



# Elite rhetoric can undermine democratic norms

Katherine Clayton<sup>a,1</sup>, Nicholas T. Davis<sup>b</sup>, Brendan Nyhan<sup>c</sup>, Ethan Porter<sup>d,e</sup>, Timothy J. Ryan<sup>f</sup>, and Thomas J. Wood<sup>g</sup>

<sup>a</sup>Department of Political Science, Stanford University, Stanford, CA 94305; <sup>b</sup>Department of Political Science, University of Alabama, Tuscaloosa, AL 35487; <sup>c</sup>Department of Government, Dartmouth College, Hanover, NH 03755; <sup>d</sup>School of Media & Public Affairs, George Washington University, Washington, DC 20052; <sup>e</sup>Institute for Data, Democracy and Politics, George Washington University, Washington, DC 20052; <sup>f</sup>Department of Political Science, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; and <sup>g</sup>Department of Political Science, The Ohio State University, Columbus, OH 43210

Edited by Lee Drutman, New America Foundation, Washington, DC, and accepted by Editorial Board Member Margaret Levi April 6, 2021 (received for review November 20, 2020)

**Democratic stability depends on citizens on the losing side accepting election outcomes. Can rhetoric by political leaders undermine this norm? Using a panel survey experiment, we evaluate the effects of exposure to multiple statements from former president Donald Trump attacking the legitimacy of the 2020 US presidential election. Although exposure to these statements does not measurably affect general support for political violence or belief in democracy, it erodes trust and confidence in elections and increases belief that the election is rigged among people who approve of Trump's job performance. These results suggest that rhetoric from political elites can undermine respect for critical democratic norms among their supporters.**

democratic norms | elite rhetoric | elections

Scholars often focus on how formal rules and laws constrain political leaders, but informal norms also play a critical role in restraining elites (1, 2). Two governing norms are thought to be critical to the stability of liberal democracy: toleration of the legitimacy of the opposition and forbearance from using state power to tilt the playing field against political rivals (3). These norms are especially important after an election, when the losing side must consent to the outcome and grant power to the winning side (4).

Little is known, however, about elites' capacity to influence popular support for democratic norms, including respect for the election process and the outcomes it produces. If democracies need the public to support the norms on which the system ultimately depends (5), then elite attacks on those norms represent a key threat to democratic stability.

Unfortunately, citizens may fail to reject or may even support democratic norm violations when they occur. Partisans can recognize and punish norm violations in hypothetical scenarios (6, 7), but the effect sizes are modest compared to other forms of scandal or misconduct (8). Moreover, forging consensus can be difficult—public evaluations are often highly polarized, and partisans demonstrate reduced concern about violations that advantage their party (9).

We thus test the extent to which elite rhetoric can erode democratic norms in the contemporary United States, where former president Donald Trump frequently challenged or disregarded standards of behavior for elected leaders. While our focus on Trump may limit the generalizability of our findings, we examine this pattern during his presidency because norm violations were such an important feature of his tenure. In response to this repeated pattern of behavior, observers voiced concerns that violations of democratic norms had become so familiar that they had become normalized or made the public desensitized (10, 11), mirroring effects that have been found after repeated exposure to norm violations or aversive stimuli in other contexts (12, 13).

We specifically consider the effects of Trump's repeated attacks on the integrity of the 2020 presidential election on belief in and support for democratic norms. Although there was no credible evidence of widespread voter fraud in the United States

(14, 15), Trump engaged in an unprecedented series of attacks on the legitimacy of the 2020 election (16) which were then amplified to an even larger audience (17, 18). These claims were an especially egregious violation of democratic norms because they target confidence in free and fair elections, which is central to citizens' understanding of democracy (19). If losers see elections as illegitimate and no longer respect their outcome, the democratic compact can unwind (4).

We expect, as prior research has shown, that exposure to claims of voter fraud will reduce confidence in elections, especially among copartisans (20). In addition, we expect that exposure to these claims will reduce support for the critical democratic norm of the peaceful transfer of power. We therefore offer the preregistered hypothesis that exposure to rhetoric challenging election legitimacy will decrease respect for electoral norms and trust and confidence in elections relative to rhetoric that does not violate democratic norms, especially as exposure increases over time (H1).

We also consider a series of preregistered research questions. First, we assess whether the effects of norm-violating rhetoric are domain specific or can “spill over” to other domains—in this case, by eroding election confidence even when the norm violations are unrelated to elections (RQ1). In addition, we compare the effects of election-related norm violations (per H1) with rhetoric violating norms unrelated to elections (RQ2).

