Democracy in America?
Partisanship, Polarization, and the Robustness of Support for Democracy in the United States

Matthew Graham and Milan W. Svolik*

First draft: March 2018, Current version: October 2018

Abstract

Is support for democracy in the United States robust enough to deter undemocratic behavior by elected politicians? In order to answer this question, we develop a model of the public as a democratic check and evaluate it using experimental data as well as a natural experiment. We conducted a series of original, nationally representative candidate-choice experiments in which some politicians adopt positions that violate key democratic principles. Respondents’ choices allow us to infer their willingness to trade-off democratic principles for other valid but potentially conflicting considerations such as political ideology, partisan loyalty, and policy preferences. We find that the viability of the U.S. public as a democratic check is strikingly limited: only a small fraction of Americans are willing to prioritize democratic principles in their electoral choices and their tendency to do so is decreasing in the strength of their partisanship, policy extremism, and in candidate platform polarization. Our findings echo classic arguments about the importance of political moderation and cross-cutting cleavages for democratic stability and highlight the dangers that political polarization represents for democracy.

*Department of Political Science, Yale University. Corresponding email: milan.svolik@yale.edu.
“It is the function of public opinion to check the use of force in a crisis, so that men, driven to make terms, may live and let live.”

Walter Lippmann, *The Phantom Public* (1925, 64)

“I could stand in the middle of 5th Avenue and shoot somebody and I wouldn’t lose voters.”

Donald Trump at a presidential campaign rally in Iowa (January 23, 2016)

# 1 Introduction

“It is nearly impossible to find an American who says that he is opposed to democracy or favors some alternative… On the contrary, nearly everyone professes to believe that democracy is the best form of government.” This is how Robert A. Dahl, writing in 1966, summarized contemporary evidence for the support for democracy in the United States (Dahl, 1966, 40). It remains conventional wisdom to this day. Research that traces its intellectual origins to Tocqueville’s *Democracy in America* finds that the United States consistently exhibits some of the highest levels of support for democracy in the world (Almond and Verba, 1963; Inglehart and Welzel, 2010; Norris, 2011).

In this paper, we show that this conventional wisdom rests on fragile foundations. We adopt an approach that, instead of asking about support for democracy directly, infers respondents’ commitment to democratic principles from their choices of candidates in hypothetical election scenarios. Each candidate is experimentally assigned attributes and platforms that approximate real-world elections and, crucially, may include positions that violate key democratic principles. In this framework, voters are said to “support democracy” when their choices reveal a preference for democratic principles over other
valid but potentially conflicting considerations such as political ideology, partisan loyalty, or favorite policies.

This research design builds on the observation that elections represent a fundamental instrument of democratic self-defense: Especially in advanced democracies, voters have the opportunity to stop politicians who violate democratic norms by defeating them at the polls. According to this line of reasoning, democracy is self-enforcing when politicians anticipate that, were they to behave undemocratically, their own supporters would punish them by voting for a competitor.

We argue that a key obstacle to the viability of such a democratic check is partisan, ideological, and policy-based polarization. In polarized societies, electoral competition often confronts voters with a choice between two valid but potentially conflicting considerations: democratic values and partisan interests. When faced with the choice between a candidate whose positions violate democratic principles but whose policies they find appealing and one who complies with democratic principles but is otherwise unappealing, a significant fraction of voters will be willing to sacrifice democratic principles in favor of electing candidate who champions their interests. In a sharply polarized electorate, even pro-democratically minded voters may act as partisans first and democrats only second.

In section 3, we formalize these intuitions and develop a model of the public as a democratic check. We extend the classic, spatial framework for voter preferences to account for undemocratic candidate positions. The latter are conceptualized as negative valence attributes: while voters may differ over policy, ideology, or partisanship, they agree that electoral competition should be democratic and prefer candidates that compete fairly. This framework implies that i) voters will be willing to trade-off democratic principles for other, potentially competing political goals, ii) centrists provide the most viable check on candidates that undermine democratic principles, and iii) elections are most likely to fail as
a democratic check in polarized societies. We further show that both voter and candidate polarization are independently detrimental to democracy. We employ this framework to guide the design of our experiment and the analysis of our data.

Our empirical results suggest that the viability of the U.S. public as a pro-democratic electoral check is surprisingly limited. We consistently find that only a small fraction of Americans are willing to prioritize democratic principles in their electoral choices when doing so goes against their partisan identification, political ideology, or favorite policies. Specifically, when confronted with a choice between a candidate who is appealing on partisan grounds but at the same time endorses an undemocratic action and one who is democratic but less appealing, our respondents overwhelmingly vote for the former. Most Americans are partisans first and democrats only second.

The following is a preview of our experimental findings, which we present in sections 4 and 5:

- **Americans value democracy, but not much**: A candidate who considers adopting an undemocratic position can expect to be punished by losing only about 11% of his overall vote share.

- **Support for democracy is highly elastic**: When the price of voting for a more democratic candidate is that candidate’s greater distance from the voter in terms of policies or partisanship, even the most centrist voters are only willing to tolerate at most a 10% increase in such a distance.

- **Centrists are a pro-democratic force**: Voters who are moderate in their policy preferences, ideology, or partisanship punish undemocratic candidates at higher rates than voter with more extreme preferences.

- **Sensitivity to the menu of manipulation varies**: Voters are most sensitive to
undemocratic positions that undermine the free press, checks and balances, and those that aim to disenfranchise opposition supporters.

- **Voters employ a partisan “double standard”:** Conservative respondents are more willing to punish undemocratic behavior by Democratic Party than Republican Party candidates and vice-versa. In most of our specifications, these effects hold equally for both Democrats and Republicans. Most Americans are “instrumental” rather than “principled” democrats.

