



Giovannoni, F., & Feltovich, N. (2022). Campaign messages, polling, and elections: theory and experimental evidence. *American Journal of Political Science*. Advance online publication.  
<https://doi.org/10.1111/ajps.12722>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY-NC

Link to published version (if available):  
[10.1111/ajps.12722](https://doi.org/10.1111/ajps.12722)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wiley at <https://doi.org/10.1111/ajps.12722>. Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Campaign Messages, Polling, and Elections: Theory and Experimental Evidence

**Nick Feltovich** Monash University  
**Francesco Giovannoni** University of Bristol

**Abstract:** *How do politicians' track records and campaign messaging interact and affect voters' welfare? We analyze this question theoretically and experimentally. In the theoretical model, which we implement in the experiment, politicians choose how much of an economy's resources to allocate to the citizenry—keeping the remainder for themselves—and then face reelection against a challenger. Both incumbents and challengers have private information about their own quality that determines the economy's level of resources. We vary whether candidates can send campaign messages and the level of variability in candidates' quality. We observe that both higher-quality variability and allowing campaigning benefit citizens by allowing them to better select and hold accountable higher-quality officials. Also, when incumbents have performed poorly or when quality variability is high, challengers' negative campaigning (criticizing the incumbent) increases and incumbents' positive campaigning (emphasizing their own strengths) decreases.*

**Verification Materials:** The data required to verify the computational reproducibility of the results, procedures, and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: <https://doi.org/10.7910/DVN/DAYSXC>.

In modern democracies, policymaking is in the hands of politicians chosen by voters through elections. Elections allow citizens to select their representatives and hold them accountable. To do so, voters rely on two main sources of information: candidates' track records and their campaign statements. The “Downsian” branch of the elections literature (Downs 1957; Lindbeck and Weibull 1987) has concentrated on the latter—before elections, candidates make binding promises and voters choose the candidate whose policies they prefer. Consequently, elections are about *selection* and voters receive their information through credible communication from candidates. Another branch (Barro 1973; Ferejohn 1986; Fox and Shotts 2009; Przeworski, Stokes, and Manin 1999) takes the opposite tack, viewing elections as a process where voters evaluate the incumbent's performance after policies have been chosen. There, elections are about *accountability*; voters rely on past

information to evaluate candidates because there is no campaigning.

Neither approach is entirely satisfying. The former assumes that campaign promises are binding and allow no role for accountability, whereas the latter assumes campaigns are irrelevant cheap talk. Later work attempts to close this gap. One solution is to introduce selection into the literature on accountability. Fiorina (1981) points out that, with rational voters, past performance should only matter to the extent that it allows voters to select higher-quality candidates (a point taken up by, e.g., Ashworth and Bueno de Mesquita 2008; Gordon, Huber, and Landa 2007). Ashworth, Bueno de Mesquita, and Friedenber (2017) demonstrate that accountability and selection can either hinder or reinforce each other, depending on whether candidates' quality and actions are substitutes or complements. Experimental studies (Konrad and Sherif 2019; Landa 2010; Woon 2012) tend to show that voters have more concerns

Nick Feltovich is Professor of Economics, Monash University, Clayton VIC 3800, Australia ([nicholas.feltovich@monash.edu](mailto:nicholas.feltovich@monash.edu)). Francesco Giovannoni is Professor of Economics, University of Bristol, Priory Road Complex, Priory Road, Bristol BS8 1TU, UK ([francesco.giovannoni@bristol.ac.uk](mailto:francesco.giovannoni@bristol.ac.uk)).

Financial support from Monash University and from the Australian Research Council (DP130101183) is gratefully acknowledged. We thank Jon Eguia, Christos Ioannou, Dimitri Landa, Santiago Sánchez-Pagés, three anonymous referees, and participants at several conferences and seminars for helpful suggestions and comments.

*American Journal of Political Science*, Vol. 00, No. 0, May 2022, Pp. 1–19

© 2022 The Authors. *American Journal of Political Science* published by Wiley Periodicals LLC on behalf of Midwest Political Science Association. DOI: 10.1111/ajps.12722

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

for accountability—over selection—than theory would predict.

However, for all its progress in introducing selection into an accountability framework, this work does not allow for campaigning. In contrast, empirical work on the question of selection versus accountability has long recognized that much of the information needed for selection comes from political communication. Indeed, observational studies have consistently found evidence that voters respond to political campaigns (for a comparative study, see Matthies 2020), although the evidence in favor of a greater role for either selection or accountability is mixed (Clarke and Stewart 1994; Erikson, MacKuen, and Stimson 2000; Healy, Persson, and Snowberg 2017). Recent work relying on natural experiments has not resolved the issue: Elinder, Jordahl, and Poutvaara (2015) find that voters tend to focus more on selection rather than past performance, whereas Avdeenko (2018) finds the opposite.

The Downsian tradition has tried to close the gap by addressing campaign credibility. Snyder and Ting (2002) argue that political parties make promises credible because candidates need to comply with their platforms in order not to incur the cost of losing party affiliation; Callander and Wilkie (2007) establish that when politicians differ in their ability and willingness to lie, more honest candidates face an electoral disadvantage but not an overwhelming one; and Schnakenberg (2016) shows that promises can be credible when the relevant information is multidimensional. With few exceptions (e.g., Aragones, Palfrey, and Postlewaite 2007), however, this literature has not viewed accountability to voters as a major reason for this perceived credibility. Yet, the empirical evidence shows that politicians sometimes keep their promises, and how much they do so depends on how accountable they are (Mansergh and Thomson 2007; Pétry and Collette 2009; Naurin, Royed, and Thomson 2019; Royed 1996).

Putting all of this together, the question remains as to how past performance and political campaigns interact, allowing voters to select representatives and hold them accountable. In particular, how does access to more information, either from track records or campaigns, affect voters' welfare? How do track records and campaigns interact to produce these outcomes? What campaign strategies generate them? We tackle these questions by introducing a theoretical model of repeated elections, where voters get information from track records and campaign messages, and where candidates may be of high or low quality. As in Harrington (1992), Aragones, Palfrey, and Postlewaite (2007), and Feltoovich and Giovannoni (2015), voters can punish politicians who

renege on past promises, or choose those who promise more, but (a) by having candidates with privately known quality, we can vary how much information is known about them, and (b) by allowing for free-form messages, we can ascertain whether candidates focus on the past or the future, and on themselves or their opponent.

We focus on two variations within the model. The first involves changing the variance in the quality distribution: higher variance should make it easier for voters to figure out the incumbent's quality. The second involves studying the model with and without campaign messages. We replicate our model in a lab experiment and test its implications. The lab allows us to study a stylized setting, like that of our model, which is focused on the informational channels of interest without confounding factors. This is difficult to accomplish with observational data or even field experiments because with those techniques, it is impossible or unethical to obtain the necessary degree of control over the relevant features of the environment.

We observe, compatibly with our theoretical results, that increased quality variance raises voters' welfare. The reason is that voters can more easily learn who the higher-quality candidates are, and once identified as such, can hold them accountable for providing good outcomes. This result is consistent with the general finding in the literature that information improves officials' responsiveness to voter concerns (see Ashworth 2012 for a discussion), although our setting allows us to identify a specific mechanism wherein information allows for more selection, which is then sustained by accountability.

More originally, we observe that campaigning can achieve much the same effect, even when quality variance is low so that voters cannot easily determine quality from politicians' track records. Voters understand that only high-quality candidates are able to credibly make ambitious campaign promises, and they select on the basis of these. Once high-quality officials are identified in this way and elected, they can be held accountable for continuing to deliver. Of course, this relies on campaign promises being credible, in that candidates only make promises they are willing to keep and voters understand this. Corazzini et al. (2014) show that campaign promises are credible even when candidates do not face reelection, and similar evidence exists in observational data (e.g., Naurin 2011). The fact that campaigns improve voters' welfare suggests that they are credible in our setting, but as further confirmation, we find that candidates tend to campaign more when the election is competitive (Erikson and Palfrey 1998 find similar results).

