Learning from the Lie: What Do Voters Infer from Anti-Democratic Proposals?*

PRELIMINARY

Patrick Liu, Daniel Markovits, and Carlo Prato†

November 3, 2026

Abstract

Why do anti-democratic candidates endure and sometimes thrive in American politics, despite broad public opposition to many anti-democratic behaviors? Existing explanations focus on voters' failure to punish anti-democratic candidates because they agree with their policy stances. We propose and empirically evaluate a different mechanism, centered on voter learning: violations of democratic rules act as a signal of a candidate's policy positions and/or strong resolve. We propose a formal model and evaluate its implications using survey experiments that manipulate both the presence of democratic violations—such as election denial and threats of violence against opposing partisans—and the amount of information voters receive about candidates' policy stances and partisan loyalty. Our results show that voters' propensity to defect from anti-democratic candidates is less responsive to their ideology when they have more information. In line with our proposed mechanism, these effects are driven by differential inferences about future behavior once in office.

*We would like to thank Arturas Rozenas, Daniel Goldstein, Alex Coppock and conference participants at the Rebecca B. Morton Conference on Experimental Political Science, MPSA, and APSA. We would like to acknowledge funding support from the Center for Political Economy at Columbia World Projects.

†Columbia University Department of Political Science.

Introduction

Across the United States, politicians have repeatedly supported positions that undermine democratic norms and institutions—all while acting with seeming electoral impunity. Calls to gerrymander congressional maps, undermine or overturn electoral outcomes, and prosecute and jail political opponents have received skeptical media coverage, condemnation from neutral observers, and occasional reprimands from courts. What puzzles observers is that while most voters do not endorse these anti-democratic positions, the electoral fortunes of these politicians rarely suffer. Despite widespread public condemnation of his attempts to thwart the certification of the 2020 election and the extremely aggressive expansion of his presidential authority, Donald Trump has continued to enjoy widespread support. Republican politicians who embraced efforts to overturn the 2020 election faced only modest penalties in general elections. In fact, some may even have benefited from such actions in primary contests (Bartels and Carnes, 2023; Malzahn and Hall, 2025). California governor Gavin Newsom, once an also-ran in his party's 2028 nominating contest, was catapulted to the top of polling averages and prediction markets by his enthusiastic embrace of overtly partisan gerrymandering.

Scholars have increasingly lamented voters' reluctance to punish anti-democratic candidates, both in the U.S. and in comparable contexts where democratic backsliding is a key concern (Levitsky and Ziblatt, 2018; Graham and Svolik, 2020; Saikkonen and Christensen, 2023; Frederiksen, 2024; Chiopris et al., 2025). The general consensus these violations constitute a valence issue to voters: an unambiguously negative signal about candidate quality, akin to tax evasion or an extramarital affair. This approach frames the electoral impunity enjoyed by candidates as a consequence of ideologically aligned voters treating anti-democratic proposals as a necessary evil—an unsavory price they are nonetheless willing to tolerate for desired policy outcomes that are part of a candidate's issue bundles. Alternatively, scholars posit that voters fail to recognize the anti-democratic nature of some actions due to

ambiguity regarding their ramifications (Clayton et al., 2021; Chiopris et al., 2025).

This perspective offers two compelling intuitions. First, because voters express broad opposition to anti-democratic behaviors in the abstract (Ahmed, 2022; Holliday et al., 2024a,b), it seems natural to assume that voters' leniency stems from practical tradeoffs induced by a multidimensional electoral context. These tradeoffs, in turn, explain why polarization undermines accountability for democratic subversion. Even if voters treat anti-democratic proposals as a negative valence attribute, they may place so much weight on high-salience policy issues that they rarely defect from their most ideologically proximate candidate. Over time, candidate features ranging from incumbency to personal scandals have gradually declined in importance as party loyalty and polarization increased (Jacobson, 2015). Violations of democratic norms may be no different.

Yet, anti-democratic positions are not merely moral transgressions. They are concrete attempts to reshape political contestation by either introducing bias in electoral competition or by weakening horizontal constraints on executive power (Grillo et al., 2024), with immediate policy ramifications that voters can often intuit. As polarization rises, the desire to either achieve particular policy ends or to prevent out-partisans from pursuing their preferred policy outcomes may generate demand for candidates willing to take more extreme action. This logic is consistent with "supply-side" explanations for the rise of right-wing populists across Western democracies. This scholarship contends that voters who perceive incumbent institutions as gridlocked and dysfunctional may instrumentally prefer candidates willing to bend rules (Berman, 2021). Indeed, despite longstanding evidence that voters decry court packing and partisan gerrymandering in the abstract (Ahmed, 2022; Holliday et al., 2024a), a majority of Democratic voters expressed enthusiasm for expanding the Supreme Court in 2020 (Sanders, 2020), and over 60 percent of Californian voters endorsed a November 2025 ballot proposition to redraw the state's congressional districts in Democrats' favor.[1] While formal

[1]The ballot proposition was justified explicitly as offsetting a Republican gerrymandering proposal in Texas: "In response to Texas' mid-decade partisan congressional redistricting, this measure temporarily requires new congressional district maps, as passed by the Legislature in August 2025" (CA Attorney General,

theories of democratic conflict foreground the strategic stakes of anti-democratic reforms, public opinion research has seldom asked whether voters back anti-democratic candidates because they expect democratic backsliding to produce partisan or ideological advantages.

Our account complements prior theories about the tradeoffs voters face when democratic norms are at risk and policy polarization is large, even as we have many empirical expectations in common existing accounts of voters who tolerate backsliding because of strong preferences over policy outcomes. It also fills in a crucial gap, examining what voters *learn* about candidates from anti-democratic behaviors and proposals. How voters learn from candidate traits is growing area of research that has yet to be applied to the setting of anti-democratic proposals. We propose two theoretically novel channels. First, voters may treat anti-democratic proposals as evidence that a candidate sincerely holds more ideologically extreme policy preferences. Second, voters may learn about a candidate's ability to implement a sincerely held policy proposal that would otherwise be blocked by institutional constraints (e.g., courts or legislative opposition).

Even assuming that citizens of varying ideological stripes make similar inferences about candidates, both policy preferences and the availability of other information about candidates alter how these inferences shift vote choice. If our posited mechanism is correct, ideologically aligned voters—those on the far ends of the ideological spectrum—will perceive anti-democratic candidates in their camp as more closely aligned with their interests, unless they are provided concrete information to the contrary. Our framework therefore accommodates the intuition that polarization facilitates democratic backsliding, while allowing candidates to derive substantial electoral benefit from attacks on democratic norms, particularly in primary contests.

To investigate these possible pathways, we combine formal methods with experimental evidence. We propose a formal model of attitude formation that traces how voters think about

2025).

the risk and opportunities associated with anti-democratic candidates. Our formalization suggests multiple channels for voter learning that vary in complexity. We then employ a pair of survey experiments involving modified conjoint designs to unpack what voters learn from anti-democratic proposals under different information environments. Our first set of results centers around false claims of voter fraud in the 2020 election and shows that (i) violations of democratic norms are associated with an electoral punishment whose magnitude decreases with the amount of additional information about candidates that voters receive, (ii) these patterns are highly contingent on ideology, and (iii) the importance of ideology decreases with the amount of additional information about candidates that voters receive.

Violations of democratic norms do not exist in a vacuum: voters infer that Republican congressional candidates who endorse false claims about voter fraud are more ideologically extreme. While the magnitude of this learning decreases among voters who learn more information about the candidates' true positions, it remains substantially large. Conservative voters are more likely to support candidates who embrace claims of widespread fraud, whereas liberal voters are more likely to punish them. For our second (planned) experiment, we outline a design intended to further explore this learning mechanism and investigate the implications of our theoretical model. Theoretically, this paper contributes to a burgeoning literature on democratic backsliding, as well as to a sparser set of papers investigating when and how voters learn about candidate positions in primary elections. We seek to clarify when and how candidates benefit from taking often unpopular positions that undermine democratic rules or norms.

The Meaning of Anti-Democratic Proposals to Voters

A fundamental point of departure between our framework and prior accounts of the electoral tradeoffs surrounding anti-democratic candidates is that we seek to explain why voters might ever affirmatively seek anti-democratic candidates. A broad range of representative surveys show that Americans across party lines oppose a host of proposed and realized violations of

democratic norms, ranging from political violence to court packing (Holliday et al., 2024a,b). However, these data disguise important heterogeneity by topic, subgroup, and time. Graham and Svolik (2020) show that voters are somewhat more tolerant toward executive aggrandizement or gerrymandering than toward positions that undermine the free press or rule of law. Using Twitter posts by Donald Trump, Clayton et al. (2021) illustrate how partisan elites can "make the case" for anti-democratic behaviors among voters, a pernicious form of elite opinion leadership that may weaken commitments to democratic norms within specific partisan camps (Lenz, 2012; Barber and Pope, 2019). In addition, as the California gerrymandering referendum exemplified, norms erode over time. Spiraling violations of norms can drive support for violations that are unpopular in the abstract (Helmke et al., 2022; Braley et al., 2023; Janssen et al., 2025). Affirmative support for anti-democratic behavior is rare, but exceptions to the rule are mounting.

A similar trend emerges when we examine the empirical evidence regarding electoral choice: despite the general pattern that voters inconsistently or weakly sanction anti-democratic behavior at the ballot box, in some contexts voters appear to embrace it. Using a simple informational intervention, Walk et al. (2024) show that admonishing Trump for violating democratic norms produces a five-percentage point shift toward Kamala Harris—a striking finding in the crowded information environment of a general election—but that effects fade to insignificance ten days after treatment. Candidate choice conjoint experiments consistently across countries find that voters tend to prioritize partisanship and policy over defending democracy in a head-to-head (Graham and Svolik, 2020; Saikkonen and Christensen, 2023; Frederiksen, 2024). Observational evidence comparing the performance of Republican election deniers to co-partisans sharing the same ballot indicates modest underperformance in general elections (Bartels and Carnes, 2023; Malzahn and Hall, 2025). Yet Bartels and Carnes offer evidence that the House Republicans who defended Trump following the January 6 Capitol attack were less likely to lose primary elections in the 2022 midterms.

Understanding why voters seem to demonstrate a growing tolerance for certain kinds of democratic norm violations requires knowing what these violations mean to voters—whether they perceive these violations as means to desired political and policy ends or whether they prefer these policies for their own sake, failing to recognize the long-term ramifications. Chiopris et al. (2025) illustrate formally how politicians with authoritarian preferences might disguise the anti-democratic nature of their proposals, fooling voters who sincerely value democracy. The potential for elite opinion leadership to accomplish precisely this is perhaps best illustrated by the recent rise in support for restrictions to voting access. A majority of Republican voters possess sincere and deeply entrenched beliefs in a false narrative that the 2020 election was rife with fraud (Graham and Yair, 2023), leading to widespread support among Republicans for voter identification laws and other proposals that would limit the franchise (Carey et al., 2022). Our experiments investigate the generalizability of this phenomenon: while in Study 1 we leverage the peculiar case of Republicans' fixation on election fraud claims, in Study 2 we cast a wider net and assess what voters infer from candidates' positions on gerrymandering and court packing propositions.

While work on democratic norms and values has extensively explored support for anti-democratic proposals, there is little research on what these proposals mean to voters. Though cross-national studies have sought to explain the correlational inputs to support for democratic norms (Claassen, 2020), democratic values largely remain a black box. A handful of papers estimate beliefs about what democratic backsliding represents to voters besides an indication of personal or moral shortcomings. Chiopris et al. (2025) show that Polish citizens often interpret anti-democratic actions as expanding the policy space available to an incumbent government and that this belief mitigates the public backlash otherwise caused by such actions. In a similar vein, Markovits (2025) shows using an incentivized prediction experiment that voters believe anti-democratic behavior skews electoral outcomes. Formalized accounts make this logic more explicit still, arguing across a range of models that electoral manipulation and executive aggrandizement are the primary positive motivations for demo-

cratic backsliding (Grillo et al., 2024; Chiopris et al., 2025). A related class of formalizations explores the mechanisms by which populist candidates gain support in part by signaling their commitment to impractical policies that are opposed by establishment or technocratic parties (Morelli et al., 2021).

Learning about Candidates for Office

Given the possibility that voters may learn seemingly orthogonal information about candidates from the observation of anti-democratic proposals, primary elections are the stage of electoral competition where that learning is likely to be most meaningful. In general, the absence of party labels magnifies the effects of many of the very same mechanisms that are minimized in general elections, ranging from electioneering efforts to incumbency advantage (Kalla and Broockman, 2018; Green et al., 2024).

In addition, recent work has established a more specific empirical basis for the greater volatility of primaries. Using panel surveys, scholars have documented that voters clearly know fewer policy stances of a candidate than in general elections. Although this knowledge increases over the course of the primary campaign, it never approaches the levels reached in general elections. This result accentuates the importance of primaries as cases where voter learning is likely to be especially salient. While much of the literature on anti-democratic behavior focuses (more or less explicitly) on general elections, we emphasize the importance of primaries as arenas where anti-democratic candidates could gain special benefit from such proposals. Existing work on primaries has emphasized that (1) absent party cues, voter knowledge is minimal; (2) voters become more informed over time; and (3) over time, voters' electoral preferences converge toward candidates who share their policy views. There is, however, substantial heterogeneity in what types of candidates create which types of informational environments. Hirano et al. (2015) argue, using polling data from 1966 through 2006 in California, that voters move toward the ideologically proximate candidate over the course of the primary campaign and become less likely to support ideologically incongruent

candidates. These scholars contend that such findings are circumstantial evidence that voters learn about primary candidates' races. In lower-profile races with minimal media coverage, however, the evidence for this learning disappears.[2] Crucially, this paper notes that lower-level candidates are unlikely to shift opinions, reducing the risk of reverse causality driven by elite opinion leadership, though there is some evidence that lower-level elites can exercise similar effects (Broockman and Butler, 2017).

Moving beyond the immediate context of primary elections, a small body of work examines how voters learn about candidates' stances on non-policy issues. Sheagley et al. (2023) is closest in approach to this paper, studying how signals of loyalty to party leaders (Trump for Republicans, Nancy Pelosi for Democrats) and ideological descriptions compete to shape co-partisan approval. They find that leader loyalty and ideological consistency have broadly comparable effects on candidates' reputations among co-partisans, while out-partisans respond more favorably to candidates who break with party leaders than to those who merely depart from their party's ideological positioning. Banda and Sievert (2024) extends this logic, showing that voters interpret opposition to abolishing the filibuster as a signal of ideological moderation—though in general ideological terms rather than with respect to specific policies or behaviors. In that setting, the relevant democratic norm had a clear substantive implication: without 60 Senate votes, progressive priorities were difficult to enact unless the filibuster's constraints on simple majorities were removed. This literature suggests that voters update their views of candidates in response to candidates' positions toward party leaders. However, the mechanisms underlying this updating—and the extent to which additional information dampens it—remain unclear. Since V.O. Key's *Southern Politics in State and Nation* (1949), scholars have argued that intra-party factions are often too fluid to sustain effective accountability, a conclusion reinforced by Hopkins and Sigler (2025). Although Banda and Sievert (2024) and Hopkins and Noel (2022) show that leader cues can shift

[2]Perhaps similarly to diminished general election accountability in incongruent media markets Bawn et al. (2012).

broad ideological perceptions, it is unclear from these papers what more specific inferences voters draw. This stands in contrast to general elections, where partisanship bundles policy positions and social identity; leader support is a more ambiguous signal when voters are choosing between two candidates from the same party.

We further expect candidate disagreement with Trump's fraud claims to be a more powerful heuristic in the eyes of voters. First, voter fraud is an unusually strongly held belief. As evidence of this belief strength, see the small magnitude of corrections in Clayton et al. (2021) and the first study of our paper, as well as work by Graham and Yair (2023), which casts doubt on voter fraud claims as examples of expressive responding or partisan cheerleading. Second, Trump himself seems to have a unique hold on the Republican electorate, showing a capability of exercising elite opinion leadership even in a more liberal direction (Barber and Pope, 2019). More broadly, we are interested in what voters learn from normatively similar attacks on democratic norms, and we emphasize the diverse nature of such attacks.

Theorizing Learning from Anti-Democratic Proposals

Drawing on existing theories of voter learning and the importance of anti-democratic behavior, we propose a formalization of how candidate proposals that violate democratic norms translate into voter updating about seemingly orthogonal features of a candidate and then into vote choice.

In our model, we allow for multiple channels through which voters can learn about a candidate from the observation of an anti-democratic proposal, emphasizing the policy rather than valence implications of proposals that have important real-world effects on political competition.

Model

Players and actions. We study an election setting with one candidate j ("she") and one voter i ("he"). The candidate has type $\theta_j = (y_j, \rho_j, \delta_j)$, where (i) $y_j \in [0, 2]$ is her ideology—her preferred policy on a positional issue—, (ii) $\rho_j \in \{\rho^l, \rho^h\}$ is her *resolve*, and (iii) $\delta_j \in \{\delta^l, \delta^h\}$ is her democratic attachment—the intrinsic value she places on preserving democracy. We denote the voter's prior as $f^0(\theta) = \Pr\{y\} \Pr\{\delta\} \Pr\{\rho\}$ with $y_j \sim \mathcal{U}[0, 2]$, $\mu_\rho^0 = \Pr(\rho_j = \rho^h)$, and $\mu_\delta^0 = \Pr(\delta_j = \delta^h)$. To isolate the effect of information, we assume no correlation between ideology, resolve, and democratic attachment. The candidate chooses whether to take an anti-democratic position $a \in \{0, 1\}$ and then chooses a policy outcome $x \in \mathbb{R}$.

The voter has an ideology $y_i \in \mathbb{R}$ and democratic attachment δ_i. He also values the candidate's valence $\tau \in \mathbb{R}$, a term capturing a set of desirable attributes (charisma, honesty) that are orthogonal to ideology. The voter chooses whether to support the candidate, $s \in \{0, 1\}$. Importantly, s is chosen before x is determined but *after* a is announced.

Timeline and information. The interaction proceeds as follows:

1. $\theta_j = (y_j, \rho_j, \delta_j)$ is drawn and privately observed by j.
2. j chooses $a \in \{0, 1\}$
3. The valence shock τ is drawn and observed by i.
4. The voter observes the announced a and information $\mathcal{I}_T$ determined by a treatment T.
5. The voter chooses $s \in \{0, 1\}$.
6. The candidate chooses $x \in \mathbb{R}$.

We consider four information environments—or treatments:

- Baseline: $\mathcal{I}_0 = \emptyset$, no additional information.

- Resolve: $\mathcal{I}_\rho = (\rho_j)$.
- Ideology: $\mathcal{I}_y = (y_j)$.
- Full information: $\mathcal{I}_1 = (y_j, \theta_j)$.

Payoffs. The candidate's payoff is given by

$$u_j = \underbrace{-(x - y_j)^2}_{\text{policy affinity}} \underbrace{-c(a;\rho_j)\mathbf{1}\{x > \overline{x}\}}_{\text{institutional sanction}} \underbrace{-\delta_j a}_{\text{dem. attachment}} \underbrace{-\gamma a}_{\text{reputation}},$$

where $\overline{x} \in (0, 2)$ denotes the policy authority associated with the office sought by the candidate and $c(a;\rho_j) > 0$ is the cost of overstepping its authority capturing the subjective penalty inflicted by courts and other institutional actors. This cost depends on the candidate's resolve—its willingness to overcome popular, judicial, and legislative opposition in pursuit of her policy goals. It also depends on whether the candidate has taken an anti-democratic position, which often weakens the ability of these actors to oppose policy changes.

