SUPPLEMENTAL APPENDIX

TWO TERMS OF ENDEARMENT?
INCUMBENT-PARTY PERFORMANCE IN U.S. PRESIDENTIAL ELECTIONS

The present study obtained Institutional Review Board (IRB) approval from the researcher's University prior to fielding the study. The survey lasted approximately 10 minutes and concluded by thanking and debriefing respondents about the general purpose of the study. Respondents participated voluntarily and were compensated for their participation in an ethical manner, and in a way consistent with existing research practice (e.g., see Berinsky, Huber and Lenz 2012). Researchers paid $1.50 per respondent for participation in the study (approximately 125% of the U.S. federal minimum wage (equal to $7.25 at the time of the study)).

Competing Interests: The author declares none.

Table A1 provides descriptive statistics for respondents in the Lucid sample. As a reference, the table also includes national benchmarks from recent US Census data.

TABLE A1. Descriptive Statistics, Lucid Sample

	Lucid Study (N=3,351)	National Benchmarks
Median HH Income	$45k-49k	$67k
Median Age	45	38.1
Female	51.66%	51%
White	71.47%	76.3%
Black	9.73%	13.4%
Hispanic	11.97%	18.5%
Democrat	43.48%	42%
Independent	21.19%	11%
Republican	35.33%	47%
Liberal	27.01%	--
Moderate	43.48%	--
Conservative	29.51%	--
Northeast	20.14%	20%
South	37.15%	34%
Midwest	18.59%	20%
West	24.11%	26%

Notes: The Lucid sample was selected to mirror U.S. Census data on Age (18-24; 25-34; 35-44; 45-54; 55-64; 65+), Race (White; Black; Hispanic; Asian; Other), Gender, and Geographic Region (West; Midwest; Northeast; South). Household Income, Age, Gender, and Race/Ethnicity national benchmarks are from most recent US Census data available. Party identification benchmarks are from Gallup (2021 data). Regional benchmarks are from Lucid's targets based upon Census data.

Survey Measures

<u>*Party Identification*</u>

Generally speaking, do you consider yourself to be a(n):

○ Strong Democrat (1)

○ Democrat (2)

○ Independent, but Leaning Democratic (3)

○ Independent (4)

○ Independent, but Leaning Republican (5)

○ Republican (6)

○ Strong Republican (7)

<u>*Incumbency Norms*</u>

The U.S. Constitution allows presidents to serve up to two (4-year) terms in office. Please rate your level of agreement with the following general statement, being sure to disregard any considerations relevant to the current U.S. president:

Regardless of their political party, as a general rule, I feel that presidents should be re-elected so that they can serve a second term in office.

- ○ Strongly Disagree (1)
- ○ Disagree (2)
- ○ Slightly Disagree (3)
- ○ Slightly Agree (4)
- ○ Agree (5)
- ○ Strongly Agree (6)

Now please rate your level of agreement with the following general statement, being sure to disregard any considerations relevant to the current U.S. president:

Regardless of their political party, as a general rule, I feel that once a political party has had control of the presidency for two full (4-year) terms, the other party should get a chance at being president.

- ○ Strongly Disagree (1)
- ○ Disagree (2)
- ○ Slightly Disagree (3)
- ○ Slightly Agree (4)
- ○ Agree (5)
- ○ Strongly Agree (6)

Had you been a voter during the time of this election, how do you think you would have voted?

○ I would have voted for the Democratic candidate (1)

○ I would have voted for the Republican candidate (2)

○ I would have voted, but not for either the Democratic candidate nor Republican candidate (3)

○ I would have abstained (not voted at all) (4)

FIGURE B1. Correlates Of Norm Endorsement
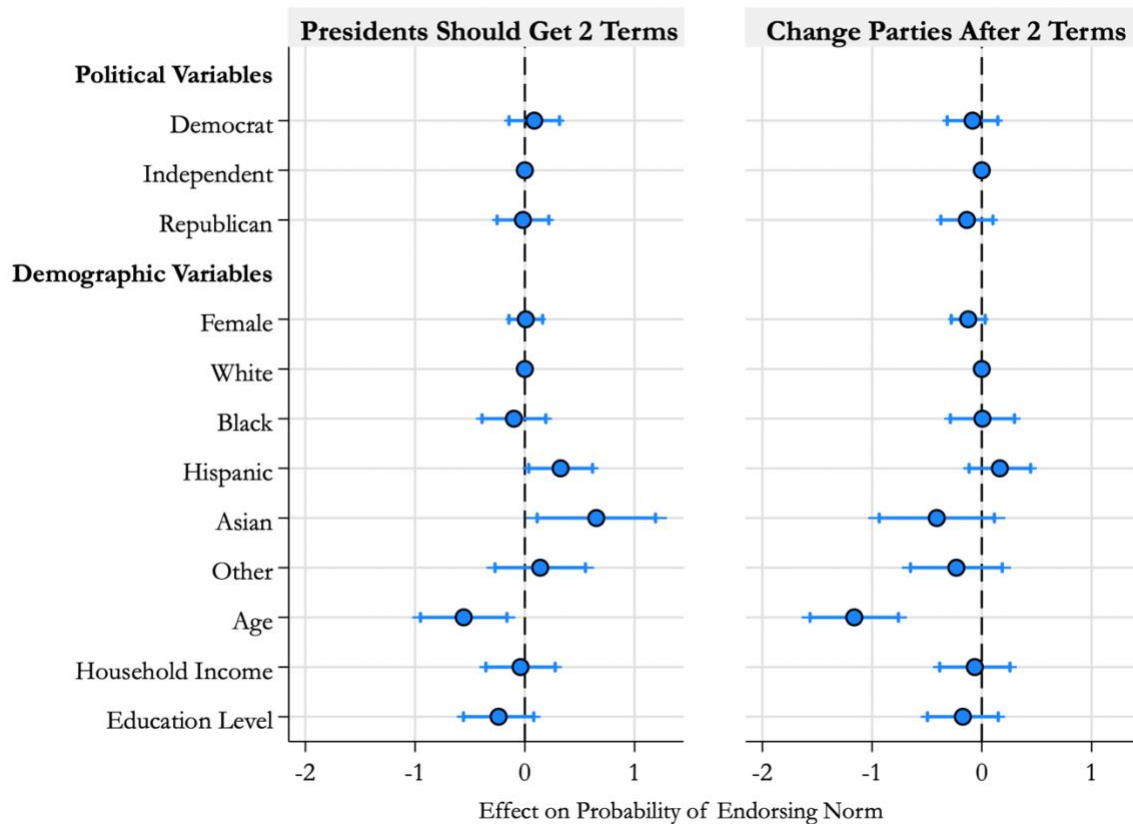
Notes: Outcomes are coded as agreement (1) vs disagreement (0) with each norm (see survey questions above). Point estimates indicate change in probability of endorsing the norm of re-electing presidents who have only served one term (left), and changing parties after a party has held office for two terms (right) with covariates at observed values. 90% (vertical caps) and 95% confidence intervals shown. Base categories appear on the vertical 0 lines. Continuous variables are recoded to range from 0 to 1. Models = logistic regression with 1,622 respondents each. Overall the models suggest that age and race are associated with endorsement of each norm, while other variables exhibit smaller (non-significant) estimated effects.

SURVEY EXPERIMENT REGRESSION MODEL OUTPUT & INVESTIGATING THE POSSIBILITY OF PRIMING

Table C1 provides the regression output underlying Figures 2 and 3 in the manuscript.

TABLE C1. Regression Output (Lucid Experiment)

	Democratic Candidate	Republican Candidate	2T vs. 1T Candidate
One-Term Incumbent	0.26**	0.19*	--
	(0.11)	(0.11)	--
Two-Term Incumbent	0.12	-0.02	-0.18*
	(0.11)	(0.11)	(0.08)
Constant	-0.28***	-0.44***	-0.14*
	(0.08)	(0.08)	(0.05)
Observations	2,014	2,010	2,678

*Notes: These are the results underlying manuscript Figures 2 and 3. The control condition is the excluded category for the first two models; one-term incumbency is the excluded category for the third model. Logistic regression model coefficients shown with standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05, ^ p<0.10 (one-tailed hypothesis tests).*

As noted in the manuscript, a subset of respondents was asked (pre-treatment) about their normative beliefs regarding incumbency in presidential elections. Specifically, 100% of control group respondents were asked these items, while only 50% of respondents in the treatment groups (i.e., a one-term or two-term incumbent-party candidate) were randomly selected to answer these items. Importantly, a series of demographic and unrelated questions were asked after the measures involving norms and before the experiment. This approach, along with asking only a subset of respondents about their normative beliefs, were deliberately taken to minimize the risk of priming respondents in a way that might augment effect sizes.