The use of free-form messages in our experiment allows us to investigate different messaging strategies. A

particularly important issue that has engaged the literature is whether candidates choose to emphasize themselves (positive campaigning) or their opponent (negative campaigning). Skaperdas and Grofman (1995) argue that candidates who are ahead in the polls are less likely to go negative, whereas Li and Li (2013) and Polborn and Yi (2006) argue that the use of positive or negative campaigning will depend on the relative standing the candidates perceive they have, so that negative campaigning will be used when a candidate perceives it is better to emphasize the opponent's negatives than her own positives. Empirically, the evidence is mixed. Theilmann and Wilhite (1998) find that negative campaigning is generally chosen by the candidate who is trailing, but they stress that this is particularly likely when the race is a "blowout" because negative campaigns can alienate voters. Conversely, Hale, Fox, and Farmer (1996) and Kahn and Kenney (1999) find that negativity is more likely in close elections.<sup>1</sup> Our analysis shows that challengers use negative campaigning when the incumbent has performed poorly, in order to emphasize these failures. When incumbents perform relatively well, challengers tend to campaign positively, to persuade voters that they can do even better.

Finally, a related question concerns the relative effectiveness of positive and negative campaigning. Lau and Rovner (2009) survey the literature and find relatively little evidence of the effectiveness of negative campaigning compared to positive campaigning. We find significant evidence of the effectiveness of positive campaigning, but weaker evidence for negative campaigning. However, voting in our environment is compulsory, whereas much of the empirical literature argues that an important rationale for negative campaigning is to depress turnout by the opponent's supporters (e.g., Ansolabehere et al. 1994; Finkel and Geer 1998; Freedman and Goldstein 1999), so we cannot give a definitive answer on this point.

## Theory and Hypotheses

There are infinitely many discrete rounds, where an official (chosen randomly in the first round, and by elections thenceforth) and  $n - 1 \geq 4$  ordinary citizens interact. The official's *quality*  $\theta_t$  (where  $\theta_t \in \{\theta^l, \theta^h\}$  and  $\theta^h > \theta^l > 0$ ) is her private information—citizens only know that both types are equally likely—and remains

<sup>1</sup>In our setting, the only cost of sending messages is the effort required to type them, so we cannot speak to the question of whether negative campaigning is more likely when candidates have less money than their opponents (Lau and Pomper 2002).

constant throughout her tenure as official. In round  $t$ , the economy receives an endowment  $n\theta_t$ , and the official chooses how much *income*  $\pi_t^c = x_t \in [0, \frac{n}{n-1}\theta_t]$  to allocate to each citizen from the endowment, keeping the remainder  $\pi_t^o = n\theta_t - (n-1)x_t$  as her *salary*.

After citizens are informed of their incomes (but not the endowment or the official's salary), a challenger is nominated, equally likely to be any of the  $n - 1$  citizens, and with quality equally likely to be  $\theta^l$  or  $\theta^h$ . The challenger's quality is private information, and no other information about the challenger is disseminated to the other voters. It is costless to run for office.

Next, a *straw poll* is run. Everyone, including candidates, simultaneously and costlessly votes, and then the results are announced. The straw poll allows us to identify the treatment effects of campaigning by comparing how voting changes between the straw poll and the election when campaigning is possible versus when it is not. We emphasize that our objective is to examine candidates' and voters' behavior in an environment where polling is present, not whether this presence changes behavior; hence, we do not vary whether there is a straw poll. Also, our implementation of polling abstracts from reality in two ways. First, everyone is polled, rather than a sample of the electorate. We do this because it is a common technique (e.g., Agranov et al. 2018) and because our electorates are small compared to their real-world counterparts, so that polling a smaller subset would result in wide margins of error. Additionally, unlike real-world polling, our polling is costless, removing the main reason for using sampling. Second, we do not allow *abstention* in polling and the election. This simplifies the setting, though we acknowledge that most elections involve voluntary rather than compulsory voting. Bhattacharya, Duffy, and Kim (2014) find that while compulsory and voluntary voting differ in how voters behave strategically (insincere voting vs. abstention), the associated electoral outcomes are similar, suggesting that allowing abstention would not materially affect our results.

In the game *without* campaign promises, the election takes place immediately after the straw poll. Again, everyone simultaneously and costlessly votes. The candidate with more votes wins and will be the official in the next round (with quality unchanged from the current round); in case of a tie, the incumbent remains in office. The election loser goes back to being an ordinary citizen.<sup>2</sup> The game *with* campaign promises allows both incumbent and challenger to simultaneously broadcast a natural-language message *after* the straw poll results are

<sup>2</sup>If, in a future round, the loser is again selected as a challenger, a new quality is randomly assigned.

announced, but *before* the election. Messages incur no costs other than the effort of composing them and do not formally bind future decisions.

The stage game (Figure 1) is infinitely repeated. All players have the same lifetime utility function:

$$U(\pi_1, \pi_2, \dots) = \sum_{t=1}^{\infty} \delta^{t-1} u(\pi_t), \quad (1)$$

where  $\delta \in [0, 1)$  is the common discount factor and  $u(\cdot)$  is a common differentiable, strictly increasing, weakly concave (risk neutral or risk averse) utility function with  $u(0) = 0$ .<sup>3</sup>

We now discuss our theoretical results informally, leaving formal exposition for the supporting information (SI; pp. 1–17). In the absence of campaign promises, there are equilibria where citizens reelect the incumbent if and only if their income exceeds a given threshold. For sufficiently low thresholds, both high- and low-quality officials will comply (pooling equilibria) and be reelected. For intermediate values of the threshold, high-quality officials will be able to comply and be reelected, whereas low-quality officials will take the maximum salary and be replaced (i.e., separating equilibria). Finally, if the required threshold is too high, all officials will take the maximal salary and be replaced (i.e., kleptocratic equilibria).

If campaign promises are possible, citizens vote based on (a) whether the incumbent provided an income consistent with her previous campaign promises and (b) the incumbent's and challenger's (credible) current promised incomes. High-quality candidates will be able to credibly promise more than low-quality candidates. If campaign promises are not possible, but we increase the quality variance of officials by increasing the

<sup>3</sup>Our utility function does not include an effort cost of campaigning. However, in the supporting information (pp. 11–15), we discuss campaigning when costs are positive but small (comparable to the effort involved in typing messages into the computer in the experiment). Also, our experiment implements the theoretical model with indefinite repetition (the game continues to the next round with probability  $\delta$ , without discounting), substituting for infinite repetition using the same value of  $\delta$  as the discount factor. This technique is well accepted (e.g., Camera and Casari 2014; Engle-Warnick and Slonim 2006) and preserves all theoretical results (Mailath and Samuelson 2006, 106).

quality of high-quality officials and reducing that of low-quality officials (leaving average quality fixed), then we make it easier to distinguish between high- and low-quality officials by increasing the distance between the thresholds the two quality types can comply with.

Thus, either increasing quality variance or allowing for campaigns makes it easier for citizens to figure out which candidates are high-quality and to choose thresholds with which only these candidates will be willing to comply (i.e., separating equilibria). Under weak additional assumptions on preferences (for details, see SI p. 10), we show citizens receive higher incomes in these separating equilibria. Thus,

*H1:* Citizens' incomes are higher when quality variance is higher.

*H2:* Citizens' incomes are higher when campaigning is possible.

Similarly, when quality variance is high, citizens are more likely to confirm only high-quality officials, thus raising citizens' expectations and making it less likely that a given income will be sufficient for reelection:

*H3:* Conditional on citizens' income, reelection is more likely under low variance than under high variance.

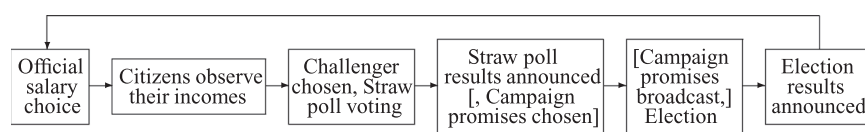
Next, we reiterate that an implication of raising quality variance is that high-quality (low-quality) officials are more (less) likely to be confirmed.