The parameter γ captures the potential penalty associated with $a = 1$.[3] The assumption intuitively captures the notion that commitments to executive aggrandizement and electoral manipulation will weaken both horizontal (courts, legislatures) and vertical (mass mobilization) checks on the candidate once elected.

The voter's payoff is given by

$$u_i = s\left[-(x - y_i)^2 + \tau - \delta_i a\right],$$

where $\tau \sim \mathcal{U}\left[-\frac{1}{2\phi}, \frac{1}{2\phi}\right]$ is the valence shock. A direct implication of this is that the net

[3] In a richer model, this cost would be endogenous to voters' support decisions. Given our focus on the public opinion implications of the model, we do not interpret the voter as a representative agent of the electorate.

ex-ante likelihood of i supporting j is a linear function of the candidate's appeal to i:[4]

$$E\{s_i^*|a, \mathcal{I}_T\} = \phi\mathbb{E}\{v_i^j|a, \mathcal{I}_T\} \equiv \phi\mathbb{E}\{-(x - y_i)^2 - \delta_i a|a, \mathcal{I}_T\}.$$

Notice that the responsiveness of a (or any other variable) affecting v_i^j is mediated by the parameter ϕ, which we refer to as the *relative salience* of the candidate's actions y and a.

Assumptions and Notation

We impose three substantive assumptions:

1. The institutional sanction decreases with a and with resolve: $c(1;\rho) < c(0;\rho)$ for all ρ and $c(a;\rho^h) < c(a;\rho^l)$ for all a.

2. Type Dominance: $\min\{c(0;\rho^h) - c(1;\rho^h), c(0;\rho^l) - c(1;\rho^l)\} > \gamma$ and $\delta^h + \gamma > \max\{c(0;\rho^h) - c(1;\rho^h), c(0;\rho^l) - c(1;\rho^l)\}$.

3. Resolve is more consequential than anti-democratic behavior: $c(1;\rho^l) > c(0;\rho^h)$.

The first assumption captures the idea that (i) all else equal, anti-democratic behavior weakens the constraints of accountability and thus reduces the expected severity of the sanctions associated with overstepping policy authority and that (ii) all else equal, resolve reduces the effectiveness of these sanctions. The second assumption ensures sufficiently large variation in candidates' democratic attachments and simplifies the description of the equilibrium. It eliminates trivial cases in which democratic attachments are inconsequential for subsequent behavior. The third assumption allows us to simplify the description of the equilibrium even under fairly weak democratic violations, but does not affect our main results.

[4]We study the presence of non-linearities in Appendix B.

Without loss of generality. we perform the following normalization

$$c(1;\rho^h)=0, \qquad c(0;\rho^l)=c^{\max}, \qquad \delta^l=0$$

so $c^{\max}$ is the largest sanction a candidate can suffer.

Finally, We impose three technical assumptions:

1. Tie-breaking: when indifferent, a candidate does not choose an anti-democratic position
2. Interior policy outcome: $c^{\max} < (2-\overline{x})^2$, which ensures that at least *some candidate* chooses to overstep her authority.
3. Arbitrarily small trembles: whenever a candidate chooses a, the voter observes a with probability $1-\varepsilon$ and $1-a$ with arbitrarily small provability. This avoids the issue of fully degenerate posteriors.

Strategies and beliefs. Pure strategies specify announcement $\widehat{a}:\theta_j \mapsto \Delta\{0,1\}$, support $\widehat{s}:(\tau,a,\mathcal{I}_T)\mapsto\Delta\{0,1\}$, and policy $\widehat{x}:(\theta_j,a,\tau,s)\mapsto\Delta\{\mathbb{R}\}$ choices. The voter's posterior is $F(\theta \mid \mathcal{I}_T)$, with marginals denoted by $F_y(\cdot \mid a,\mathcal{I}_T)$, $\mu_\rho=\Pr(\rho=\rho^h \mid a,\mathcal{I}_T)$, and $\mu_\delta=\Pr(\delta=\delta^h \mid a,\mathcal{I}_T)$.

Analysis

Equilibrium policies

In the Appendix, we show that once elected, a candidate will enact her ideal point *unless* her ideology falls within the *restraint window* $[\overline{x},\overline{x}+\overline{R}(\delta,\rho)]$, with

$$\overline{R}(\delta,\rho)=\begin{cases}\sqrt{\gamma+c(1;\rho)} & \delta=\delta^l,\\ \sqrt{c(0;\rho)} & \delta=\delta^h.\end{cases}$$

$\overline{R}$ captures the range of ideology types that are deterred by the threat of institutional sanctioning and, instead of enacting their ideal policy $y_j > \overline{x}$, choose not to overstep their authority.

The width of the restraint window is driven by institutional factors (via $c(\cdot,\cdot)$), public opinion (via γ), and by the candidate's resolve (ρ_j) and democratic attachment (δ_j). Notice that under the assumptions, they are ordered as follows:

$$\overline{R}(\delta^h,\rho^l) > \overline{R}(\delta^l,\rho^l) > \overline{R}(\delta^h,\rho^h) > \overline{R}(\delta^l,\rho^h).$$

Proposition 1. *In equilibrium,*

$$\widehat{a}(\theta_j) = \mathbf{1}\{\delta_j = \delta^l\}\mathbf{1}\{y_j > \overline{x} + \overline{R}(\delta_j,\rho_j)\},$$
$$\widehat{x}(\theta_j) = y_j + (\overline{x} - y_j)\mathbf{1}\{y_j \in [\overline{x}, \overline{x} + \overline{R}(\delta_j,\rho_j)]\}.$$

Our main characterization result implies that high-attachment candidates never take anti-democratic positions: even when they anticipate overstepping their authority when sufficiently extreme, the strength of their democratic attachment more than offsets the value of reducing the institutional sanction. Low-attachment candidates (δ^l), instead, face a different calculation and will *always* choose to take anti-democratic positions when they anticipate subsequent overstepping. Figures 1–4 illustrate.

Treatment effects

We now turn to the effect of observing an anti-democratic position on the likelihood of voter support. The expected support as a function of a is given by

$$\mathbb{E}\{v_i^j(x,a) \mid a, \mathcal{I}_T\} = -\delta_i a - \mathcal{D}_i(a,\mathcal{I}_T),$$

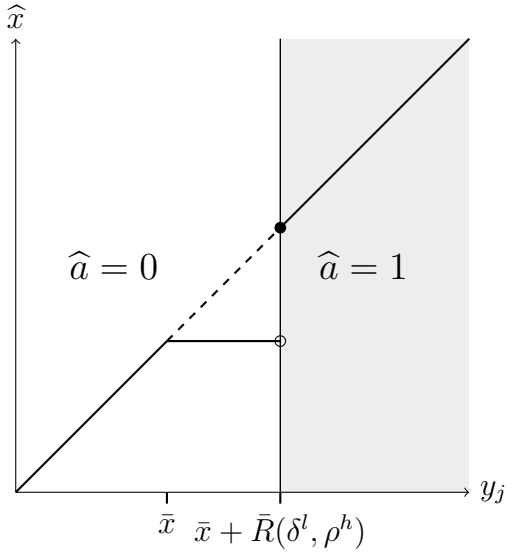


Figure 1: Translating Politician's Preferences Into Policy (δ^l, ρ^h)

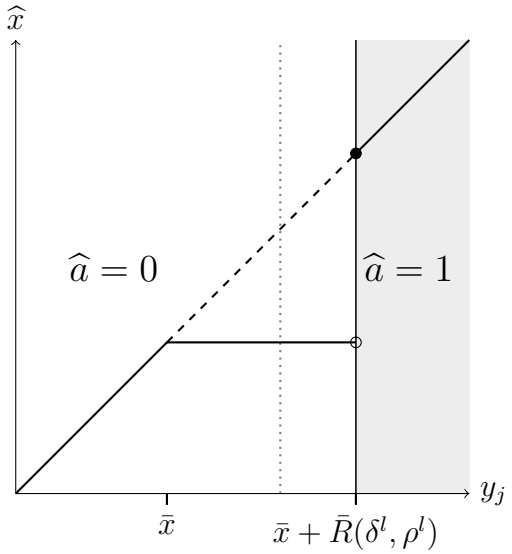


Figure 2: Translating Politician's Preferences Into Policy (δ^l, ρ^l)

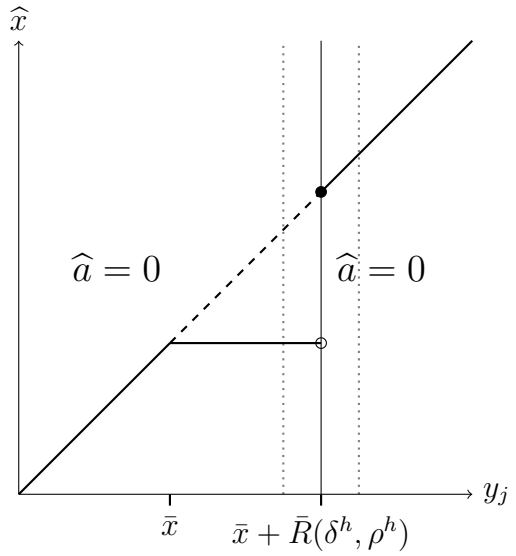


Figure 3: Translating Politician Preferences Into Policy (δ^h, ρ^h)

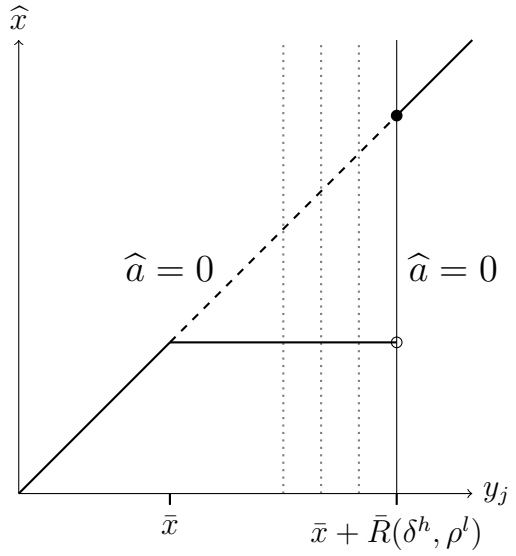


Figure 4: Translating Politician Preferences Into Policy (δ^h, ρ^l)

where

$$\mathcal{D}_i(a, \mathcal{I}_T) = \int_\theta (\widehat{x}(\theta) - y_i)^2 \, dF(\theta \mid \mathcal{I}_T)$$

is the expected policy disagreement between i and j given information $(a, \mathcal{I}_T)$. The change in the incentive to support associated with $a = 1$ under treatment T is thus given by

$$Z_i(\mathcal{I}_T) \equiv \mathcal{D}_i(0, \mathcal{I}_T) - \mathcal{D}_i(1, \mathcal{I}_T) - \delta_i.$$

Under a benchmark of full information ($T = 1$), $Z_i(\theta) = -\delta_i$ and is thus independent of the voter's ideology y_i.[5]

Voter ideology and the anti-democratic penalty. We can write the expected policy disagreement as

$$\mathcal{D}_i(a, \mathcal{I}_T) = y_i^2 - 2y_i \mathbb{E}\{x \mid a, \mathcal{I}_T\} + \mathbb{E}\{x^2 \mid a, \mathcal{I}_T\},$$

so that

$$\begin{aligned} Z_i(\mathcal{I}_T) + \delta_i = \mathcal{D}_i(0, \mathcal{I}_T) - \mathcal{D}_i(1, \mathcal{I}_T) = & 2y_i[\mathbb{E}\{x \mid a = 1, \mathcal{I}_T\} - \mathbb{E}\{x \mid a = 0, \mathcal{I}_T\}] \\ & + \mathbb{E}\{x^2 \mid a = 0, \mathcal{I}_T\} - \mathbb{E}\{x^2 \mid a = 1, \mathcal{I}_T\}. \end{aligned}$$

The following proposition implies that, after observing $a = 1$, the voter's expectations about policy weakly shift in a more extreme direction:

$$\mathbb{E}\{x \mid a = 1, \mathcal{I}_T\} \geq \mathbb{E}\{x \mid a = 0, \mathcal{I}_T\}.$$

and the same inequality holds for $\mathbb{E}\{x \mid a, \mathcal{I}_T\}$. Intuitively, democratic subversion reduces the institutional sanctions associated with the pursuit of extreme policy outcomes. Notice that this result holds despite the absence of primitive correlation between ideology, resolve, and democratic attachments. As a result, the expected change in support $Z_i(\mathcal{I}_T)$ is affine in

[5]Notice that we would also have $Z_i = -\delta_i$ if the voters did not observe *anything* about the candidate.

y_i.

Proposition 2. *For all treatments T, $Z_i(\mathcal{I}_T)$ can be written as*

$$Z_i(\mathcal{I}_T) = -\delta_i + z_0(\mathcal{I}_T) + z_1(\mathcal{I}_T)y_i$$

with $z_0(\mathcal{I}_T) \leq 0$ and $z_1(\mathcal{I}_T) \geq 0$. Moreover, when $T \in \{0, \rho\}$ or $T = y$ and y intermediate, we have $z_0(\mathcal{I}_T) < 0$ and $z_1(\mathcal{I}_T) > 0$, and there exists a cutoff $y^\dagger(\mathcal{I}_T) > 0$ such that $Z_i(\mathcal{I}_T) \geq -\delta_i$ iff $y_i \geq y^\dagger(\mathcal{I}_T)$.

The first part of the proposition implies that voters on the opposing ideological side of the candidate should lower their propensity to support her upon seeing $a = 1$, *even in the absence of democratic attachments* (i.e., even when $\delta_i = 0$). The reason is that they expect more extreme policy outcomes, which they intrinsically dislike. It also implies that the change in support associated with $a = 1$ is a linear function of the voter's ideology y_i. The slope coefficient, in particular, is proportional to the expected policy shift associated with observing an anti-democratic position:

$$\mathbb{E}\{x \mid a = 1, \mathcal{I}_T\} - \mathbb{E}\{x \mid a = 0, \mathcal{I}_T\}.$$

The second part says that, unless the voter receives enough information to fully predict policy, there exists a unique ideology cut-point such that observing $a = 1$ increases policy-affinity-based support if and only if the voter's ideology is above that cut-point. In other words, observed changes in support underestimate democratic attachments among the candidate's co-partisans and overestimate them among the candidate's ideological opponents. Figure 5 below illustrates.

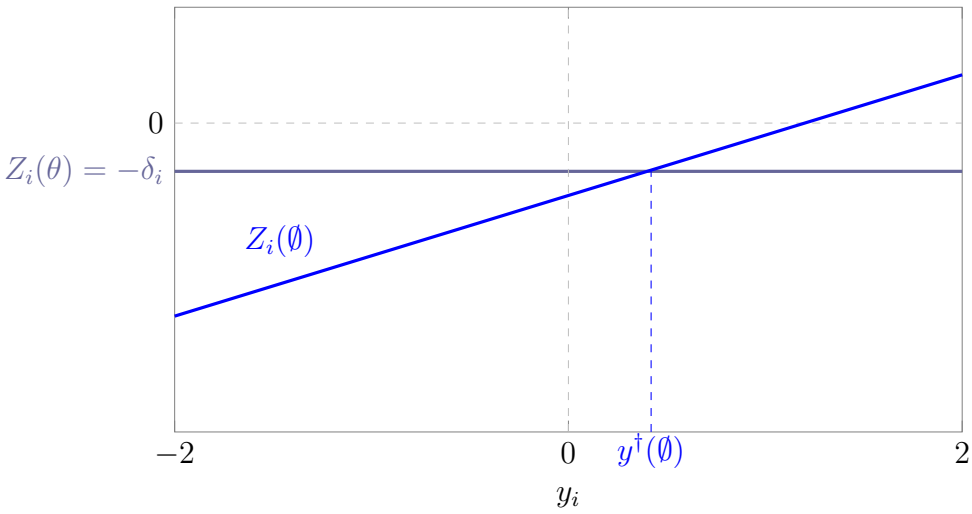


Figure 5: Baseline treatment effect. Change in support associated with an anti-democratic position, by voter ideology.

The effect of additional information. How does the effect of observing $a = 1$ vary with the amount of additional information voters receive? In this subsection, we consider two possible treatments: additional information about resolve ($T = \rho$) and additional information about the candidate's ideology ($T = y$). While naive Bayesian intuition suggests that additional information should always weaken the informational content of $a = 1$, this *does not* imply that the effect of observing $a = 1$ is less responsive to voter ideology.

Proposition 3. *The responsiveness of the anti-democratic penalty to voter ideology is larger when the candidate is known to be low-resolve than when she is known to be high-resolve:* $z_1(\rho^h) < z_1(\rho^l)$.

The proposition says that, relative to the baseline treatment, the effect of observing $a = 1$ is *more responsive* to voter ideology y_i after learning that j is low-resolve and *less responsive* after learning that j is high-resolve. Intuitively, if a candidate is low-resolve and still chooses $a = 1$, she must be very extreme. Conversely, even a relatively moderate high-resolve candidate would want to choose $a = 1$, because the reduction in the institutional sanction associated with exceeding $\overline{x}$ is very valuable for her. The key implication of this result is that more information does not necessarily "flatten" Z_i. Figure 6 illustrates.

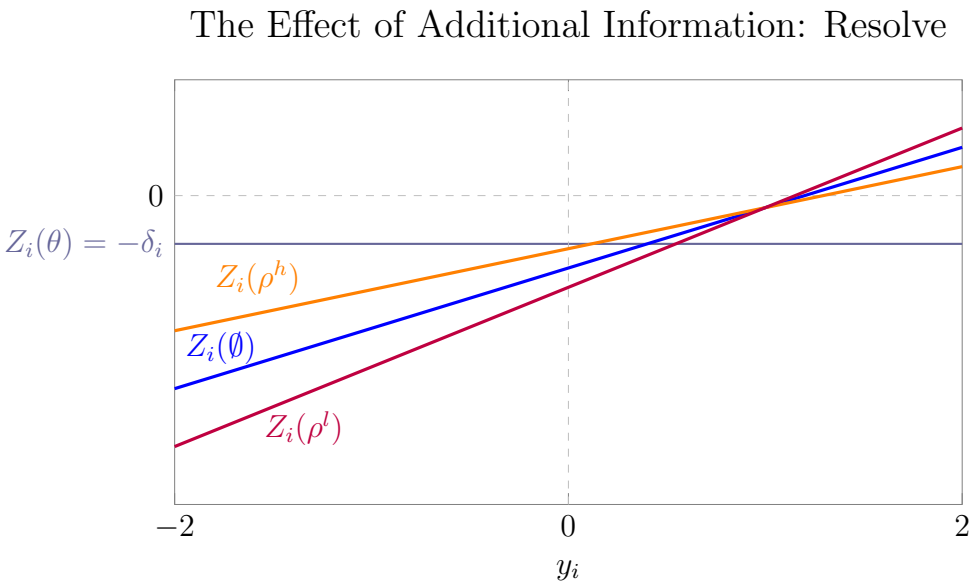


Figure 6: How additional information about resolve affects the treatment effect. Change in support associated with an anti-democratic position and knowledge of ρ_j, by voter ideology.

The effect of revealing a candidate's ideology, instead, flattens Z_i in the limit as the probability of not observing a goes to zero.