## Significance

Democracies depend on candidates and parties affirming the legitimacy of election results even when they lose. These statements help maintain confidence that elections are free and fair and thereby facilitate the peaceful transfer of power. However, this norm has recently been challenged in the United States, where former president Donald Trump has repeatedly attacked the integrity of the 2020 US election. We evaluate the effect of this rhetoric in a multiwave survey experiment, which finds that exposure to Trump tweets questioning the integrity of US elections reduces trust and confidence in elections and increases beliefs that elections are rigged, although only among his supporters. These results show how norm violations by political leaders can undermine confidence in the democratic process.

Author contributions: K.C., N.T.D., B.N., E.P., T.J.R., and T.J.W. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. L.D. is a guest editor invited by the Editorial Board.

Published under the PNAS license.

<sup>1</sup> To whom correspondence may be addressed. Email: kpc14@stanford.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2024125118/-/DCSupplemental>.

Published June 2, 2021.

We next consider two possible mechanisms for these effects. If people observe a norm not being followed, then they may begin to see compliance as both optional and less normatively desirable, especially if norm violations occur frequently (21). Perceptions of norms of behavior for elected leaders may change if people repeatedly observe attacks on election integrity, a cognitive process we refer to as normalization. We thus ask if our treatments affect perceptions of democratic norms among past political leaders (RQ3).

In contrast, a second process we consider is the tendency for aversive stimuli to evoke weaker psychological responses as exposure levels increase. This process, which we call desensitization, explains why repeated exposure to violence or trauma might numb the fear, anxiety, and physiological arousal that such stimuli initially provoke (13). We accordingly test whether exposure to norm-violating rhetoric reduces emotional reactivity to this rhetoric (RQ4).

Exposure to rhetoric claiming that elections are illegitimate, or more general norm-violating rhetoric, may also threaten normative commitments to peace and democracy. RQ5 and RQ6 seek to measure these attitudes by estimating effects on responses to a general index of political violence questions and a broad measure of support for democracy (22). We note, however, that people do not always connect related ideas about politics that are linked by abstract principles (24, 25). In addition, responses to abstract questions about support for violence or democracy may not reflect actual behavior in the real-world contexts in which citizens might engage in violent actions or challenge democratic processes.

Finally, we examine whether views of the figure in question (i.e., Trump approval) and/or partisan identification affect how people respond to norm violations (RQ7), a pattern that has been observed in prior studies of voter fraud beliefs (20).

Our results indicate that attacks on election integrity do not measurably affect our broad measures of support for political violence or belief in democracy (although we cannot rule out their effects on specific violent actions or support for democratic principles, especially among small groups of extremists). However, exposure to Trump's rhetoric erodes trust and confidence in elections and increases the belief that elections are rigged among people who approve of Trump's job performance. We also find suggestive evidence that people become desensitized to norm-violating rhetoric over time. Overall, these results imply that rhetoric from political elites can undermine support for critical democratic norms among their supporters.

### Experimental Design

Our four-wave panel experiment was preregistered on October 7, 2020 (<https://osf.io/a4tds>) and fielded from October 7–24, 2020 on Amazon Mechanical Turk. Participants were first invited to participate in a baseline survey (wave 1) measuring demographic characteristics and pretreatment attitudes (see *SI Appendix* for all survey instruments and study stimuli). Five days later, participants who completed wave 1 were recontacted for wave 2, which was open for 3 d before closing. Participants who completed wave 2 were then eligible to participate in wave 3 (open for 4 d after wave 2) and, subsequently, wave 4 (open for 5 d after wave 3). All participants were then debriefed in two separate messages to ensure they were not misled or discouraged from participating in the upcoming election.

The experimental intervention took place in waves 2 and 3 (wave 1 was a baseline survey, and wave 4 was an end-line survey). Participants who accepted the invitation to the wave 2 survey were block randomized within groups defined by wave 1 measures of political interest (median split), Trump approval, and support for respecting electoral outcomes (median split).

Respondents each viewed 20 tweets from President Trump in wave 2 and 20 in wave 3. Ten tweets that were unrelated to elections and did not violate democratic norms were fixed across conditions in each wave. The other 10 tweets in the treatment waves were randomized by condition as shown in Table 1.

*SI Appendix, Table S2* shows that respondent characteristics are balanced by condition. (See *Materials and Methods* for further details on the sample population, which is politically and demographically diverse, highly attentive, and had very low levels of attrition.)

The experimental treatments are images of real tweets by Donald Trump, excluding their original date stamp and any additional media (see the survey instrument in *SI Appendix* for an example). We selected our experimental stimuli using a protocol specified in *SI Appendix* in which 261 candidate tweets—including many previously identified as election norm violations by subject matter experts—were rated by respondents on Lucid for whether they follow or depart from past practices by American presidents and whether their topic is US elections or some other topic. The two placebo conditions included tweets that respondents considered normal behavior for American presidents (either election-related or not), while the two norm violation conditions included tweets that respondents rated as departing from past practices by American presidents (either election-related or not). The set of tweets for all conditions by survey wave is available in *SI Appendix, Table S1*.