*H4:* Reelection will be more likely for high-quality officials, and less likely for low-quality officials, when quality variance is higher.

The importance of the straw poll is that it allows us to identify the impact of campaign messages. Because the noncandidate citizens have identical preferences and constitute a majority, there always exist equilibria where they vote in the straw poll in the same way they would vote, absent messages, in the election.<sup>4</sup> Thus, if voting changes between the straw poll and the election when

<sup>4</sup>The straw poll is cheap talk, so other equilibria exist, including babbling equilibria where the poll is ignored. As there is no incentive to vote insincerely (unlike, e.g., in Agranov et al. 2018), we focus on equilibria with sincere voting.

**FIGURE 1 Structure of Rounds**



Notes: Arrows indicate the order of events within each round.

campaigning is possible, we will attribute these differences to the messages that are sent. With nonzero effort costs, campaigning should be more prevalent when polling indicates a competitive election. We define *polling bias* as the challenger's straw-poll vote share minus her election vote share and *polling error* as the absolute value of the polling bias.

Our analysis in the SI (pp. 11–15) shows that when messages have positive but vanishingly small costs, high-quality incumbents will campaign and low-quality incumbents will not, whereas challengers will campaign if promises are broken by the incumbent, but not otherwise. We classify messages according to the following (nonexclusive, nonexhaustive) categories. A message

- contains *positive campaigning* if it refers to the sender's behavior and/or quality.
- contains *negative campaigning* if it refers to the opponent's behavior and/or quality.<sup>5</sup>
- is about *quality* if it refers to either candidate's quality.
- is *retrospective* if it refers to either candidate's previous performance and *prospective* if it refers to future performance. In both cases, we say the message is about *performance*.

As messages are cheap talk, there is no game-theoretic prediction regarding their content. However, our analysis suggests that candidates make credible promises and, if elected, are held accountable for them; this suggests some natural message strategies. When citizens' incomes are low, promises are less likely to have been kept, and challengers will emphasize that through negative and retrospective messages. When incomes are high, challengers' messages will avoid mentioning incumbent performance and thus be positive and prospective. Also, when quality variance is higher, a given income is less likely to be satisfactory to citizens (Hypothesis 3), implying more negative and retrospective campaigning from challengers. Hence,

*H5:* When campaign messages are possible, the probability of negative and retrospective campaigning by challengers is (a) decreasing in the income given to citizens and (b) higher in the high-variance treatment conditional on citizens' income.

For incumbents, the reverse predictions apply, except that we do not expect to see negative campaigning be-

<sup>5</sup>In the experiment, candidates were allowed to campaign positively about the opponent or negatively about themselves, but this almost never happened.

cause challengers do not have a track record that can be attacked. Therefore, we present our final hypothesis:

*H6:* When campaign messages are possible, the probability of positive and retrospective campaigning by incumbents is (a) increasing in the income given to citizens and (b) lower in the high-variance treatment conditional on citizens' income.

## Experimental Procedures

The experimental sessions took place at Monash University (see Table 1) and were programmed in z-Tree (Fischbacher 2007). Subjects were primarily undergraduate students (no exclusion criteria). Subjects received written instructions, which were also read aloud.<sup>6</sup> Questions were answered individually by the experimenter.

Subjects were randomly assigned to groups at the beginning of each supergame; these stayed fixed throughout the supergame. Subjects were visually isolated and asked not to communicate, and no identifying information was provided about other group members.

A round began with all subjects' screens displaying their group's population; the official was also shown the endowment and prompted to choose a salary. Citizens were informed of their income, but not the endowment, official's quality, or salary. One citizen was privately informed of nomination as challenger and his or her quality.<sup>7</sup> Then the straw poll was held.

In the Messages cells, candidates could write messages ("campaign announcements") after the poll vote counts and winner were announced. Messages had to be 0–140 characters, in English, and without identifying information. Messages were displayed to all group members and then the election took place (in the Election-only treatment, this happened after the straw-poll results were announced). After the election winner and vote counts were announced, subjects were informed whether the supergame would continue or end (so all decisions were made before knowing this).

<sup>6</sup>Sample instructions are in the SI (pp. 20–21). Additional materials are available from the authors.

<sup>7</sup>Quality was randomly drawn for challengers, even those who had lost an earlier election. This is intended to keep our setting comparable to the existing literature (e.g., Persson, Roland, and Tabellini 1997), where infinite population means a zero probability of being renominated after an election loss. Our experiment necessarily uses finite populations, but redrawing new candidates' qualities, and other design features such as maintaining subject anonymity, let us simulate infinite populations.

**TABLE 1 Experiment Information**

	Campaign Messages?	Quality Heterogeneity	Group Size			Sessions	Subjects
			5	6	7		
MH (messages, high variance)	Yes	High	8	2	4	5	80
ML (messages, low variance)	Yes	Low	7	1	6	5	83
EH (election only, high variance)	No	High	6	2	3	4	63
EL (election only, low variance)	No	Low	6	1	4	4	64

Notes: Table 1 shows the parameters we used in the experiment for each treatment cell.

After the sixth supergame, subjects completed a lottery task (Eckel and Grossman 2008) and a demographic questionnaire. They were privately paid the sum of payoffs (salary for officials, income for other citizens) from the last round of each supergame plus the lottery task, in Australian dollars (AUD), and a 5 AUD ( $\approx$  3.75 USD) participation payment. Sessions typically lasted 60–120 minutes. Total earnings averaged about 48 AUD, ranging from 23.05 to 112.05.

## Experimental Results

Many of the results reported below relate to the hypotheses; we note when this is the case. (This is especially important in experiments—like ours—that were not pre-registered.) Other results do not relate to the hypotheses but are interesting in their own right. Because of the risk of false positives, our emphasis is on results connected to hypotheses, but we include the additional results for completeness.

### Economic Conditions and Election Outcomes

We observe (Table 2) substantial rents to being in office; on average, officials receive more than 6 AUD per round (over 20% of the endowment) more than individual citizens. This difference is confirmed by a Wilcoxon signed-ranks test (pooled MH, ML, EH, and EL cells,  $p < .001$ ).<sup>8</sup>

Low-quality officials take a significantly larger fraction of the economy's endowment than high-quality officials in every cell (pooled MH, ML, EH, and EL,  $p < .001$ ). Since the endowment is smaller when the offi-

cial is low quality, the effect of officials' quality on salary choices in dollars depends on quality variance (and hence the difference between low and high endowments): low-quality officials' dollar salaries are higher under low variance and lower under high variance ( $p \approx .012$  and  $p \approx .055$  for pooled MH/EH and pooled ML/EL, respectively) compared to high-quality officials. The effect of officials' quality on citizens' incomes is unambiguous: higher (in dollars or fractions of the endowment) when the official is high quality (pooled cells,  $p < .001$  for both comparisons).

Additionally, both officials and citizens earn more under high variance (robust rank-order test [RRO], pooled E and M cells,  $p \approx .036$  and  $p \approx .027$  for dollar salaries and dollar incomes, respectively). Campaigning appears to benefit citizens at the expense of officials, consistent with Hypothesis 2, with lower salaries in the M treatment than the E treatment, though these differences are not always significant (pooled H and L cells,  $p \approx .036$  and  $p \approx .073$  for salaries and incomes, respectively).

Next, we estimate tobit models, using the group-round as the unit of observation. Our main explanatory variables are dummies for the messages treatment, the high-variance treatment, high official quality, and all interactions of these. We include a constant term, the supergame and round numbers, dummies for group sizes of 5 and 7 (baseline = 6), and the official's post-experiment lottery choice.

Table 3 shows the main results. We focus on marginal effects (MEs)—either average or conditional on particular variables' values—accompanied by standard errors (SEs), clustered by session.<sup>9</sup> The main results from Table 2 are confirmed, and all are significant. In particular,

<sup>8</sup>Our nonparametric tests err conservatively by using session-level data and two-tailed rejection regions. When not stated explicitly, it is understood that we continue to use the test mentioned most recently. See Siegel and Castellan (1988) for descriptions of these tests and Feltoovich (2005) for critical values of the robust rank-order test.