- **Platform polarization is bad for democracy:** The larger the difference between the candidates’ policy platforms, the weaker the punishment for undemocratic behavior.

- **Primaries rather than general elections may be the most viable check on undemocratic behavior:** Voters punish undemocratic platforms more severely in contests with candidates from the same party than they do in contests with candidates from different parties.

- **Strong partisans punish undemocratic behavior by abstaining:** The stronger a respondent’s preference for a candidate, the more likely she is to abstain rather than defect when that candidate adopts an undemocratic position.

- **Americans have a solid understanding of what democracy is and what it is not:** A majority of our respondents correctly distinguish real-world undemocratic practices from those that are consistent with democratic principles.

- **Most voters are partisans first and democrats only second:** Only about 40% of our respondents are willing to support the more democratic candidate when doing so goes against their partisanship. Only independents and partisan “leaners” support
the more democratic candidate enough to defeat the undemocratic candidate regardless of his partisan affiliation.

These results obtain in spite of the fact that – just as in earlier studies – our respondents profess a strong commitment to democracy when measured by conventional, direct-questioning techniques. For instance, when asked to rate on a scale from 1 to 10 “how important [it is for them] to live in a country that is governed democratically,” more than 80% of our respondents given an answer of 8 to 10, with a median answer of 9. This is statistically indistinguishable from answers to the same question in major surveys like the Americas Barometer and the World Values Survey. Put differently, conventionally measured support for democracy appears to be mostly cheap talk. In turn, our existing knowledge about the support for democracy in the United States is of alarmingly limited use when it comes to answering a key question: When can we realistically expect the American public to serve as a check on the authoritarian temptations of elected politicians?

We move from analyzing hypothetical election scenarios to a real-world election in section 6, where we examine a natural experiment that occurred during the 2017 special election for Montana’s only seat in the U.S. House of Representatives. On the day before the election, one of the two major candidates physically attacked a journalist, which we interpret as a negative public signal about his respect for a free press. Crucially, only in-person voters saw this signal before they could cast a ballot; absentee voters, who in Montana make up about half of registered voters, had already cast their ballots. This allows us to adopt a difference-in-differences empirical strategy that compares precinct-level vote shifts between absentee and in-person voters to infer their willingness to punish the candidate for the attack on the journalist. Our findings are consistent with both our theoretical expectations and experimental results: Montanans value a free press, but not enough for most partisans to abandon their favored candidate.
2 Related Research


- Polarization and American Politics: Abramowitz and Saunders (2008), Hetherington and Weiler (2009), Levendusky (2009), McCarty, Poole, and Rosenthal (2008).

3 A Model of the Public as a Democratic Check

Consider a model according to which two sets of candidate attributes determine voters' electoral choices: i) voters' preferences over positional issues, which may include policy positions, ideology, and partisan affiliation, and ii) voters' support for democracy. We think of voters' support for democracy as a valence issue. In other words, while voters may differ over policy, ideology, or partisanship, they all agree that electoral competition should be democratic and prefer candidates that compete fairly.
Formally, voter $i$’s payoff from candidate $j$ is

$$u_i(X_j, M_j) = -\sum_K \alpha_k (x_{ik} - x_{jk})^2 - \delta M_j,$$

where $x_{ik}$ is voter $i$’s favorite policy on issue $k$, $x_{jk}$ is candidate $j$’s platform on issue $k$, and $\alpha_k$ is the weight that $i$ attaches to that issue. Meanwhile, $M_j$ is candidate $j$’s democracy platform where $M$ is increasing in how undemocratic $j$’s platform is (i.e. $M$ stands for “manipulation.”) Thus we may think of all voters’ ideal democracy platforms as being zero, resulting in the quadratic formulation above. The term $\delta$ is the weight that $i$ attaches to fair democratic competition – in effect, the intensity of $i$’s support for democracy.

Normalizing both policy and democracy weights so that they sum to one yields

$$\delta = 1 - \sum_K \alpha_k.$$ 

A key implication of this model is that voters who hold intense policy preferences (a large $\alpha_k$) or extreme policy preferences ($x_{ik}$ to the left or right of both candidates’ platforms) are willing to tolerate undemocratic behavior by their favored candidate.

Formally, assuming only a single policy issue, $i$ votes for candidate 1 as long as $u_i(X_1, M_1) \geq u_i(X_2, M_2)$, or equivalently as long as

$$x_{ik} \leq \frac{x_{1k} + x_{2k}}{2} - \frac{\delta(M_1 - M_2)}{2\alpha_k(x_{2k} - x_{1k})},$$

where we are assuming that candidate 1’s policy platform is to the left of candidate 2’s platform.

Call the voter whose ideal policy point $x_{ik}$ barely satisfies the above inequality the *swing voter*. Note that the first term on the right-hand side of this inequality is the midpoint
between the two candidates’ policy platforms, i.e. it separates the electorate into those who are policy-wise closer to candidate 1 and those who are closer to candidate 2. The swing voter is in turn located either to the left or to the right of this midpoint, depending on whether it is candidate 1 or candidate 2 who adopts an undemocratic platform.

When candidate 1 adopts an undemocratic platform, the swing voter is located to the left of the midpoint \( \frac{x_{1k} + x_{2k}}{2} \). Voters to the left of the midpoint between the two candidate’s platforms but to the right of the swing voter favor candidate 1 based on their policy preferences, yet are sufficiently put off by his undemocratic platform to vote for candidate 2 instead. By contrast, voters whose ideal policies are to the left of the swing voter are willing to tolerate candidate 1’s undemocratic platform as their support for democracy is outweighed by their proximity to his policies. The converse holds when candidate 2 adopts an undemocratic platform.