**Outcome Measures.** We assess the following preregistered outcome measures (exact wording in *SI Appendix*):

- 1) “Trust and confidence in elections” is a factor score combining responses to questions asking whether respondents trust elections, are confident votes will be counted as intended, and are confident the ballot count will be run fairly.
- 2) “Respect for electoral norms” is measured as agreement that “An important part of democracy is to accept election losses peacefully”; “Elections in the United States are rigged in favor of [other party (randomized for pure independents)]”; and “Sometimes regular people need to be a little violent to make sure votes are counted correctly” (six-point scales). (These items did not scale well in a factor analysis, so we analyze them separately per our preregistration.)
- 3) “Support for political violence” is a factor score combining responses to questions about how often the following are acceptable (23): sending threatening and intimidating messages to [other party] leaders, harassing ordinary [other party] persons on the Internet, using violence to advance political goals, and using violence if the [other party] wins the 2020 election (party names randomized for pure independents).
- 4) “Support for democracy” is a factor score combining responses to questions asking whether it would be a good or bad way of governing the United States to have a strong leader who does not have to bother with Congress and elections, to have the army rule, or to have a democratic political system. (“Having experts, not government, make decisions according to what they think is best for the country” did not load well with the other questions

**Table 1. Randomization scheme**

Condition	Probability	Norm violation	Election
Nonelection placebo	$P = \frac{1}{6}$		
Election placebo	$P = \frac{1}{6}$		✓
General norm violation	$P = \frac{1}{3}$	✓	
Election norm violation	$P = \frac{1}{3}$	✓	✓

**Table 2. Main effects of exposure to norm violations**

	Trust in elections	Accept election	Elections rigged	Election violence	Political violence	Support democracy
Election norm violations	−0.001 (0.026)	−0.030 (0.032)	0.059 (0.048)	0.030 (0.043)	−0.040 (0.034)	−0.025 (0.031)
General norm violations	−0.017 (0.025)	−0.024 (0.032)	0.067 (0.048)	0.134 (0.043)	0.021 (0.036)	0.036 (0.033)
Difference in effects						
Election – general	0.016 (0.026)	−0.006 (0.032)	−0.008 (0.048)	−0.105 (0.044)	−0.061 (0.034)	−0.061 (0.032)
Control variables	✓	✓	✓	✓	✓	✓
N	2,137	2,137	2,137	2,137	2,001	2,001

The  $p$  values are as follows: \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.005$  (two-sided; adjusted to control the FDR per ref. 26 with  $\alpha = 0.05$ ). Cell entries are OLS coefficients, with robust standard errors in parentheses. All models control for pretreatment variables selected as most prognostic via lasso regression (see preregistration for details and list of candidate variables). Outcome variables for first four models calculated as mean of nonmissing values for each respondent across waves 2 to 4 (see *SI Appendix* for results by wave). Support for political violence and democracy were measured in wave 4. The marginal effects of the treatments on support for political violence and democracy (fifth and sixth columns, “Difference in effects” rows) were not preregistered and are thus exploratory; we include these estimates for presentational consistency.

in a factor analysis and was therefore excluded per our preregistration.)

- 5) “Past practices” is the perceived accuracy of the following description of how things generally work in American politics: “Presidential candidates accept the outcome of elections even if they narrowly lose” (four-point scale).
- 6) “Emotional reactions” are measured as the mean levels of anger/outrage (anger), anxiety/fear (anxiety), and enthusiasm/happiness (enthusiasm) that respondents reported feeling after exposure to stimulus tweets (four-point scale).

## Results

We first evaluate whether we can pool the nonelection and election placebo conditions. Across 16 preregistered models, we never reject the null of no difference in means between these conditions (see *SI Appendix, Table S3*). We therefore pool them and treat the combined set of respondents as the reference category in the models below.

We estimate ordinary least squares (OLS) regressions with HC2 robust standard errors. Each model includes a set of prognostic covariates chosen using a lasso variable selection procedure (see preregistration for details) and fixed effects for the blocks from our block randomization procedure. We also separately control the false discovery rate (FDR) for main effects and for subgroup effects (26). Except where specifically noted, all main effects and subgroup marginal effects below and in *SI Appendix* incorporate these adjusted  $P$  values ( $\alpha = 0.05$ ).

We first evaluate the main effects of exposure to election norm violation and general (nonelection) norm violation tweets relative to the pooled placebo group among our full sample. Table 2 reports tests of H1, RQ1, RQ2, RQ5, and RQ6. Results for the mean value across waves of the trust and confidence in elections and election norm outcomes are reported in the first four columns (see *SI Appendix, Tables S4 and S5* for results by wave). The last two columns report outcomes measured in wave 4 only. All scale outcomes (trust in elections, political violence, and support for democracy) are standardized factor scores; the support for election norms (accept election, elections rigged, and election violence) items are measured on six-point agree/disagree scales.