<sup>9</sup>Clustering by group instead of session, using linear models, and leaving out the official's risk tolerance do not materially change the results. We omit some variables from the tables to save space because they are insignificant or their effects are unimportant to us; however, we leave them in the regressions themselves. Details of all regressions, including those not reported in the main text, are available from the authors.

**TABLE 2 Salaries and Incomes**

Cell	Official Quality	Salary Choice		Citizen Income	
		Dollars	Fraction of Endowment	Dollars	Fraction of Endowment
MH	8	12.190	0.273	7.064	0.156
	3	8.129	0.488	1.860	0.108
	Pooled	11.058	0.333	5.612	0.143
ML	6	8.720	0.247	5.454	0.156
	5	9.636	0.332	4.068	0.145
	Pooled	9.003	0.273	5.026	0.152
EH	8	13.777	0.303	6.781	0.151
	3	10.406	0.617	1.398	0.083
	Pooled	12.881	0.386	5.349	0.133
EL	6	9.906	0.284	5.224	0.155
	5	12.144	0.416	3.560	0.128
	Pooled	10.793	0.336	4.565	0.144

Notes: Officials' salary choices and corresponding citizen incomes, for each cell and official quality, and for each cell (pooled qualities), all rounds.

*Result 1:* Citizens' incomes are higher in the high-variance treatment than the low-variance treatment.

*Result 2:* Citizens' incomes are higher in the messages treatment than the election-only treatment.

Result 1 provides support for Hypothesis 1. Holding constant the average quality of politicians overall, increasing variance raises the average quality of officeholders since those with below-average quality get voted out. This leads to larger resource endowments under high variance, and hence higher citizen incomes.

Result 2 provides support for Hypothesis 2. Campaigning allows high-quality candidates to make generous but credible campaign promises. As with Result 1, voters can better select higher-quality officials, leading to higher citizen incomes. The credibility of campaign promises here is consistent with the experimental literature (e.g., Blume, Lai, and Lim 2020), which finds cheap-talk communication is often informative, despite the lack of compelling game-theoretic justifications for informativeness.

We next examine voting behavior. Table 4 shows two measures of election outcomes: incumbents' vote share and reelection frequency. We observe,

*Result 3:* Low-quality officials fare worse in elections, and high-quality officials fare better, under high variance, though differences are not always significant.

The intuition behind this result is straightforward. Higher-quality variance makes it easier for voters to recognize high-quality candidates, and this strengthens the incentive to elect them over low-quality officials. These differences, consistent with Hypothesis 4, are significant in three of the four cases (when we pool message and election-only treatments), and just miss significance in the fourth.

Figure 2 displays relationships between citizens' incomes and incumbents' electoral success in each cell. Both measures improve with citizens' incomes and have roughly S-shaped curves consistent with retrospective voting, and specifically threshold voting strategies (see SI p. 1). A positive correlation between incomes and incumbent election success is seen in every session of every cell, and in both the straw poll and the election; the probability of this occurring if there were actually zero or negative correlation is vanishingly small.

Table 5 shows associated probit regression results. The dependent variables are (a) an individual's straw-poll vote for the incumbent, (b) the incumbent winning the poll, (c) an individual's election vote for the incumbent, and (d) incumbent reelection. The unit of observation is the subject-round for (a) and (c) (dropping candidates from the sample) and the group-round for (b) and (d). Our main explanatory variables are the two treatment dummies used previously, the income of a representative citizen, and all interactions, along with the supergame and round numbers, group-size dummies, and a constant term.



TABLE 3 Tobits: Salary and Income

	Official Salary		Citizen Income	
	Dollars	Fraction of Endowment	Dollars	Fraction of Endowment
Messages (dummy)	-1.642* (0.731)	-0.051* (0.022)	0.318* (0.160)	0.012* (0.005)
High variance (dummy)	2.186** (0.693)	0.068** (0.022)	0.379** (0.139)	-0.017** (0.005)
High quality (dummy)	-0.282 (0.600)	-0.186** (0.025)	2.867** (0.121)	0.044** (0.006)
Supergame	-0.488** (0.157)	-0.010* (0.004)	0.108** (0.033)	0.003* (0.001)
Round	0.169 (0.099)	0.007* (0.003)	-0.038 (0.021)	-0.002* (0.001)
Official's risk tolerance	0.514** (0.144)	0.015** (0.005)	-0.121** (0.032)	-0.004** (0.001)
Group-size dummies?	Yes	Yes	Yes	Yes
N	1,555	1,555	1,555	1,555

Notes: Average marginal effects, session-clustered standard errors.

†  $p < .1$ , \*  $p < .05$ , \*\*  $p < .01$ .

We observe an increasing relationship between citizens' income and incumbents' election success: not connected to our hypotheses, but clearly consistent with retrospective voting, previous experimental results, and Figure 2.

We also observe that for a given citizen's income, allowing campaigning favors incumbents, as they receive significantly higher shares of both the straw vote and

the election vote; however, this success yields only an insignificant boost in poll and election victories.<sup>10</sup> Finally, high variance favors the challenger, consistent with Hypothesis 3:

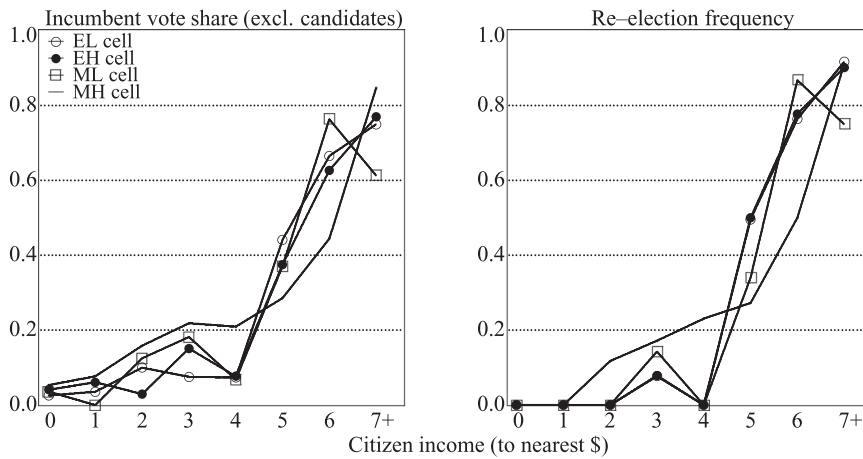
<sup>10</sup>Finding a significant effect in vote counts but not election wins is not paradoxical due to our use of subject-level data for individual votes and group-level data (entailing smaller sample sizes and larger standard errors) for outcomes.

TABLE 4 Election Outcomes

	Official Quality	Incumbent Vote Share			Reelection Frequency		
		High Variance	Low Variance	Significantly Different?	High Variance	Low Variance	Significantly Different?
Message (M) treatment	High	0.718	0.632	$p \approx .044$	0.787	0.742	$p > .20$
	Low	0.203	0.379	$p \approx .008$	0.103	0.281	$p > .20$
Election-only (E) treatment	High	0.614	0.557	$p \approx .114$	0.755	0.648	$p \approx .114$
	Low	0.181	0.323	$p \approx .028$	0.031	0.255	$p \approx .028$
Pooled M and E	High	0.669	0.597	$p \approx .025$	0.772	0.698	$p \approx .086$
	Low	0.193	0.348	$p < .001$	0.071	0.266	$p \approx .006$

Notes: Incumbent vote share and reelection frequency, by cell and official quality (all rounds). The  $p$ -values for differences between high/low variance are based on RRO tests.

**FIGURE 2 Election Outcomes by Citizen Income**



Notes: Each line shows the association in one cell between current-round citizen income and incumbent election success.

*Result 4:* Incumbents’ electoral outcomes are worse under high variance than under low variance, controlling for citizens’ incomes.