Proposition 4. *(i.) If* $y_j \notin \left[\overline{x} + \overline{R}(\delta^l, \rho^h),\ \overline{x} + \overline{R}(\delta^h, \rho^l)\right]$, *then* $Z_i(y_j) = -\delta_i$ *and* $z_1(y_j) = 0 < z_1(\emptyset)$;

(ii.) If $y_j \in \left(\overline{x} + \overline{R}(\delta^l, \rho^h),\ \overline{x} + \overline{R}(\delta^h, \rho^l)\right)$ *then* $z_1(y_j) \in (0, z_1(\emptyset))$, *provided* $c^{\max}$ *is sufficiently small.*

Intuitively, knowledge of a candidate's ideology makes her policy choice $\widehat{x}$ *extremely* predictable. Outside the interval

$$\left[\overline{x} + \overline{R}(\delta^l, \rho^h),\ \overline{x} + \overline{R}(\delta^h, \rho^l)\right],$$

knowing y_j is sufficient to predict $\widehat{x}$, so the policy component of $Z_i(y_j)$ vanishes and $Z_i(y_j)$ coincides with the full-information benchmark $Z_i(\theta) = -\delta_i$.

Inside that interval, the voter knows that the policy outcome will either be $\overline{x}$ or y_j. Thus, the posterior collapses to a Bernoulli: $\widehat{x} \in \{\overline{x}, y_j\}$. Under the regularity condition

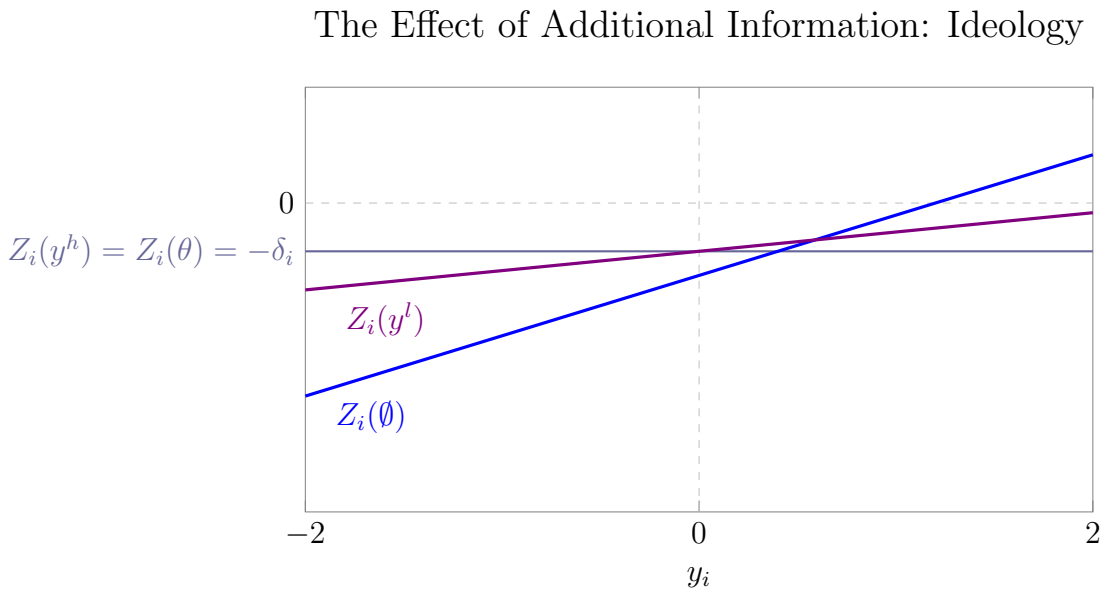


Figure 7: How additional information about ideology affects the treatment effect. Change in support associated with an anti-democratic position and knowledge of y_j, by voter ideology. The figure displays two possible values: $y^h > \overline{x}+\overline{R}(\delta^h, \rho^l)$ and $y^l \in [\overline{x}+\overline{R}(\delta^l, \rho^h), \overline{x}+\overline{R}(\delta^h, \rho^l)]$.

in Proposition 4, this makes Z_i flatter than in the baseline case: $z_1(y_j) < z_1(\emptyset)$. Figure 7 illustrates.

Alternative Mechanisms

We briefly discuss the robustness of our predictions. First, we consider a simpler mechanism: the anti-democratic action does not affect the cost of the sanction, but voters simply believe that type dimensions are correlated: more extreme candidates have lower democratic attachments and/or higher resolve. Second, we consider the possibility of non-linearity in the support function.

Direct Signaling via Type Correlation.

We assume that the action $a = 1$ does not, in the eyes of voters, affect the subsequent cost of overstepping one's authority: the institutional sanction is $c(\rho_j)$, with $c(\rho^h) \in (0, c(0; \rho^h))$ and $c(\rho^l) \in (c(1; \rho^l), c^{\max})$. We also assume any pattern of correlation between democratic

attachment, resolve, and ideology. Can we empirically distinguish between this setting and our baseline model? The answer is generally *yes*.

To see this, notice that in this case, in equilibrium, the choice of a cannot be a signal of resolve *nor ideology*. It can only respond to δ_j. To see this, observe that

$$\widehat{x}(\theta_j) = y_j + (\overline{x} - y_j)\mathbf{1}\left\{y_j \in \left[\overline{x}, \overline{x} + \sqrt{c(\rho_j)}\right]\right\}.$$

The optimal choice of x therefore does not depend on the previous choice of a. As a result,

$$\begin{aligned} V_j(a;\theta_j) = &-(\gamma + \delta_j)a - \mathbf{1}\left\{y_j \in \left[\overline{x}, \overline{x} + \sqrt{c(\rho_j)}\right]\right\}(y_j - \overline{x})^2 \\ &- \mathbf{1}\left\{y_j \in \left[\overline{x} + \sqrt{c(\rho_j)}, 2\right]\right\}c(\rho_j), \end{aligned}$$

which implies

$$\widehat{a}(\theta_j) = \mathbf{1}\{\gamma + \delta_j < 0\}.$$

As a result, in this version of the model, $a = 1$ can only occur if it is electorally rewarded ($\gamma < 0$) or if the candidate has an intrinsic taste for it ($\delta_j < 0$): under the assumptions, neither is possible.

Only by allowing $\gamma + \delta^l < 0$ this model can produce $a = 1$. Is it enough? Not really. Suppose that, in addition, $\gamma + \delta^h < 0$. In this case, we have $\widehat{a}(\theta_j) = 1$ for all θ_j, which implies, again, that voters cannot directly infer anything about $theta_j$ from $a = 1$.

Proposition 5. *Suppose a does not affect the institutional sanction, so that the sanction is $c(\rho_j)$. Unless $\gamma + \delta^l < 0 \leq \gamma + \delta^h$, observing $a = 1$ does not change the voter's belief about the implemented policy. Hence, for every treatment T, $Z_i(\mathcal{I}_T) = -\delta_i$..*

Let's now consider the starkest case, which involves $\gamma + \delta^l < 0 < \gamma + \delta^h$. We have

$$\widehat{a}(\theta_j) = \mathbf{1}\{\delta_j = \delta^l\}.$$

Further assume that $F(y|\delta^h, \rho) \succeq_{\text{FOSD}} F(y|\delta^h, \rho)\ \forall \rho \in \{\rho^l, \rho^h\}$ and $F(y|\delta, \rho^h) \succeq_{\text{FOSD}} F(y|\delta, \rho^l)$ $\forall \delta \in \{\delta^l, \delta^h\}$, so low-attachment and high-resolve candidates are on average more extreme. In this case observing $a = 1$ does shift expected policy outcomes and thus $Z_i(\mathcal{I}_T)$ is affine and weakly increasing in y_i, as in the baseline model.

Can we empirically distinguish this model from the baseline? The answer is yes. To see this, consider $T = \rho$, and consider the effect of additionally disclosing ρ. In general, it is no longer true that $z_1(\rho^h) < z_1(\rho^l)$. In fact, under fairly mild assumptions, the opposite obtains.

$$\begin{aligned}\frac{z_1(\rho)}{2} = &\int_{y \notin \left[\overline{x}, \overline{x}+\sqrt{c(\rho)}\right]} \left[F(y \mid \delta^h, \rho) - F(y \mid \delta^l, \rho)\right] dy \\ &+ \sqrt{c(\rho)} \left[F(\overline{x} + \sqrt{c(\rho)} \mid \delta^h, \rho) - F(\overline{x} + \sqrt{c(\rho)} \mid \delta^l, \rho)\right]\end{aligned}$$

We assume a form of complementarity between democratic attachment and resolve: specifically, the gap

$$D(y; \rho) \equiv F(y \mid \delta^h, \rho) - F(y \mid \delta^l, \rho).$$

is such that $D(y; \rho^h) \geq D(y; \rho^l)$ (strictly for some interval in $[0, 2]$) and weakly decreasing for $y \in \left[\overline{x} + \sqrt{c(\rho^h)}, \overline{x} + \sqrt{c(\rho^l)}\right]$. The first condition says that the increasing in extremism associate with low attachment is stronger among high-resolve candidates. The second is a regularity condition that says that the ideological separation between low-attachment and high-attachment types is not concentrated in the marginal restraint region where high-resolve candidates would overstep but low-resolve candidates would still be restrained.

Proposition 6. *Under the assumptions,* $z_1(\rho^h) > z_1(\rho^l)$

Non-linearities in support.

We want to formalize two observationally equivalent ideas: (i) most voters are so partisan/ideologically rigid that only a small subset of them is swayed by observing anti-

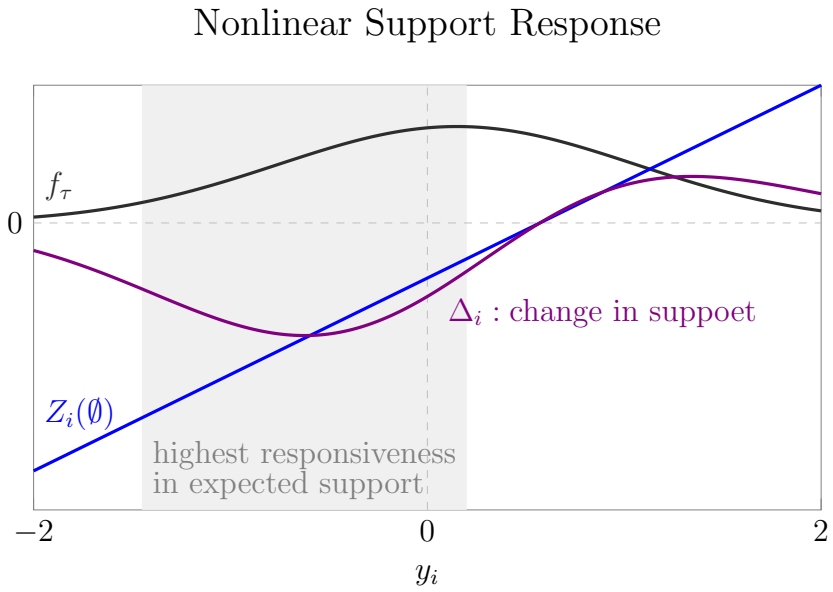


Figure 8: Nonlinear support responses. The blue line shows the informational effect $Z_i(\emptyset)$. The dark gray curve shows the local density of the valence shock near the relevant indifference threshold, plotted as a function of voter ideology. The violet curve shows the observed change in support $\Delta_i == S_i(1,\emptyset) - S_i(0,\emptyset) = F_\tau\left(-\mathcal{D}_i(1,\emptyset) - \delta_i\right) - F_\tau\left(-\mathcal{D}_i(0,\emptyset)\right)$, which is attenuated when the local density of the valence shock is small. The shaded region marks the range of voters for whom observed support is most responsive.

democratic positions, and (ii) voters are not as sophisticated as in the baseline model and do not actually update on policy.

In Appendix B, now consider a version of the model in which support is nonlinear in the candidate's expected appeal. Specifically, we assume that the valence shock τ be drawn from a continuous distribution F_τ, with density f_τ, symmetric, and single-peaked around zero.

The analysis shows that under fairly conventional assumptions, the presence of non-linearities should increase the responsiveness to $a = 1$ of voters on the opposite side of the ideological spectrum and decrease the responsiveness of voters on the the same side of the candidate, as illustrated in Figure 8.

Motivated Reasoning.

A more radical departure from our model is the idea that voters do not update their beliefs about y from their available information, but they choose them *after* having made a choice to support the candidate or not—to maximize consistency.

In this scenario, the causal chain is $a \rightarrow s \rightarrow X$, where X is a random variable denoting the voter's belief about the implemented policy.

Study 1

Our model generates a series of testable empirical implications—among them, that certain types of anti-democratic actions and proposals will boost rather than erode electoral support, that this electoral premium is concentrated among ideologically aligned voters holding extreme policy preferences, and that these effects are highly contingent on the amount of additional information provided to voters. Our proposed mechanism is that these effects stem from voters inferring candidates' policy behavior once in office. Study 1 offers initial empirical support for these intuitions. Study 2 then addresses competing interpretations, contrasts alternative types of anti-democratic proposals, and unpacks the mechanisms governing how voters learn from anti-democratic behavior.

Study 1 centered around a violation of democratic norms principally associated with President Donald Trump: demonstrably false claims about *widespread* election fraud, increasingly weaponized to justify restricting ballot access through proposals such as voter identification laws. Because (i) sincere belief in this "Big Lie" is common among Republican voters and (ii) Republican elites are uniquely positioned to correct these false beliefs (Graham and Yair, 2023; Carey et al., 2025), understanding the electoral incentives for right-wing candidates to promote voter fraud claims is a particularly consequential testing ground for our theory. Recruitment efforts of survey participants were therefore focused on Republicans and

Republican-leaning independents, who comprise 2,442 respondents in our sample. The remaining subjects were 1,479 Democrats and 746 pure independents. We fielded this factorial experiment between August 8 and 18, 2024, recruiting a total of 4,667 respondents from the online recruitment platform CloudResearch Connect.

Respondents were randomly assigned to view one of four pairs of candidate profiles. Each comparison represented a hypothetical Republican congressional primary contest featuring a Candidate A who supports voter fraud claims about the 2020 presidential election and a Candidate B who denies those claims.[6]

While the two candidates were identical along all other attributes, experimental conditions differ by the amount of information provided about these attributes. In the pure control condition, voters only learned the candidates' biographical information (two white males, the most common demographic profile among Republican members of Congress, with prior experience as a state legislators) and their views on election fraud. The other three conditions included additional information, listing either their conservative immigration policy ("Supports building a wall"), evidence of their partisan loyalty ("Campaigned hard for other Republican candidates"), or both. The conditions are summarized below:

1. Biography + Election Fraud Position
2. Biography + Election Fraud Position + Issue Position (Immigration)
3. Biography + Election Fraud Position + Partisan Loyalty
4. Biography + Election Fraud Position + Issue Position + Partisan Loyalty

Respondents were asked to choose between three options: abstention, voting for Candidate A, voting for Candidate B. They were then asked about how each candidate would behave once in office, rating the likelihood that each candidate would (i) cooperate with Democrats,

[6]Candidate A's listed "Views on the 2020 Election" were that he "Believes the 2020 election was stolen from Donald Trump and that Joe Biden was not legitimately elected president," whereas Candidate B "Believes the 2020 election was fairly won by Joe Biden and that no convincing evidence suggests that the 2020 election was stolen." Appendix D.1 contains the full outcome measures and treatment stimuli.

(ii) support a national abortion ban, and (iii) support raising tariffs. We refer to these measures as "learning" outcomes, in line with the idea that informational treatments should affect voters' behavioral inferences about candidates. Our final outcomes are (i) the between-candidate difference in voters' propensity to support and (ii) the between-candidate difference in voters' expectations about behavior in office across all three measures.

While comparable to a conjoint experiment, this design was limited to a single profile because of (i) direct comparability with our theory and (ii) the complexity of the outcome measures. Our design resembles the conjoint experiments in Fowler (2020), who randomize the number of candidate attributes to assess how information about issue positions mitigates the effect of partisanship. Fowler interprets his main finding—a reduction in partisan voting associated with expanded revelation of issue positions—as evidence that partisanship acts as a signal of candidate issue positions. By the same logic, finding that exposure to information about immigration positions and party loyalty diminishes the penalty suffered by Republican primary candidates who defend voter fraud claims would suggest that voters use anti-democratic stances to make behavioral inferences about candidates. Rather than simply suggesting it, the learning outcomes allow us to provide direct support for this interpretation.

Estimation and Hypotheses

We begin by assessing the effect of support for voter fraud claims on vote choice and learning outcomes in the control condition.This comparison is descriptive and does not have the interpretation afforded by AMCEs in standard conjoint designs, since other attributes are not randomized. However, it establishes a baseline for the electoral penalty associated with these claims that we can benchmark against the evidence of an electoral punishment (albeit modest) of anti-democratic behavior documented by (Graham and Svolik, 2020).[7]

As described in our pre-registration plan (Appendix F.1), we then evaluate the additional

[7]The important exception is support for voting restrictions, which receives support from Republican voters and donors (Carey et al., 2022).

effects of providing issue position and party loyalty information by regressing all outcomes on two binary indicators that capture whether issue position and party loyalty information were provided *in addition to biographical information and election fraud position.*

As is standard in factorial experiments, the sum of the treatment coefficients and the interaction term enables us to estimate the ATE for the provision of both types of information. To improve precision, we include pre-treatment measures of respondent's ideology, education, age, gender, and race. For our estimation, we use OLS and HC2 standard errors:

$$Y_i = \beta_0 + \beta_1 \text{Issue}_i + \beta_2 \text{Loyalty}_i + \beta_3 \text{Issue}_i \text{Loyalty}_i + \boldsymbol{X}_i' \boldsymbol{\gamma} + \epsilon_i, \tag{1}$$

where i indexes respondents, $\boldsymbol{X}_i$ is the vector of pre-registered covariates, and Y_i is one of the outcomes described above. For each respondent, the voting choice outcome takes values in $\{-1, 0, 1\}$, where 1 indicates a vote for the candidate supporting fraud claims, -1 for the candidate opposing fraud claims, and 0 for a decision not to vote. As a result, higher values indicate increased net favorability for the anti-democratic position. For each of our learning outcomes, Y_i is constructed by taking the difference in perceived likelihoods between candidates. The outcomes are ordered so that a higher value indicates the candidate supporting Trump's fraud claims is perceived as less likely to compromise with Democrats, more likely to support an abortion ban, or more likely to support raising tariffs.

This leads us to the following pre-registered hypothesis.

Hypothesis 1: The magnitude of the effect of supporting voter fraud claims on candidate choice and learning outcomes decreases with each additional piece of information.