As Table 2 indicates, we find virtually no evidence that exposure to election-related or general norm violations substantially affects trust in elections, respect for election norms, support for political violence, or support for democracy among the full sam-

ple of respondents.\* We also find no significant differences in effects between the election norm violation and general norm violation treatments.

However, these null results may reflect countervailing effects among different subgroups—specifically, effects may vary by approval of President Trump or partisanship per RQ7. Following our preregistration, we therefore allow our treatment effect estimates to vary by whether respondents approve of Trump. We present these marginal effect estimates for Trump approvers and disapprovers in graphical form in Fig. 1. The interaction models from which these estimates are derived, which show that treatment effects often vary significantly by Trump approval, are presented in *SI Appendix*.<sup>†</sup>

Fig. 1 first plots how the effect of exposure to norm-violating rhetoric on trust and confidence in elections varies by Trump approval (see *SI Appendix, Table S7* for full results). Unlike in Table 2, we present marginal effects for the election norm violation and general norm violation conditions by wave as well as the mean across waves.

As the figure indicates, Trump’s election norm violations decrease trust and confidence in elections among people who approve of him by 0.24 standard deviations, on average, across waves ( $P < 0.005$ ). By contrast, exposure to the election norm violation tweets actually increases trust and confidence in elections by 0.11 standard deviations, on average ( $P < 0.01$  after  $P$  values are adjusted to control the FDR), among Trump disapprovers, mirroring the observational trend observed from 2014 to 2016 among supporters of Hillary Clinton (28). This result is consistent with literature that finds citizens often adopt political beliefs that rationalize their partisan preferences (29). More broadly, it suggests that reactions to norm violations may be

\*These null effects are fairly precise. We conducted a series of exploratory equivalence tests (27) for  $P < 0.05$  assuming unequal variances. We can rule out effects outside of the following bounds: trust in elections: [−0.07, 0.097] for general norm violation (general) vs. control and [−0.083, 0.083] for election norm violation (election) vs. control; accept elections: [−0.066, 0.073] for general vs. control and [−0.076, 0.065] for election vs. control; elections rigged: [−0.308, 0.056] for general vs. control and [−0.201, 0.055] for election vs. control; election violence: [−0.241, −0.045] for general vs. control and [−0.094, 0.095] for election vs. control; political violence: [−0.13, 0.052] for general vs. control and [−0.009, 0.159] for election vs. control; support democracy: [−0.114, 0.064] for general vs. control and [−0.038, 0.133] for election vs. control.

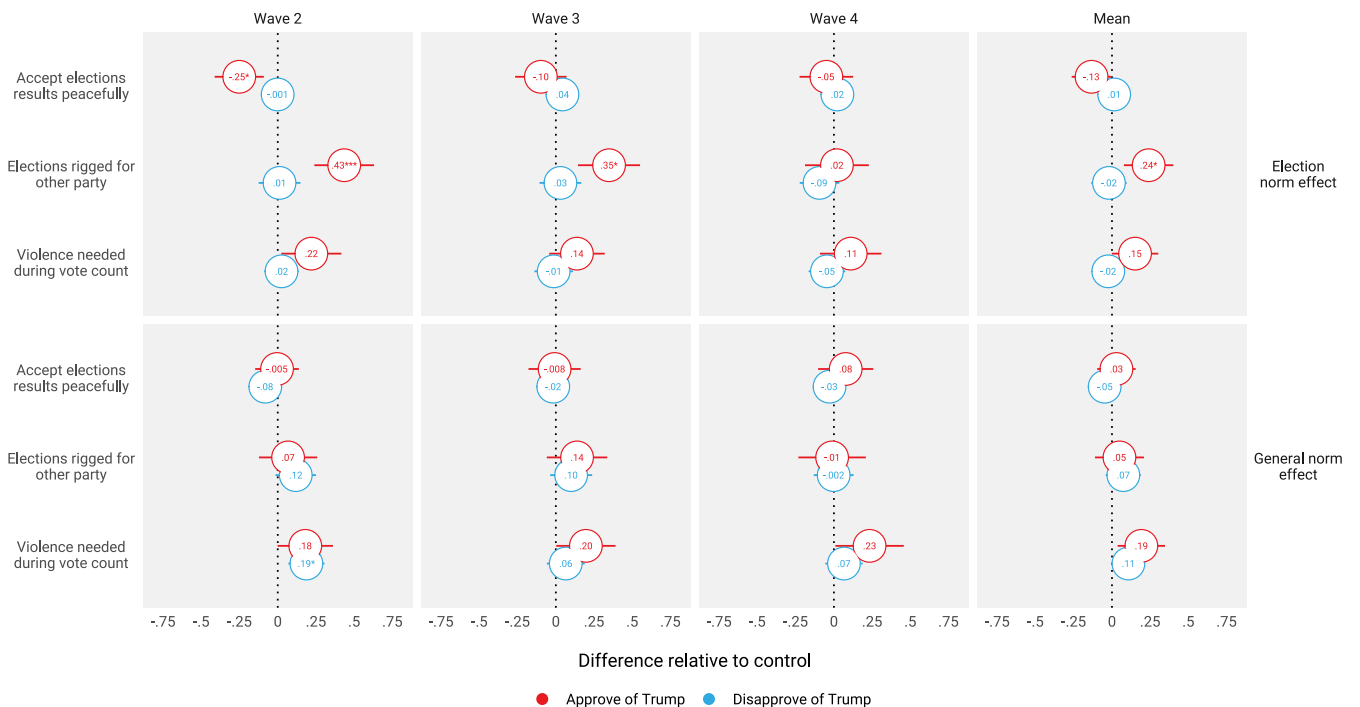
<sup>†</sup>Results when the treatments are instead interacted with an indicator for whether the respondent identifies with or leans toward the Republican Party are generally very similar; we thus do not discuss them further here but present the results in tabular form in *SI Appendix*.