The distance between the midpoint \( \frac{x_{1k} + x_{2k}}{2} \) and the swing voter corresponds to the segment of the electorate that candidate 1 loses due to his undemocratic platform. It is increasing in the size of the support for democracy parameter \( \delta \) and decreasing in the weight \( \alpha_k \) that voters attach to the single policy issue and the polarization of the candidates’ platforms \( (x_{2k} - x_{1k}) \). That is, both intense policy preferences and the polarization of platforms make voter more willing to sacrifice democratic principles.

Denote the expected “unfair” vote gain from an undemocratic platform for a candidate by \( \mu_j \). We may say that the public successfully checks undemocratic behavior, if \( \mu_j \) is smaller than the fraction of voters that either candidate loses due to his undemocratic platform. The latter are the voters located on the interval

\[
\left[ \frac{x_{1k} + x_{2k}}{2}, \frac{x_{1k} + x_{2k}}{2} + \frac{\delta(M_1^2 - M_2^2)}{2\alpha_k(x_{2k} - x_{1k})} \right].
\] (3)
Democracy first voters are between the two swing voters:

\[ x_1^s, x_2^s = \frac{x_{1k} + x_{2k}}{2} \pm \frac{\delta}{2\alpha_k(x_{2k} - x_{1k})}. \] (4)

Because this interval is located in the middle of the policy space under most circumstances (e.g. in the case when candidate platforms straddle the electorate’s median on issue \( k \)), it is less electorally consequential when the distribution of voters is “polarized,” i.e. with a weak center and heavy extremes. That is, elections are more likely to fail as a democratic check in polarized societies.

## 4 Candidate-Choice Experiment

In order to evaluate our theoretical framework, we adopt two empirical strategies: an original, nationally-representative candidate-choice survey experiment and a natural experiment. The candidate-choice experiment was designed to examine a key mechanism in our framework: even voters who value democratic principles may trade off those principles for partisan ends when confronted with a choice between the two. The experiment allows us to examine this mechanism at the level at which it is hypothesized to operate – that of the individual voter. By modelling one of the most essential actions that voters perform – the choice between two candidates in an election – the conjoint-based design that we introduce below allows us probe our respondents’ willingness to trade off democratic principles for partisan interests without alerting them to it.\(^1\)

In the candidate-choice experiment, each respondent made a series of choices between two candidates for a state legislature. Candidates were described by experimentally

\(^1\)Our candidate-choice experiment belongs to a broader category of survey-experimental techniques known as conjoint experiments (Hensher, Rose, and Greene, 2015; Hainmueller, Hopkins, and Yamamoto, 2015).
manipulated attributes typically seen in real-world elections: age, gender, race, profession, years of experience, partisan affiliation, two policy platforms, and a “democracy” position. This last attribute is the focus of our analysis; we therefore describe its design and assignment below. We introduce most of the remaining attributes throughout the paper. In the appendix, we outline the design and assignment of all attributes and present an example a candidate-choice scenario as seen by our respondents.

Each candidate was assigned a democracy position that was either “undemocratic” – an action or statement by the candidate that violates a key democratic principle – or a democratically neutral, “generic” position. The undemocratic positions were:

1. “Supported a redistricting plan that gives [own party]s [2 or 10] extra seats despite a decline in the polls.”
2. “Supported a proposal to reduce the number of polling stations in areas that support [opposite party]s.”
3. “Said the [own party] governor should rule by executive order if [opposite party] legislators don’t cooperate.”
4. “Said the [own party] governor should ignore unfavorable court rulings by [opposite party]-appointed judges.”
5. “Said the [own party] governor should prosecute journalists who accuse him of misconduct without revealing sources.”
6. “Said the [own party] governor should ban far-[left or right] group rallies in the state capital.”

Above, [party] refers to a candidate’s randomly assigned political party (Democrat or Republican); [opposite party] denotes the complement.

---

2 Whether 2 or 10 extra seats were listed was randomized and equally likely.
3 Candidates who were Democrats were advocating to ban far-right group rallies and vice versa.
In designing these undemocratic positions, we employed the following criteria:

**Conceptual validity:** The undemocratic positions capture violations of key democratic principles. Following classic scholarship on democratization (Dahl, 1971), this includes measures that undermine electoral fairness (items 1 and 2 above), checks and balances (items 3 and 4), and civil liberties (items 5 and 6).\(^4\)

**Contextual realism and partisan balance:** The undemocratic positions or their milder variants approximate practices that have been used by politicians to subvert the democratic process in the United States and can be plausibly adopted by both major parties. This is why the above undemocratic positions are situated at the state level— the level at which most attempts to subvert the democratic process for partisan gain occur in the United States and have historically been attempted by both major parties. In the appendix, we provide real-world examples of each undemocratic position.

A key feature of such attempts in the United States and around the world is the use of ostensibly legal, incremental, and complementary measures. This has several consequences. First, to be implemented, most measures must be conducted by or in conjunction with the executive. This is why most of our undemocratic positions refer to actions that the candidate suggests a co-partisan governor take. Second, any single measure may allow for a partisan interpretation according to which it is consistent with some—often more majoritarian—conception of democracy or corrects for an existing deficiency in the democratic process. For instance, proponents of stricter voter ID laws respond to accusations of voter suppression by claiming such measures are needed to prevent voter fraud, and proponents of gerrymandering may claim they are correcting an already existing and unfair status quo. Jointly and in their political context, however,

\(^4\)For recent perspectives on how to conceptualize and measure democracy, see Boix, Miller, and Rosato (2013), Cheibub, Gandhi, and Vreeland (2010), and Coppedge et al. (2011).
such measures result in an uneven playing field that favors their proponents (Levitsky and Way, 2010b; Schedler, 2002; Waldner and Lust, 2018). In turn, we deploy multiple versions of undemocratic positions, assign each respondent to see each version at least once, and analyze their effects both jointly and individually.