Nevertheless, the possibility of a priming effect is something that can be investigated empirically. As such, Figure C1 shows the effect of changing from a one-term party-incumbent candidate to a

FIGURE C1. Investigating the Possibility of Priming for 2T vs. 1T Contrast
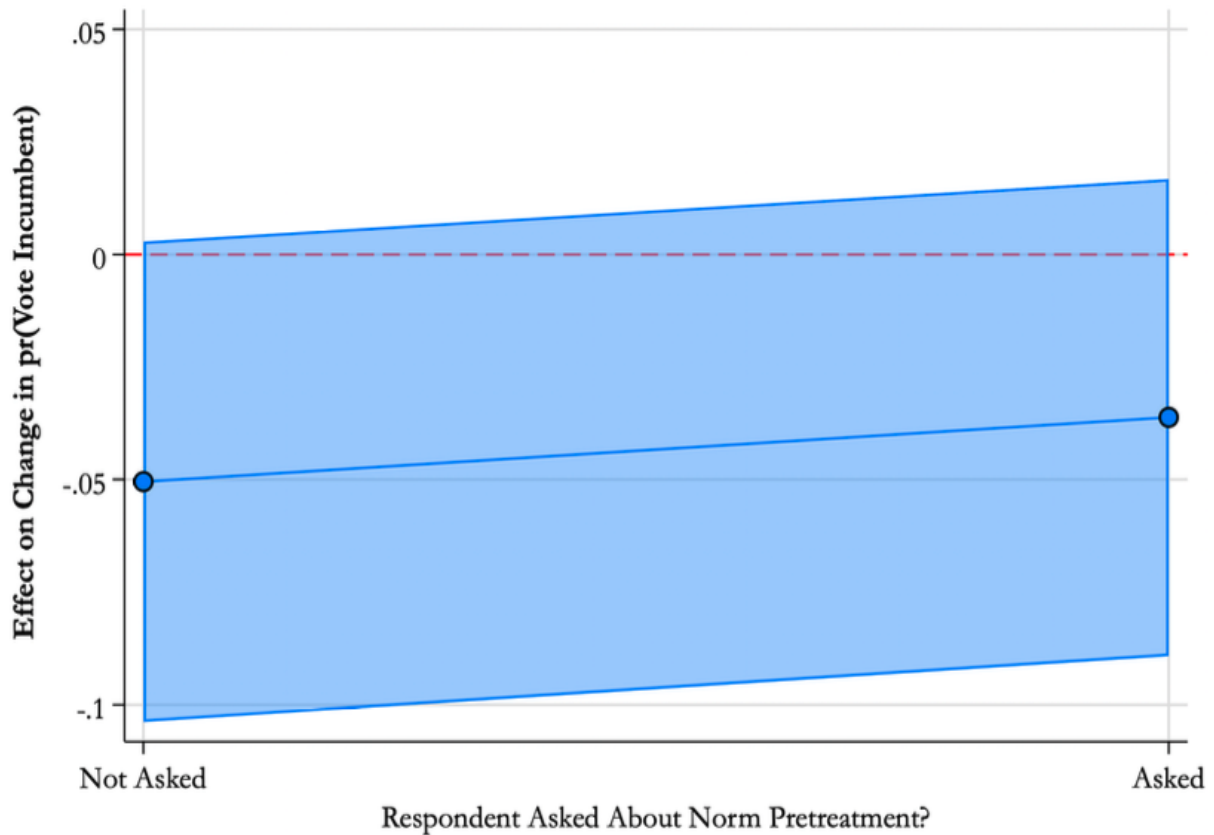
Notes: Models are logistic regression with an interaction specified between treatment assignment and a whether pre-treatment norm questions were asked. 95% CIs shown. N=2,678.

two-term party incumbent candidate (see *y*-axis).[1] The *x*-axis separates this effect by whether a respondent was or was not asked the normative belief questions pre-treatment. As reported in the manuscript, this effect is negative. However, for our purposes here, what Figure C1 reveals is that the estimated negative effect is very similar regardless of whether a respondent was asked the pre-treatment normative belief question or not (the effect among those not asked (asked) rounds to 5 (4) percentage points; the interaction term between treatment and being asked was also non-significant (p=.72)). Thus, there is no evidence here that a priming effect might explain the results reported in the paper (indeed, the slightly smaller effect among those "Asked" is opposite what we would expect had priming occurred.

---

[1] Because all control group respondents were asked about their normative beliefs, this analysis cannot be performed using a treatment vs. control contrast. However, the 2T vs. 1T contrast is a strong effect that should still exhibit a priming effect if indeed there was one.

TREATMENT EFFECTS BY RESPONDENTS' PARTY IDENTIFICATION & NORMATIVE BELIEFS

Does Party Moderate Treatment Effects?

As the Lucid experiment features partisan candidates, it is reasonable that we might observe noticeably stronger or weaker treatment effects depending upon the party identification of the respondent. As such, the results featured in manuscript Figures 3 and 4 were re-analyzed via allowing the treatment to interact with respondents' party identification (Democrat, Independent, or Republican (leaners were coded as partisans)).

The results for the Democratic candidate are featured in Figure D1, which plots the estimated treatment effect of 1) the 1-term (1T) vs. Control contrast, and 2) the 2-term (2T) vs. Control contrast, for each partisan group. First, echoing the results shown in the manuscript, it is clear that the 2T vs. Control contrast (see red point estimates) exhibits noticeably weaker effects compared to the 1T vs. Control contrast (see blue point estimates). Second, treatment estimates tend to be smallest for Democratic respondents, likely because these respondents were inclined to vote for the Democratic candidate regardless of incumbency status. Third, it is notable that while the largest estimate is for Independents (i.e., those who might be most inclined to vote based upon incumbency status), we nevertheless observe a sizable 1T effect among Republican respondents.
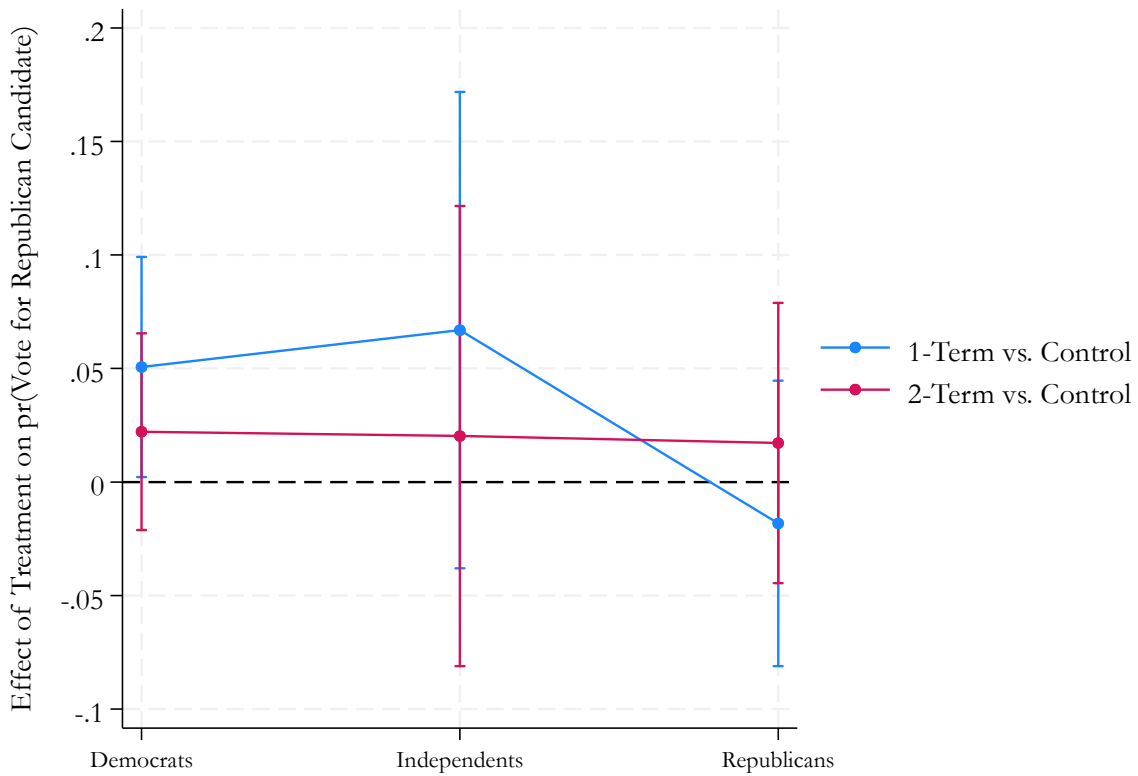
FIGURE D1. Treatment Effects on pr(Voting for Democratic Candidate) by Party
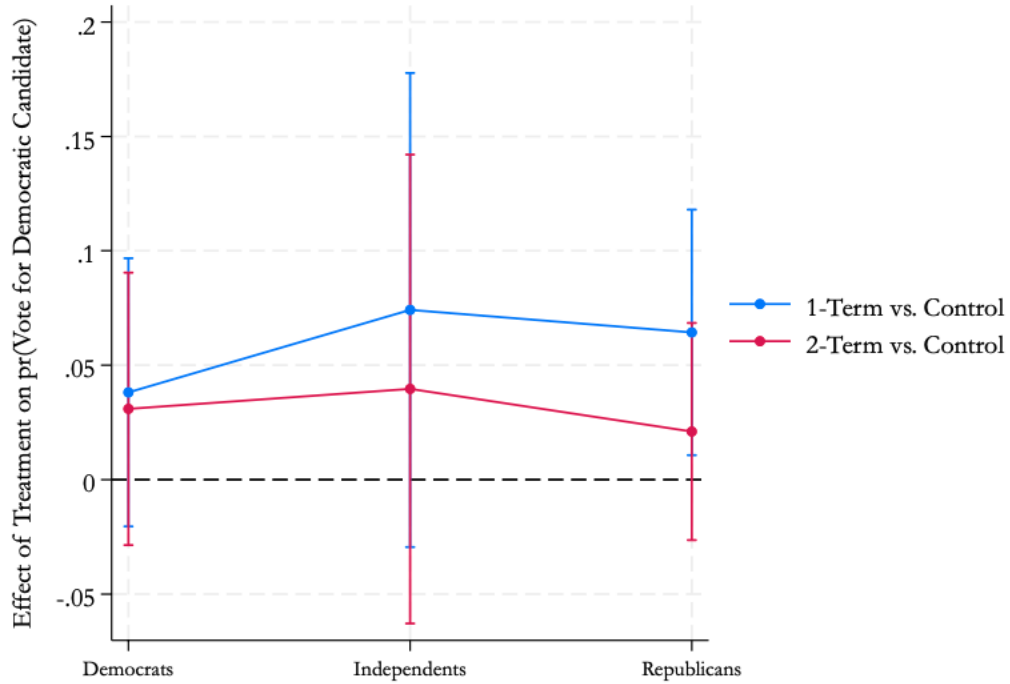
Notes: Models are logistic regression with an interaction specified between treatment assignment and a trichotomous measure of party identification (leaners coded as partisans). 95% CIs shown. N=1,344.