As mentioned, citizens have raised expectations under high variance (Result 1), so they are less likely to reelect at any given income.

Table 6 shows reelection frequencies broken down by cell and incumbent–challenger quality pair. Voters appear successful at inferring incumbents’ quality, with high-quality incumbents typically 50 percentage points more likely than low-quality incumbents to be reelected. These differences are insignificant for individual cells due to the use of session-level data (Wilcoxon,  $p = .125$  for in-

**TABLE 5 Probits: Poll and Election Outcomes**

	Straw Vote for Incumbent (Excluding Candidates)	Incumbent Wins Straw Poll	Vote for Incumbent (Excluding candidates)	Reelection
Messages (dummy)	0.056** (0.017)	0.024 (0.023)	0.057** (0.018)	0.019 (0.024)
High variance (dummy)	-0.037** (0.014)	-0.055* (0.024)	-0.045** (0.015)	-0.051* (0.021)
Citizen income (\$)	0.111** (0.004)	0.151** (0.010)	0.118** (0.006)	0.162** (0.012)
Supergame	(0.006) (0.006)	(0.008) (0.008)	(0.005) (0.005)	(0.007) (0.007)
Round	0.008** (0.003)	0.009** (0.002)	0.008** (0.002)	0.007** (0.002)
Group-size dummies?	Yes	Yes	Yes	Yes
N	5,910	1,555	5,910	1,555

Notes: Average marginal effects, session-clustered standard errors.  
<sup>†</sup>  $p < .1$ , \*  $p < .05$ , \*\*  $p < .01$ .

cumbents facing either high- or low-quality challengers in either EH or EL;  $p = .0625$  for the same comparisons in either MH or ML), though we note that these represent the lowest two-sided  $p$ -values possible given the sample sizes of four (EH and EL cells) and five (MH and ML).

Table 6 highlights an important difference between increasing variance and allowing campaign messages. In the former, voters are better able to evaluate the incumbent's performance and to discern her quality, but they still have no information about challengers. In the latter, voters may discern challengers' quality via their campaign messages. As evidence of this, voters are more likely to return incumbents facing low-quality challengers: the difference is small and insignificant when the incumbent herself is high quality ( $p > .20$  and  $p = .125$  in the MH and ML cells, respectively), but it is larger when the incumbent is low quality (the minimum  $p$ -value of .0625 in either MH or ML). By contrast, voters in the election-only cells are generally not more likely to reelect when the challenger is low quality.<sup>11</sup>

A final result comes from comparing elections between two high-quality candidates versus elections between two low-quality candidates. In both MH and ML cells, incumbents are more often reelected in (H, H) elections than in (L, L) elections ( $p = .0625$  for both comparisons).<sup>12</sup> Thus, although campaigning matters for electoral success, track records matter even more.

<sup>11</sup>In our theoretical analysis, we argue for equilibria with credible communication, where low-quality incumbents are replaced regardless of opponent quality. Table 6 suggests that this usually happens, but under low variance (ML cell), voters sometimes do reelect low-quality incumbents who face a low-quality challenger. Importantly, the overwhelming majority of voters replace low-quality incumbents in favor of high-quality challengers when given the opportunity.

<sup>12</sup>Pooling low- and high-variance cells further strengthens the results in Table 6. Fixing the challenger's quality, comparing high- and low-quality incumbents yields  $p$ -values of .008 and .002 (election-only and messages treatments, respectively). Comparing incumbents of a given quality facing low- versus high-quality challengers in the messages treatment yields  $p$ -values of .027 and .002, whereas in the election-only treatment, the corresponding differences are still insignificant ( $p \approx .16$  and  $p > .20$  for low- and high-quality incumbents, respectively). Comparing (H, H) and (L, L) elections in the messages treatment yields a  $p$ -value of .002.

**TABLE 6** Reelection Frequencies

Candidate Qualities (Incumbent, Challenger)	Cell			
	MH	ML	EH	EL
(H, L)	0.821	0.801	0.782	0.633
(H, H)	0.752	0.677	0.724	0.664
(L, L)	0.196	0.409	0.025	0.187
(L, H)	0.017	0.127	0.036	0.321

Notes: Reelection frequencies contingent on cell and incumbent/challenger qualities (all rounds).

## Polling and Election Outcomes

Straw-poll votes are cheap talk, with no strict incentive to vote sincerely. Also, voters may change their minds between the straw poll and the election. Nonetheless, straw polls forecast the election result well, with the straw-poll winner winning the subsequent election 92.1% of the time in the election-only treatment and 88.7% of the time in the messages treatment.

Figure 3 shows, for every election, a plotted point whose coordinates are the incumbent's poll and election vote shares. Overall, there is a pro-incumbent polling bias that is small (2.633 and 2.029 percentage points in the election-only and messages treatments, respectively) but significant (Wilcoxon,  $p \approx .008$  and  $p \approx .010$ , respectively). It does not differ significantly by cell (RRO,  $p > .20$  for all pairwise comparisons).

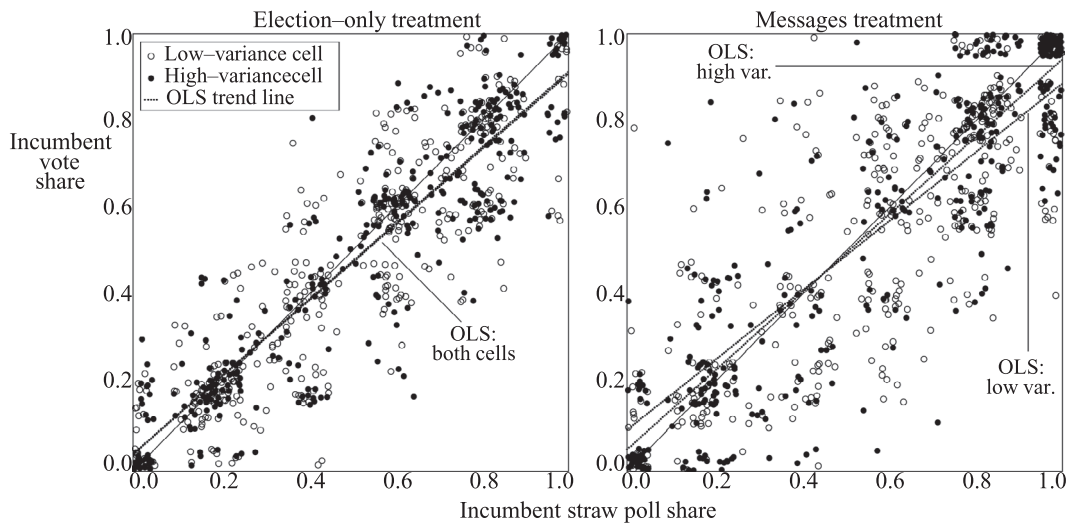
By contrast, polling error (the absolute value of polling bias) is substantially and significantly higher in the messages treatment: 11.486 percentage points versus 7.690 in the election-only treatment ( $p \approx .032$  for MH vs. EH or ML vs. EL,  $p \approx .001$  for pooled messages vs. pooled election-only).

## Campaign Messages: Quantitative Analysis

Differences in polling error between the message and election-only treatments (see the previous section), and in challenger quality's effect on reelection rates (Table 6), suggest a role for campaigning: candidates convey information that voters view as informative. Figure 4 shows how message lengths and frequencies of blank (zero-character) messages depend on polling. Messages take effort to compose, with their length a proxy for the effort required. If the poll vote is close, both candidates have a chance of winning, so both should expend more effort, sending more and longer messages.