**Fig. 1.** Marginal effects on trust and confidence in elections by Trump approval; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.005$  (two-sided; adjusted to control the FDR per ref. 26 with  $\alpha = 0.05$ ). Outcome measures are factor scores combining responses to questions asking whether respondents trust elections (seven-point scale) and are confident that votes nationwide will be counted as intended and that election officials will manage counting fairly (four-point scales). Bars represent 95% CIs (not shown if CI is smaller than circle indicating the point estimate; note that these intervals do not incorporate the FDR correction and so significance cannot be assessed visually). See *SI Appendix, Table S7* for exact wording and full results.

conditional upon attitudes toward the individual in question. Fig. 1 shows a similar but weaker pattern for general norm violation tweets. Three effect estimates indicate that exposure to these statements reduces trust in elections among Trump approvers using unadjusted  $P$  values, but none remain statistically significant after our preregistered FDR adjustment (see *SI Appendix, Table S7*).

Similarly, Fig. 2 shows that exposure to rhetoric violating election norms sometimes reduces respect for those norms among Trump approvers (see *SI Appendix, Tables S9 and S10*). Most notably, beliefs that elections are rigged increase by 0.43 points on a six-point scale in wave 2 ( $P < 0.005$ ) and by 0.24 points, on average, across waves 2 to 4 ( $P < 0.05$ ). Election norm violations also decrease willingness to accept election results peacefully



**Fig. 2.** Marginal effects on democratic norms by Trump approval; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.005$  (two-sided; adjusted to control the FDR per ref. 26 with  $\alpha = 0.05$ ). Mean agreement or disagreement with three separate statements on election-related democratic norms (six-point scale) by wave (first through third columns) and across waves (fourth column). Bars represent 95% CIs (not shown if CI is smaller than circle indicating the point estimate; note that these intervals do not incorporate the FDR correction and so significance cannot be assessed visually). See *SI Appendix, Tables S9–S12* for exact wording and full results.

among Trump approvers, but only in wave 2 ( $P < 0.05$ ). However, the election norm violation condition has no measurable effect on beliefs that violence is needed for votes to be counted correctly across waves or overall. Similarly, no measurable effects are found for rhetoric violating election norms among Trump disapprovers or for the general norm violation condition among either group.

This pattern of heterogeneous effects by Trump approval does not extend to support for political violence or democracy. The election norm violation treatment does not measurably affect these outcomes among either Trump approvers or disapprovers (see *SI Appendix, Table S13*). We underscore, however, that our findings do not indicate that norm-violating rhetoric has no effect on support for political violence or democracy. Caution is required in extrapolating these findings beyond the bounds of the survey context in which they were measured (a caveat we return to and expand upon in *Conclusion*).

Finally, we investigate whether repeated exposure to norm violations creates normalization (RQ3) and/or desensitization (RQ4). We test for normalization by examining treatment effects on perceptions of past respect for democratic norms. We find no significant effects of the treatments on beliefs that past candidates failed to respect narrow losses overall or by Trump approval (see *SI Appendix, Tables S15 and S16*). By contrast, evidence of desensitization is mixed (see *SI Appendix, Tables S18, S19, S21, and S22*). Self-reported anger and anxiety both decrease between waves 2 and 3 among people exposed to either type of norm-violating rhetoric (by 0.07 to 0.08 for anger and 0.06 to 0.08 for anxiety, on four-point scales). However, these declines are significant for only one subgroup of respondents in one treatment condition after we apply our preregistered adjustment to the  $P$  values to control the FDR: decreased anger in response to general norm violation tweets among Trump disapprovers ( $-0.14$ ,  $P < 0.005$ ). Additionally, prior exposure to election or general norm violation tweets decreases both anger and anxiety in response to novel election norm violation tweets in wave 4 (by 0.14 to 0.16 for anger and 0.14 for anxiety, on four-point scales). However, although these contrasts are statistically significant under classical hypothesis tests, none remain significant after our FDR correction. As with the change in reactions between waves 2 and 3, we instead observe only a single significant subgroup effect (anxiety decreases by 0.20 among Trump disapprovers after exposure to general norm violations,  $P < 0.05$ ). Desensitization thus appears to be a more likely consequence of repeated exposure to norm violations than normalization, but our results are not conclusive.