This is a direct implication of Propositions 3 and 4. In the pre-analysis plan, we also specified a stronger version of the hypothesis: the effect of supporting voter fraud claims on candidate choice and learning outcomes will decrease to *zero* when two additional pieces of information are provided. This hypothesis was pre-registered *before* the model was written

and only follows from the model under additional assumptions.[8]

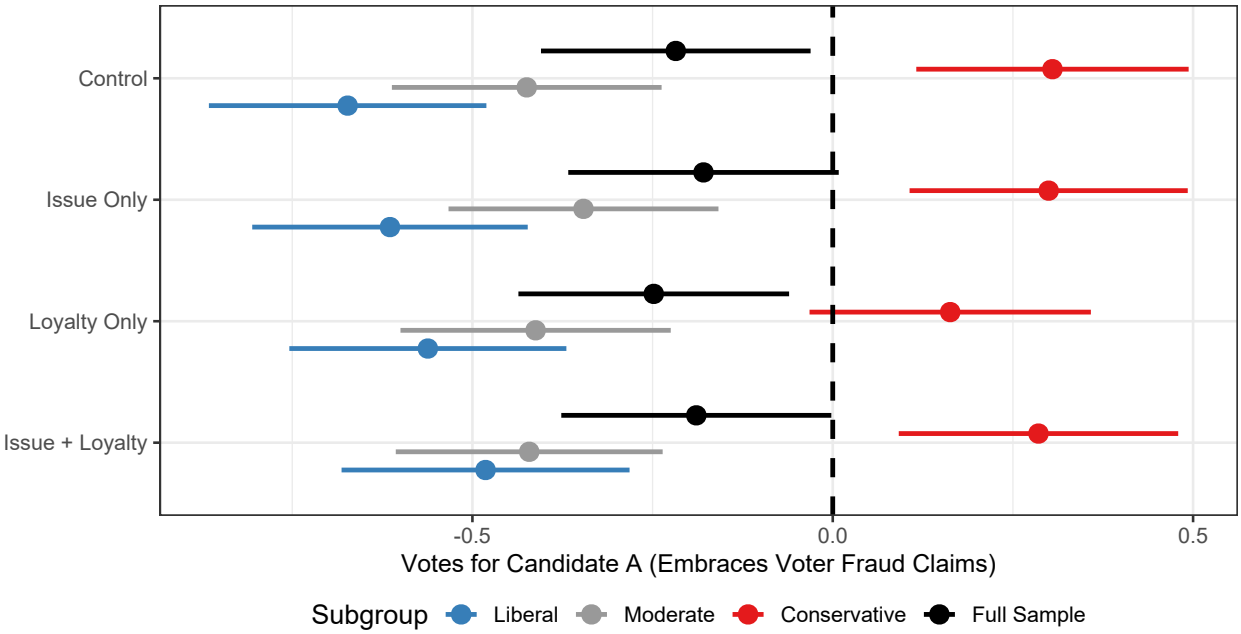
In line with the implications of our model, we also decompose the effects of information by respondent ideology by re-estimating equation (1) for each of three sub-samples based on self-reported ideology (liberal, moderate, conservative). Because our three-point ideology scale overlaps substantially with a three-point partisanship measure, we proceed with some caution when interpreting these results. This is because for self-identified Democratic voters participation in a Republican primary involves additional considerations of crossover voting in the opposing party's primary (Cohen and Markovits, 2025).

Results

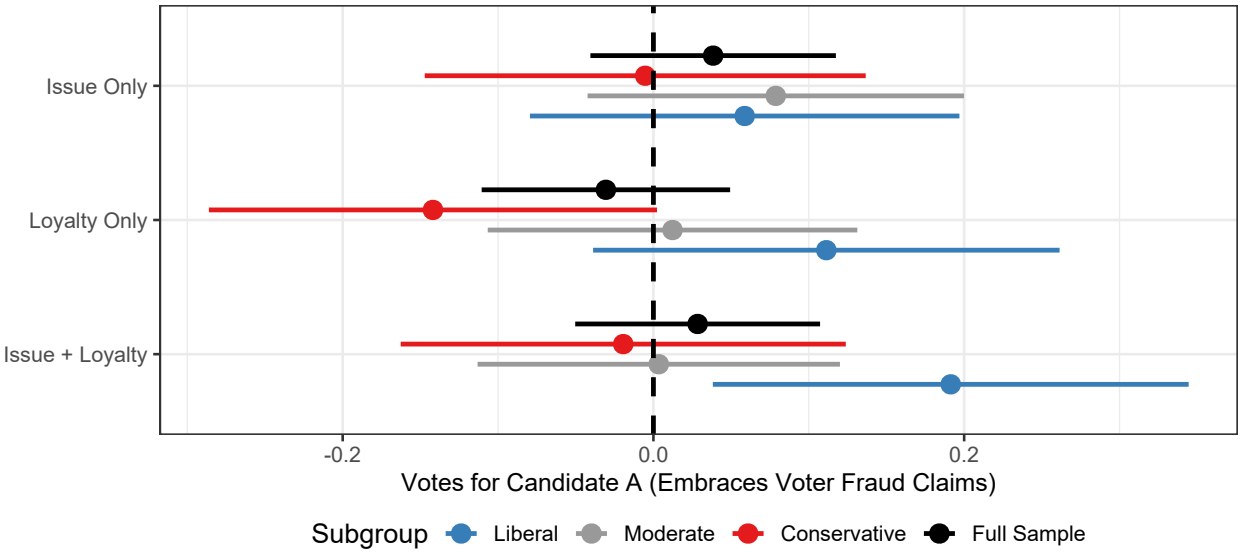
We begin by examining how voter fraud positions affect candidate support. The top panel of Figure 9 shows the estimated marginal means for our vote choice outcome across conditions, with full sample estimates in black as well as estimates decomposed by ideology. The Control group results shown in the top row suggest that voters of distinct ideological (or partisan) hues respond to positions on voter fraud in divergent ways. In the absence of more specific issue or loyalty information, self-reported liberals and moderates punish the candidate who embraces claims of voter fraud. Yet the same candidate is rewarded electorally by conservatives, with a net vote margin of around 30.5 percentage points. These results offer initial evidence consistent with our model predictions, suggesting anti-democratic positions sometimes beget an electoral advantage from ideologically aligned voters with sufficiently extreme preferences. We do note that, as Figure C1 in the appendix indicates, a similar pattern obtains when we decompose our effects by party over ideology, meaning these effects could be driven by partisanship rather than perceived policy alignment.

Next, we examine how the provision of additional information alters these patterns. The

[8]Specifically, we need to assume either (i) no noise in the observation of a candidate's policy positions ($\varepsilon = 0$, which implies that immigration is the only salient dimension of candidate ideological differentiation), (ii) or that voters make no assumptions about anti-democratic policies affecting the ease of implementing extreme policies (c independent of a), a possibility explored in the extension called "An Alternative Theory."



(a) Estimated Marginal Means



(b) Average Treatment Effects

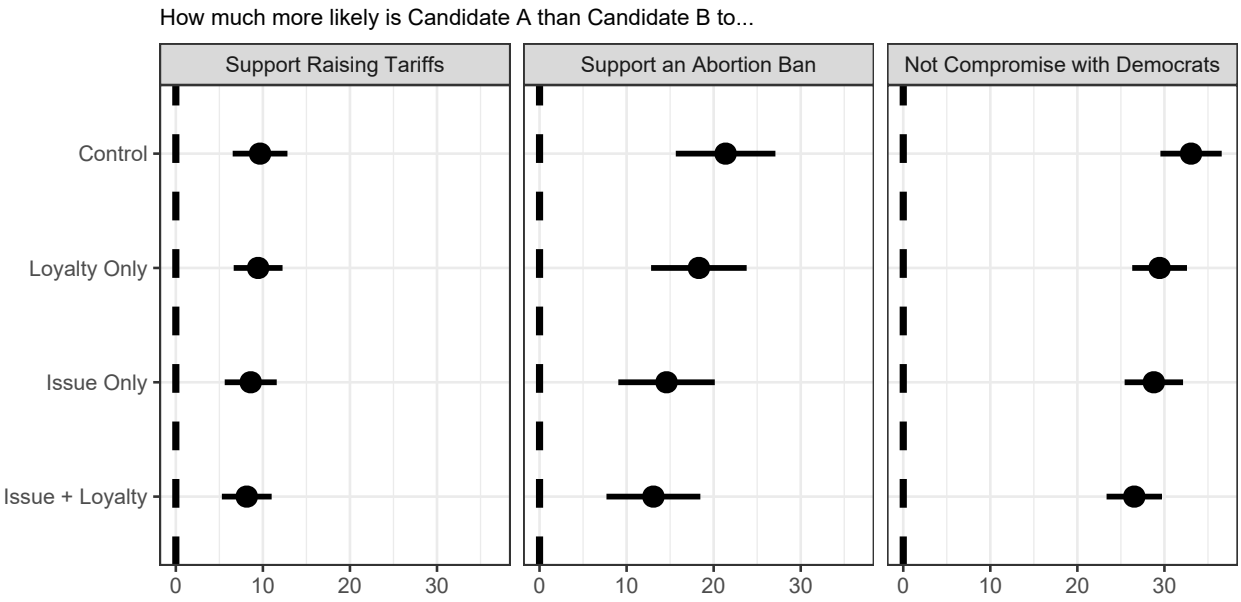
Figure 9: Net votes for Candidate A, by information condition and ideology. Error bars indicate 95% confidence intervals.

bottom panel of Figure 9 captures our estimated treatment effects. Information that both candidates are equally invested in the campaigns of fellow Republicans reduced the electoral premium that conservatives award to Candidate A. We detect no effects on vote choice from

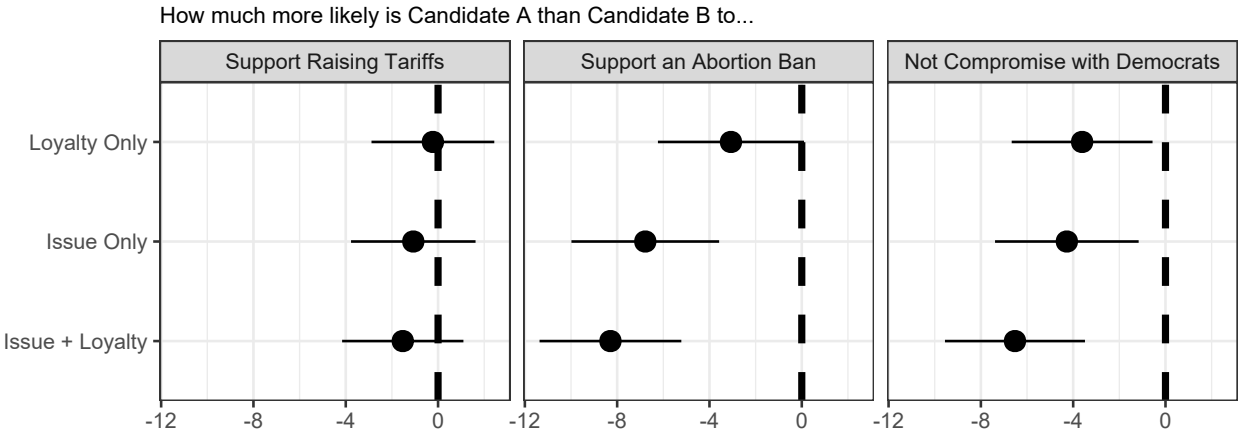
the provision of other sorts of information. Meanwhile, liberals became less supportive of the candidate who rejects fraud claims when they learned he had an identical party-stereotypical record to his anti-democratic opponent on both the issue and party loyalty dimensions. We find directionally similar but not-quite-significant effects from the provision of issue and party loyalty information alone.

Next, we evaluate our learning outcomes, highlighting descriptive evidence on the size of each gap at baseline and then how the gap changes in the presence of randomized information. As the top panel of Figure 10 illustrates, voters appear to learn a great deal from a candidate's position on voter fraud. In the absence of other policy information, respondents believed that a candidate who supports fraud claims would be 9.68 percentage points more likely to support raising tariffs, 21.4 percentage points more likely to support a national abortion ban, and 33.1 percentage points less likely to cooperate with Democrats. Consistent with our expectations and prior work, voters perceive candidates who support Trump's fraud claims as fundamentally distinct in terms of policy views from co-partisans who do not support those claims. These findings suggest that voters mentally bundle candidates' ideology with either their attitudes toward the democratic process and their alignment with Trump (Banda and Sievert, 2024; Hopkins and Noel, 2022).

These gaps narrow when respondents receive additional pro-stereotypical information, but the reductions are modest. As shown in Figure 10, the cooperation and abortion gaps shrink only modestly, by between 3 and 7 percentage points (less than 0.15 standard deviations) in response to a single piece of party-stereotypical information. We also see strong diminishing returns: combining pro-stereotypical policy and loyalty information barely outperforms—and is not statistically distinguishable from—either component alone in reducing perceived support for a national abortion ban or willingness to cooperate with Democrats in Congress. These results cohere with the weak form but not the strong form of our hypothesis. While we cannot extrapolate to the provision of (for example) a dozen pieces of policy information,



(a) Estimated Marginal Means



(b) Average Treatment Effects

Figure 10: Perceived gaps between Candidates A and B, by information condition. Error bars indicate 95% confidence intervals. Full model results in Appendix E.1.

our finding suggests that it may be difficult to fully bridge the gap in policy perceptions of candidates who support versus oppose voter fraud.

Figure 11 shows that voters of varying ideological hues make directionally similar inferences about candidates from their positions on voter fraud. We do, however, uncover meaningful differences in the degree of learning, which attribute to differential levels of pre-treatment knowledge about Republican candidates between the different ideological sub-groups. At

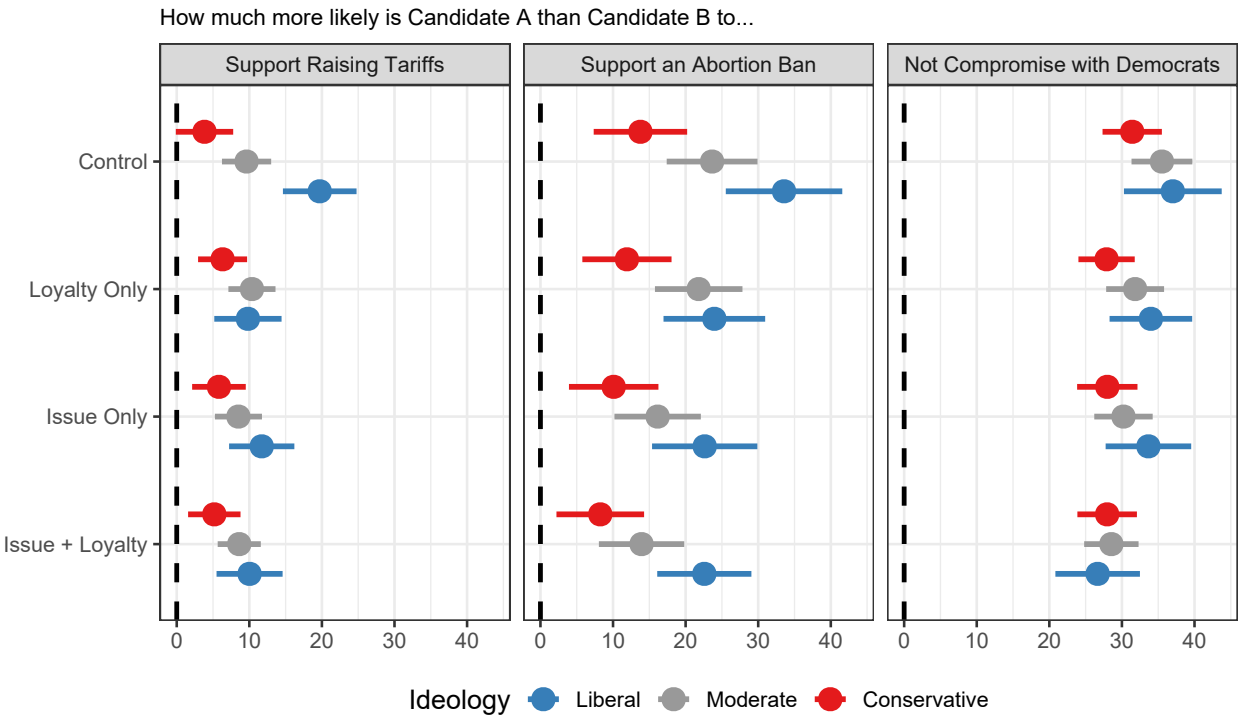


Figure 11: Perceived gaps between Candidates A and B, by information condition and self-reported ideology. Error bars indicate 95% confidence intervals.

baseline, conservatives treated the candidate's voter fraud position as less informative about their positions on tariffs or abortions than did moderates or liberals. This does not extend to perceptions about the candidate's likelihood of compromising with Democrats, where liberals, moderates, and conservatives appear equally likely to perceive embracing voter fraud claims as a signal of strong party loyalty. We further observe weaker effects from the randomized provision of party-stereotypical information on conservatives' candidate perceptions compared to liberals' candidate perceptions. This pattern is illustrated more clearly by the effects plot in Appendix Figure C2.

A central intuition of our formal model is that voters perceive candidates who violate democratic norms as likely to pursue more ideologically extreme policies. An alternative explanation for the learning we observe could be that respondents instead inferred how strongly candidates who support voter fraud claims were committed to supporting Trump as the party leader and backing his policy platform. Yet our results preclude this com-

peting interpretation. First, despite Trump having publicly denounced the possibility of a national abortion ban at the time of survey fielding, respondents inferred the candidate who embraces Trump's voter fraud claims would be substantially more likely to support an abortion ban. This perceived link between ideological moderation and opposition to voter fraud claims could still stem from many same-party critics of Trump—including prominent Never-Trumpers and Maine Senator Susan Collins—being social moderates. However, a second aspect of our results suggests voters draw more ambiguous inferences from voter fraud positions about policies where their party has yet to stake out a clear position.

Tariffs are a policy domain in which Trump has differed noticeably from many other Republicans, especially in August 2024 when our survey was fielded. If learning was driven by perceived loyalty to Trump, then the perceived "tariff gap" should be larger than the gaps observed for more standard policy areas. Instead, we find the opposite. The perceived tariff gap is smaller than the perceived abortion gap, and the effects of additional information on this gap are inconsistent and imprecisely estimated, even if in the expected direction.

That voters successfully infer positions on one issue from their knowledge of another—and that voters draw more ambiguous conclusions about issues where the parties have yet to clearly polarize at the elite level—suggests voters possess a form of "what goes with what" knowledge about candidate ideology (Converse, 1964). Our experiment was designed to mimic the sort of low-information environments in which primary voters may be operating (Hirano et al., 2015; Lupia, 1994), where party cues are absent, where voters must extrapolate from sparse known facts about elite behavior to make guesses about how candidates will perform in office, and where violations of democratic norms may receive disproportionate media coverage.

Limitations

One limitation of this initial design lies with the restricted randomization of issue and loyalty information: Candidates A and B are always shown to have identical positions. Thus, for an ideologically down-the-line conservative voter, the candidate who *opposes* Trump's fraud allegations can never exceed the ideological conservatism of Candidate A. This may explain why we see relatively little movement in either support for Candidate A or perceived ideological differences between the candidates among self-identified conservatives—especially if conservatives hold more confident priors about the ideological positions of candidates who support voter fraud claims. In contrast, for ideological liberals for whom participation in a Republican primary was an unusual choice, we see sharper shifts in net support. As Candidate B becomes relatively less attractive to these voters, they shift from supporting Candidate B to abstaining, drawing votes away from the pro-democracy candidate. Our subsequent study will evaluate this possibility by fully randomizing the issue positions taken by Candidates A and B, enabling respondents to sometimes observe a pro-democratic candidate who takes more extreme policy positions than their anti-democracy opponent and vice versa.

Study 2

In our first experiment, we investigated how voters learn from a specific violation of democratic norms: support for Trump's claims that the 2020 election was marred by fraud. In our second experiment, we both seek to generalize from this one behavior to a broader set of anti-democratic proposals and to allow for testing our full theoretical model. To accommodate this expanded scope and theoretical perspective, we use a more complex design while replicating some of the key features of our initial experiment.

Respondents ($N = 3,000$) will view 8 pairs of candidates, with partisans and independents

who lean toward a party assigned to view candidates competing in their party's primary for governor. To increase external validity, the sample will be restricted to self-reported registered voters in the 36 states with gubernatorial elections in 2026. Moreover, whereas the first experiment held attribute levels the same across candidates, here we fully randomize all attributes except the anti-democratic proposal in order to evaluate with greater generality how our theorized learning mechanisms operate in the context of complex electoral tradeoffs. On top of these randomized attribute levels, we further randomize the number of attributes seen by respondents, much like in Study 1.

Our main hypotheses center around three randomized dimensions: Type of Anti-Democratic Proposal (2 conditions), Number of Issue Attributes (3 conditions), and Number of Non-Policy Background Attributes (2 conditions). The result is a 2x3x2 factorial design, wherein the first randomization occurs within-subjects but between-trials while the second and third randomizations occur between-subjects.

