffering consequences for sharing their 2024 election views, as well as how these effects vary with other treatments and covariates. We define the following base model, where T3 indicates whether the respondent receives the social media prime, then define for each hypothesis the relevant outcome and relevant interaction terms to be added:

$$\text{Outcome}_i = \beta_0 + \beta_1 \text{T3}_i + \eta \mathbf{\Phi}_i + \epsilon_i \tag{11}$$

- H6a: Respondents primed to think about sharing their video to social media, rather than solely about sharing their video privately to acquaintances, will perceive a higher likelihood of being sanctioned as a result (Outcome = sanction_mean; $\beta_1 > 0$).
- H6b: Disaggregating concerns about career, interpersonal relationships, and hostility from strangers, respondents primed to think about sharing their video to social media

will perceive a higher likelihood of experiencing career-related consequences and hostility from strangers but not a higher likelihood of damaging relationships with people they care about ($\beta_1 > 0$ for Outcome = sanction1 and Outcome = sanction3 but not Outcome = sanction2).

- H6c: The ATE of the social media prime on perceived likelihood of sanctions will be smaller in magnitude for respondents who were randomly assigned to the co-partisan opponent and out-partisan targets (indicated by the dummy CD13_i) rather than any other possible pair of targets for the video-sharing CDs (Model 9 with $\beta_2\text{CD13}_i + \beta_3(\text{T3}_i \times \text{CD13}_i)$: $\beta_3 < 0$ for Outcome = sanction_mean).[17]

For H6d-H6e, we use a variation of model 9 that includes three interaction terms:

$$\begin{aligned}\text{Outcome}_i = \beta_0 + \beta_1\text{T3}_i + \beta_2(\text{T3}_i \times \text{PartyMinority}_i) + \beta_3(\text{T3}_i \times \text{NetworkOut}_i) \\ + \beta_4(\text{T3}_i \times \text{NetworkCoOpponent}_i) + \eta\mathbf{\Phi}_i + \epsilon_i\end{aligned} \tag{12}$$

- H6d: The CATE of the social media prime on perceived likelihood of sanctions will be larger for respondents with more disagreeable networks, as public political commentary implicates entire networks (Outcome = sanction_mean: $\beta_3 > 0$ and $\beta_4 > 0$).

- H6e: The CATE of the social media prime on perceived likelihood of sanctions will be larger for respondents who hold the minority view in their party regarding their party's 2024 presumptive nominee (Outcome = sanction_mean: $\beta_2 > 0$).

Observational Hypotheses

The following hypotheses and research questions use pre-treatment covariance to make population level claims about self-reported features of self-censorship. We first report rates of intra- and inter-party self-censorship:

- ORQ1: What proportion of respondents (in total and disaggregated by party minority status and party) will report avoiding political discussion with some co-partisans

[17] Such respondents were already primed to think of both same-party and out-party opponents, whereas the remaining respondents were only primed to think of one type of opponent prior to the social media prime. We conceptualize the social media video-sharing assignment as priming *exposure to the rest of the network* not already made salient to the respondent.

they know (censor_co)? How does this proportion compare to censor_climate and censor_out?

For the remaining questions and hypotheses, we will report results from both a specification with no additional covariates (other than the right-hand side variables described by the hypothesis/question) and a specification including the full vector of covariates defined earlier in our pre-analysis plan.

- OH1a: Compared to party majorities, party minorities will be more likely to report avoiding political discussions with some co-partisans they know (censor_co).
- OH1b: The extent to which party majorities and minorities differ on censor_co will depend on party affiliation, such that members of the Democratic majority and minority differ less than members of the Republican majority and minority.

Next, we investigate the extent to which perceptions of the current political climate writ large overlap or fail to overlap with experiences of intra-party self-censorship.

- ORQ2: Will censor_climate predict censor_out better than it predicts censor_co?
- OH2: Republican respondents will be more likely than Democrats to agree that the political climate prevents them from saying things they believe (censor_climate), but this inter-party asymmetry will be smaller for censor_co and censor_out.

We further examine to what extent perceptions of a hostile political climate are influenced by real exposure to potential hostility (as a party minority or through self-reported personal networks) or other factors (such as partisan media rhetoric). Regressing censor_climate on party, party_minority, network_co_opponent, and network_out, we ask:

- ORQ3a: Conditioning on party, will party minority status, exposure to co-partisan opponents in one's personal network, or exposure to out-partisans in one's personal network predict perceptions of a hostile political climate? Or will party alone predict perceptions?

Exploratory Analyses

1. To unpack the mechanisms behind our hypothesis tests for H1 and what concerns may be evoked by different targets, we will analyze how the targets to which the respondent was assigned (T2 randomization) affect the perceived likelihood of sanction1, sanction2, and sanction3 for respondents not assigned to the social media prime (controlling for T1 randomization). We hypothesize that assignment to CD1 and CD3 will increase concerns about sanctions overall and that CD3 exposure will increase concerns about hostility from strangers and careers.

2. Following our analysis of H2a, we will examine whether the ask-tell ATE on our mechanism check is higher for the "All Americans" outcome than the "People you know" outcome.

3. Following our analysis of H2b, we will examine separately the ask-tell ATE on sanction1, sanction2, and sanction3.

4. Following our analysis of H4b, we will examine how the ask-tell ATE on the group discussion CD varies with economic versus social issues.