Neutral presentation: The undemocratic positions are presented in a manner that avoids negative connotations or normatively leading language. For instance, positions 1 and 2 are instances of gerrymandering and voter suppression, respectively, but we intentionally avoided employing such language. Put differently, we want respondents to figure out on their own whether a position violates a democratic principle or not.

When not assigned to hold an undemocratic position, a candidate held one of seven democratically neutral, “generic” positions. For instance, one of these positions read: “Served on a committee that establishes the state legislature’s schedule for each session;” we list the remaining six in the appendix. The purpose of such generic positions was to prevent candidates who were not assigned to hold an undemocratic position from appearing visually conspicuous and to create a balance in the cognitive effort needed to distinguish a candidate who adopted an undemocratic platform from one that did not.

Each respondent made 16 distinct candidate choices of which 11 were based on the following experimental design: In four randomly chosen scenarios, both candidates adopted one of the democratically neutral, “generic” positions. Throughout, we treat these as our control scenarios and label them $D^+ \text{ v. } D^+$. In seven randomly chosen scenarios, one of the candidates adopted one of our undemocratic positions while the other held a neutral position. We refer to these as our treatment scenarios and label them $D^- \text{ v. } D^+$. Whether the undemocratic position was held be the candidate visually presented on the

---

5The remaining five scenarios featured designs intended to provide extensions and robustness checks of our core design. We introduce them in section 5 and discuss in detail in the appendix.
left or right was randomly assigned. To simplify the presentation and analysis of our findings, we reshape our data so that candidate 2 always holds a neutral position \((D^+)\) and, depending on the experimental condition, candidate 1 varies between \(D^+\) and \(D^-\).

The candidate-choice experiment was embedded in a nationally representative survey of American voters that took place in August-September 2018.\(^6\) The 1,692 respondents made a total of 21,151 candidate choices.

### 4.1 Democratic Principles versus Policy Preferences

We begin our analysis of the candidate-choice experiment by examining Americans’ willingness to trade off democratic principles for preferred economic and social policies. Each candidate proposed a platform on one economic and one social policy area. Economic policies concerned either state funding for education or income taxes, while social policies concerned either immigration or the legalization of marijuana. Policy areas were randomly assigned but identical across the two candidates in each candidate-choice scenario; each candidate was independently and randomly assigned to propose one of four possible positions per policy area, ranging from typical liberal to typical conservative positions.

Before being presented with our candidate-choice scenarios, each respondent was asked to express their support on a 0-100 proximity scale for each of the 16 policy positions that the candidates might adopt. This allows us to identify each respondent’s ideal policy position on each policy area and, following the theoretical framework in section 3, compute the squared distance between a respondent’s ideal policy and each candidate’s platform for

---

\(^6\)The survey was implemented via LUCID. The first wave, which asked questions about partisanship, policy preferences, and support for democracy took place on August 28-29, 2018; the primary focus of the second wave, which took place between September 4-25, 2018, was the candidate-choice experiment. A pilot survey, implemented via Amazon Mechanical Turk, took place in March 2018. The appendix benchmarks our sample against demographic data from the US Census Bureau and partisan and attitudinal questions from the ANES.
Figure 1: Fraction voting for candidate 1 by the difference in respondents’ policy proximity to candidate 1 (vertical bars denote 95% confidence intervals)

each of our candidate-choice scenarios. The results that follow are robust to alternative measures of policy proximity between respondents and candidates; we present these in the appendix.

Figure 1 plots the fraction of respondents voting for candidate 1 as a function of the difference in policy proximity to the respondent between candidate 1 and candidate 2. In order to simplify the interpretation of our findings, we normalized the horizontal axis to the unit interval [−.5, .5], with 0 corresponding to cases when the two candidates are equally proximate to the respondent and .5 corresponding to cases when candidate 1 is a full scale closer to the respondent than candidate 2 on both policy areas. We treat the $D^+ v. D^+$ scenario (black solid line), when both candidates adopt neutral democracy
positions but differ across other attributes, as our control condition; we treat the $D^- \text{ v. } D^+$ scenario (red dashed line) as our treatment condition.

The $D^+ \text{ v. } D^+$ control scenario provides an initial plausibility check of our design. Consistently with our spatial framework, the closer candidate 1 is to respondents’ ideal policies relative to candidate 2, the more likely respondents vote for candidate 1. Specifically, the fraction of respondents voting for candidate 1 increases from 0% when candidate 1 is a full scale less proximate to the respondent than candidate 2 to 86% in the opposite case. Furthermore, when the two candidates are equally proximate to respondents, the latter act accordingly: the fraction of policy-wise indifferent respondents voting for candidate 1 is statistically indistinguishable from 50%. Respondents’ ratings of the proximity of each candidate’s policy positions thus appear to be a good predictor of their actual choices in our candidate-choice experiment.

Figure 1 also provides an initial estimate of whether and how much Americans value democracy. Because the only systematic difference between our control and treatment condition is candidate 1’s democracy position, we can interpret a change in the fraction of voters who support candidate 1 as a measure of effective support for democracy. Overall, this change amounts to a 11.15% decline in candidate 1’s vote share when he adopts an undemocratic position, with the 95% confidence interval of (9.62, 12.69). Put differently, a candidate who considers adopting one of our undemocratic platforms can expect to be defeated by a margin of about 22%.