Type of Anti-Democratic Proposal For every hypothetical election we present to respondents, one candidate takes an anti-democratic position ($a = 1$) while the other candidate does not ($a = 0$). Though restricting the randomization of this attribute means we do not estimate the average component marginal effect (AMCE) of anti-democratic proposals on candidate support or inferred issue positions, the independent and full randomization of all *other* attributes along with the random assignment of the anti-democratic position to Candidate A versus B means our quantity is nonetheless a causal one. This approach estimates what Abramson et al. (2023) describe as the *average feature choice probability* (AFCP): the average probability that a feature ($a = 1$) is selected in direct head-to-head contests with some baseline feature ($a = 0$). The added benefit of a restricted randomization here is power: since our subsequent analyses treat the perceived differences between the $a = 1$ candidate and $a = 0$ candidate as our task-level outcomes, comparisons between two $a = 0$ or two

$a = 1$ candidates would provide no outcome measure.[9]

To evaluate the dual mechanisms in our model, we randomize the type of anti-democratic behavior revealed for the $a = 1$ candidate between five possible anti-democratic actions: redrawing congressional maps to the advantage of their party, ignoring court rulings made by judges appointed by out-partisans, freezing legally mandated funding and directing state agencies to ignore related legislative mandates, attacking a journalist, or threatening violence against an out-partisan opponent. The first three are proposals with overt consequences for the types of policies the candidate will be able to deliver as well as obstruct once in office (y_i), whereas the latter two are behaviors that suggest contempt for the press or for the opposition with fewer direct policy implications.[10] As we discussed in the section "An Alternative Theory," the two behavioral forms of a do not affect $c(\rho_j)$. To test the alternative theory mechanism in our model, we directly test for Proposition 5, assessing if the effect of revealing this type of anti-democratic behavior on candidate support varies with ideology.

Number of Issue Attributes Each respondent will see 0, 1, or 2 pieces of information about the candidates' issue positions. We perform this randomization between-subjects so that each respondent observes the same number of issue attributes across all conjoint tasks, in order to avoid priming respondents to wonder at the absence of an issue position that was displayed in a previous profile. Those assigned to observe 1 issue attribute will be randomly assigned either to observe the candidates' positions on abortion or on border agents. Those assigned to observe 2 attributes will see both policy positions.

Number of Background Attributes In addition to anti-democratic proposals and issue attributes, candidate profiles will include rows for gender, race, and political party (always

[9] In affect, we are restricting randomizations to those that directly estimate the AFCP rather than those that only allow indirect comparisons contributing to an AMCE

[10] The funding freeze detail is phrased differently depending on respondents' partisanship. Republicans are told the candidate pledges to defund diversity-related programs and ignore related legislative mandates, while Democrats are told the candidate pledges to defund armed school resource officers and ignore state laws permitting concealed carry.

same-party). We then randomly assign half of our respondents to see one additional non-policy background attribute. This Prior Experience attribute is randomized across three possible levels: No Political Experience (baseline), State Attorney General, or Congressman/Congresswoman. This attribute will help to benchmark our results. While the addition of any background attributes will dilute the effects of anti-democratic proposals on candidate support, we expect that the provision of policy support will moderate the effect of a to much greater degree than providing information about prior experience, consistent with voters inferring policy positions from anti-democratic proposals. [11]

Estimation and Hypotheses

We assess how anti-democratic proposals alter candidate perceptions and support at the profile level. Borrowing notation from Hainmueller et al. (2014), respondents indexed by $i \in \{1, \ldots, N\}$ are shown $K = 8$ choice tasks each containing $J = 2$ alternatives. Thus, outcomes represented by Y_{ijk} are measured at the profile level, such as whether the candidate received a vote. We begin with the following specification,

$$Y_{ijk} = \beta_0 + \beta_1 \text{Anti-Democratic Proposal}_{ijk} + \beta_2 \text{Anti-Democratic Behavior}_{ijk} + \boldsymbol{X}_i'\boldsymbol{\gamma} + \epsilon_i, \tag{2}$$

where the first binary term Anti-Democratic Proposal$_{ijk}$ indicates if the candidate endorsed partisan gerrymandering, ignoring court orders from judges appointed by out-partisans, or freezing legally mandated funding in defiance of legislative mandates, and the second binary term Anti-Democratic Behavior$_{ijk}$ indicates if the candidate either attacked a journalist or threatened violence toward a political opponent. These are mutually exclusive attributes. We also include respondent covariates $\boldsymbol{X}_i$ (for precision) and cluster at the respondent level.

Our main interest lies with the approximate one-sixth of respondents assigned to receive no

[11] Note for discussant, we are considering eliminating this randomization because we are confident from the results in Bansak et al (2021)

additional information about either issue positions or background attributes. Subsetting our analysis to these respondents, our model predicts that anti-democratic proposals will alter candidate perceptions in several ways: increasing either the extremity of policies that voters believe will be accomplished under their administration or the extremity of issue positions that voters believe are sincerely held by the candidate, and decreasing the likelihood of policy ends desired by the opposing party. We further compare β_1 and β_2 to assess whether anti-democratic proposals produce both stronger inferences about candidates and increase vote choice further relative to anti-democratic behaviors with less immediate policy ramifications.

Proposition 5 in particular implies that the effect of our two anti-democratic behaviors on candidate support should not vary with respondent ideology. Here, rather than self-reported vote choice, we measure respondents' preferences on the particular issues for which we provide candidate stances and ask respondents to make inferences. We construct an index of these policy preferences but also evaluate the robustness of our findings using policy preferences on individual issues.[12] We modify equation 2 as follows:

$$\begin{aligned} Y_{ijk} = \beta_0 + \beta_1 \text{Anti-Dem Proposal}_{ijk} + \beta_2 \text{Anti-Dem Behavior}_{ijk} + \beta_3 \text{Covariate}_i \\ + \beta_4 \text{Anti-Dem Proposal}_{ijk} \text{Covariate}_i + \beta_5 \text{Anti-Dem Behavior}_{ijk} \text{Covariate}_i + \boldsymbol{X}_i' \boldsymbol{\gamma} + \epsilon_i, \end{aligned} \tag{3}$$

Let Ideology_i stand in for Covariate_i. A larger value indicates farther-right preferences for self-reported Republicans and farther-left preferences for Democrats. Proposition 3 implies that β_3 should be positive, where Proposition 5 implies that β_5 should be indistinguishable from 0.

Our subsequent analyses involve the following respondent-task specification, again clus-

[12] We do so in order to account for the potential multidimensionality of citizen preferences, as discussed by Broockman (2016). Following his approach, we ask respondents to select which of seven policies comes closest to their preferred policy in each of four domains: two randomized issues (abortion and border agents) along with the two issues embedded in our learning outcomes (climate and tax policy). On each domain, the middle position is the status quo, the positions to the left and right are respectively the Democratic and Republican party platforms, and the two leftmost and two rightmost positions are positions more extreme than the party mainstream.

tering at the respondent level:

$$Y_{ik} = \beta_0 + \beta_1 \text{Issues}_{ik} + \beta_2 \text{Background}_{ik} + \beta_3 \text{Anti-Democratic Position}_{ik} + \boldsymbol{X}_i' \boldsymbol{\gamma} + \epsilon_i, \quad (4)$$

where Issues_{ik} and Background_{ik} capture the randomized issue and background conditions to which respondents were assigned, and $\text{Anti-Democratic Position}_{ik}$ captures the type of anti-democratic proposal or behavior for the $a = 1$ candidate. We estimate separate regressions for those tasks where $a = 1$ took the form of gerrymandering, ignoring courts, or defying legislative mandates versus attacking journalists or threatening violence toward opponents. We expect that perceived differences between the $a = 0$ candidate and the $a = 1$ candidate who supports gerrymandering, ignoring courts, or defying legislative mandates will diminish with the provision of one or two pieces of issue information and that the effect of providing issue information will exceed that of providing a background attribute. We interpret this latter pattern as evidence that our empirical strategy in fact uncovers our modeled mechanism—anti-democratic proposals signal future policy ends, such that imputing information about candidate issue positions mitigates the effects of $a = 1$—and that the provision of issue information does not merely reduce the initial gap as a measurement artifact.

Results

This section will be populated when the second study is fielded in August of 2026.

1 Study 3

Our third experiment will use a voter-file-linked survey to estimate the effects of anti-democratic position-taking on what voters learn about real-world candidates appearing on the ballot in their states and on their subsequent vote choices. Because these respondents will not be compensated and will be recruited via text message, we will use a sparser experimental design to test our theories in the field. Our goal is to extend the basic intuitions of the

model to a more realistic electoral setting in which respondents evaluate actual candidates and receive information about positions those candidates have genuinely taken.

Respondents will be recruited via text message in New Hampshire in the days leading up to the September 8 primary. All respondents will be asked to evaluate four candidates running in the New Hampshire primary, with each candidate evaluated individually.[13] For each candidate, we will randomize whether the candidate's true position on gerrymandering is revealed. This allows us to estimate how learning about a candidate's position on an anti-democratic practice affects both what voters infer about that candidate and their willingness to support them. Because we are using real candidates, however, we are unable to randomize over the full range of possible positions without deceiving participants. Of note, our goal is to field a more comprehensive version of this design in the lead-up to the November midterm election, though we will test the design in the days immediately preceding the New Hampshire primary on September 8.

In addition, to investigate a mechanism that is present in our model but not tested using our convenience samples in Studies 1 or 2, we will randomize the candidate's level of resolve, which we define — consistent with our model — as the candidate's willingness to pursue a policy in the face of opposition. We are careful to operationalize resolve in a way that avoids directly implicating either anti-democratic actions or the candidate's ideology (for example by emphasizing the politician's enthusiasm for partisan conflict). Instead, the goal is to capture, using realistic language about a candidate, the type of personal or political trait that allows a politician to persevere in pursuing a preferred course of action despite opposition from political opponents. This design allows us to test whether voters interpret information about anti-democratic position-taking differently depending on how willing they believe a candidate is to follow through on their stated positions.

A final novel feature of this third design, relative to experiments 1 and 2, is the random

[13]We do this instead of using a paired-candidate, conjoint-style design in order to minimize the informational challenges for respondents.

assignment of a video in which a prominent party elite makes the case for gerrymandering. Democratic respondents will be assigned either a video featuring AOC or a placebo video, while Republican respondents will be assigned either a video featuring Trump or a placebo video. This component allows us to examine whether elite arguments in favor of gerrymandering alter how voters interpret candidates' positions on the issue and how those positions affect candidate evaluations and vote choice. Consistent with the third experiment's goal of incorporating realism, this type of argument is common in how partisan elites make the case for gerrymandering.

Discussion

Why do voters support anti-democratic politicians? While existing research suggests that voters weakly oppose proposals that violate democratic norms, it offers few explanations for how such proposals might generate affirmative electoral benefits for candidates and does not explore the full set of meanings that voters may attribute to anti-democratic proposals. Specifically, neither empirical nor formal work investigates how anti-democratic proposals serve as signals of other candidate traits, especially in low-information contexts. This omission leaves the literature ill-equipped to explain why ambitious politicians have so enthusiastically endorsed—and seemingly benefited from—violations of democratic norms across parties and levels of government, both in the United States and in comparable democracies.

We propose a set of mechanisms through which heterogeneous violations of democratic norms can produce electoral benefits for politicians; with these benefits explained in large part by what voters learn about candidates from observing an anti-democratic proposal in a sparse information environment. These mechanisms can both reduce the punishment associated with unpopular anti-democratic actions and amplify the rewards attached to popular attacks on democratic rules, thereby reshaping the electoral incentives facing candidates who undermine democratic institutions.

Crucially, our framework flexibly accounts for the diversity of democratic violations (Graham and Svolik, 2020; Ahmed, 2022; Holliday et al., 2024a), with some approximating valence issues in their unpopularity and others achieving majority support within important segments of the population, such as primary electorates. Even isolated instances where backsliding appears popular as a standalone issue, such as support for gerrymandering among Democratic voters in California, cannot be fully explained if we do not understand what voters consider as the policy implications of anti-democratic behavior

This paper also contributes to a literature on information and voter learning in primaries, showing the power of anti-democratic proposals in low-information environments. Given the dearth of general election competition in most US states and an overwhelming majority of congressional districts (Hirano and Jr, 2019), primaries are the key arena which determine the fortunes of anti-democratic candidates. Our results show that the learning effects of anti-democratic proposals are magnified in contexts where little is known about the candidates, even before accounting for the extent to which media is likely to amplify anti-democratic proposals. These results suggest that politicians in lower-salience elections, or those about whom the public has weak knowledge of policy platforms, have a greater incentive to magnify attacks on democratic rules and norms. As a literature on how public opinion constraints attack on vertical and horizontal checks grows, more attention should be paid to how politicians predict the public will react to democratic violations (Christenson and Kriner, 2019; Nyhan and Titiunik, 2024).

Finally, our model and empirical results allow for two mechanisms through which voters can learn about ideology through an anti-democratic proposal. The first is by learning about how attacks on democracy norms erode policy constraints. The second is through assumptions about the correlation of anti-democratic proposals and particular policy stances in the population of same-party candidates for office that is observed over time by primary voters. This latter mechanism allows voters to have only very general priors about the stances

of candidates and allows learning to happen even if voters do not make specific inferences about the effects of anti-democratic policies on ideological outcomes.

What are the implications of our results for political competition in America today? Crucially, we predict heterogeneity in how anti-democratic proposals affect candidates in high versus low-information contexts, in ways that interact with the composition of primary electorates to create heterogeneous incentives. In low-salience elections where we might expect little voter learning and a highly ideological electorate (Hirano et al., 2015; Hirano and Jr, 2019), candidates may have strong incentives to endorse anti-democratic proposals, while in other contexts they may face incentives to adjure or downplay the same proposals.

References

Scott F Abramson, Korhan Kocak, Asya Magazinnik, and Anton Strezhnev. Aggregation, Interpretation, and Estimation of Preferences in Conjoint Experiments, April 2023. URL `https://osf.io/preprints/socarxiv/xjre9_v1/`.

Amel Ahmed. Is the American Public Really Turning Away from Democracy? Backsliding and the Conceptual Challenges of Understanding Public Attitudes. *Perspectives on Politics*, pages 1–12, July 2022. ISSN 1537-5927, 1541-0986. doi: 10.1017/S1537592722001062.

Kevin K. Banda and Joel Sievert. How Filibuster Rhetoric Informs Perceptions of Politicians. *Legislative Studies Quarterly*, 49(3):673–693, 2024. ISSN 1939-9162. doi: 10.1111/lsq.12445.

Michael Barber and Jeremy C. Pope. Does Party Trump Ideology? Disentangling Party and Ideology in America. *American Political Science Review*, 113(1):38–54, February 2019. ISSN 0003-0554, 1537-5943. doi: 10.1017/S0003055418000795.

Larry M. Bartels and Nicholas Carnes. House Republicans were rewarded for supporting Donald Trump's 'stop the steal' efforts. *Proceedings of the National Academy of Sciences*, 120(34):e2309072120, August 2023. doi: 10.1073/pnas.2309072120.

Kathleen Bawn, Martin Cohen, David Karol, Seth Masket, Hans Noel, and John Zaller. A Theory of Political Parties: Groups, Policy Demands and Nominations in American Politics. *Perspectives on Politics*, 10(3):571–597, 2012. ISSN 1537-5927. doi: 10.1017/S1537592712001624.

Sheri Berman. The Causes of Populism in the West. *Annual Review of Political Science*, 24 (Volume 24, 2021):71–88, May 2021. ISSN 1094-2939, 1545-1577. doi: 10.1146/annurev-polisci-041719-102503.

Alia Braley, Gabriel S. Lenz, Dhaval Adjodah, Hossein Rahnama, and Alex Pentland. Why

voters who value democracy participate in democratic backsliding. *Nature Human Behaviour*, 7(8):1282–1293, August 2023. ISSN 2397-3374. doi: 10.1038/s41562-023-01594-w.

David E. Broockman. Approaches to Studying Policy Representation. *Legislative Studies Quarterly*, 41(1):181–215, 2016. ISSN 1939-9162. doi: 10.1111/lsq.12110.

David E. Broockman and Daniel M. Butler. The Causal Effects of Elite Position-Taking on Voter Attitudes: Field Experiments with Elite Communication. *American Journal of Political Science*, 61(1):208–221, 2017. ISSN 1540-5907. doi: 10.1111/ajps.12243. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12243`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12243.

CA Attorney General. Proposition 50. Technical report, 2025. URL `https://vig.cdn.sos.ca.gov/2025/special/pdf/prop50.pdf`.

John Carey, Katherine Clayton, Gretchen Helmke, Brendan Nyhan, Mitchell Sanders, and Susan Stokes. Who will defend democracy? Evaluating tradeoffs in candidate support among partisan donors and voters. *Journal of Elections, Public Opinion and Parties*, 32 (1):230–245, January 2022. ISSN 1745-7289. doi: 10.1080/17457289.2020.1790577.

John M. Carey, Brian Fogarty, Marília Gehrke, Brendan Nyhan, and Jason Reifler. Prebunking and credible source corrections increase election credibility: Evidence from the US and Brazil. *Science Advances*, 11(35):eadv3758, August 2025. doi: 10.1126/sciadv.adv3758.

Caterina Chiopris, Monika Nalepa, and Georg Vanberg. A Wolf in Sheep's Clothing: Citizen Uncertainty and Democratic Backsliding. *The Journal of Politics*, 87(4):1272–1287, October 2025. ISSN 0022-3816. doi: 10.1086/734253.

Dino P. Christenson and Douglas L. Kriner. Does Public Opinion Constrain Presidential Unilateralism? *The American Political Science Review*, 113(4):1071–1077, 2019. ISSN 0003-0554.

Christopher Claassen. In the Mood for Democracy? Democratic Support as Thermostatic Opinion. *The American Political Science Review*, 114(1):36–53, 2020. ISSN 0003-0554.

Katherine Clayton, Nicholas T. Davis, Brendan Nyhan, Ethan Porter, Timothy J. Ryan, and Thomas J. Wood. Elite rhetoric can undermine democratic norms. *Proceedings of the National Academy of Sciences*, 118(23):e2024125118, June 2021. doi: 10.1073/pnas.2024125118.

Hayley Cohen and Daniel B Markovits. Crossover Voting in Congressional Primaries. 2025.

Philip E. Converse. The Nature of Belief Systems in Mass Publics. In D. E. Apter, editor, *Ideology and Discontent*, pages 206–261. University of Michigan Press, Ann Arbor, 1964.

Anthony Fowler. Partisan Intoxication or Policy Voting? *Quarterly Journal of Political Science*, 15(2):141–179, April 2020. ISSN 15540634. doi: 10.1561/100.00018027a.

Kristian Vrede Skaaning Frederiksen. Do Partisanship and Policy Agreement Make Citizens Tolerate Undemocratic Behavior? *The Journal of Politics*, 86(2):766–781, April 2024. ISSN 0022-3816, 1468-2508. doi: 10.1086/726938. URL `https://www.journals.uchicago.edu/doi/10.1086/726938`.

Matthew H. Graham and Milan W. Svolik. Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States. *American Political Science Review*, 114(2):392–409, May 2020. ISSN 0003-0554, 1537-5943. doi: 10.1017/S0003055420000052.

Matthew H. Graham and Omer Yair. Expressive Responding and Trump's Big Lie. *Political Behavior*, July 2023. ISSN 1573-6687. doi: 10.1007/s11109-023-09875-w. URL `https://doi.org/10.1007/s11109-023-09875-w`.

Donald P. Green, Thomas Leavitt, and Daniel Markovits. Challenges that Proprietary Research Poses for Meta-Analysis. In Janet M. Box-Steffensmeier, Dino P. Christenson, and Valeria Sinclair-Chapman, editors, *Oxford Handbook of Engaged Methodological Pluralism*

in Political Science, page 0. Oxford University Press, October 2024. ISBN 978-0-19-286828-2. doi: 10.1093/oxfordhb/9780192868282.013.21.