The results for the Republican candidate are featured in Figure D2. First, we again observe that the 2T vs. Control contrast (see red point estimates) tends to exhibit weaker effects compared to the 1T vs. Control contrast (see blue point estimates). Second, mirroring the previous graph, the weakest results tend to be for Republican respondents (again, likely because these respondents were inclined to vote for the Republican candidate irrespective of the treatment). Third, we again see that the largest estimate is for Independents and that a sizable 1T effect exists among Democratic respondents.

FIGURE D2. Treatment Effects on pr(Voting for Republican Candidate) by Party

Notes: Models are logistic regression with an interaction specified between treatment assignment and a trichotomous measure of party identification (leaners coded as partisans). 95% CIs shown. Lucid data. N=1,335.

Lastly, Figure D3 reports, for each partisan group, the effect of the 2T vs. 1T contrast on the probability of voting for the incumbent. Here we observe remarkably similar results across partisan categories. Compared to one-term incumbents, candidates whose party has occupied the presidency for the two previous terms receive between 4 and 5 percentage-points less in terms of vote share, regardless of the party of the respondent.

FIGURE D3. Effect of 2T vs. 1T on pr(Voting for Incumbent-Party Candidate) by Party
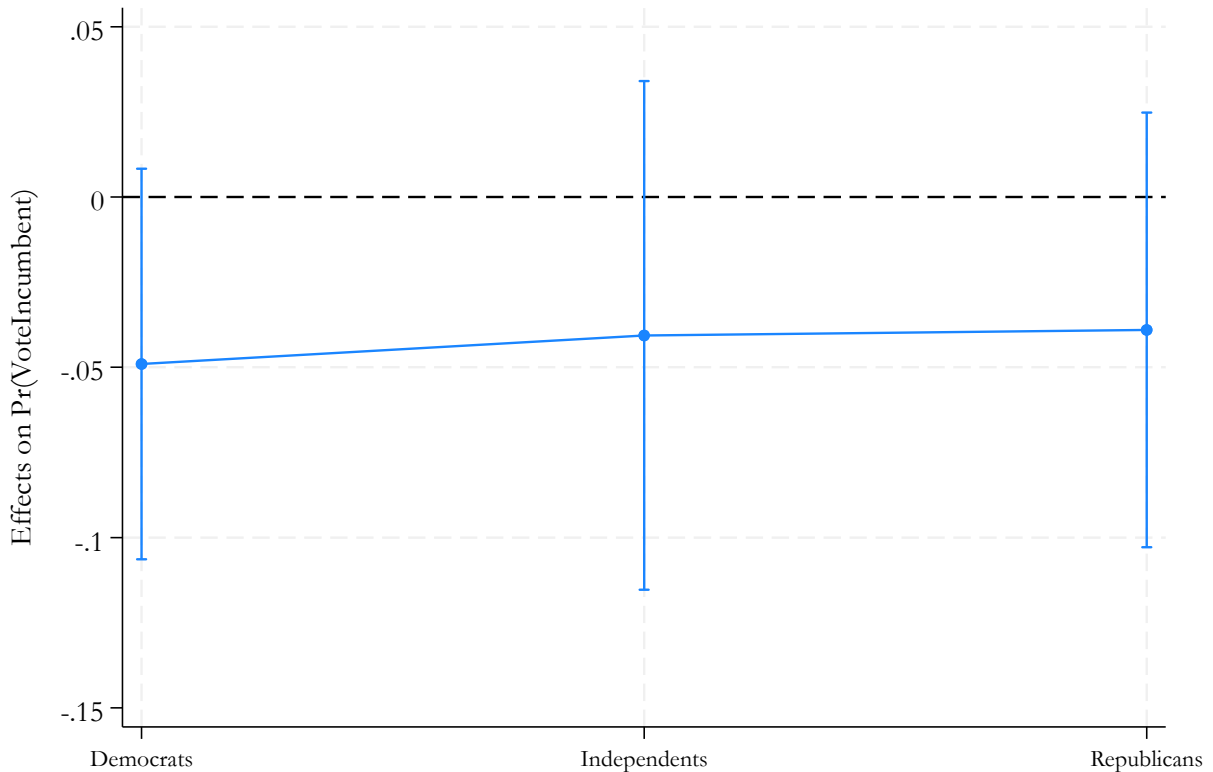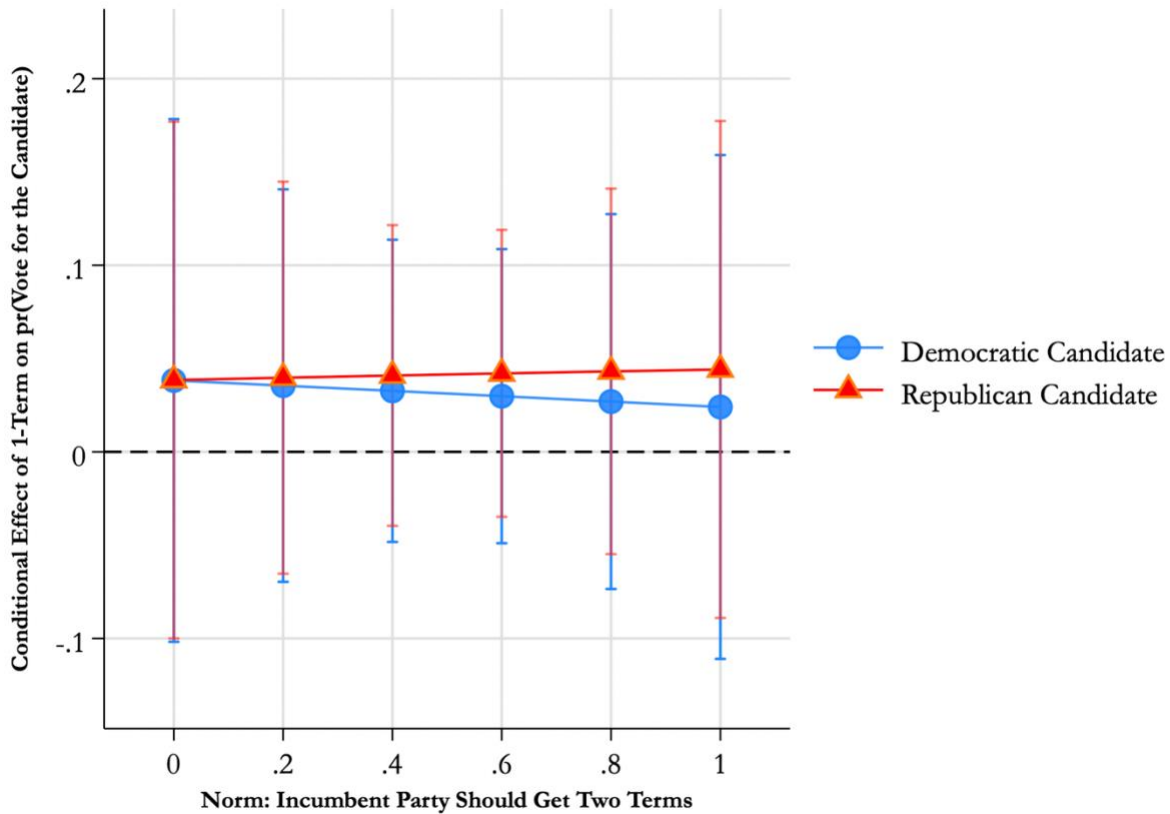

Notes: Models are logistic regression with an interaction specified between treatment assignment and a trichotomous measure of party identification (leaners coded as partisans). 95% CIs shown. Lucid data. N=2,678.