## Conclusion

While Donald Trump's attacks on democratic norms prompted concern from journalists, scholars, and everyday citizens, the causal effect of such rhetoric on public attitudes toward democracy is not known. We present a study estimating the effects of exposure to norm-violating rhetoric from a multiwave experiment conducted during the waning days of the 2020 US presidential election. We find no evidence that support for a battery of general questions on political violence or support for democracy change after repeated exposure to these statements.

We urge caution, however, in interpreting these results. Our findings should not be understood to exonerate Trump for inciting violence, including during the January 6 insurrection. First, decades of scholarship in political science tell us that citizens often fail to draw connections between abstract principles and specific political attitudes and behavior (24, 25). Second, Trump supporters may refrain from endorsing violence in their survey responses but still act to support it elsewhere. Finally, it is also possible that Trump's rhetoric incites violent, antidemocratic actions among a small number of people whose extreme prefer-

ences are rare in a sample like ours, but who can still coordinate to wreak havoc on democracy—a limitation of survey research on this topic.

Moreover, we find compelling evidence that exposure to norm violations has other pernicious effects among Trump's supporters. Among people who approve of his performance in office, repeated exposure to norm-violating rhetoric about electoral fraud erodes trust and confidence in elections and increases beliefs that elections are rigged.

Our study has important limitations. While we strove for realism in the design of our treatments, participants nonetheless encountered Trump's tweets in the context of an online survey rather than the way they would on Twitter or in other settings in which they are exposed to political news and information. The effects of Trump's tweets likely also vary by whether they are reinforced or countered by other information, a design variant that should be evaluated in future research. Twitter, for example, flagged some of Trump's claims about fraud after the election for including disputed or misleading information, which may shape users' reactions to such content (30). Second, we conducted our experiment in a saturated news environment in which many respondents had presumably already been exposed to Trump's statements multiple times via other means. The effects of additional exposure, including potential normalization or desensitization, may therefore have been limited, especially given Internet use levels in our sample (31). Third, we focused on norm-violating rhetoric from Trump alone, but future research should seek to understand how norm violations by other politicians affect public opinion. Finally, although treatment effect heterogeneity by sample type is frequently overstated (32), our study should be replicated in a representative sample if acceptable levels of attrition can be achieved.

Nonetheless, our study offers causal estimates of the effects of Trump's antidemocratic rhetoric on the mass public's commitment to democracy. Norms are typically thought to constrain the behavior of elites (1). As we show here, however, when elites strategically violate norms, their supporters respond accordingly. Just as elites can shape policy views along partisan lines (33), elite rhetoric can shape normative beliefs in core democratic values such as confidence in elections and support for peaceful transfers of power. These findings do not indicate that elites can erode democratic norms easily or that the effects of norm violations are uniform across the entire population. At least for a politician's supporters, however, support for democratic norms appears to be more fragile than previously assumed. These dynamics represent a potential threat to the acceptance of unfavorable election results.

## Materials and Methods

**Internal Review Board Approval.** The study was approved by the Committee for the Protection of Human Subjects at Dartmouth College (ID: STUDY00032100; MOD00010368). All participants provided informed consent prior to participating in the study.

**Participant Sample.** Participants for this study were recruited from a pool of approximately 3,000 people who previously took part in an unrelated study conducted on Mechanical Turk by some of the authors. Although online convenience samples have notable limitations, results from studies conducted with Mechanical Turk panelists mirror those obtained from nationally representative samples (32, 34–36). Using Mechanical Turk is essential for conducting this study due to the theoretical importance of measuring the effects of repeated exposure to the treatment in question over time. Respondent retention rates for multiwave surveys on Mechanical Turk substantially exceed even those observed in benchmark surveys like the American National Election Study (37). As a result, we greatly reduce the risk of posttreatment bias due to differential attrition between conditions, which otherwise plagues survey experiments of this type (38). We also note that the pattern of results we observe showing significant effects for some outcomes but not others

suggests that the null findings we do observe are not driven by respondent inattention.