Edoardo Grillo, Zhaotian Luo, Monika Nalepa, and Carlo Prato. Theories of Democratic Backsliding. *Annual Review of Political Science*, 27(Volume 27, 2024):381–400, July 2024. ISSN 1094-2939, 1545-1577. doi: 10.1146/annurev-polisci-041322-025352.

Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments. *Political Analysis*, 22(1):1–30, January 2014. ISSN 1047-1987, 1476-4989. doi: 10.1093/pan/mpt024.

Gretchen Helmke, Mary Kroeger, and Jack Paine. Democracy by Deterrence: Norms, Constitutions, and Electoral Tilting. *American Journal of Political Science*, 66(2): 434–450, 2022. ISSN 1540-5907. doi: 10.1111/ajps.12668. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12668. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12668.

Shigeo Hirano and James M. Snyder Jr. *Primary Elections in the United States*. Cambridge University Press, September 2019. ISBN 978-1-108-66624-4. Google-Books-ID: 3NisDwAAQBAJ.

Shigeo Hirano, Gabriel S. Lenz, Maksim Pinkovskiy, and James M. Snyder Jr. Voter Learning in State Primary Elections. *American Journal of Political Science*, 59 (1):91–108, 2015. ISSN 1540-5907. doi: 10.1111/ajps.12093. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12093. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12093.

Derek E. Holliday, Shanto Iyengar, Yphtach Lelkes, and Sean J. Westwood. Uncommon and nonpartisan: Antidemocratic attitudes in the American public. *Proceedings of the National Academy of Sciences*, 121(13):e2313013121, March 2024a. doi: 10.1073/pnas.2313013121.

Derek E. Holliday, Yphtach Lelkes, and Sean J. Westwood. The July 2024 Trump assassination attempt was followed by lower in-group support for partisan violence and increased group unity. *Proceedings of the National Academy of Sciences*, 121(49):e2414689121, December 2024b. doi: 10.1073/pnas.2414689121.

Daniel J. Hopkins and Hans Noel. Trump and the Shifting Meaning of "Conservative": Using Activists' Pairwise Comparisons to Measure Politicians' Perceived Ideologies. *American Political Science Review*, 116(3):1133–1140, August 2022. ISSN 0003-0554, 1537-5943. doi: 10.1017/S0003055421001416. URL https://www.cambridge.org/core/journals/american-political-science-review/article/trump-and-the-shifting-meaning-of-conservative-using-activists-pairwise-comparisons-to-measure-politicians-perceived-ideologies/EC8511D8EB16CD947D9B5DC8201F0515.

Gary C. Jacobson. It's Nothing Personal: The Decline of the Incumbency Advantage in US House Elections. *The Journal of Politics*, 77(3):861–873, July 2015. ISSN 0022-3816. doi: 10.1086/681670.

Lisa Janssen, Anna Kern, and Hannah Werner. 'When they go low, we kick them' : affective polarization and its effects on the support for reciprocal democratic transgressions. In *Working Paper*, 2025. URL https://files.osf.io/v1/resources/pwcz2_v1/providers/osfstorage/68484094d7e96b4d055395bb?format=pdf&action=download&direct&version=2.

Joshua L. Kalla and David E. Broockman. The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments. *The American Political Science Review*, 112(1):148–166, 2018. ISSN 0003-0554.

Gabriel S. Lenz. *Follow the Leader?: How Voters Respond to Politicians' Policies and Performance*. University of Chicago Press, October 2012. ISBN 978-0-226-47213-3. Google-Books-ID: H_ownvbnWCgC.

Steven Levitsky and Daniel Ziblatt. *How Democracies Die*. Penguin, New York, 2018.

Arthur Lupia. Shortcuts Versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections. *American Political Science Review*, 88(1):63–76, March 1994. ISSN 1537-5943, 0003-0554. doi: 10.2307/2944882.

Janet Malzahn and Andrew B. Hall. Election-Denying Republican Candidates Underperformed in the 2022 Midterms. *American Political Science Review*, 119(3):1536–1541, August 2025. ISSN 0003-0554, 1537-5943. doi: 10.1017/S0003055424001084.

Daniel Markovits. Beliefs about Democratic Backsliding. *Working Paper*, 2025.

Massimo Morelli, Antonio Nicolò, and Paolo Roberti. A Commitment Theory of Populism, 2021. URL `https://papers.ssrn.com/abstract=3982036`.

Brendan Nyhan and Rocío Titiunik. Public opinion alone won't save democracy. *Science*, 386(6719):268–270, October 2024. doi: 10.1126/science.ads7371. URL `https://www.science.org/doi/10.1126/science.ads7371`.

Inga A.-L. Saikkonen and Henrik Serup Christensen. Guardians of Democracy or Passive Bystanders? A Conjoint Experiment on Elite Transgressions of Democratic Norms. *Political Research Quarterly*, 76(1):127–142, March 2023. ISSN 1065-9129. doi: 10.1177/10659129211073592. URL `https://doi.org/10.1177/10659129211073592`.

Linley Sanders. Washington Examiner/YouGov Poll: By 47% to 34% voters oppose court packing, October 2020. URL `https://yougov.com/en-us/articles/32408-supreme-cort-packing-poll`.

Geoffrey Sheagley, Logan Dancey, and John Henderson. A Good Partisan? Ideology, Loyalty, and Public Evaluations of Members of Congress. *Legislative Studies Quarterly*, 48(3):657–679, 2023. ISSN 1939-9162. doi: 10.1111/lsq.12399.

Erin Walk, Derek E Holliday, Yphtach Lelkes, and Sean Westwood. Anti-Democratic Con-

demnation has Limited and Inconsistent Effects on Vote Choice and Democratic Attitudes, August 2024. URL https://osf.io/qacmr_v1/.

Appendix

A Proofs: Baseline Model **1**

B Proofs: Extensions **8**

C Exploratory Analyses **11**

D Study Details **13**

D.1 Study 1 Conjoint Design and Outcomes 13

D.2 Study 1 Descriptives 14

D.3 Study 2 Pre-Treatment Battery 15

E Full Model Results **18**

E.1 Study 1 18

F Pre-Analysis Plans **20**

F.1 Study 1 Pre-Registration (AsPredicted #185908) 20

A Proofs: Baseline Model

Throughout the proofs, let $X \equiv \widehat{x}(\theta_j)$ denote the policy outcome as a random variable. The tremble in the observation of a is used only to make all histories well-defined. All posterior statements below are evaluated in the limit as $\varepsilon \to 0$. Thus, whenever a history has positive probability in the limiting model, the posterior is the usual Bayesian posterior. When a history is reached only because of the observation tremble, the posterior is the limiting posterior induced by the tremble.

For compactness, write

$$\Delta(\rho) \equiv c(0;\rho) - c(1;\rho).$$

By assumption, $\Delta(\rho) > \gamma$ for both $\rho \in \{\rho^h, \rho^l\}$, and $\delta^h + \gamma > \max\{\Delta(\rho^h), \Delta(\rho^l)\}$. Also define

$$\overline{R}(\delta,\rho) = \begin{cases} \sqrt{\gamma + c(1;\rho)} & \delta = \delta^l, \\ \sqrt{c(0;\rho)} & \delta = \delta^h. \end{cases}$$

Candidate's policy choice. Fix a and ρ_j. If $y_j \le \overline{x}$, the candidate chooses $x = y_j$. If $y_j > \overline{x}$, choosing $x = y_j$ gives payoff $-c(a;\rho_j)$, while choosing $x = \overline{x}$ gives payoff $-(y_j - \overline{x})^2$. Hence the candidate chooses $x = \overline{x}$ iff

$$(y_j - \overline{x})^2 \le c(a;\rho_j).$$

Therefore

$$\widehat{x}(a;\theta_j) = y_j + (\overline{x} - y_j)\mathbf{1}\left\{ y_j \in \left[\overline{x}, \overline{x} + \sqrt{c(a;\rho_j)} \right] \right\}.$$

Notice that, under the assumptions, $\widehat{x}(a;\theta_j) = y_j$ for sufficiently large y_j.

Candidate's announcement choice. Let $V_j(a;\theta_j) \equiv u_j(\widehat{x}(a;\theta_j), a, \theta_j)$. For $y_j > \overline{x}$, the net value of choosing $a = 1$ rather than $a = 0$ is

$$V_j(1;\rho_j,\delta_j,y_j) - V_j(0;\rho_j,\delta_j,y_j) = -\gamma - \delta_j + \left[\min\{(y_j - \overline{x})^2, c(0;\rho_j)\} - \min\{(y_j - \overline{x})^2, c(1;\rho_j)\} \right].$$

The bracketed term is bounded above by $\Delta(\rho_j)$. Hence, by type dominance, high-attachment types never choose $a = 1$, since $\Delta(\rho_j) - (\gamma + \delta^h) < 0$.
Low-attachment types have $\delta_j = \delta^l = 0$. For them, $a = 1$ is strictly optimal iff

$$\min\{(y_j - \overline{x})^2, c(0;\rho_j)\} - \min\{(y_j - \overline{x})^2, c(1;\rho_j)\} > \gamma.$$

Since $\Delta(\rho_j) > \gamma$, this is equivalent to $(y_j - \overline{x})^2 > \gamma + c(1;\rho_j)$. By tie-breaking, equality leads to $a = 0$. Therefore

$$\widehat{a}(\theta_j) = \mathbf{1}\{\delta_j = \delta^l\}\mathbf{1}\{y_j > \overline{x} + \overline{R}(\delta_j,\rho_j)\}.$$

Proof of Proposition 1. The characterization of $\widehat{a}$ follows from the comparison above. It remains to characterize the induced policy.

If $\delta_j = \delta^h$, then $\widehat{a}(\theta_j) = 0$. Hence the effective restraint window is the $a = 0$ window:

$$\widehat{x}(\theta_j) = y_j + (\overline{x} - y_j)\mathbf{1}\left\{y_j \in \left[\overline{x}, \overline{x} + \sqrt{c(0;\rho_j)}\right]\right\}.$$

Using the definition of $\overline{R}$, this is

$$\widehat{x}(\theta_j) = y_j + (\overline{x} - y_j)\mathbf{1}\{y_j \in [\overline{x}, \overline{x} + \overline{R}(\delta^h, \rho_j)]\}.$$

If $\delta_j = \delta^l$, then the candidate chooses $a = 0$ exactly when

$$y_j \leq \overline{x} + \sqrt{\gamma + c(1;\rho_j)}.$$

Because $\gamma < c(0;\rho_j) - c(1;\rho_j)$, this threshold lies inside the $a = 0$ restraint window:

$$\gamma + c(1;\rho_j) < c(0;\rho_j).$$

Thus all low-attachment types who choose $a = 0$ implement $X = \overline{x}$. If instead

$$y_j > \overline{x} + \sqrt{\gamma + c(1;\rho_j)},$$

then the candidate chooses $a = 1$, and this same inequality implies

$$(y_j - \overline{x})^2 > c(1;\rho_j),$$

so she implements $X = y_j$. Hence

$$\widehat{x}(\theta_j) = y_j + (\overline{x} - y_j)\mathbf{1}\{y_j \in [\overline{x}, \overline{x} + \overline{R}(\delta^l, \rho_j)]\}.$$

Combining the high- and low-attachment cases gives

$$\widehat{x}(\theta_j) = y_j + (\overline{x} - y_j)\mathbf{1}\{y_j \in [\overline{x}, \overline{x} + \overline{R}(\delta_j, \rho_j)]\}.$$

□

Proof of Proposition 2. For any treatment T, we have

$$\mathcal{D}_i(a, \mathcal{I}_T) = y_i^2 - 2y_i\mathbb{E}[X \mid a, \mathcal{I}_T] + \mathbb{E}[X^2 \mid a, \mathcal{I}_T].$$

Therefore

$$Z_i(\mathcal{I}_T) = -\delta_i + z_0(\mathcal{I}_T) + z_1(\mathcal{I}_T)y_i,$$

where

$$z_1(\mathcal{I}_T) = 2\{\mathbb{E}[X \mid a = 1, \mathcal{I}_T] - \mathbb{E}[X \mid a = 0, \mathcal{I}_T]\}$$

and

$$z_0(\mathcal{I}_T) = \mathbb{E}[X^2 \mid a = 0, \mathcal{I}_T] - \mathbb{E}[X^2 \mid a = 1, \mathcal{I}_T].$$

Thus $Z_i(\mathcal{I}_T)$ is affine in y_i. Moreover, since $X \geq 0$, if $X \mid a = 1, \mathcal{I}_T$ first-order stochastically dominates $X \mid a = 0, \mathcal{I}_T$, then $z_1(\mathcal{I}_T) \geq 0$ and $z_0(\mathcal{I}_T) \leq 0$, with strict inequalities whenever the dominance is strict.

First consider $T = 0$. Observing $a = 1$ implies $\delta_j = \delta^l$ and

$$y_j > \overline{x} + \overline{R}(\delta^l, \rho_j),$$

so $X = y_j$. Hence, conditional on each ρ, $X \mid a = 1, \rho$ is drawn from the upper tail

$$(\overline{x} + \overline{R}(\delta^l, \rho), 2].$$

By contrast, without conditioning on a, implemented policy is weakly below the candidate's ideal point, and strictly below it on a positive-measure restraint window. Thus $X \mid a = 1, \rho$ first-order stochastically dominates $X \mid \rho$. Since this holds for each ρ, $X \mid a = 1$ first-order stochastically dominates X. Since X is a nontrivial mixture of $X \mid a = 1$ and $X \mid a = 0$, it follows that

$$X \mid a = 1 \succ_{\text{FOSD}} X \mid a = 0.$$

Therefore $z_1(\emptyset) > 0$ and $z_0(\emptyset) < 0$.

Now consider $T = \rho$. For each $\rho \in \{\rho^l, \rho^h\}$, observing $a = 1$ implies $\delta_j = \delta^l$,

$$y_j > \overline{x} + \overline{R}(\delta^l, \rho),$$

and $X = y_j$. The same first-order stochastic dominance argument gives

$$X \mid a = 1, \rho \succ_{\text{FOSD}} X \mid a = 0, \rho.$$

Therefore $z_1(\rho) > 0$ and $z_0(\rho) < 0$.

Now consider $T = y$. If

$$y \leq \overline{x} + \overline{R}(\delta^l, \rho^h),$$

then no type with ideology $y_j = y$ chooses true $a = 1$. The observation tremble may generate the history $a = 1$, but in the limit the voter knows that the true action was $a = 0$. Moreover, all types with ideology y implement the same policy. Hence $X \mid a = 1, y$ and $X \mid a = 0, y$ have the same limiting distribution, so $z_1(y) = z_0(y) = 0$ and $Z_i(y) = -\delta_i$.

If

$$y \geq \overline{x} + \overline{R}(\delta^h, \rho^l),$$

then all types with ideology $y_j = y$ implement $X = y$, regardless of ρ_j, δ_j, or a. Hence again $z_1(y) = z_0(y) = 0$ and $Z_i(y) = -\delta_i$.

Finally, suppose

$$y \in \left(\overline{x} + \overline{R}(\delta^l, \rho^h), \overline{x} + \overline{R}(\delta^h, \rho^l)\right).$$

Observing $a = 1$ implies $X = y$. Observing $a = 0$, the candidate either implements $X = y$

or is restrained and implements $X = \overline{x}$. Let

$$\pi(y) \equiv \Pr(X = \overline{x} \mid y, a = 0).$$

Then

$$Z_i(y) = -\delta_i + \pi(y)\left[2y_i(y - \overline{x}) + \overline{x}^2 - y^2\right].$$

Since $y > \overline{x}$ and $\pi(y) > 0$ in the intermediate region, this gives

$$z_1(y) = 2(y - \overline{x})\pi(y) > 0$$

and

$$z_0(y) = (\overline{x}^2 - y^2)\pi(y) < 0.$$

Therefore there is a unique cutoff

$$y^\dagger(y) = \frac{y + \overline{x}}{2} > 0$$

such that $Z_i(y) > -\delta_i$ iff $y_i > y^\dagger(y)$. More generally, whenever $z_1(\mathcal{I}_T) > 0$, the cutoff is

$$y^\dagger(\mathcal{I}_T) = -z_0(\mathcal{I}_T)/z_1(\mathcal{I}_T) > 0.$$

□

Proof of Proposition 3. Recall that

$$z_1(\rho) = 2\{\mathbb{E}[X \mid a = 1, \rho] - \mathbb{E}[X \mid a = 0, \rho]\}.$$

For a fixed ρ, write

$$R^l_\rho \equiv \overline{R}(\delta^l, \rho) = \sqrt{\gamma + c(1; \rho)} \qquad \text{and} \qquad R^h_\rho \equiv \overline{R}(\delta^h, \rho) = \sqrt{c(0; \rho)}.$$

Conditional on $a = 1$ and ρ, we have $\delta_j = \delta^l$, $y_j > \overline{x} + R^l_\rho$, and $X = y_j$. Therefore

$$\mathbb{E}[X \mid a = 1, \rho] = \mathbb{E}[y_j \mid y_j > \overline{x} + R^l_\rho] = \frac{\overline{x} + R^l_\rho + 2}{2}.$$

Now consider $\mathbb{E}[X \mid a = 0, \rho]$. Conditional on ρ, the event $a = 0$ includes all high-attachment types and the low-attachment types with $y_j \leq \overline{x} + R^l_\rho$. Thus the conditioning mass is

$$(1 - \mu^0_\delta)\frac{\overline{x} + R^l_\rho}{2} + \mu^0_\delta.$$

For low-attachment types with $a = 0$, implemented policy equals $X = y_j$ for $y_j < \overline{x}$ and $X = \overline{x}$ for $y_j \in [\overline{x}, \overline{x} + R^l_\rho]$. Hence their contribution to the numerator is

$$\int_0^{\overline{x}} y\,\frac{1}{2}\,dy + \int_{\overline{x}}^{\overline{x}+R^l_\rho} \overline{x}\,\frac{1}{2}\,dy = \frac{1}{2}\left(\frac{\overline{x}^2}{2} + \overline{x}R^l_\rho\right).$$

For high-attachment types, implemented policy equals $X = y_j$ for $y_j < \overline{x}$, equals $X = \overline{x}$ for $y_j \in [\overline{x}, \overline{x} + R_\rho^h]$, and equals $X = y_j$ again for $y_j > \overline{x} + R_\rho^h$. Hence their contribution to the numerator is

$$\int_0^{\overline{x}} y \, \frac{1}{2} \, dy + \int_{\overline{x}}^{\overline{x}+R_\rho^h} \overline{x} \, \frac{1}{2} \, dy + \int_{\overline{x}+R_\rho^h}^{2} y \, \frac{1}{2} \, dy = \frac{1}{2} \left(2 - \frac{(R_\rho^h)^2}{2} \right).$$

Multiplying numerator and denominator by 2, we obtain

$$\mathbb{E}[X \mid a = 0, \rho] = \frac{(1 - \mu_\delta^0) \left(\frac{\overline{x}^2}{2} + \overline{x} R_\rho^l \right) + \mu_\delta^0 \left(2 - \frac{(R_\rho^h)^2}{2} \right)}{(1 - \mu_\delta^0)(\overline{x} + R_\rho^l) + 2\mu_\delta^0}.$$

Therefore

$$z_1(\rho) = \overline{x} + R_\rho^l + 2 - \frac{(1 - \mu_\delta^0) \left(\overline{x}^2 + 2\overline{x} R_\rho^l \right) + \mu_\delta^0 \left(4 - (R_\rho^h)^2 \right)}{(1 - \mu_\delta^0)(\overline{x} + R_\rho^l) + 2\mu_\delta^0}.$$

The derivative of $z_1(\rho)$ with respect to R_ρ^h is

$$\frac{2\mu_\delta^0 R_\rho^h}{(1 - \mu_\delta^0)(\overline{x} + R_\rho^l) + 2\mu_\delta^0} \geq 0.$$

The derivative of $z_1(\rho)$ with respect to R_ρ^l is

$$\frac{(1 - \mu_\delta^0)^2 \left(2\overline{x} R_\rho^l + (R_\rho^l)^2 \right) + (1 - \mu_\delta^0)\mu_\delta^0 \left(4R_\rho^l + 4 - (R_\rho^h)^2 \right) + 4(\mu_\delta^0)^2}{\left[(1 - \mu_\delta^0)(\overline{x} + R_\rho^l) + 2\mu_\delta^0 \right]^2},$$

which is strictly positive because $R_\rho^l > 0$ and $(R_\rho^h)^2 = c(0; \rho) \leq c^{\max} < 1$. Thus $z_1(\rho)$ is strictly increasing in R_ρ^l and weakly increasing in R_ρ^h. By the maintained ordering of restraint windows,

$$R_{\rho^l}^l > R_{\rho^h}^l \qquad \text{and} \qquad R_{\rho^l}^h > R_{\rho^h}^h.$$

Therefore

$$z_1(\rho^l) > z_1(\rho^h).$$

□

Proof of Proposition 4. First suppose

$$y_j \notin \left[\overline{x} + \overline{R}(\delta^l, \rho^h), \overline{x} + \overline{R}(\delta^h, \rho^l) \right].$$

If

$$y_j < \overline{x} + \overline{R}(\delta^l, \rho^h),$$

then no type with ideology y_j chooses true $a = 1$, and all types with ideology y_j implement the same policy. The observation tremble may generate the history $a = 1$, but in the limit as $\varepsilon \to 0$, observing $a = 1$ only reveals that the observation of a was mistaken. It does not

change the voter's belief about the implemented policy. If instead

$$y_j > \overline{x} + \overline{R}(\delta^h, \rho^l),$$

then all types with ideology y_j implement $X = y_j$, regardless of a. In both cases,

$$\mathcal{D}_i(0, y_j) = \mathcal{D}_i(1, y_j),$$

so $Z_i(y_j) = -\delta_i$ and $z_1(y_j) = 0$.