Are Americans willing to trade off democratic principles in exchange for more appealing policies? Figure 1 allows us to address this question by partitioning our experimental electorate into two theoretically distinct subsets of voters anticipated by our theory: “democracy first” and “policy first” voters. A majority of the former supports the more democratic candidate regardless of their treatment condition. These respondents lie in the
interval at the center of Figure 1 between the intersection of the $D^-$ v. $D^+$ line with the 0.5 horizontal line and its mirror image along that line. This interval corresponds to the values ($-.1, .1$) on the horizontal axis and its limits are the empirical counterpart of the two swing voters in our model. A majority of “democracy first” respondents vote for the more democratic candidate even when doing so goes against their policy interests. By contrast, voters to the left and right of this interval are “policy first” voters: a majority of them supports the more policy-wise proximate candidate, even if that candidate adopts an undemocratic position.

To get further insights into the trade-offs between democratic principles and policies that our respondents are willing to make, consider the following counterfactual: Start with the $D^+ v. D^+$ control scenario in which both candidates adopt neutral democracy positions. Suppose that candidate 1 switches to an undemocratic position. How much more attractive would candidate 1’s policies have to become to compensate for his switching from $D^+$ to $D^-$? Figure 1 allows us to approximate such a marginal rate of substitution between democratic principles and policy proximity. For a respondent who is initially indifferent between the two candidates, this marginal rate of substitution is about 10 points on the policy proximity scale: to compensate for his less democratic position, candidate 1 has to propose policies that will makes him 10 points more attractive on the policy proximity scale. Alarmingly, candidate 1 can accomplish this by shifting just one position closer to the respondent on a single policy area. Put differently, support for democracy is highly elastic with respect to voters’ policy preferences.

We gain additional insights into the robustness of support for democracy by examining differences in the severity with which respondents punish candidate 1 for adopting an undemocratic platform. Figure 2 plots the fraction of respondents in each policy proximity category that defect from candidate 1 after he adopts an undemocratic position.
Consistent with our theoretical predictions, we see that “policy centrists” — respondents who are indifferent between the two candidates’ policies — defect from candidate 1 at a rate of almost 30% and that this rate declines as we move toward “policy extremists” — respondents who have a strong policy-based preference for one of the two candidates.

Alarmingly, the defection rate in the two most extreme subgroups on each side of Figure 2 is statistically indistinguishable from 0.\(^7\) Policy extremists are willing to trade off democratic principles for favorite policies.\(^8\)

---

\(^7\)Our finding that defection rates are either positive or statistically indistinguishable from 0 is consistent with our theoretical assumption that undemocratic platforms can be thought of as a negative valence attribute and implies that the civic virtue parameter \(\delta\) in our model is positive.

\(^8\)The conclusions in this paragraph hold up when we limit attention to “experimentally generated” policy
These differences in willingness to punish undemocratic behavior are consistent with our arguments about the pernicious consequences of polarization for democracy. Our representative sample allows us to simulate counterfactual electorates with increasing levels of policy polarization by varying the ratio of “policy centrists” to “policy extremists.” As suggested by Figure 2, an electorate consisting entirely of “policy centrists” would result in a resounding defeat of a candidate who would adopt an undemocratic platform. By contrast, an undemocratic candidate would have a 60-40 chance of prevailing in an electorate consisting entirely of the most extreme subgroups on each side of Figure 2.

4.2 Does Partisanship Trump Civic Virtue?

Are voters willing to prioritize democratic principles over partisan loyalties? To address this question, consider first contests between candidates from different parties. Figure 3 plots the fraction of respondents voting for a Republican Party candidate as a function of our respondents’ party identification on the conventional 7-point scale. As in the previous section, the control condition Rep $D^+$ v. Dem $D^+$ refers to scenarios when both candidates adopt a neutral democracy position. As expected, stronger partisans vote for their party in greater proportion, with independents breaking about evenly for the two parties.

The two treatment conditions, Rep $D^+$ v. Dem $D^-$ and Rep $D^-$ v. Dem $D^+$, correspond to the scenarios in which the Democratic and Republican Party candidate, respectively, adopts one of our undemocratic positions (blue dashed and red dotted lines in Figure 3, respectively.) Overall, a candidate who adopts an undemocratic position is penalized by the defection of 9.63% and 11.32% voters in the Rep $D^+$ v. Dem $D^-$ and Rep $D^-$ v. Dem $D^+$ scenarios, respectively. Both effects are statistically different from 0 centrists and extremists. That is, because policy platforms are randomly assigned, each respondent has a positive probability of finding herself either in between the two candidates’ platforms or to one side of both candidates’ platforms. See the appendix for details.
Figure 3: Different party contests: Fraction voting for a Republican Party candidate (with a p-value of $2.2 \times 10^{-16}$) but not statistically different from each other. At the aggregate level, voters punish undemocratic behavior by both parties and they do so evenly.

Are Americans willing to vote across party lines to punish a candidate for adopting an undemocratic position? Figure 3 allows us to address this question by examining whether a majority of each of the six partisan subgroups votes for the $D^+$ candidate when a candidate of their party supports an undemocratic position. We see that independents who “lean” toward one of the parties defect from an undemocratic co-partisan in large enough numbers to defeat that candidate. Respondents who identify as “Democrat” or “Republican” only break even. And a majority of “strong” partisans from either party would rather elect a candidate that violates democratic principles than vote across party lines.
This is the result of two forces, both of which are consistent with our theory. First, strong partisans need to defect from undemocratic candidates from their party at higher rates to compensate for their initially high levels of support for those candidates in the control condition. Second, strong partisans tend to do exactly the opposite: they are more lenient on violations of democratic principles by candidates from their party. Figure 4 shows this tendency by plotting the fraction of respondents that defect from the $D^-$ candidate by that candidate’s partisanship. We see that independents and partisan leaners punish undemocratic positions at about the same rate, regardless of the $D^-$ candidate’s partisanship. By contrast, partisans and strong partisans punish candidates from the other party 1.2 to 6.4 times as severely as those from their party. Strong partisans are willing to punish their own for violating democratic principles – but they hold candidates from the other party to a much higher standard.