How Much do Treatment Effects Vary by Pre-Treatment Normative Beliefs?

Given the pre-registered analyses, as well as the pre-treatment measure of respondents' normative beliefs regarding incumbency, we can explore the extent to which the effects observed in the experiment vary based respondents' normative beliefs. Logistic models were run—one for the Democratic candidate and one for the Republican candidate—which specified an interaction between treatment assignment (1-term party incumbent versus Control) and the 1T normative belief measure. In Figure D4, we observe that the effect of changing from the Control condition to a 1-term party incumbent (see y-axis) does not vary a great deal depending on respondents' pre-treatment normative beliefs about 1-term incumbents. Regardless of whether the one-term party

FIGURE D4. Moderating Effects of 1-Term Advantage Norm on 1T vs. Control Contrast
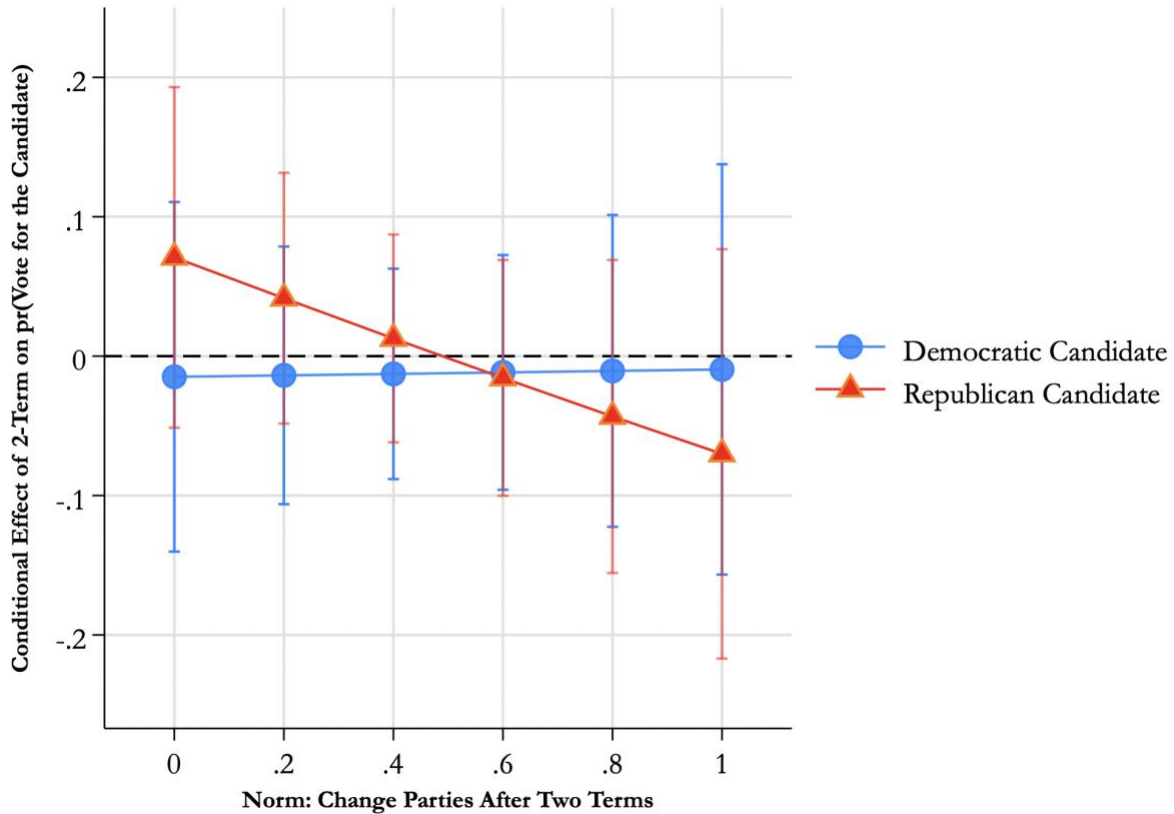

Notes: Models are logistic regression with an interaction specified between treatment assignment and the norm involving whether one-term incumbents should be re-elected. The latter is coded to range from 0 to 1 for interpretive ease, with higher values indicating greater agreement that one-term party candidates should be given a second term. 95% CIs shown. N=657 and 658 for the Democratic and Republican candidate models, respectively.

incumbent was a Democrat or Republican, the slopes are quite flat. Both interactions terms are non-significant (p=.90 for the Democrat and p=.98 for the Republican) and the total estimated change in probability of voting for the candidate is small (about -1.4 percentage-points for the Democratic candidate and 0.6 percentage points for the Republican candidate). As such, this indicates that effects were reasonably similar across levels of normative beliefs.

Next, logistic models were run—again, one for the Democratic candidate and one for the Republican candidate—which specified an interaction between treatment assignment (2-term party incumbent versus Control) and the 2T normative belief measure. The results are presented
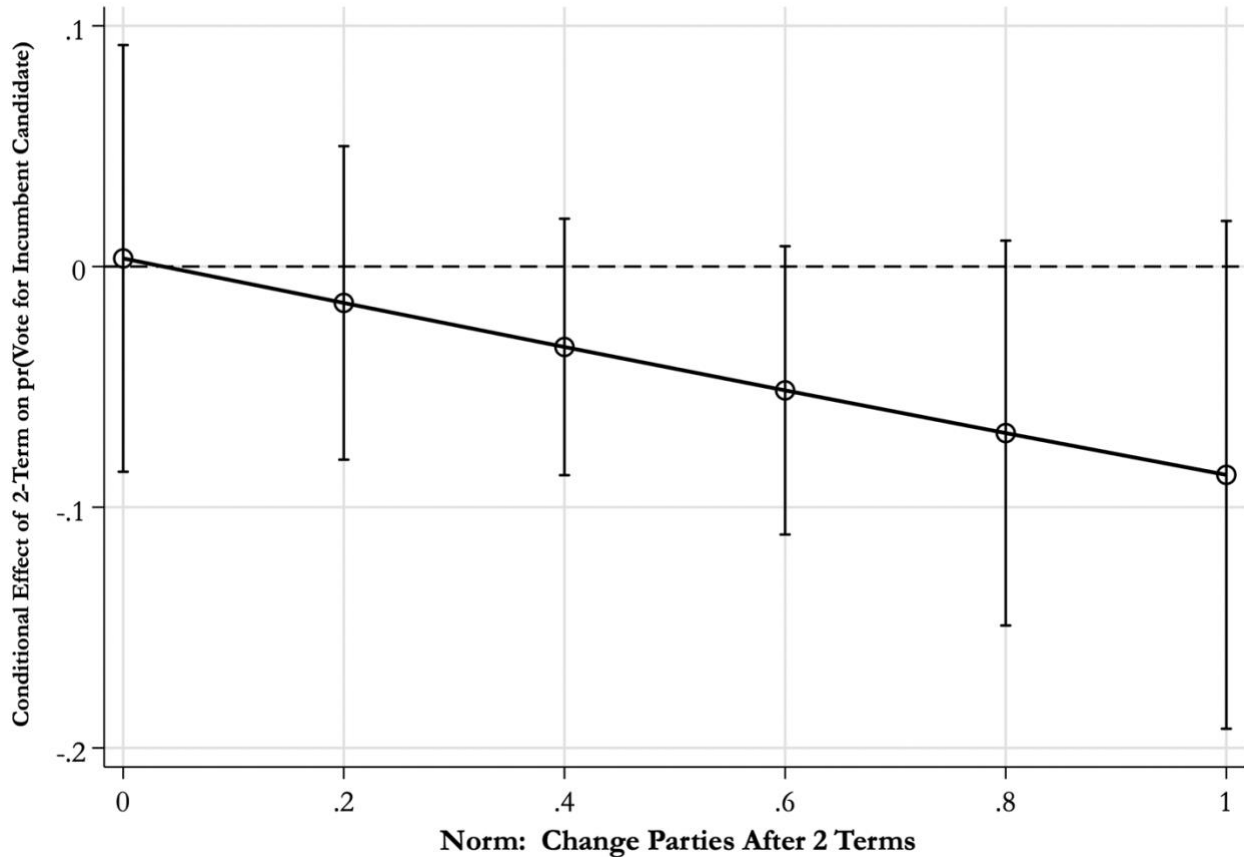
FIGURE D5. Moderating Effects of 2-Term Disadvantage Norm on 2T vs. Control Contrast