Candidates behave consistently with this logic. Pooling the two kinds of uncompetitive election (favoring incumbents and favoring challengers), we find messages are significantly shorter and more often blank than in

**FIGURE 3 Incumbent Polling versus Election Results**



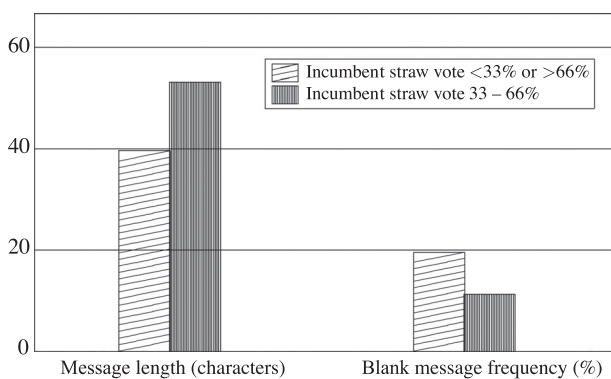
*Notes:* Each circle represents one group-round’s straw poll and election incumbent vote share. Lines represent ordinary least squares trends (labeled) and the 45-degree line (unlabeled) where the poll is correct. For visibility, individual points were “jittered” by adding noise (uniform over  $[-0.05, 0.05]$ ) to both coordinates. The trend lines and statistical analysis are based on the original,unjittered points.

close elections (Wilcoxon,  $p \approx .002$  for message lengths and  $.004$  for blank messages). Also, challengers send longer messages than incumbents. This difference is significant whether we pool close and uncompetitive elections ( $p \approx .014$  for length,  $p \approx .008$  for blank message frequency) or examine them separately ( $p$ -values of  $.027$ – $.084$  for the four comparisons).

Table 7 reports tobit results with either polling bias or polling error as the dependent variable. In Models 9 and 11, explanatory variables are the message and high-variance dummies, citizen income, and all interactions, along with a constant term, the supergame and

round numbers, and group-size dummies. Models 10 and 12 drop the message dummy and its interactions, add the candidates’ message lengths and their products with the high-variance dummy, and use only the messages–treatment data. Consistent with our earlier discussion (e.g., Table 5 and Figure 3), allowing campaigning benefits the incumbent and raises polling error overall. Message length also matters: longer messages increase the candidate’s likelihood of winning (though for challengers, the effect is insignificant, with a  $p$ -value of  $.114$ ) and increase polling error.

**FIGURE 4 Message Effort**



*Notes:* Measure of message effort (length in characters, percentage frequency of blank messages), contingent on whether the straw vote indicates a close election.

### Campaign Messages: Content Analysis

We now examine the *content* of messages. We hired three research assistants (RAs) who were not otherwise involved with this study. Each was given the experiment instructions and a data file that included—for each group-round—the session, supergame and round numbers, group size and (representative) citizen’s income, straw-poll result, and incumbent and challenger message texts. They were not told our research questions, nor did they have access to candidates’ private information (e.g., qualities or salary choices). They received a fixed wage and worked independently.

The RAs coded each message as (a) positive about the candidate, (b) negative about the opponent, (c) about own or opponent quality, (d) about past performance

**TABLE 7 Tobits: Effects of Messages on Polling**

	Polling Bias		Polling Error	
Messages treatment (dummy)	0.107** (0.022)		0.068** (0.016)	
High-variance setting (dummy)	-0.023 (0.021)	-0.016 (0.040)	-0.035* (0.016)	-0.037 (0.021)
Citizen income	-0.007 (0.006)	-0.009 (0.006)	0.010** (0.003)	0.007* (0.003)
Incumbent message length (in 100s)		0.120** (0.042)		0.035 (0.020)
Challenger message length (in 100s)		-0.075 (0.047)		0.091** (0.027)
Supergame	-0.009 (0.007)	-0.008 (0.008)	-0.008 (0.004)	-0.007 (0.005)
Round	-0.001 (0.005)	-0.005 (0.009)	-0.003 (0.002)	-0.006 (0.003)
Sample	All groups	Messages treatment	All groups	Messages treatment
N	1,555	808	1,555	808

Notes: Average marginal effects, session-clustered standard errors.

†  $p < .1$ , \*  $p < .05$ , \*\*  $p < .01$ .

(including the official's current salary), and (e) about future decisions the candidate would take if elected.<sup>13</sup> Here, we call these *positive campaigning*, *negative campaigning*, *quality statement*, *performance statement*, and *campaign promise*, but these terms were not used in discussions with the coders. We classified a message as satisfying a criterion if at least two of the coders classified it that way.

Figure 5 shows that associations between message content and straw-poll results are often nonmonotonic. Exceptions exist; notably, challengers' negative campaigning uniformly decreases as the incumbent's straw vote improves (as bad polling likely reflects low incumbent quality or opportunistic salary choices). But of the 10 message types, four are significantly more likely in close elections (positive campaigning and quality statements by the challenger, promises by both candidates; Wilcoxon,  $p < .050$ ), and for one more the difference just

misses significance (incumbent performance statements,  $p \approx .084$ ).

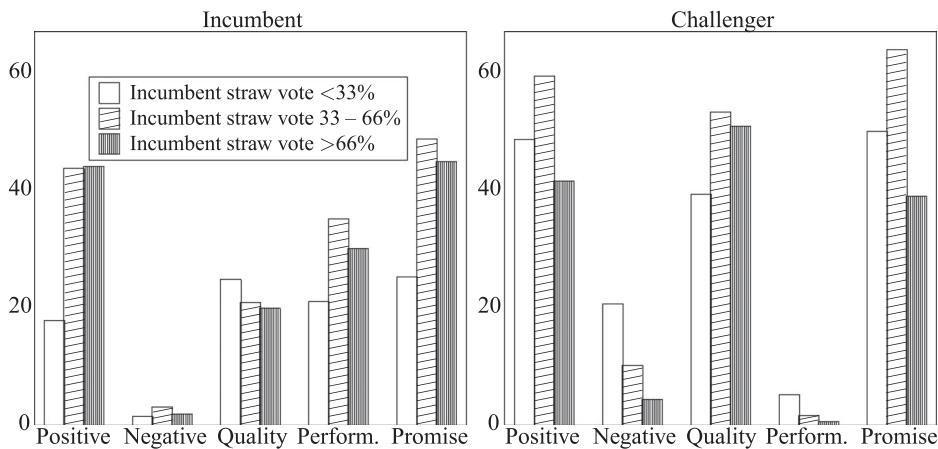
Differences across candidate and message types are also illuminating. Positive campaigning is far more common than negative campaigning, even for challengers.<sup>14</sup> Challengers rarely make statements about performance (though negative campaigning may implicitly concern incumbent performance), but they compensate by making more quality statements and more promises. Incumbents make statements about quality and performance roughly equally often, though performance messages become more common as their polling improves.

Table 8 reports probit results with these message classifications as dependent variables. The main explanatory variables are dummies for the high-variance treatment and a high-quality candidate, their product, the incumbent straw-poll share and its square (to allow for nonmonotonic relationships), and citizens' incomes.

<sup>13</sup>Additional features of campaign messages may matter to voters, such as the level and tone of a message or the use of particular wording. Advances in machine learning may facilitate more thorough analyses of natural-language messages and their effects in the lab and in real-life elections.

<sup>14</sup>Given the lack of scope for negative campaigning by incumbents (the challenger has no track record to criticize), its rarity is unsurprising, but it may be reassuring as verification that subjects understood the setting.

**FIGURE 5 Percent of Messages in Close Elections and Landslides**



Notes: Frequency of message types, contingent on whether the straw vote indicates a close election, a landslide for the challenger, or a landslide for the incumbent.

Additional variables include a dummy for incumbency and its product with these variables (to allow differing effects on incumbents and challengers), the supergame and round numbers, group-size dummies, and a constant. The sample was all candidate-rounds in the messages treatment.

Table 8 shows the average MEs of the incumbency dummy, supergame, and round, and those of other variables separately for incumbents and challengers.<sup>15</sup> Nearly all of the incumbency dummy’s MEs are negative, with the exception of a positive effect on performance messages, suggesting that incumbents campaign less than challengers across the board. Though we had not anticipated this result, it has a natural explanation: when messages are costly, incumbents rely on track records, which challengers cannot do. (The exception—performance messages—is probably also a consequence of incumbents running on their records.)