Now suppose

$$y_j \in \left(\overline{x} + \overline{R}(\delta^l, \rho^h), \overline{x} + \overline{R}(\delta^h, \rho^l)\right).$$

Observing $a = 1$ implies $X = y_j$. Observing $a = 0$, the candidate either implements $X = y_j$ or is restrained and implements $X = \overline{x}$. Let

$$\pi(y_j) \equiv \Pr(X = \overline{x} \mid y_j, a = 0).$$

Then

$$Z_i(y_j) = -\delta_i + \pi(y_j)\left[2y_i(y_j - \overline{x}) + \overline{x}^2 - y_j^2\right],$$

and therefore

$$z_1(y_j) = 2(y_j - \overline{x})\pi(y_j).$$

It remains to compute $\pi(y_j)$. If

$$y_j \in \left(\overline{x} + \overline{R}(\delta^l, \rho^h), \overline{x} + \overline{R}(\delta^h, \rho^h)\right),$$

then all types consistent with $a = 0$ are restrained, so $\pi(y_j) = 1$. If

$$y_j \in \left(\overline{x} + \overline{R}(\delta^h, \rho^h), \overline{x} + \overline{R}(\delta^l, \rho^l)\right),$$

then the restrained $a = 0$ types are exactly the low-resolve types, while the unrestrained $a = 0$ types are high-resolve and high-attachment. Hence

$$\pi(y_j) = \frac{1 - \mu_\rho^0}{1 - \mu_\rho^0(1 - \mu_\delta^0)}.$$

If

$$y_j \in \left(\overline{x} + \overline{R}(\delta^l, \rho^l), \overline{x} + \overline{R}(\delta^h, \rho^l)\right),$$

then $a = 0$ implies $\delta_j = \delta^h$, and among high-attachment types only low-resolve candidates are restrained. Hence $\pi(y_j) = 1 - \mu_\rho^0$. Therefore

$$z_1(y_j) = \begin{cases} 2(y_j - \overline{x}) & y_j \in \left(\overline{x} + \overline{R}(\delta^l, \rho^h), \overline{x} + \overline{R}(\delta^h, \rho^h)\right), \\ 2(y_j - \overline{x})\dfrac{1 - \mu_\rho^0}{1 - \mu_\rho^0(1 - \mu_\delta^0)} & y_j \in \left(\overline{x} + \overline{R}(\delta^h, \rho^h), \overline{x} + \overline{R}(\delta^l, \rho^l)\right), \\ 2(y_j - \overline{x})(1 - \mu_\rho^0) & y_j \in \left(\overline{x} + \overline{R}(\delta^l, \rho^l), \overline{x} + \overline{R}(\delta^h, \rho^l)\right). \end{cases}$$

Thus $z_1(y_j) > 0$ throughout the intermediate region.

It remains to show that $z_1(y_j) < z_1(\emptyset)$ when $c^{\max}$ is sufficiently small. Because

$$c(1;\rho^h) = 0, \qquad c(0;\rho^l) = c^{\max},$$

and because sanctions decrease with resolve and with a, every sanction term is bounded above by $c^{\max}$. Moreover, type dominance implies

$$\gamma < \min\{c(0;\rho^h) - c(1;\rho^h), c(0;\rho^l) - c(1;\rho^l)\} \leq c^{\max}.$$

Therefore every restraint window is $O(\sqrt{c^{\max}})$:

$$\overline{R}(\delta,\rho) = O(\sqrt{c^{\max}}) \qquad \text{for all } (\delta,\rho).$$

Since the intermediate ideology region has width $O(\sqrt{c^{\max}})$, and since $\pi(y_j) \leq 1$, we have

$$\sup_{y_j} z_1(y_j) \leq 2\overline{R}(\delta^h,\rho^l) = 2\sqrt{c^{\max}} \to 0 \qquad \text{as } c^{\max} \to 0.$$

By contrast, the baseline slope is bounded away from zero. To see this, note that, as $c^{\max} \to 0$, all restraint windows converge to zero. Hence

$$2\mathbb{E}[X \mid a = 1] \to \overline{x} + 2.$$

Also, conditional on $a = 0$, low-attachment types who choose $a = 0$ have $y_j \leq \overline{x} + o(1)$, while all high-attachment types choose $a = 0$. Thus

$$2\mathbb{E}[X \mid a = 0] \to \frac{(1-\mu_\delta^0)\overline{x}^2 + 4\mu_\delta^0}{(1-\mu_\delta^0)\overline{x} + 2\mu_\delta^0}.$$

Therefore

$$\lim_{c^{\max}\to 0} z_1(\emptyset) = \overline{x} + 2 - \frac{(1-\mu_\delta^0)\overline{x}^2 + 4\mu_\delta^0}{(1-\mu_\delta^0)\overline{x} + 2\mu_\delta^0} = \frac{2\overline{x}}{(1-\mu_\delta^0)\overline{x} + 2\mu_\delta^0}.$$

Since $\overline{x} > 0$, this limit is strictly positive. Hence there exists $\bar{c} > 0$ such that, for all $c^{\max} < \bar{c}$,

$$z_1(\emptyset) > \sup_{y_j} z_1(y_j).$$

Therefore

$$0 < z_1(y_j) < z_1(\emptyset)$$

for every

$$y_j \in \left(\overline{x} + \overline{R}(\delta^l,\rho^h), \overline{x} + \overline{R}(\delta^h,\rho^l)\right),$$

as claimed. □

B Proofs: Extensions

Proof of Proposition 6. Recall that

$$D(y;\rho) = F(y \mid \delta^h, \rho) - F(y \mid \delta^l, \rho)$$

By the assumed FOSD ordering of ideology with respect to democratic attachment, $D(y;\rho) \geq 0$ for all y and ρ. Since the implemented policy is

$$\widehat{x}(y;\rho) = y + (\overline{x} - y)\mathbf{1}\left\{y \in [\overline{x}, \overline{x} + \sqrt{c(\rho)}]\right\},$$

we have

$$\frac{z_1(\rho)}{2} = \mathbb{E}[\widehat{x}(y;\rho) \mid \delta^l] - \mathbb{E}[\widehat{x}(y;\rho) \mid \delta^h].$$

Using integration by parts,

$$\frac{z_1(\rho)}{2} = \int_0^{\overline{x}} D(y;\rho)\, dy + \sqrt{c(\rho)} D\Big(\overline{x} + \sqrt{c(\rho)}; \rho\Big) + \int_{\overline{x}+\sqrt{c(\rho)}}^{2} D(y;\rho)\, dy.$$

Recall $\sqrt{c(\rho^h)} < \sqrt{c(\rho^l)}$. Under the assumptions,

$$D(y;\rho^h) \geq D(y;\rho^l) \quad \text{for all } y,$$

with strict inequality on a set of positive measure, and $D(y;\rho^l)$ is weakly decreasing on $\left[\overline{x} + \sqrt{c(\rho^h)}, \overline{x} + \sqrt{c(\rho^l)}\right]$. Then

$$\frac{z_1(\rho^h)}{2} \geq \int_0^{\overline{x}} D(y;\rho^l)\, dy + \sqrt{c(\rho^h)} D\Big(\overline{x} + \sqrt{c(\rho^h)}; \rho^l\Big) + \int_{\overline{x}+\sqrt{c(\rho^h)}}^{2} D(y;\rho^l)\, dy.$$

Subtracting $z_1(\rho^l)/2$ from the right-hand side gives

$$\sqrt{c(\rho^h)} D\Big(\overline{x} + \sqrt{c(\rho^h)}; \rho^l\Big) - \sqrt{c(\rho^l)} D\Big(\overline{x} + \sqrt{c(\rho^l)}; \rho^l\Big) + \int_{\overline{x}+\sqrt{c(\rho^h)}}^{\overline{x}+\sqrt{c(\rho^l)}} D(y;\rho^l)\, dy.$$

Since $D(y;\rho^l)$ is weakly decreasing on this interval,

$$\int_{\overline{x}+\sqrt{c(\rho^h)}}^{\overline{x}+\sqrt{c(\rho^l)}} D(y;\rho^l)\, dy \geq \left(\sqrt{c(\rho^l)} - \sqrt{c(\rho^h)}\right) D\Big(\overline{x} + \sqrt{c(\rho^l)}; \rho^l\Big).$$

Therefore the difference $\frac{z_1(\rho^h) - z_1(\rho^l)}{2}$ is at least

$$\sqrt{c(\rho^h)}\left[D\Big(\overline{x} + \sqrt{c(\rho^h)}; \rho^l\Big) - D\Big(\overline{x} + \sqrt{c(\rho^l)}; \rho^l\Big)\right],$$

which is weakly positive because $D(y;\rho^l)$ is weakly decreasing. Hence $z_1(\rho^h) \geq z_1(\rho^l)$. □

Non-linearities

Let the valence shock τ be drawn from a continuous distribution F_τ, with density f_τ, symmetric and single-peaked around zero. Candidate behavior is unchanged. The probability that voter i supports candidate j after observing $(a, \mathcal{I}_T)$ is

$$S_i(a, \mathcal{I}_T) = \Pr\left(\tau \geq \mathcal{D}_i(a, \mathcal{I}_T) + \delta_i a\right) = F_\tau\left(-\mathcal{D}_i(a, \mathcal{I}_T) - \delta_i a\right),$$

by symmetry of F_τ. Consider the baseline treatment. The change in support induced by observing $a = 1$ is

$$\Delta_i \equiv S_i(1, \emptyset) - S_i(0, \emptyset) = F_\tau\left(-\mathcal{D}_i(1, \emptyset) - \delta_i\right) - F_\tau\left(-\mathcal{D}_i(0, \emptyset)\right).$$

A Taylor expansion around $-\mathcal{D}_i(0, \emptyset)$ gives

$$\Delta_i \approx f_\tau\left(-\mathcal{D}_i(0, \emptyset)\right)\left[\mathcal{D}_i(0, \emptyset) - \mathcal{D}_i(1, \emptyset) - \delta_i\right] + \frac{1}{2} f'_\tau\left(-\mathcal{D}_i(0, \emptyset)\right)\left[\mathcal{D}_i(0, \emptyset) - \mathcal{D}_i(1, \emptyset) - \delta_i\right]^2$$

Equivalently, by the mean value theorem, there exists ξ_i between $-\mathcal{D}_i(0, \emptyset)$ and $-\mathcal{D}_i(1, \emptyset) - \delta_i$ such that $\Delta_i = f_\tau(\xi_i)[\mathcal{D}_i(0, \emptyset) - \mathcal{D}_i(1, \emptyset) - \delta_i]$. Therefore, if

$$\sup_{t \in [\min\{-\mathcal{D}_i(0,\emptyset), -\mathcal{D}_i(1,\emptyset) - \delta_i\}, \max\{-\mathcal{D}_i(0,\emptyset), -\mathcal{D}_i(1,\emptyset) - \delta_i\}]} f_\tau(t) \leq \eta,$$

then $|\Delta_i| \leq \eta |D_i(0, \emptyset) - \mathcal{D}_i(1, \emptyset) - \delta_i|$.

Thus, even when $D_i(0, \emptyset) - \mathcal{D}_i(1, \emptyset) - \delta_i \neq 0$, observed support can be arbitrarily insensitive to anti-democratic behavior if the density of the valence shock is small near the relevant support threshold. This provides a way to formalize the idea that many voters can be effectively rigid: information may change their expected utility, but not their observed support, because they are far from the margin of indifference.

Finally, notice that under these assumptions responsiveness is concentrated among voters for whom the interval

$$[-\mathcal{D}_i(0, \emptyset), -\mathcal{D}_i(1, \emptyset) - \delta_i]$$

passes through a region where f_τ is large. Equivalently, responsiveness is large when the voter is close to the support cutoff $\mathcal{D}_i(a, \emptyset) - \delta_i a$ of the valence shock τ under at least one of the two informational states and when

$$|\mathcal{D}_i(0, \emptyset) - \mathcal{D}_i(1, \emptyset) - \delta_i|$$

is large.

To see how this varies with ideology, observe that

$$\mathcal{D}_i(0, \emptyset) - \mathcal{D}_i(1, \emptyset) = \mathbb{E}[X^2 \mid a = 0, \emptyset] - \mathbb{E}[X^2 \mid a = 1, \emptyset] + 2y_i \left\{\mathbb{E}[X \mid a = 1, \emptyset] - \mathbb{E}[X \mid a = 0, \emptyset]\right\}.$$

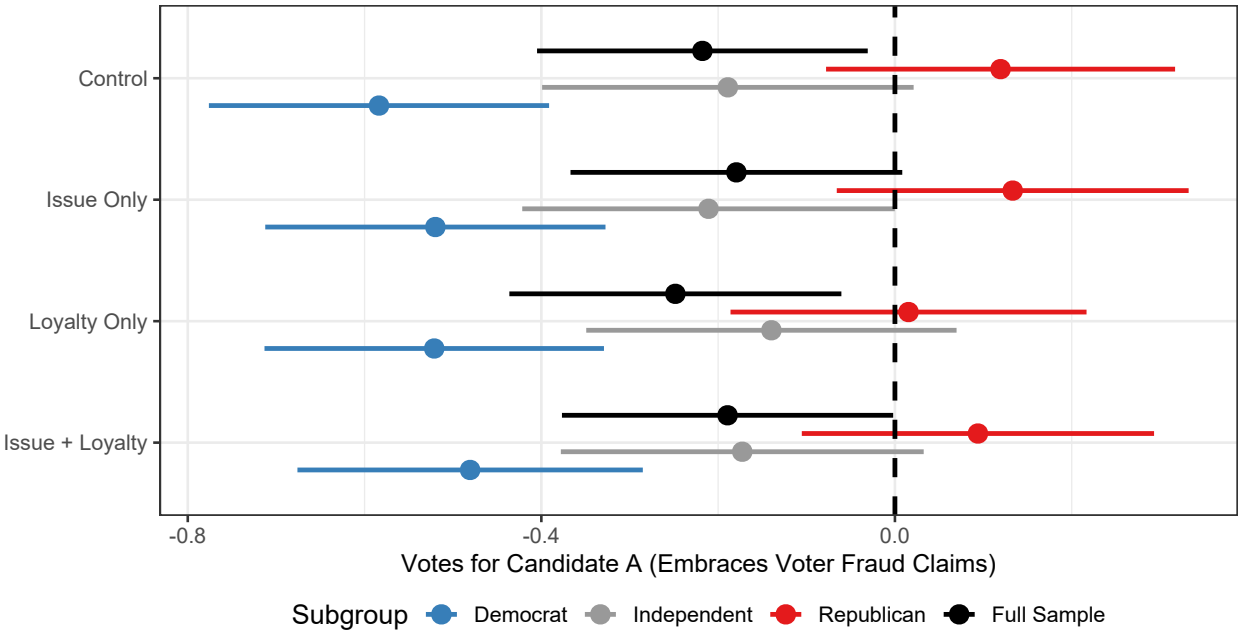
Since anti-democratic behavior is associated with a more extreme implemented policy,

$$\mathbb{E}[X \mid a = 1, \emptyset] > \mathbb{E}[X \mid a = 0, \emptyset],$$

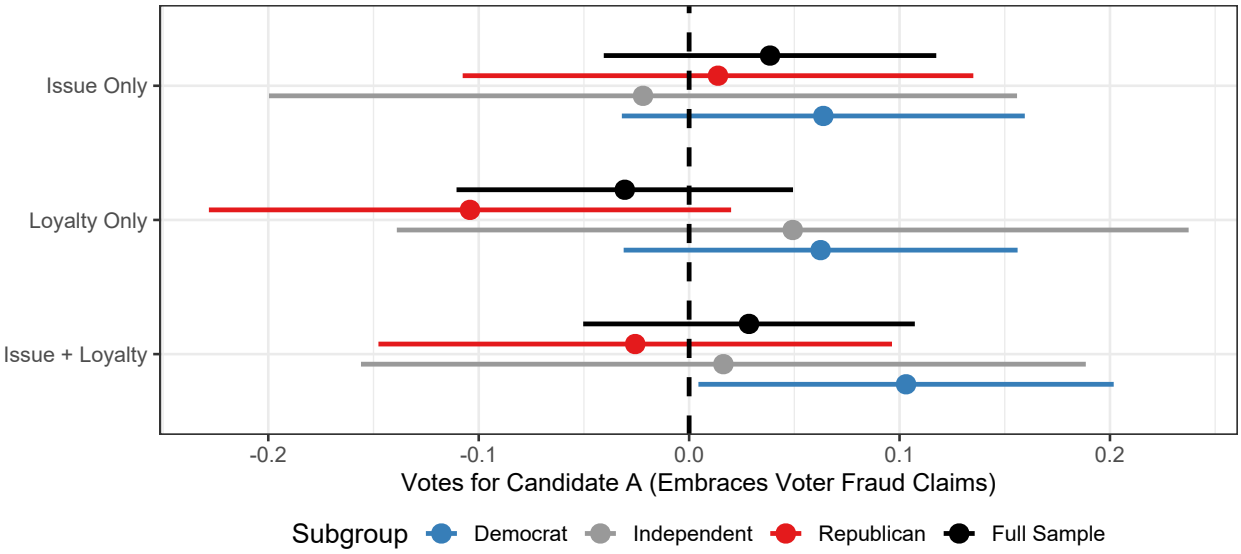
this difference is increasing in y_i. Hence, it is negative for sufficiently left-wing voters and positive for sufficiently right-wing voters. When $\delta_i > 0$ is large, the democratic cost reinforces the negative response among voters for whom the anti-democratic position worsens expected policy (those with y_i sufficiently low) and weakens the negative response among voters for whom the anti-democratic position improves expected policy (those with y_i sufficiently large). As a consequence, observed responsiveness will be concentrated among voters (i) on the opposite side of the ideological spectrum from the candidate and (ii) those sufficiently moderate on the candidate's side.