Notes: Models are logistic regression with an interaction specified between treatment assignment and the norm involving whether other party should be elected after one party has held office for two terms. The latter is coded to range from 0 to 1 for interpretive ease, with higher values indicating greater agreement that other party should be elected after incumbent party candidate has served two terms. N=676 and 665 for the Democratic and Republican candidate models, respectively.

in Figure D5. First, for the Democratic candidate, the slope is again quite flat, exhibiting a negligible estimated increase in effect size (about a half of a percentage point) from low agreement with the 2T norm to high agreement (the interaction term p-value = .97). For the Republican candidate, however, the estimated change in effect size across the *x*-axis is far more substantial (approximately 14 percentage points). This is consistent with the notion of a two-term disadvantage: after a party's candidate has served two terms in office (versus the number of terms not being specified), their party performs poorest among those who believe that the presidency should change party hands once one party has served two terms. That said, the confidence interval for each point estimate overlaps with zero (the interaction term itself is also non-significant at conventional levels (p=.22)).

FIGURE D6. Moderating Effects of 2-Term Disadvantage Norm on 2T vs. 1T Contrast

Notes: Models are logistic regression with an interaction specified between treatment assignment and the norm involving whether other party should be elected after one party has held office for two terms. The latter is coded to range from 0 to 1 for interpretive ease, with higher values indicating greater agreement that other party should be elected after incumbent party candidate has served two terms. 95% CIs shown. N=1,344.

Finally, we can examine whether the effect of going from a one-term to a two-term incumbent party candidate on vote choice varies depending on the norm regarding two-term candidates (i.e., that the other party should have a chance to take office after one party has already served two terms). Logistic regression model results are shown in Figure D6 (these results average over the party of the incumbent). We observe that a normative belief that the presidency should change party hands after two terms tends to result in a more *negative* effect of going from a one-term party candidate to a two-term party candidate. In other words, respondents who strongly endorse this belief were even less likely to vote for the candidate whose party had already served two terms. At maximum belief in this norm, the effect is approximately an 8.6 percentage-point decrease in the probability of voting for the incumbent party candidate. That said, the 95% confidence intervals again always overlap with 0, and the interaction term itself is non-significant (p=.27).

The overall results, therefore, suggest that the treatments operated fairly similarly regardless of pre-treatment normative attitudes toward incumbency. Although some substantively meaningful differences are observable in Figures D5 and D6, other effects seem quite homogeneous. In

addition, statistically, there is too much uncertainty in the models to confidently reject the null hypothesis of homogeneous effects.

Why are the effects not more heterogeneous depending upon normative beliefs? Two points are worth highlighting. First, because the moderator (i.e., the normative belief) was not randomized, it may be correlated with other omitted factors that are also associated with the outcome. To the extent this is the case, some degree of statistical bias could be obscuring true heterogeneous effects. Second, and more importantly, it is worth stressing that since the manipulation (i.e., number of presidential terms) was isolated, little else could be driving the main effects reported in the manuscript beyond normative beliefs about incumbency in presidential elections. In other words, the narrowness of the manipulation allows for greater confidence that normative beliefs are the key mechanism. Thus, while the lack of a strong interaction here suggests that the effect was similar regardless of ostensible, general beliefs about incumbency, in the real world, wherein numerous additional types of information are present (party cues, economic indicators, candidate quality, etc.), such normative beliefs may play a much larger, more easily detectable role in determining presidential vote choice.

To measure respondent inattentiveness, and confirm attentiveness to the manipulation itself, a factual manipulation check (FMC) was asked after the outcome measure (see Kane and Barabas 2019). Overall, 71.29% of the sample answered the FMC (shown below) correctly. While this constitutes a clear majority, it is nevertheless indicative of a substantial amount of inattentiveness to the experiment.

To confirm that the manipulation was attended to, a chi-squared test was performed. This test examined whether responses to the FMC significantly covaried with treatment assignment. As shown in Table E1, this is clearly the case: within each experimental condition, between 65% and 80% of respondents answered the FMC in a way that indicates they were significantly attentive to the key manipulated details of the experiment (p<.001).

Factual Manipulation Check

Which of the following statements accurately reflects the scenario you just read about?

○ There was no mention of whether either candidate had been president before (1)

○ The Democratic candidate was the incumbent president (2)

○ The Republican candidate was the incumbent president (3)

○ The Democratic Party had held the presidency for the previous 8 years (two terms) (4)

○ The Republican Party had held the presidency for the previous 8 years (two terms) (5)

TABLE E1. Responses to FMC Significantly Covary with Treatment Assignment

Response to FMC	Experimental Condition					
	Control	*1TD*	*1TR*	*2TD*	*2TR*	*Total*
Control	**80.53**	13.41	12.39	16.57	13.78	27.39
1Term Democrat (1TD)	5.35	**70.64**	7.10	7.61	4.30	19.01
1Term Republican (1TR)	4.16	7.00	**72.81**	3.88	7.11	18.83
2 Term Democrat (2TD)	5.35	6.26	4.08	**67.31**	9.63	18.53
2 Term Republican (2TR)	4.61	2.68	3.63	4.63	**65.19**	16.23
Total	100.00	100.00	100.00	100.00	100.00	100.00

Notes: Column percentages are shown. The share answering correctly in each condition appears in bold text. Pearson Chi2 = 5659.59. p<.001. Lucid data. Total n=3,351.

ESTIMATING TREATMENT EFFECTS ACROSS LEVELS OF ATTENTIVENESS

Given the possibility of substantial respondent inattentiveness, respondents were shown—at the start of the survey and prior to the experiment—a mock vignette (shown below) as well as several mock vignette checks (MVCs) afterward (see Kane, Velez and Barabas 2023). The first MVC was used to exclude respondents from continuing the survey. The following two MVCs (shown below) were used to construct an additive scale, with higher scores indicating higher respondent attentiveness.

Per the findings of Kane et al. (2023), the attentiveness scale was interacted with the treatment condition to examine whether treatment effects were stronger among those who (likely) paid greater attention to the experimental vignette and outcome measures. Figure F1, which reexamines the results of Figure 3 in the manuscript, demonstrates this to be the case: those who were attentive show noticeably larger effects, with those who were least attentive (i.e., answered the two MVCs above incorrectly) showing near-zero effects. Specifically, the results from the manuscript's left panel of Figure 2 are re-estimated to be 12.7 and 9.1 percentage points for the Democratic and Republican candidates, respectively ($p<.05$ in both cases). The treatment effect in the right panel of Figure 3 is re-estimated to be 7.1 percentage points ($p<.05$).

Figure F2 includes the analyses featured in Figure 4 of the manuscript, with comparisons between 2T candidates and the Control condition. Though there is a tendency for the more attentive to exhibit slightly more positive effects (see dotted lines), averaging across the Republican and Democratic candidates, these changes in estimated effects are relatively flat compared to the changes observed in Figure F1 and not statistically distinguishable from zero.

Mock Vignette

A Passage from a Recent Magazine Article:
More than 125 scientific societies and journal publishers are urgently warning lawmakers not to move forward with a rumored policy that would make all research supported by federal funding immediately free to the public. In three separate letters, they argue such a move would be costly, could bankrupt many scientific societies that rely on income from journal subscriptions, and would harm the scientific enterprise. Lawmakers won't comment on whether they are considering a policy that would change publishing rules, and society officials say they have learned no details. But if the rumor is accurate, the order would represent a major change from current U.S. policy, which allows publishers to withhold federally-funded research from the general public for up to 1 year.