Because Mechanical Turk overrepresents political liberals (39) and we expected heterogeneous treatment effects, we adopted a recruitment strategy that would maximize our ability to compare people of different political leanings. We first conducted extensive screening prior to the study to recruit a substantial number of Republicans/conservatives. We also limited recruitment to respondents who previously identified as a Democrat or Republican or said they leaned toward a major party, excluding so-called pure independents (note that  $N = 33$  respondents identified as pure independents in wave 1 of our study). Finally, we screened out bots and low-effort respondents with an open-ended text question. Respondents whose answers did not meet the criteria suggested in prior research were deemed ineligible (40), as were those who sped too quickly through screening surveys.

The resulting sample provides high-quality survey responses (96% correct on an attention check in wave 1) and represents a wide range of political and demographic groups (see *SI Appendix, Table S2*), including Trump approvers (31.6%) and Republicans (39.3%). Additionally, our sample is externally valid in that it is made up disproportionately of people who frequently use the Internet—precisely the group that is most likely to encounter norm-violating rhetoric on a platform like Twitter. Our respondent pool therefore constitutes a valid sample for testing our hypotheses (although replication on a representative sample would, of course, be desirable).

A total of 2,477 participants completed the wave 1 baseline survey. Those who completed wave 1 were then invited to wave 2, the wave in which

participants were assigned to treatment. In total, 2,151 people completed wave 2, the first treatment wave. Wave 2 participants were then invited to wave 3, the second treatment wave ( $N = 1,960$ ), and wave 4, the end line survey ( $N = 2,013$ ). Participants were paid \$1.50 per wave completed, plus an additional \$2 bonus if they completed all four waves. To reduce the risk of bias due to differential attrition, we include all respondents who completed wave 2 in our analysis regardless of whether they completed wave 3 and/or 4. However, attrition was exceptionally low; 91.1% and 93.6% of wave 2 participants took part in waves 3 and 4, respectively. We find little evidence of differential attrition across treatment conditions overall or by wave (see *SI Appendix* for details).

**Data Availability.** Data files and scripts necessary to replicate the results in this article have been made available at the Open Science Framework (<https://osf.io/a4tds>).

**ACKNOWLEDGMENTS.** We are grateful to Bright Line Watch, the Stanford Center for American Democracy, and the Department of Political Science at the University of Alabama for financial support and to John Carey, Mia Costa, Eugen Dimant, Matthew Graham, Gretchen Helmke, Michael Herron, Yusaku Horiuchi, Dean Lacy, Michael Neblo, Katy Powers, Carlos Prato, Mitch Sanders, Serge Severenchuk, Paul M. Sniderman, Udi Sommer, Sue Stokes, and Shun Yamaya for helpful comments. This research is also supported by the John S. and James L. Knight Foundation through a grant to the Institute for Data, Democracy & Politics at The George Washington University. All mistakes are our own.