Thus, we should expect, on average, a stronger response to anti-democratic positions among voters on the opposite side of the ideological spectrum from the candidate.

C Exploratory Analyses



(a) Estimated Marginal Means



(b) Average Treatment Effects

Figure C1: Reanalysis of Figure 9 using party over ideology. Net votes for Candidate A, by information condition and party. Error bars indicate 95% confidence intervals.

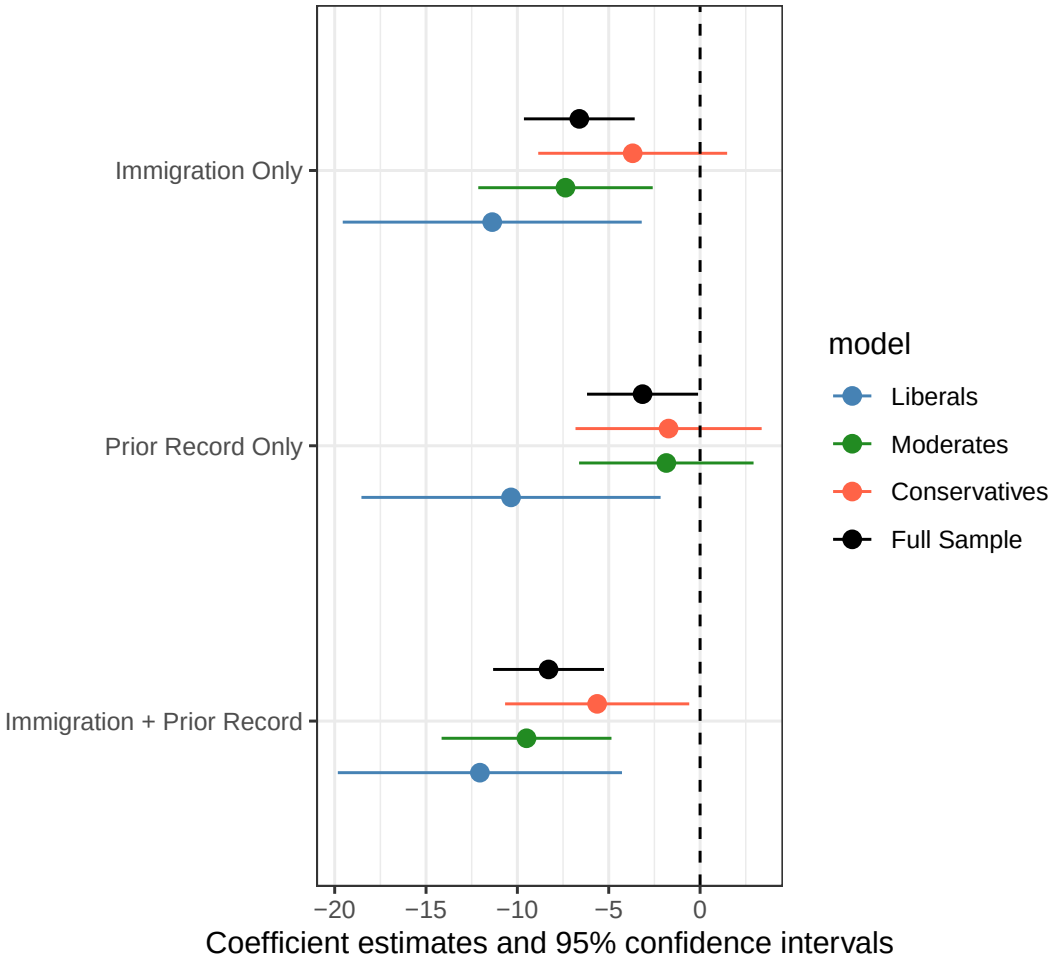


Figure C2: Learning about Abortion Policy, by Ideological Subgroup

D Study Details

D.1 Study 1 Conjoint Design and Outcomes

We want to know how voters feel about candidates running in Republican primaries.

Below is some information about the candidates in a Republican congressional primary election.

	Candidate A	**Candidate B**
Demographics	White Male	White Male
Views on the 2020 Election	Believes the 2020 election was stolen from Donald Trump and that Joe Biden was not legitimately elected president	Believes the 2020 election was fairly won by Joe Biden and that no convincing evidence suggests that the 2020 election was stolen.
Prior Experience	State Legislator	State Legislator
Political Record	Campaigned hard for other Republican Candidates	Campaigned hard for other Republican Candidates
Immigration Policy	Supports building a wall	Supports building a wall

Which candidate would you vote for? [*Candidate A*, *I would not vote*, *Candidate B*]

How likely are the following things to be true for **Candidate A**, should he win the election?

Please answer on the scale below where 0 = Not at all likely; 50 = Neither likely nor unlikely, 100 = Completely likely

- He will try to compromise with Democrats [*0–100 integer scale*]
- He will support a national abortion ban [*0–100 integer scale*]
- He will support raising tariffs [*0–100 integer scale*]

How likely are the following things to be true for **Candidate B**, should he win the election?

Please answer on the scale below where 0 = Not at all likely; 50 = Neither likely nor unlikely, 100 = Completely likely

- He will try to compromise with Democrats [*0–100 integer scale*]
- He will support a national abortion ban [*0–100 integer scale*]
- He will support raising tariffs [*0–100 integer scale*]

D.2 Study 1 Descriptives

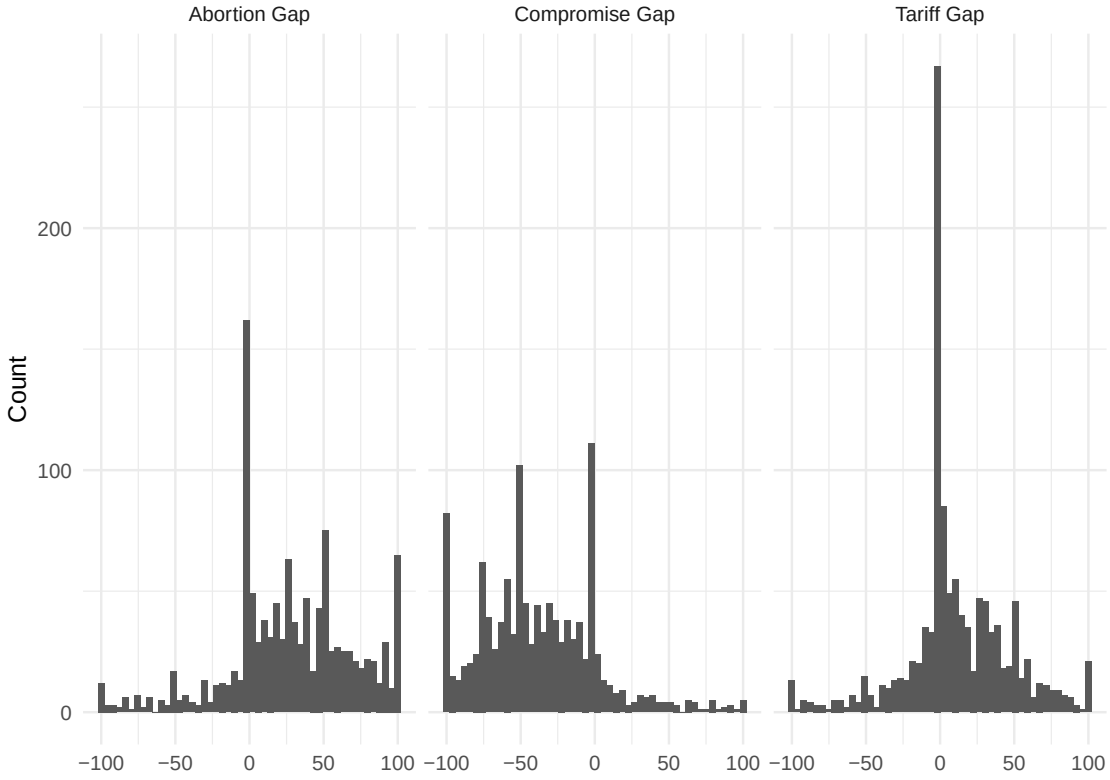


Figure D1: Perceived Gaps Between A and B in the Absence of Extraneous Information

	Control		Immigration Only		Loyalty Only		Both	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Rep	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
conservative	0.4	0.5	0.4	0.5	0.4	0.5	0.4	0.5
born	1983.0	13.1	1982.8	13.2	1982.4	13.6	1982.9	13.2
female	0.5	0.5	0.5	0.5	0.6	0.5	0.5	0.5

Table D1: Study 1 Balance Table

D.3 Study 2 Pre-Treatment Battery

Each respondent completes a set of pre-treatment policy items measuring their positions across four domains: taxes, abortion, immigration enforcement, and the environment. Following Broockman (2016), each item offers seven ordered response options spanning a single left-to-right dimension. Option 3 corresponds to the mainstream position of the political left, option 5 to the mainstream position of the political right, and option 4 to the status quo; options 1 and 2 describe positions successively further to the left, and options 6 and 7 positions successively further to the right. For each item we report the question stem, the response options, and sources for these proposals.

Taxes Which statement comes closest to describing your views on **taxes**?

1. Raise the top federal income tax rate to 70% on income over $10 million to fund major new social programs, and sharply increase taxes on wealth.
2. Raise the top federal income tax rate to roughly 45–50% on high earners, and add a new annual tax on the wealth of the richest households.
3. Raise the top federal income tax rate from 37% to 39.6% on income over $400,000 ($450,000 for married couples), while not raising taxes on anyone earning less.
4. Keep federal income tax rates where they are now, including the current top rate of 37%, and maintain current levels of federal spending.
5. Decrease all individuals' income tax rates, especially high earners who pay the most in taxes now, accomplished by decreasing government services.
6. Move to a completely flat income tax system where all individuals pay the same percentage of their income in taxes, accomplished by decreasing government services.
7. Move to a flat consumption tax where all individuals pay the same percentage of their purchases in taxes, banning the income tax, even if this means the poor pay more in taxes than the rich. Significantly decrease government services in the process.

Sources.

– **1–2.** Proposal by Rep. Ocasio-Cortez. Tax Policy Center: `https://taxpolicycenter.org/taxvox/about-rep-ocasio-cortezs-70-percent-tax-rates`

– **3.** 2024 Democratic Party Platform: `https://www.presidency.ucsb.edu/documents/2024-democratic-party-platform`; CNBC: `https://www.cnbc.com/2023/03/09/president-bidens-proposed-2024-budget-calls-for-top-39point6percent-tax-rate.html`

– **4.** Current rates under the 2017 law and extended via the 2025 One Big Beautiful Bill Act (OBBBA). Tax Foundation (TCJA rate schedule): `https://taxfoundation.org/research/all/federal/big-beautiful-bill-house-gop-tax-plan/`

– **5.** 2024 Republican Party Platform (sought to make the 2017 tax cuts permanent, accomplished in 2025 via the OBBBA): `https://www.presidency.ucsb.edu/documents/2024-republican-party-platform`; CBPP: `https://www.cbpp.org/research/federal-tax/ten-questions-on-house-republicans-upcoming-tax-bill`

– **6–7.** Proposal to abolish the income tax and replace with a national sales tax (FairTax Act, H.R. 25). Congress.gov: `https://www.congress.gov/bill/119th-congress/house-bill/25`

Abortion Which statement comes closest to describing your views on **abortion**?

1. Abortions should always be legal, and the government should pay for all abortions.
2. Abortions should always be legal, and the government should help women pay for abortions when they cannot afford them.
3. Abortions should be legal in the first two trimesters at least, and the government should require private insurers to cover abortions.
4. Abortions should be legal in the first two trimesters, though the government should not play any role in financially supporting abortions.
5. Abortion should only be legal if the life of the mother is in danger or in cases of rape and incest.
6. Abortion should only be legal if the life of the mother is in danger.
7. Abortion should always be illegal.

Sources.

– TK

Immigration Enforcement Which statement comes closest to describing your views on ICE (Immigration and Customs Enforcement) and **CBP** (Customs and Border Protection)?

1. Abolish ICE and sharply limit Border Patrol to rescue and safety duties. Deportations should be extremely rare.
2. Cut ICE and CBP substantially. End workplace and community raids. Use detention only in exceptional cases. Emphasize legal status for most undocumented immigrants.
3. Reduce aggressive enforcement and focus removals on people convicted of serious crimes and recent unlawful entrants, while expanding lawful pathways and alternatives to detention.
4. Maintain current ICE and CBP levels. Enforce immigration laws but prioritize people with serious criminal records and recent unauthorized entrants.

5. Expand ICE and CBP staffing and resources significantly. Resume construction of the border wall and carry out large-scale deportation of those who entered illegally.

6. Implement a mass deportation program using military assets to assist ICE and CBP, conduct frequent workplace raids, penalize sanctuary cities, and invoke emergency powers to restrict asylum.

7. Indefinitely suspend all immigration processing and seal the border completely. Authorize the use of maximum force to stop crossings, and detain and deport all undocumented residents immediately without judicial review.

Sources.

– TK

Pollution and the Environment Which statement comes closest to describing your views on pollution and the environment?

1. The government should institute a carbon tax or cap-and-trade system that would significantly decrease carbon emissions over the next several decades.

2. The government should institute a carbon tax or cap-and-trade system that would keep carbon emissions at or just below their current levels.

3. The government should subsidize solar, wind, battery, and other clean energy through tax incentives and set stricter efficiency standards, while discouraging the highest-emitting energy sources.

4. The government should enact regulations encouraging energy efficiency and subsidize the use and development of solar, wind, and nuclear energy.

5. The government should lift restrictions on oil, natural gas, and coal production to maximize energy output, and repeal clean-energy mandates and subsidies.

6. The government should allow further oil drilling offshore and on federal lands, prioritizing energy independence and low prices over environmental concerns.

7. The government should both allow and subsidize increased domestic production of fossil fuels (coal, oil, and gas).

Sources.

– TK

E Full Model Results

E.1 Study 1

Table E1: Estimated Marginal Means of Perception Gap by Treatment Condition and Issue

Model	Treatment Condition	Expected Mean	Std. Error	95% CI
Tariffs	Control	9.676	1.605	(6.531, 12.822)
	Loyalty Only	9.452	1.431	(6.646, 12.258)
	Issue Only	8.603	1.524	(5.615, 11.591)
	Issue + Loyalty	8.147	1.459	(5.288, 11.007)
Abortion	Control	21.372	2.918	(15.650, 27.093)
	Loyalty Only	18.303	2.800	(12.814, 23.792)
	Issue Only	14.590	2.826	(9.049, 20.131)
	Issue + Loyalty	13.080	2.750	(7.688, 18.472)
Compromise	Control	33.058	1.796	(29.537, 36.579)
	Loyalty Only	29.446	1.604	(26.300, 32.591)
	Issue Only	28.784	1.712	(25.428, 32.140)
	Issue + Loyalty	26.536	1.623	(23.353, 29.719)

Table E2: Effects of information provision on learning outcomes

	Tariffs	Abortion	Compromise
Intercept	−0.173	−10.359	11.504***
	(4.128)	(16.172)	(3.324)
Prior Record	−0.224	−3.069+	−3.613*
	(1.357)	(1.614)	(1.557)
Immigration	−1.073	−6.782***	−4.274**
	(1.376)	(1.632)	(1.587)
Immigration x Prior Record	−0.232	1.559	1.365
	(1.854)	(2.162)	(2.156)
Conservative	−3.995***	−7.868***	−2.805*
	(1.056)	(1.220)	(1.196)
Liberal	3.641**	6.902***	1.369
	(1.306)	(1.578)	(1.673)
Other/Missing Ideology	−0.587	−5.057	−4.214
	(3.039)	(3.446)	(3.270)
Age	0.176***	−0.096*	−0.087*
	(0.037)	(0.044)	(0.042)
Female	−7.210***	−0.087	2.472*
	(0.930)	(1.081)	(1.083)
White	−0.941	−0.583	−2.060
	(1.161)	(1.350)	(1.365)
Education: K-12	4.302	44.000*	29.177***
	(6.402)	(17.367)	(7.319)
Education: High school, GED	7.666+	42.240**	29.788***
	(3.964)	(16.164)	(3.045)
Education: Some college	8.103*	40.625*	31.812***
	(3.844)	(16.124)	(2.831)
Education: Associate's	8.852*	45.252**	28.125***
	(3.950)	(16.158)	(3.013)
Education: Bachelor's	9.571*	44.894**	32.431***
	(3.781)	(16.099)	(2.707)
Education: Master's or higher	9.986*	45.710**	33.209***
	(3.881)	(16.141)	(2.925)
Num.Obs.	4652	4652	4652
R2	0.024	0.033	0.012

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

F Pre-Analysis Plans

F.1 Study 1 Pre-Registration (AsPredicted #185908)

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

No, no data have been collected for this study yet.

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

In a candidate choice experiment, does the penalty borne by hypothetical Republican elites for taking positions against voter fraud claims decrease in magnitude when information that they are a strong conservative on immigration is provided? Does this effect occur when information that the elite supports playing hardball with Democrats is provided? When both of these pieces of information are revealed, does the initial penalty decrease to zero?

We offer two specific hypotheses:

Hypothesis 1 (weak): The effect of voter fraud on vote choice and beliefs about the candidate will decrease with each additional piece of information.

Hypothesis 2 (strong): The effect of voter fraud on vote choice and beliefs about the candidate will decrease to zero when two additional pieces of information are provided.

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

The key dependent variables are (i) vote choice between the two candidates or abstaining, and (ii) the differences between ratings given to two hypothetical candidates, one of whom opposes voter fraud claims and the other of whom embraces voter fraud claims (voter fraud skeptic rating - voter fraud supporter rating). Respondents will rate both candidates from 0 to 100 in terms of their likelihood of engaging in three actions: cooperating with Democrats on policy, supporting a national abortion ban, or raising tariffs to protect American industry.

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

Respondents will be assigned to one of four conditions at random. We will indicate whether the respondent is provided information about the elite's view on immigration using the binary variable "issue" and whether they see information about the elite's view on congressional compromise using the binary variable "party." In the first condition, respondents will see descriptions of two hypothetical Republican candidates who differ only in terms of views on 2020 voter fraud but not any other demographic information (issue = 0; party = 0). Respondents will rate the candidates in a few dimensions. In the second condition, respondents will see the same information but also be provided information about the two candidates' issue position on immigration, which will be the same (issue = 1; party = 0). In the third condition, respondents will see information about the candidates' party loyalty positions (same across candidates) but not about immigration (issue = 0; party = 1). In the fourth condition, respondents will see both pieces of information (issue = 1; party = 1).

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

Main analysis: We plan to regress all outcomes on "issue" and "party" and on pre-treatment measures of ideology, education, age, gender and race (white=1) with HC2 robust standard errors (using OLS regression). We hypothesize that the difference in ratings will be smaller for respondents in conditions 2, 3, and 4 than for respondents in condition 1 and that respondents in conditions 2, 3, and 4 will be more likely to support the candidate who rejects voter fraud claims on average than those in condition 1. We hypothesize further that differences in vote choice and ratings between the two candidates will be indistinguishable from 0 in condition 4.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

We plan to analyze only the responses of those who pass an attention check.

7) How many observations will be collected or what will determine sample size?

No need to justify decision, but be precise about exactly how the number will be determined.

Exact sample size is unknown because this in embedded in a larger survey but at least 4000 and up to 8000 respondents are expected.