Mock Vignette Checks

Regarding the rumored change in policy that was discussed, the magazine excerpt indicated that:

○ Lawmakers won't comment on whether they are considering it (1)

○ Legal scholars stated the change in policy would be challenged in courts (2)

○ Journal publishers have already begun preparing for the change in policy (3)

○ Scientific researchers are divided in terms of their support for the policy (4)

○ All of the above (5)

○ None of the above (6)

According to the magazine excerpt you just read, current policy allows federally-funded research to be withheld from the general public for up to:

○ 1 month (1)

○ 6 months (2)

○ 1 year (3)

○ 3 years (4)

○ 5 years (5)

○ None of the above (6)

FIGURE F1. Larger Effects Among the More Attentive (see MS Figure 3)
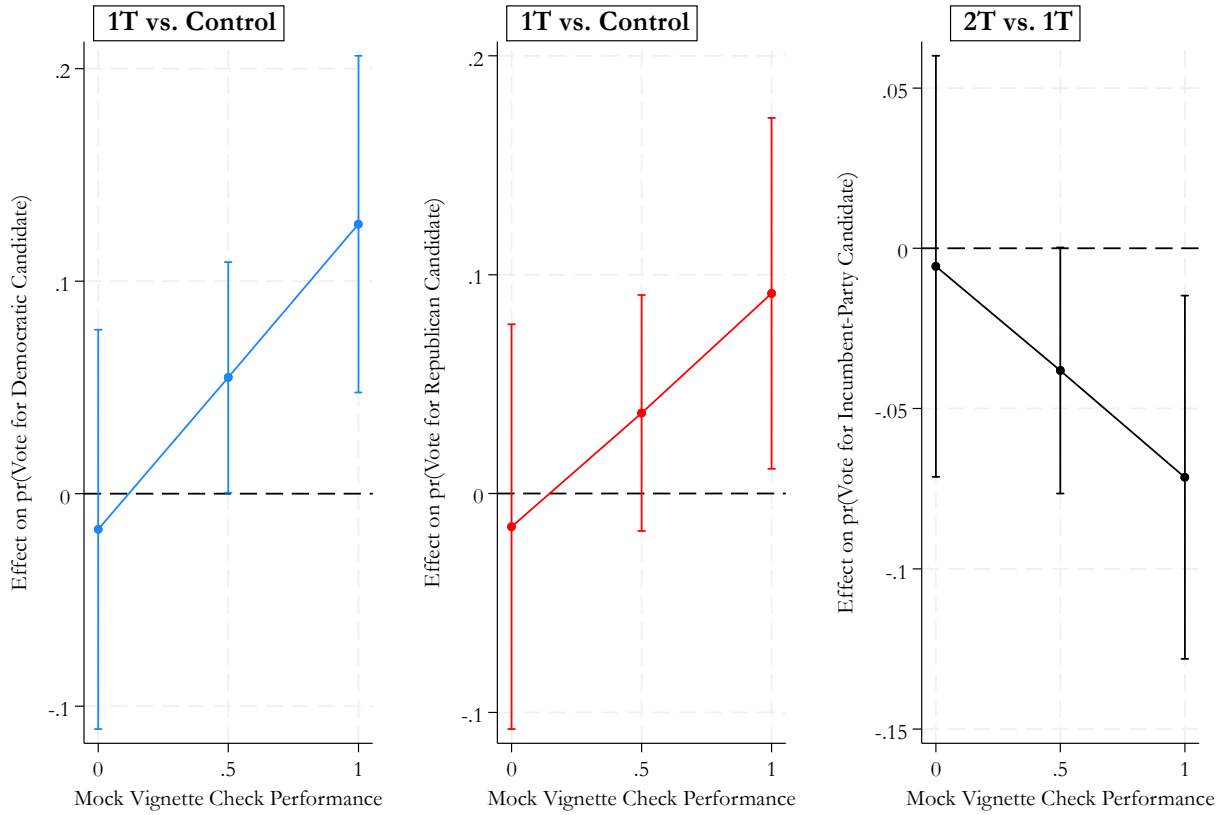
Notes: Models are logistic regression with an interaction specified between treatment assignment and the additive attentiveness scale. Vertical axis displays effect of a specific treatment (see title of each panel) on predicted probability of voting for a particular incumbent-party candidate (Democratic or Republican). 95% CIs shown. Lucid data. Total n=3,351.

FIGURE F2. No Evidence of a Two-Term Disadvantage Across Levels of Attentiveness
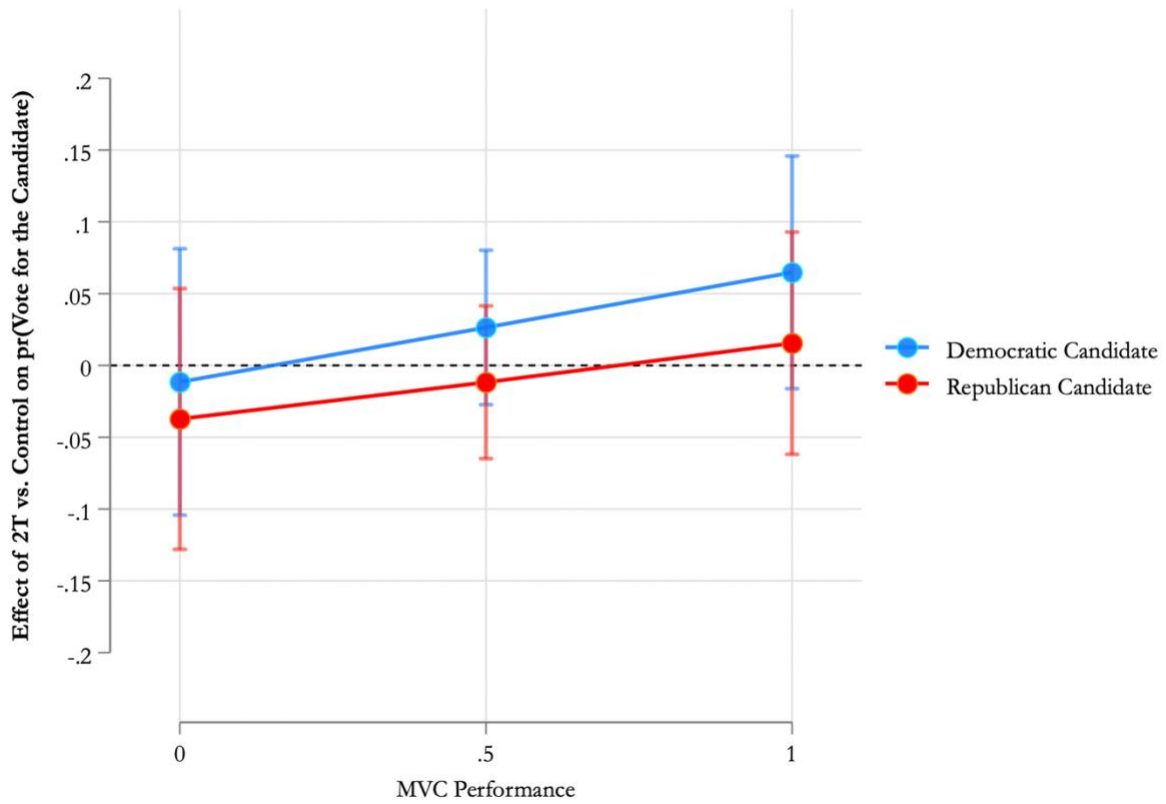
Notes: Models are logistic regression with an interaction specified between treatment assignment and the additive attentiveness scale. Vertical axis displays effect of two-term treatment (vs. Control) on predicted probability of voting for a particular incumbent-party candidate (Democratic or Republican). 95% CIs shown. Lucid data. Total n=3,351.

The manuscript features analyses using Voter Study Group panel data. The dataset is the "September-November 2020 Voter Full Data Set", which features survey data on the same set of respondents between 2011 (the baseline year) and 2020 (respondents in 2011 were asked about their vote choice in 2008). The data are available at: https://www.voterstudygroup.org/data. Table G1 describes each variable used in the analysis, as well as its corresponding survey item in the VSG data set and details regarding any recoding of the original variable. All continuous variables were recoded to range between 0 and 1 for the analyses featured in the manuscript. Table G2 features the regression output underlying Figure 5 in the manuscript.

TABLE G1. VSG Survey Items and Coding Details

Description	Survey Item	Coding Details
Vote choice in 2008	*presvote08_2011*	Coded as voting for Obama (1) or not (0). Non-voters excluded.
Vote choice in 2012	*presvote_2012*	Coded as voting for Obama (1) or not (0). Non-voters excluded.
Vote choice in 2016	*presvote_2016*	Coded as voting for Trump (1) or not (0). Non-voters excluded.
Vote choice in 2020	*presvote_2020Nov*	Coded as voting for Trump (1) or not (0). Non-voters excluded.
Respondent's party (baseline year)	*pid7_2011*	Leaners coded as partisans. "Not sure" coded as independents. "Skipped" and "Not asked" excluded.
Political interest (baseline year)	*polinterest_2011*	Reverse coded (so that higher values indicate greater interest). "Not sure" "Skipped" and "Not asked" excluded.
Gender (baseline year)	*gender_2011*	Males coded as 0, females coded as 1. "Skipped" and "Not asked" excluded.
Racial identification (baseline year)	*race_2011*	Because they constituted less than a combined 5% of the total sample, Asians, Native Americans, "Mixed", "Other" and Middle Eastern were coded as "Other". "Skipped" and "Not Asked" excluded.
Age (baseline year)	*birthyr_2011*	Subtracted birth year from 2011. "Skipped" and "Not asked" excluded.
Family income (baseline year)	*faminc_2011*	The value of 31 ("150,000 or more") was changed to 12 to create a continuous variable. "Prefer not to say", "Skipped" and "Not asked" excluded.
Highest level of education completed	*educ_2011*	"Skipped" and "Not asked" excluded.