- G. Helmke, S. Levitsky, Informal institutions and comparative politics: A research agenda. *Perspect. Polit.* 2, 725–740 (2004).
- J. R. Azari, J. K. Smith, Unwritten rules: Informal institutions in established democracies. *Perspect. Polit.* 10, 37–55 (2012).
- S. Levitsky, D. Ziblatt, *How Democracies Die* (Crown, New York, NY, 2018).
- C. J. Anderson, A. Blais, S. Bowler, T. Donovan, O. Listhaug, *Loser's Consent: Elections and Democratic Legitimacy* (Oxford University Press, Oxford, United Kingdom, 2005).
- B. R. Weingast, The political foundations of democracy and the rule of the law. *Am. Polit. Sci. Rev.* 91, 245–263 (1997).
- J. Carey *et al.*, Who will defend democracy? Evaluating tradeoffs in candidate support among partisan donors and voters. *J. Elections, Public Opin. Parties*, 10.1080/17457289.2020.1790577 (2020).
- M. H. Graham, M. W. Svobik, Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States. *Am. Polit. Sci. Rev.* 114, 392–409 (2020).
- S. J. Basinger, Scandals and congressional elections in the post-watergate era. *Polit. Res. Q.* 66, 385–398 (2013).
- J. M. Carey, G. Helmke, B. Nyhan, M. Sanders, S. Stokes, Searching for bright lines in the Trump presidency. *Perspect. Polit.* 17, 699–718 (2019).
- A. Jentleson, Trump is normalizing the unthinkable—Booing him is a civic responsibility. *GQ*, 1 November 2019. <https://www.gq.com/story/the-necessity-of-booming-trump>. Accessed 11 May 2021.
- B. Klaas, After three years of Trump, we've lost our ability to be shocked. *Washington Post*, 14 January 2020. <https://www.washingtonpost.com/opinions/2020/01/14/after-three-years-trump-weve-lost-our-ability-be-shocked/>. Accessed 11 May 2021.
- C. Bicchieri, *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms* (Oxford University Press, Oxford, United Kingdom, 2016).
- E. B. Foa, T. M. Keane, M. J. Friedman, J. A. Cohen, *Effective Treatments for PTSD: Practice Guidelines from the International Society for Traumatic Stress Studies* (Guilford, 2010).
- L. C. Minnite, *The Myth of Voter Fraud* (Cornell University Press, Ithaca, NY, 2011).
- E. Kiely, L. Robertson, R. Rieder, D. Gore, The president's trumped-up claims of voter fraud *FactCheck.org* (2020). <https://www.factcheck.org/2020/07/the-presidents-trumped-up-claims-of-voter-fraud/>. Accessed 11 May 2021.
- R. Ballhaus, J. Palazzolo, A. Restuccia, Trump and his allies set the stage for riot well before January 6. *Wall Street Journal*, 6 January 2021. <https://www.wsj.com/articles/trump-and-his-allies-set-the-stage-for-riot-well-before-january-6-11610156283>. Accessed January 8, 2020.
- Y. Benkler *et al.*, Mail-in voter fraud: Anatomy of a disinformation campaign (Research Publ. 2020-6, Berkman Center, 2020).
- G. Pennycook, D. Rand, Examining false beliefs about voter fraud in the wake of the 2020 presidential election. *Harvard Kennedy School Misinf. Rev.* 2, 1–19 (2021).
- N. T. Davis, K. Goidel, Y. Zhao, The meanings of democracy among mass publics. *Soc. Indic. Res.* 153, 849–921 (2021).
- B. Albertson, K. Guiler, Conspiracy theories, election rigging, and support for democratic norms. *Res. Pol.* 7, 2053168020959859 (2020).
- C. Bicchieri *et al.*, Social proximity and the erosion of norm compliance (Tech. Rep. 13864, Institute of Labor Economics, 2020).
- L. Drutman, J. Goldman, L. Diamond, Democracy maybe: Attitudes on authoritarianism in America. *Democracy Fund Voter Study Group* (2020). <https://www.voterstudygroup.org/publication/democracy-maybe>. Accessed 11 May 2021.
- L. Mason, N. Kalmoe, What you need to know about how many Americans condone political violence — and why. *The Washington Post*, 11 January 2021. <https://www.washingtonpost.com/politics/2021/01/11/what-you-need-know-about-how-many-americans-condone-political-violence-why/>. Accessed 21 May 2021.
- P. E. Converse, *The Nature of Belief Systems in Mass Publics* (University of Michigan Press, Ann Arbor, MI, 1964).
- R. E. Lane, *Political Ideology: Why the American Common Man Believes What He Does* (Free Press of Glencoe, Glencoe, IL, 1962).
- Y. Benjamini, A. M. Krieger, D. Yekutieli, Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93, 491–507 (2006).
- D. Lakens, Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Soc. Psychol. Personal. Sci.* 8, 355–362 (2017).
- B. Sinclair, S. S. Smith, P. D. Tucker, "It's largely a rigged system": Voter confidence and the winner effect in 2016. *Polit. Res. Q.* 71, 854–868 (2018).
- B. E. Lauderdale, Partisan disagreements arising from rationalization of common information. *Pol. Sci. Res. Method.* 4, 477–492 (2016).
- K. Conger, How Twitter policed Trump during the election. *NY Times*, 6 November 2020. <https://www.nytimes.com/2020/11/06/technology/trump-twitter-labels-election.html>. Accessed 11 May 2021.
- J. N. Druckman, T. J. Leeper, Learning more from political communication experiments: Pretreatment and its effects. *Am. J. Polit. Sci.* 56, 875–896 (2012).
- A. Coppock, Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Pol. Sci. Res. Method.* 7, 613–628 (2019).
- G. S. Lenz, *Follow the Leader? How Voters Respond to Politicians' Policies and Performance* (University of Chicago Press, Chicago, IL, 2013).
- J. J. Horton, D. G. Rand, R. J. Zeckhauser, The online laboratory: Conducting experiments in a real labor market. *Exp. Econ.* 14, 399–425 (2011).
- A. J. Berinsky, G. A. Huber, G. S. Lenz, Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Polit. Anal.* 20, 351–368 (2012).
- K. J. Mullinix, T. J. Leeper, J. N. Druckman, J. Freese, The generalizability of survey experiments. *J. Exp. Pol. Sci.* 2, 109–138 (2015).
- K. Gross, E. Porter, T. J. Wood, Identifying media effects through low-cost, multiwave field experiments. *Polit. Commun.* 36, 272–287 (2019).
- J. M. Montgomery, B. Nyhan, M. Torres, How conditioning on posttreatment variables can ruin your experiment and what to do about it. *Am. J. Polit. Sci.* 62, 760–775 (2018).
- Y. Krupnikov, A. S. Levine, Cross-sample comparisons and external validity. *J. Exp. Pol. Sci.* 1, 59–80 (2014).
- C. Kennedy *et al.*, *Assessing the Risks to Online Polls from Bogus Respondents* (Pew Research Center, 2020).