Additionally, the negative ME of income for challengers, in the “negative campaigning” column, indicates, supporting Hypothesis 5(a),

**Result 5:** As citizens’ incomes increase, challengers’ negative campaigning decreases.

Clearly, the lower incomes are, the more likely it is that the incumbent is low quality or promises were broken. In either case, challengers will emphasize that through negative campaigning.

By contrast, there is no evidence that citizens’ income impacts positive campaigning or whether

<sup>15</sup>Our quadratic specification for the straw vote makes its overall ME hard to interpret, but Figure 6 in the SI (p. 18) shows how polling influences candidates’ messaging.

challengers’ messages are retrospective or prospective. Additionally, high-quality challengers are more likely to campaign positively, mention quality, and make promises than low-quality challengers.<sup>16</sup>

The natural converse of Result 5 is that higher incomes should lead to more positive campaigning by the incumbent. Indeed, consistent with Hypothesis 6(a), and with intuition similar to Result 5’s,

**Result 6:** As citizens’ incomes increase, positive campaigning by the incumbent increases.

Additionally, high-quality incumbents are *less* likely to mention quality or performance than low-quality incumbents.

We can also address the remaining parts of Hypotheses 5 and 6.

**Result 7:** Challengers’ negative and performance messaging are unaffected by quality variance.

<sup>16</sup>In our experiment, messages coded as “campaign promises” are arguably more precise about future intended behavior than other messages. Previous research has found that while precise messages can be viewed as signals of competence (Frenkel 2014), ambiguity is common and often not punished by voters (Kartik, Van Weelden, and Wolton 2017; Tomz and Van Houweling 2009). We find that higher-quality challengers send more precise messages (Table 8), consistent with Frenkel. Nonetheless, both precise and imprecise messages are sent (Figure 5), and the next section provides only equivocal evidence that voters prefer precise messages (e.g., Table 9), consistent with Kartik, Van Weelden, and Wolton and Tomz and Van Houweling. We additionally observe more precision in close elections (Figure 5), under low variance (Table 8), and from challengers (Figure 5), suggesting that candidates expect precision to be rewarded, and that precision can partially substitute for other information such as quality and performance. These findings, while unhyphothesized, are worthy of further investigation.

**TABLE 8** Probits: Factors Affecting Message Characteristics

	Positive Campaigning	Negative Campaigning	Quality Message	Performance Message	Campaign Promise
<b>Average ME</b>					
Incumbent	-0.139** (0.030)	-0.078** (0.011)	-0.277** (0.050)	0.282** (0.027)	-0.085 (0.048)
Supergame	-0.023 (0.014)	0.007** (0.002)	0.021* (0.009)	-0.006 (0.004)	-0.018* (0.009)
Round	-0.009* (0.004)	0.007** (0.002)	-0.009** (0.003)	-0.006 (0.004)	-0.013** (0.005)
<b>ME conditional on incumbent</b>					
High quality (dummy)	-0.041 (0.038)	-0.018 (0.019)	-0.120** (0.053)	-0.118** (0.043)	-0.018 (0.041)
High-variance setting (dummy)	-0.113* (0.056)	-0.009 (0.011)	0.043 (0.079)	-0.035 (0.062)	-0.055 (0.101)
Citizen income	0.061** (0.017)	0.000 (0.003)	0.018 (0.013)	0.017 (0.019)	0.012 (0.014)
<b>ME conditional on challenger</b>					
High quality (dummy)	0.359** (0.022)	-0.022 (0.014)	0.211** (0.036)	-0.003 (0.009)	0.119** (0.029)
High-variance setting (dummy)	0.034 (0.049)	-0.024 (0.026)	0.108 (0.057)	0.002 (0.012)	-0.059 (0.079)
Citizen income	-0.003 (0.010)	-0.015* (0.007)	0.008 (0.010)	-0.001 (0.002)	-0.002 (0.013)
Group-size dummies	Yes	Yes	Yes	Yes	Yes
Incumbent straw vote share	Yes	Yes	Yes	Yes	Yes

Notes: Average marginal effects, session-clustered standard errors.

<sup>†</sup>  $p < .1$ , \*  $p < .05$ , \*\*  $p < .01$ .

**Result 8:** Incumbents are less likely to campaign positively in the high-variance setting.

Result 7 does not support Hypothesis 5(b), suggesting that any impact of candidate heterogeneity on challengers' messaging occurs via citizens' incomes. By contrast, Result 8 strongly confirms Hypothesis 6(b) and suggests that at any given income, incumbents recognize that citizens are less likely to be satisfied when variance is high, and hence are less likely to campaign positively.

The results of this section parallel campaign patterns in real elections: challengers attack incumbents who have performed poorly, whereas successful in-

cumbents focus on their positives. Consider modern U.S. presidential elections. In the 1980 campaign, challenger Ronald Reagan famously attacked incumbent Jimmy Carter's economic record in a presidential debate, asking viewers, "Are you better off than you were four years ago?" In the 1984 election, Reagan was a popular president who mostly campaigned positively, focusing on his strengths and his record on the economy and in the Cold War, as exemplified, respectively, by his "Morning in America" and "Bear in the Woods" television advertisements. By contrast, in 2020, the incumbent president Donald Trump—burdened by a poorly handled pandemic response—concentrated his

campaign messaging on claims about challenger Joe Biden's quality.

### Voter Responses to Campaign Messages

We examine voters' responses to campaigning with four probits. The dependent variable is either a vote for the incumbent or reelection, so the sample is all subject-rounds, excluding candidates, or all group-rounds, respectively. We estimate separate models for the election-only and messages treatments. For the election-only probits, the right-hand variables are a constant, the incumbent straw-vote share and its square (allowing for nonlinear effects), and dummies for the incumbent winning the straw vote, group sizes, and, when applicable, a straw-poll vote for the incumbent. For the messages probits, we include those variables, message-characteristic dummies, their products with the incumbency dummy, and the products of all of these with the incumbent straw-vote share and its square (to allow these effects to depend on the polling).

The main results are shown in Table 9 (see also Figure 7, SI p. 19). Individuals typically vote in the election as they did in the straw poll. There is an apparent bandwagon effect, with voters more likely to vote for a better-polling candidate, though winning the poll outright carries little additional benefit. Finally, campaign messages appear to work:

*Result 9:* Positive campaigning benefits the candidate, both for incumbents and challengers. There is weak evidence that negative campaigning and campaign promises benefit the candidate.

Though not hypothesized, the intuition behind this result is clear. Positive campaigning improves that candidate's election performance, even after controlling for the poll (which should incorporate other information voters had before polling, such as income); this is likely due to candidates' credible signaling of quality or future performance. Relatedly, campaign promises and negative campaigning work for challengers, though not significantly (but with the expected sign) for incumbents.<sup>17</sup>

<sup>17</sup>We also examined the possibility that voter responses to messages depend on how often they receive those messages, using regressions like those in Table 9, but including the total number of each kind of campaign message (e.g., positive campaigning by the incumbent) received by the voter in all previous campaigns (including previous supergames, and previous rounds of the current supergame under a different official), along with interactions with the message variables themselves (e.g., the number of previous positive messages from the incumbent multiplied by an indicator for a positive incumbent message in the current round). Most

## Discussion

Elections allow voters to select representatives and hold them accountable. The information voters rely on comes primarily from candidates' track records and from their campaigns. The existing literature has, for the most part, examined these two sources separately. In the Downsian tradition, voters select politicians based on their campaign promises, and since these are binding, track records do not matter. In the accountability tradition, voting is based on candidates' past performance, with no role for campaigns. In this article, we introduce a dynamic model—with each period allowing for the whole cycle of policymaking, campaigning, and elections—and test its predictions in the lab. This setting allows us to address the question of how political campaigns and past policy performance interact to allow voters to select higher-quality politicians and hold them accountable.