TABLE G2. Regression Output Underlying Manuscript Figure 4

	Vote for 1-Term Incumbents	Vote Against 2-Term Incumbents
Democrat	0.01	-0.97***
	(0.32)	(0.22)
Republican	-1.02**	-1.32***
	(0.38)	(0.24)
Political Interest	-0.84*	-0.63*
	(0.35)	(0.30)
Female	0.56*	-0.08
	(0.25)	(0.19)
Black	-1.11^	-1.03*
	(0.60)	(0.52)
Hispanic	0.37	0.26
	(0.44)	(0.36)
Other Race	0.14	-0.30
	(0.45)	(0.41)
Age	-0.81	1.23*
	(0.70)	(0.58)
Family Income	-0.41	-0.31
	(0.46)	(0.36)
Educational Attainment	-1.56**	-0.87*
	(0.48)	(0.36)
Constant	-1.43**	-1.36**
	(0.55)	(0.44)
Observations	2,527	2,527

Notes: *These are the results underlying manuscript Figure 4. The dependent variables are as follows: "Vote For 1-Term Incumbents" = voting for Obama in 2012 and Trump in 2020; "Vote Against 2-Term Incumbents" = voting for Obama in 2008 and for Trump in 2016. Logistic regression model coefficients shown with standard errors in parentheses. *** $p<0.001$, ** $p<0.01$, * $p<0.05$, ^ $p<0.10$ (two-tailed hypothesis tests).*

APPENDIX H:
OSF PRE-REGISTRATION STUDY INFORMATION

Study Information
Hypotheses

H1: A one-term incumbent presidential candidate will attract more votes than a non-incumbent candidate. H2: A presidential candidate whose party has held the presidency for two terms will attract fewer votes than a candidate whose party has not held the presidency for the previous two terms.

Design Plan
Study type

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials. Blinding

- For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Is there any additional blinding in this study?

No response
Study design

Experiment involves exposure to one of five conditions featuring experimental vignettes. The vignettes purport to feature details about a presidential election from the past, but no candidates are named. What is manipulated between conditions is how many terms a particular candidate (Democrat or Republican) has been elected (no terms in office, one term in office, or whose party has held the presidency for the previous two terms). Respondents are then asked which candidate they would vote for (options include abstention).

No files selected
Randomization

Randomization occurs within the survey programming. All randomization is at the individual-level.

Sampling Plan
Existing Data

Registration prior to creation of data
Explanation of existing data

No response
Data collection procedures

Adult U.S. population, nationally representative on age, race/ethnicity and geographic region.

No files selected

Sample size

Approximately 3,333.
Sample size rationale

Arbitrary constraint (maximum size available given existing funds).
Stopping rule

No response

Variables
Manipulated variables

Manipulated variables are the incumbency status of each candidate (as well as the party of the candidate).

No files selected
Measured variables

Predictors: gender Do you identify as: o Woman (1) o Man (2) o Transgender (3) o Non-binary / non-conforming (4) pid7 Generally speaking, do you consider yourself to be a(n): o Strong Democrat (1) o Democrat (2) o Independent, but Leaning Democratic (3) o Independent (4) o Independent, but Leaning Republican (5) o Republican (6) o Strong Republican (7) ideo Thinking about politics these days, how would you describe your own political viewpoint? o Very Liberal (1) o Liberal (2) o Moderate (3) o Conservative (4) o Very Conservative (5) o Not Sure (6) religion What is your religious identification? o Jewish (1) o Christian-Protestant (2) o Christian-Catholic (3) o Other Christian Denomination (4) o Muslims (5) o Agnostic / Atheist (6) o Hindu (7) o Buddhism (8) o Other Non-Christian Denomination (9) race Would you describe yourself as: o Non-Hispanic White (1) o Non-Hispanic Black (2) o Hispanic (3) o Asian (4) o American Indian / Alaskan Native (5) o Other Race (6) Regardless of their political party, as a general rule, I feel that presidents should be re-elected so that they can serve a second term in office. o Strongly Disagree (1) o Disagree (2) o Slightly Disagree (3) o Slightly Agree (4) o Agree (5) o Strongly Agree (6) Regardless of their political party, as a general rule, I feel that once a political party has had control of the presidency for two full (4-year) terms, the other party should get a chance at being president. o Strongly Disagree (1) o Disagree (2) o Slightly Disagree (3) o Slightly Agree (4) o Agree (5) o Strongly Agree (6) Outcome: Within each experimental condition, the outcome is measured using the following item: Had you been a voter during the time of this election, how do you think you would have voted? o I would have voted for the Democratic candidate (1) o I would have voted for the Republican candidate (2) o I would have voted, but not for either the Democratic candidate nor Republican candidate (3) o I would have abstained (not voted at all) (4) As the interest is in voting for the incumbent, this variable will be recoded to reflect selecting the incumbent (1) vs. not the incumbent (0).

No files selected
Indices

N/A

No files selected

Analysis Plan
Statistical models

Regression analyses will be performed to test main hypotheses, conduct robustness and exploratory analyses. The first set of analyses will compare, for each partisan group, the effect of each treatment (1-term and 2-term) condition against the control condition in predicting a vote for the incumbent. A second set of analyses will combine partisans together to compare voting for the incumbent in the a) 1-term condition, vs. the b) 2-term condition. Robustness tests will test whether pre-treatment attitudes about 1-term incumbents and 2-term incumbents (measured for a random 50% of the sample) significantly moderate the effect of each treatment on the choice to vote for the incumbent. Exploratory analyses will examine demographic and political predictors of the pre-treatment incumbency attitudinal measures. I will also investigate the possibility of attenuated effects due to inattentiveness. This will be done by interacting the treatment with a pre-treatment measure of inattentiveness.

No files selected
Transformations

No response
Inference criteria

Given the one-tailed hypotheses, one-tailed p-values will be reported. $p<=.05$ will be the cutoff for statistical significance; $p<=.10$ will be treated as marginally statistically significant.
Data exclusion

No data will be excluded. However, pre-treatment "Mock Vignette Checks" will be used to gauge inattentiveness.
Missing data

Listwise deletion (though from previous experience, missing data is minimal).
Exploratory analysis

Exploratory analyses will examine: --demographic and political predictors of the pre-treatment incumbency attitudinal measures. --whether treatment effects are substantially stronger when interacting pre-treatment attitudes regarding incumbency with the treatment --whether effects substantially change when interacting a pretreatment measure of attentiveness with the treatment

Other
Other

No response

This section supplements the analyses presented in the manuscript using all presidential elections in the post-Civil War era (1868-2020). Beyond providing more data points to analyze, the pre-1952 period features, relative to the post-1952 period, many more instances of parties occupying the presidency for three or more terms.

Using publicly available data from The American Presidency Project[2], I coded each election since 1868 as whether the election featured a candidate whose party had (at the time of the election) been occupying the presidency for one term, two terms, three terms, etc. In total, 16 elections featured a candidate whose party had served one term, 12 elections featured a candidate whose party had served two consecutive terms, and 11 elections featured a candidate whose party had served three or more consecutive terms.[3] I then calculated the two-party (Democratic and Republican) vote share for each candidate. Figure I1 below displays the two-party vote share of each candidate in each election. (Note that, because it is the two-party vote share, the vote share (s) for one party is equal to 1-s for the opposing party, in each election. In other words, the parties' respective vote shares will be symmetrical around the vertical .5 line.)

---

[2] The data are publicly available at the following link: https://www.presidency.ucsb.edu/statistics/elections
[3] Of these 11 elections, 5 elections had a candidate whose party served three consecutive terms, 3 elections had a candidate whose party served four consecutive terms, 2 elections had a candidate whose party served five consecutive terms, and 1 election (1884) featured candidate whose party had served six consecutive terms.