We gain additional insights into such partisan double standard when we examine contests in which both candidates come from the same party. Figure 5 plots the fraction of respondents voting for candidate 1 in contests between either two Democrats or two Republicans. When both candidates adopt a neutral democracy platform (Dem $D^+$ v. Dem $D^+$ and Rep $D^+$ v. Rep $D^+$ in a black solid line.) that fraction is (by construction) 0.5. Now consider the punishment for a Democrat who adopts an undemocratic position (Dem $D^-$ v. Dem $D^+$ in a blue dashed line): its severity is increasing as we move rightward along the horizontal axis, away from partisan supporters and toward partisan opponents. The reverse holds when a Republican candidate adopts an undemocratic position (Rep $D^-$ v. Rep $D^+$ in a red dotted line.) This occurs in spite of the fact that voters are choosing between candidates from same party. Put differently, voters employ a double standard even when the partisanship of the winning candidate is preordained!
These findings are consistent with our theoretical predictions about the relationship between partisan polarization and the electorate’s willingness to punish candidates that violate democratic principles. Strong partisans defect from $D^-$ candidates at lower rates than moderates and independents, and they do so because they are reluctant to punish their own. Paralleling our conclusions from the preceding section, an electorate consisting of only strong partisans would provide a weak check on candidates who violate democratic principles. Furthermore, in electorates with a large enough partisan imbalance – as in the majority of legislative districts in the United States – a candidate of the majority party can adopt a $D^-$ position and get away with it. Precisely when the proper functioning of an electoral democratic check requires the defection of a large fraction of co-partisans, the
Figure 5: Same party contests: Fraction voting for candidate 1
latter are inclined to be forgiving.

4.3 Resisting the Menu of Manipulation

When examining our respondents’ willingness to punish candidates that violate democratic principles, we have so far pooled all democracy positions into two groups, neutral and undemocratic positions. We now examine the differences in Americans’ willingness to tolerate the distinct ways in which the individual undemocratic positions violate democratic principles and interpret them in light of several benchmarks.

Figure 6 summarizes the effects of candidates’ democracy positions as well as any other attributes on a candidate’s vote share. The dots mark coefficient estimates from a linear model that regresses candidate choices on all experimentally manipulated attributes, with bars representing the associated 95% confidence intervals. (Dots without confidence intervals represent baseline categories.) Following Hainmueller, Hopkins, and Yamamoto (2015), we interpret these coefficients as the expected change in a candidate’s vote share caused by the corresponding attribute, relative to the relevant baseline category and averaging over all other attribute levels.

Consider first the coefficients associated with the democratically neutral, “generic” positions. All seven are individually (and jointly) statistically indistinguishable from 0, implying that they do not affect a candidate’s vote share. This validates our design and interpretation of these attributes as not only democratically neutral but also more generically inconsequential.

Figure 6 also demonstrates a considerable variation in the effect of the individual undemocratic platforms on a candidate’s vote share. While all undemocratic platforms effect a candidate’s vote share negatively, the magnitude of that effect ranges from 7.8% to 14.2%. Respondents punish most severely candidates who want to prosecute journalists,
close polling stations, and ignore court rulings. These undemocratic positions result in the
defection of 12-14% voters who would have otherwise voted for that candidate.
Respondents are least sensitive to candidates who endorse gerrymandering (by 2 seats) and suggest that the governor ignore the legislature and rule by executive order.9

To put the magnitude of these effects in context, compare the effect of these
undemocratic positions to that of other positional and valence candidate attributes.
Consistent with our discussion in sections 4.1 and 4.2, the two main positional attributes –
a candidate’s party and policy platforms – have a greater impact on a respondent’s
candidate choice than any of the undemocratic positions.

Of the attributes assigned in the core 11 of the total 16 candidate choices that our
respondents made, the most naturally interpretable as valence are candidate age, years of
experience, and profession. From among the nine professions, only military service comes
close to being statistically significant but its effect is an order of magnitude smaller than
that of any of the undemocratic positions. Due to space constraints, Figure 6 omits
candidates’ age and years of experience. With a few exceptions, effects of these attributes
are also close to zero and not statistically significant.10

To help us further interpret the effect undemocratic positions, we included in 2 of the
16 choices that our respondents made two negative valence attributes intentionally
unrelated to democracy. According to the first, the candidate “was convicted of
underpaying federal income taxes;” according to the second, the candidate “was reported
to have had multiple extramarital affairs.” Estimates associated with these two attributes
appear at the bottom of Figure 6 and are labelled $V^-$. We see that, voters punish
candidates for extramarital affairs and underpaying taxes more severely than they punish

---

9In the appendix, we also present estimates that differentiate by the respondent’s partisanship. Consistent
with our earlier findings, we find few difference between supporters of the two parties.
10We present a complete set of results in the appendix.
Figure 6: Effects of candidate attributes and democracy positions on a candidate’s vote share. Dots represent coefficient estimates based on a linear regression model; bars represent 95% confidence intervals.
them for undermining democratic principles.

5 Extensions and Robustness Checks

5.1 Do Americans Know What Democracy Is (And Is Not)?

One potential objection to our conclusions so far is that Americans may simply have a poor understanding of what democracy is or what it is not. In order to evaluate this objection, we included in our survey a mix of democratic and undemocratic practices and asked each respondent to evaluate them. We list these in Figure 7. The battery of practices was introduced by the statement “Countries around the world differ in how democratic they are. We sampled the following practices from around the world. How democratic do you think each one is?” Respondents were asked to rate each statement on a scale where 1 means “not at all democratic” and 10 means “completely democratic.” The mean ratings reported in Figure 7 have been normalized to a 0-1 scale.