5. For respondents asked to send a video to a co-partisan opponent, the reference target (the other target to which the respondent was randomized) may affect the CD1 outcome. We intend to examine this possibility but do not hold strong priors about the direction in which this effect will run. We consider our approach to presenting multiple options at once as reflecting real world choices. Individuals may consider whether to speak with an individual in the context of speech with alternative people in their lives. The relative cost of these actions may depend on the reference individuals.

6. We will examine whether the order of the randomized targets (CD1-CD3) affects price outcomes.

7. We will examine how our results concerning T1 (the ask-tell intervention) change in response to filtering out respondents whose mean "ask'' estimates fall below 12% and who therefore would receive an upward rather than downward shock to their tolerance meta-perceptions.

8. We will examine how our results concerning heterogeneity by party_minority vary if we filter out respondents who answered that they would not vote to the 2024 vote intention CHIP50 core items.

9. We will examine how our results involving CD0-CD3 (video-sharing to a private target) vary if we filter out respondents who respond that their network includes "None'' of the following people: co-partisans, out-partisans, or co-partisans who are hesitant to vote for the party leader.

Analysis Details

Our estimands of interest will be covariate-adjusted (through the `estimatr` package in R) average treatment effects between our conditions. Covariates are specified above in our Covariates and Hypotheses sections. Our main conditional average treatment effects (CATEs) of interest are how video-sharing CD prices and the effect of our ask-tell treatment differ between party majorities and minorities. In addition, we will explore CATEs by party and network heterogeneity using Causal Forests from the `grf` package in R. When not otherwise specified, we will use the procedures described here in the Green Lab SOP.

As specified above in the Hypotheses section, we will use Bonferroni multiple comparisons corrections for RQ3 and H5b. All other hypotheses assess substantively different phenomena and define a single statistical test. For this reason, we reserve additional multiple comparison corrections for appendix analyses.

Power Analysis

To conclude, we briefly describe a basic power analysis for our main hypotheses of interest. We simulate power for several main hypotheses and for what we suspect is our least well-powered interaction model. We first show our power to distinguish between the different target CD prices. We make a guess as to the standard deviation of outcomes from a pre-treatment pilot.[18] Our pilot results suggested treatment demands for co-partisans ally, co-partisan opponents and out-partisans had means of $88, $102, and $129 respectively and standard deviations of $101, $106, and $107 respectively.

[18]This pilot included different ceilings than the main study but is useful as a reference in the absence of a more compelling prior.

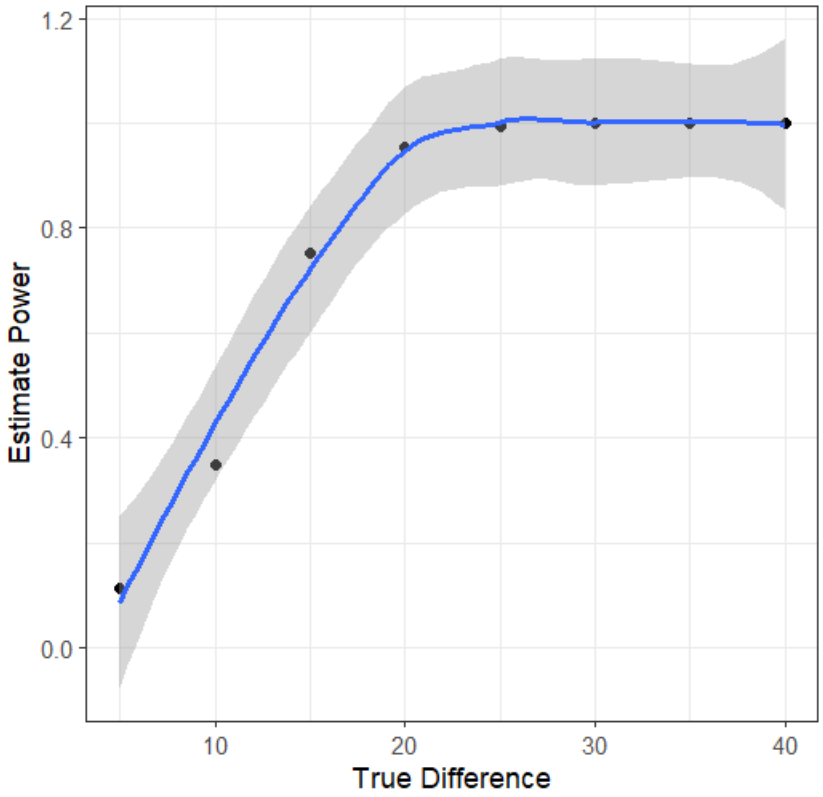

Figure A12: Power Curve for H5b

We first simulate power for H1b, which explores the difference in compensation demands between co-partisan opponents and allies. The first analysis will cluster standard errors at the respondent level, so our *effective N* for this comparison is less than 20,000 units. We find that we are nonetheless well-powered for this analysis across a range of estimated effect sizes (simulated from $10 to $50 differences).

We then simulate our investigation into H5a which explores co-partisan opponent prices for party minorities compared to party majorities. We believe this outcome is where we are most poorly powered because it conditions on a specific target and explores sub-group effects by party minority status, which we estimate will be about $\frac{1}{5}$ of our sample. The power curve below shows that we are poorly powered if we assume a small difference but become well-powered as the true estimand of the difference between sub-groups rises.