FIGURE I1. Two-Party Vote Share in Post-Civil War Presidential Elections
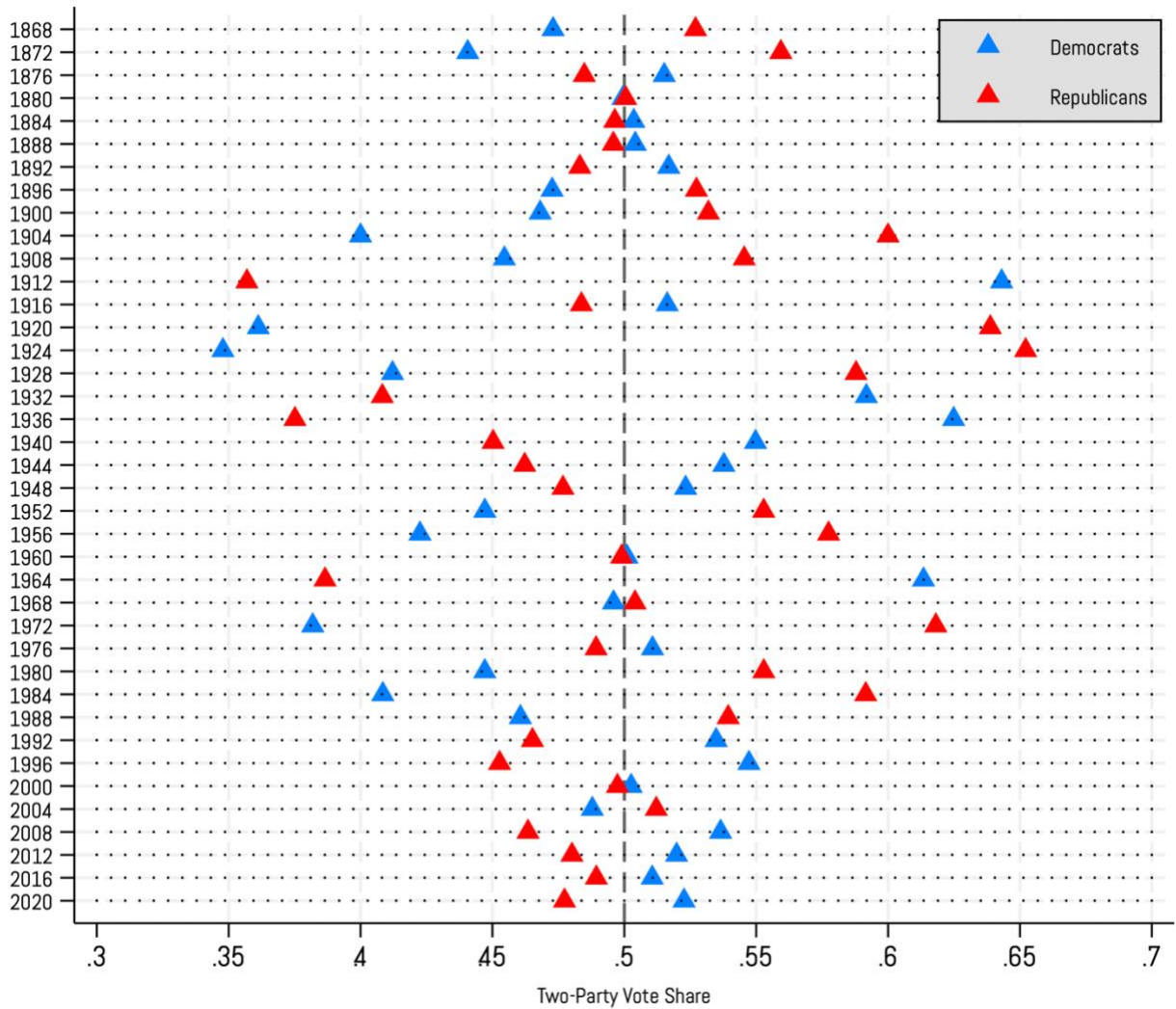
Note: Data are from The American Presidency Project.

"Cost of Ruling" in the U.S.

The results featured above and in Figure 1 of the manuscript are consistent with the wealth of evidence that finds a cost of ruling effect in democracies around the world (Nannestad and Paldam 2002; Wlezien 2017). That is, the longer that a party holds office for consecutive terms, the worse it tends to do in terms of vote share. Figure 1 certainly suggests that this phenomenon is occurring in the U.S., which is consistent with Wlezien (2017). Nevertheless, there are several points worth noting.

First, as reported in the manuscript, incumbent-party candidates do *not* tend to do any worse, in terms of two-party vote share, in their first reelection attempt compared to when they were first elected. For example, in Wilson's (1916), FDR's (1936), Eisenhower's (1956), Reagan's (1984), Clinton's (1996), and Bush's (2004) reelection, the incumbent's two-party vote share was *higher* than it was in the initial election. Overall, in only 31% of applicable cases do we observe a decrease in two-party vote share between the initial election and first reelection attempt. This adds some nuance to the "cost of ruling" effect in the US, perhaps suggesting that a one-term incumbent-party

26

advantage (1TA) safeguards a party's electoral chances after one term (e.g., because of a normative belief in the public that presidents should, *ceteris paribus*, be afforded two terms). After that, however, "cost of ruling" begins to take effect, resulting in a steady decline in vote share as parties occupy the presidency for two, three, four, etc. full terms.

Second, a methodological complication worth noting is that the final value in a party's number of consecutive terms must be a losing election (i.e., <50% of vote share); otherwise it would not have been the party's last consecutive term. Thus, at least to some extent, the vote share for candidates whose party has served an unusually large number of consecutive terms (e.g., five terms) will likely tend to be lower than 50%.

A third point worth stressing is that the two-term vote share, particularly in the post-1952 period, is not well below 50%. As noted in the manuscript, while the vote share is below what one-term incumbent-party candidates receive, it is not evidence of the absolute disadvantage (against the opposing candidate) implied by the notion that voters—*on average*—become disillusioned with a party after eight years, nor that the party has pushed policies too far from the median voter (e.g., see Nannestad and Paldam 2002). Thus, while such elections may represent a decline in vote share for the ruling party compared to the preceding two elections (per the cost of ruling effect), such elections are nevertheless, in expectation, virtually equivalent to a coin flip.

To summarize, the poorer performance of candidates whose party has served two terms (relative to one) and three or more terms (relative to one or two) is indeed consistent with the "cost of ruling" effect. Yet, the (on average) *improvement* in vote share of one-term candidates (compared to their initial election), and roughly *even* chances of winning after two terms, potentially hint at a slightly weaker cost of ruling effect in the U.S. relative to elsewhere in the world. Aside from the existence of a normative belief that presidents should serve two terms (which can help explain the 1TA over opponents, and the increase in vote share from the preceding election), what other factors might constrain the "cost of ruling" in the U.S.?

Though this question is beyond the scope of the present study, there are several reasons future research might consider. First, given the high levels of group-based sorting into the two parties (e.g., Kane, Mason, and Wronski 2021), and high levels of inter-party animosity and partisan polarization (Druckman and Levendusky 2019; Mason 2018), this likely serves to limit the extent to which much of the public can be persuaded vote for the opposing party. For example, disparate issue-based groups in the U.S. may find it relatively more difficult to form viable coalitions against the party-in-power. That is, groups alienated by the party-in-power may be less likely to find support among groups that normally support the opposition party. Thus, with a public highly polarized along partisan lines, it may be more difficult for the president to lose votes to the opposing party which, in effect, could limit the magnitude of the cost of ruling effect.

Compounding this possibility, the high levels of polarization likely place a limit on the vote share that *can possibly be won* in the first term. This lower ceiling in terms of potential vote share would naturally mean that the decrease in vote share over successive terms will tend to be smaller (because it already begins at a lower level).

Along similar lines, a relatively stable economy in the U.S. (compared to other democracies) should also lead to a weaker cost of ruling effect: more economic volatility should increase the likelihood that voters will want to change the party in power. Hence, less volatility should weaken this impulse and assist parties with remaining in power longer (see Nannestad and Paldam 2002).

Finally, the existence of term limits (per the 22nd Amendment) may affect how much of a cost of ruling effect can be expected within the U.S. Though a president may perform poorly over eight years, and/or steer policy far away from that of the median voter's preferences, the party then

introduces a new candidate. Thus, while much of the public may grow jaded with the sitting president (per the cost of ruling effect) over the course of eight years, they are not asked to vote for this president again. They are asked to vote for a *new* candidate. Of course, whether voters blame the party's new candidate for the party's sitting president's actions is by no means certain, though this is undoubtedly a campaigning strategy that can be employed by the opposition party.

REFERENCES

Druckman, James N., and Matthew S. Levendusky. 2019. "What Do We Measure When We Measure Affective Polarization?" *Public Opinion Quarterly* 83 (1): 114–22.

Kane, John V., Lilliana Mason, and Julie Wronski. 2021. "Who's at the Party? Group Sentiments, Knowledge, and Partisan Identity." *The Journal of Politics* 83 (4): 1783–99.

Mason, Lilliana. 2018. *Uncivil Agreement: How Politics Became Our Identity*. Chicago, IL: University of Chicago Press.

Nannestad, Peter, and Martin Paldam. 2002. "The Cost of Ruling: A Foundation Stone for Two Theories." In *Economic Voting*, edited by Han Dorussen and Michael Taylor, 1st ed., 17–44. London: Routledge.

Wlezien, Christopher. 2017. "Policy (Mis) Representation and the Cost of Ruling: U.S. Presidential Elections in Comparative Perspective." *Comparative Political Studies* 50 (6): 711–38.