As Figure 7 shows, most Americans subscribe to the same liberal conception of democracy that political scientists do. The average rating is at or above .5 for items that we consider consistent with democracy (these are labelled $D^+$) and below .5 for undemocratic items ($D^-$). A closer examination of the distribution of these ratings that we present in the appendix shows some understandable partisan differences – as when conservative respondents rate item 3 (on welfare) as less democratic than liberal respondents. Crucially for our research design, items 7-10 and 12, which parallel the undemocratic treatments that we adopt in our candidate-choice experiment, consistently receive some of the lowest ratings. Put differently, most of our respondents knew that they were voting for a less democratic candidate when they did so in our candidate-choice experiment.
<table>
<thead>
<tr>
<th>Statement</th>
<th>Type</th>
<th>Mean rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Journalists frequently disagree with the president’s policies.</td>
<td>D+</td>
<td>0.8</td>
</tr>
<tr>
<td>2. The country’s legislature passed a law that bans sitting presidents from running for re-election.</td>
<td>D+</td>
<td>0.59</td>
</tr>
<tr>
<td>3. A candidate promised an increase in welfare benefits to attract voters.</td>
<td>D+</td>
<td>0.57</td>
</tr>
<tr>
<td>4. About half of the country’s registered voters do not vote in legislative elections.</td>
<td>D+</td>
<td>0.5</td>
</tr>
<tr>
<td>5. The legislature changed the electoral system from proportional to majoritarian representation.</td>
<td>N</td>
<td>0.49</td>
</tr>
<tr>
<td>6. The military overthrew a corrupt government.</td>
<td>D−</td>
<td>0.48</td>
</tr>
<tr>
<td>7. Candidates from the incumbent party use government resources when campaigning.</td>
<td>D−</td>
<td>0.38</td>
</tr>
<tr>
<td>8. Parties redraw legislative districts in order to secure control of the legislature.</td>
<td>D−</td>
<td>0.31</td>
</tr>
<tr>
<td>9. The president refused to implement a ruling by the country’s highest court.</td>
<td>D−</td>
<td>0.31</td>
</tr>
<tr>
<td>10. The government cut school funding in districts that supported the opposition.</td>
<td>D−</td>
<td>0.3</td>
</tr>
<tr>
<td>11. State media only report favorably about the governing political party.</td>
<td>D−</td>
<td>0.3</td>
</tr>
<tr>
<td>12. The country’s president closed down the legislature and is governing by executive order.</td>
<td>D−</td>
<td>0.27</td>
</tr>
<tr>
<td>13. Opposition candidates are banned from criticizing the government.</td>
<td>D−</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Figure 7: Respondents’ rating of how democratic real-world practices are (mean ratings on 0-1 scale)
5.2 Voter versus Candidate Polarization

When examining the consequences of polarization for voters’ willingness to punish undemocratic candidates, we have so far focused on the polarization of the voters’ policy preferences and partisanship. Our experimental design also allows us to study the independent effect of the polarization in candidates’ platforms. Because candidates’ policies were independently and randomly assigned, we can examine how the distance in candidates’ policy platforms affects voters’ willingness to punish undemocratic candidates. Figure 8 shows the effect of the mean distance in candidates’ policy platforms (across the two issue areas) on the fraction of respondents defecting from a $D^-$ candidate. Consistently with our theoretical framework, we see that as candidates’ policy platforms become more polarized, voters become more reluctant to punish candidates the violate democratic principles.

5.3 From an Experimental to the Real-World Party-Policy Alignment

Throughout, we based our analysis on an experimental design that assigns candidates’ policies and partisanship independently of each other. The key advantage of this design is that it allows us to separate the causal effect of candidates’ policy positions from the effect of the candidates’ partisan affiliation. The downside is that such independent assignment may result in candidates with untypical combinations of partisanship and policies (e.g. a Democrat who favors eliminating the state income tax.)

We now restrict attention to candidate choice scenarios that are more typical of the real-world combinations of candidate policy platforms and partisanship. Figure 9 shows the equivalent of Figure 3, but now discarding those candidate choice scenarios in which Democratic Party candidates adopted the rightmost of the four possible policy positions.
Figure 8: The effect of the mean distance in candidates’ policy platforms on the fraction of respondents defecting from a $D^-$ candidate
Figure 9: Different party contests when partisanship and policies align: Fraction voting for a Republican Party candidate and vice-versa for Republican Party candidates. Comparing the two figures, we see that no subset of partisans in the electorate is now willing to vote across party lines to punish a candidate for adopting an undemocratic position. This gets worse when we impose even greater alignment on candidates’ policies and partisanship. Consistent with our theoretical arguments, when candidate policy platforms and partisanship compound – just as they do in the real world – the viability of the public as a democratic check declines.
Figure 10: Turnout rates by the difference in respondents’ candidate ratings

5.4 Exit versus Choice: Abstention as a Democratic Check

After seeing a profile of two candidates, each respondent was first asked which candidate they preferred and then whether they would vote in an election that pitted the two candidates against each other. In our discussion so far, we have focused on the first of these questions – the respondents’ vote choices. The second question allows us to examine the possibility that the primary manner by which voters punish candidates who adopt undemocratic positions is not defection but rather abstention. Figure 10 supports this hypothesis: The stronger a respondent’s preference for a candidate, the more likely she is to abstain rather than defect when that candidate adopts an undemocratic position.
6 The 2017 Montana Natural Experiment

Depending on whether they voted in-person or by absentee ballot, voters in the 2017 special election for Montana’s single U.S. House seat saw two very different races. Both races pitted the Republican Greg Gianforte against the Democrat Rob Quist. Absentee voters, who in Montana make up about 50% of registered voters, saw a small-government Republican with business credentials compete against a Democrat former musician who supported mainstream liberal positions. All three major newspapers in Montana endorsed the Republican Greg Gianforte.

In-person voters saw an altogether different race. Before a final campaign rally on the day before the election, Gianforte physically attacked a journalist who repeatedly asked an inconvenient question. The attack dominated the news coverage that evening and lead the three major newspapers in Montana to rescind their endorsement of Gianforte on the morning of election day. Gianforte nonetheless won by a 5.6% vote margin.

The timing of Gianforte’s blow-up offers a real-world, quasi-experimental opportunity to evaluate the theoretical framework presented in section 3 and to corroborate the experimental results presented in section 4. We can interpret Gianforte’s attack on the journalist as a public signal about his respect for a free press or, at a minimum, as a negative signal about some other valence characteristic – e.g. the (lack of a) character fit for elected office.

---

11 The special election was held to fill the U.S. House seat vacated by Ryan Zinke, who became the Secretary of the Interior in President Trump’s administration.


13 This is, in fact, how one major newspaper, The Billings Gazette, interpreted it in its election-day editorial: “First, we hope that Republican party members and leaders call this for what it appears to be, an inexcusable act. We hope that partisan politics has not eroded our decency to the point where leaders and supporters feel the need to defend the indefensible... This incident is not Montana. It’s not America. It’s not who we are, and attacking literally those with whom we disagree cannot be justified, tolerated or explained away. We must adopt zero tolerance for such behavior if freedom of expression means anything.” The Billings Gazette Editorial Board, May 24, 2017.
Figure 11: Differences in precinct-level vote shifts for the Republican U.S. House candidate between November 2016 and May 2017

This allows us to adopt a difference-in-differences empirical strategy that compares the (relative) shift among absentee and in-person voters for the Republican U.S. House candidate between the November 2016 general election and the May 2017 special election. Even though absentee voters may be different from in-person voters, those differences cancel out when we use each group as its own, pre-treatment benchmark. That is, we think of vote shifts among absentee voters as a control that reflects what would have happened if no voters observed the negative valence shock; vote shifts among in-person voters reflect the causal effect of the negative valence shock.¹⁴

¹⁴As the Billings Gazette wrote, “To the voters who have not voted yet, we simply urge you to evaluate each candidate very carefully and make the best choice. To those who have voted: Unfortunately, Montana does not allow those who voted early to reconsider and vote again. We’re one of the few states that does not.
Our theory anticipates that, at the level of individual precincts, such
difference-in-differences should be largest in moderate precincts and decreasing in either
major party’s vote share. In the context of Montana’s partisan make-up, this implies that
the precinct-level relative decline in the Republican vote share should be largest in
moderate Republican precincts and decreasing as the strength of the Republican Party in a
precinct grows. This is because only moderate partisans are willing to punish the
Republican candidate for his attack on the journalist by either abstaining or voting for a
Democrat; for strong partisans, party loyalty trumps valence considerations.

This is precisely what we observe in Figure 11. The figure plots differences between
precinct-level Republican vote shares between November 2016 and May 2017 by the voting
method: absentee vote shifts are shown as circles, in-person vote shifts as diamonds.
Positive differences between shifts associated with the two voting methods are highlighted
by upward-facing red arrows, negative differences by downward-facing blue arrows.
Consistent with our theoretical expectations, differences in vote shifts are negative and
largest in moderate precincts; they decrease in magnitude and some even become positive
as we move to the right along the horizontal axis. This trend is statistically significant.
Montanans value a free press, but not enough for most hardline Republicans to vote for a
Democrat.

---

This would seem to be the best reason we should urge our state leaders to change that law.” The Billings

15That is, some hardline Republican precincts appear to have rewarded rather than punish Gianforte’s
attack on the journalist.
7 Conclusion: Support for Democracy or Cheap Talk?

Most of our knowledge about support for democracy in the United States (and around the world) is based on direct questions, as in “How important is it for you to live in a country that is governed democratically?” This approach invites socially desirable answers and – by design – fails to capture respondents’ willingness to trade off democratic values for other political ends. Our finding that only a small fraction of Americans are willing to prioritize democratic values in their electoral choices raises questions about the behavioral relevance of these conventional measures.

Consider our respondents’ answers to the question mentioned above. When asked to rate the importance of living in a country that is governed democratically on scale from 1 to 10, where 1 stands for “not at all important” and 10 for “absolutely important,” more than 80% of our respondents answer 8-10 and more than 95% answer 5-10. The distribution of these answers is shown in Figure 12. The mean answer of 8.6 and the median of 9 are statistically indistinguishable from answers to the same question in the 6th (last) wave of the World Values Survey (2011).\footnote{\footnotetext{For a recent reassessment of the robustness of support for democracy in the United States and Europe that is based on such conventional measures, see the debate between Alexander and Welzel (2017), Foa and Mounk (2017), Norris (2017), and Voeten (2017).}}

Nonetheless, in the different-party scenarios discussed in section 4.2, only 21% of our respondents are willing to punish an ideologically more proximate candidate who endorses an undemocratic action by defecting to an ideologically more distant candidate. That is, fewer than 1 out of 5 liberals are willing to punish a Democratic Party candidate for adopting an undemocratic platform by voting for a Republican Party candidate, and vice versa.
Figure 12: Conventionally measured support for democracy: “How important is it for you to live in a country that is governed democratically?” (1 stands for “not at all important” and 10 for “absolutely important”)

Crucially, this fraction is about the same among the more than 80% of our respondents who rate the importance of “living in a country that is governed democratically” as between 8 and 10 on a 1 to 10 scale. In other words, conventional measures of support for democracy perform poorly as predictors of a respondent’s vote for the more democratic candidate in our candidate-choice experiment and mask a remarkable willingness of Americans to put partisanship ahead of democratic principles.
References