In particular, we investigated the effect of changing (a) how easy it is for voters to determine responsibility for past performance and (b) whether campaigning is possible. We find, compatibly with earlier field experiments (Berry and Howell 2007; Snyder and Strömberg 2010), that more transparency on past performance increases voter welfare. A more novel result is that campaigning also increases voter welfare, by facilitating selection even when it is difficult for voters to evaluate past performance. This is because the possibility of campaigning forces candidates to promise future incomes that are as high as they can credibly promise to deliver, and high-quality candidates can promise more.

Our setting also allows us to assess whether and how campaigns change electoral outcomes. We find that allowing campaigning gives incumbents a small overall advantage and, consistent with previous studies (Jacobson 2013; Squire 1995), longer campaign messages benefit the candidate sending them and are chosen when polling indicates a close election. A large literature has argued that campaigning benefits challengers because it allows them to make themselves known to voters (e.g., Jacobson 1990; Kahn and Kenney 1999). We find that low-quality incumbents facing high-quality challengers are less likely to survive when campaigning is allowed, but they fare better against low-quality challengers. Conditionally on campaigns being allowed, however, we do

notably, the effect of incumbents' negative campaigning decreases with the number of negative messages the voter has received; it is positive when no previous negative messages were received, but it becomes significantly negative when three or more were received. We treat this finding with caution, as we observe no similar effect for messages received within the current supergame only, or only those received under the current incumbent.



**TABLE 9** Probits: Noncandidates' Responses to Messages

	Vote for Incumbent (Election-Only)	Reelection (Election-Only)	Vote for Incumbent (Message)	Reelection (Message)
Voted for incumbent in straw poll (dummy)	0.630** (0.023)		0.252** (0.021)	
Incumbent straw vote share	0.248*** (0.037)	0.998** (0.134)	0.442** (0.059)	0.692** (0.056)
Incumbent wins straw vote (dummy)	0.018 (0.018)	-0.016 (0.030)	0.005 (0.025)	0.006 (0.030)
<b>Incumbent messages</b>				
Positive (dummy)			0.033* (0.015)	0.055* (0.022)
Negative (dummy)			0.020 (0.058)	0.044 (0.041)
Promise (dummy)			0.025 (0.018)	0.002 (0.018)
<b>Challenger messages</b>				
Positive (dummy)			-0.088*** (0.025)	-0.106** (0.025)
Negative (dummy)			-0.059* (0.030)	-0.051 (0.047)
Promise (dummy)			-0.053 (0.031)	-0.096** (0.035)
Group-size dummies?	Yes	Yes	Yes	Yes

Notes: Average marginal effects, session-clustered standard errors.

Other variables: Quality, performance, message length (both candidates), and group size.

†  $p < .1$ , \*  $p < .05$ , \*\*  $p < .01$ .

not find that challengers have a higher return on campaigning than incumbents (Erikson and Palfrey 1998 observe a similar result). Our results also address positive and negative campaigning. The literature does not provide decisive arguments supporting negative campaigning (see Lau, Sigelman, and Rovner 2007). We similarly find that negative campaigning is infrequent compared to positive campaigning and is typically done by challengers when the incumbent has performed poorly.

We close by discussing some limitations of our study. One set of limitations is shared with most laboratory experiments: the sources of its advantages can also raise questions about external validity. For example, our subject population of university students is reasonably

representative of voters.<sup>18</sup> However, it is less representative of politicians, who face multiple rounds of selection and may therefore differ in economic traits (e.g., risk attitudes) and psychological traits (e.g., extraversion, overconfidence, willingness to lie) from the general population. Some of these may affect levels of behavior (e.g., if real officials have different risk attitudes, this may affect citizens' incomes), though it is unclear whether qualitative differences between treatments would be affected. Some have the potential to impact treatment effects (e.g.,

<sup>18</sup>Snowberg and Yariv (2021) report evidence that while students, other convenience samples (Mechanical Turk participants), and representative samples sometimes behave differently on average, differences across treatments are similar across these groups.

if real officials are less reluctant to lie, the effect of allowing campaigning may be weakened), though the ability of voters to punish lies in future elections may discipline officials' worst excesses.

Recent research suggests that the lab's abstractness on its own does not reduce external validity.<sup>19</sup> However, other differences between lab and field may matter. While our financial stakes are high by the standards of lab experiments, real politicians face much higher stakes, possibly making them behave in a more risk-averse way than they would have over smaller stakes. Higher stakes might also reduce noise in behavior (Camerer and Hogarth 1999). Finally, subjects in lab experiments do not have the experience and know-how of real-life campaign operations, nor is the lab suitable for distinguishing between campaigning through different media, suggesting that a more granular analysis of campaign messages is better suited to observational or field work.

A second set of limitations arises from design choices we made, but these could be chosen differently in future studies. Here, all citizens receive the same income, so their preferences over the official's choice are perfectly aligned. If officials made ideological decisions (e.g., on a political left–right spectrum), or if officials could favor some citizens over others, this “solidarity” among citizens would be lost. Also, our policy space is unidimensional; multiple policy dimensions would allow officials to satisfy voters on some dimensions but not on others. Finally, our setting does not allow for abstention and cannot address the issue of turnout. Changing these three factors could have many effects, but one would be a likely increase in negative campaigning, as candidates seek to motivate their own supporters and suppress turnout by the opponent's base. Similarly, challengers would be able to focus on the dimensions where incumbents disappointed. Future work concentrating on the effects of negative campaigning might include one or more of these features.

Our setting also had no candidate-to-voter communication except for one-shot messages. In real elections, campaigns have a dynamic component and candidates react to each other's campaigns; they can use different methods; voters talk to each other (in pairs or larger groups) and receive messages from third parties (e.g., media, partisan groups). For example, negative campaigning directly from candidates may be perceived as distasteful, but the same message may be more palatable

<sup>19</sup>For example, Duch and Stevenson (2013) find that subjects' ability to recognize agenda-setting power in the lab positively correlates with their ability to recognize it in a real legislature.

coming from a third party (e.g., the 2004 “swift-boat” campaign against John Kerry) or in response to attacks from her opponent. Allowing for this diversity in communication may lead to more negative campaigning, but also a greater consensus among voters.

Our experiment did not vary the cost of campaigning, other than infinite (election-only) versus nominal (message). Intermediate treatments with positive but finite costs of campaigning could look at questions of whether changing costs (e.g., by policies such as public matching funding) primarily affect campaigning on the extensive or intensive margin, and whether they affect the kinds of messages that are used.

A final extension concerns polling. In our setting, the straw poll is used as an identification device. While its impact was not one of our research questions, it is a topic worth investigating. A future experimental study could vary whether voters are polled before the election. We conjecture that compared to the no-polling case, allowing polling serves to coordinate voters' voting strategies and to reduce noise in candidates' campaigning strategies (without polling, candidates cannot tailor their messages to how they are faring in the polls). Future work could also vary what fraction of voters are polled: all of them (as in our study) or a randomly or self-selected sample. We speculate that polling a sample rather than the population would reduce the effects of polling compared to no polling.

## References

- Agranov, Marina, Jacob K. Goeree, Julian Romero, and Leat Yariv. 2018. “What Makes Voters Turn Out: The Effects of Polls and Beliefs.” *Journal of the European Economic Association* 16(3): 825–56.
- Ansolabehere, Stephen, Shanto Iyengar, Adam Simon, and Nicholas Valentino. 1994. “Does Attack Advertising Demobilize the Electorate?” *American Political Science Review* 88(4): 829–38.
- Aragones, Enriqueta, Thomas Palfrey, and Andrew Postlewaite. 2007. “Political Reputations and Campaign Promises.” *Journal of the European Economic Association* 5(4): 846–84.
- Ashworth, Scott. 2012. “Electoral Accountability: Recent Theoretical and Empirical Work.” *Annual Review of Political Science* 15(1): 183–201.
- Ashworth, Scott, and Ethan Bueno de Mesquita. 2008. “Electoral Selection, Strategic Challenger Entry, and the Incumbency Advantage.” *Journal of Politics* 70(4): 1006–25.
- Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Friedenberg. 2017. “Accountability and Information in Elections.” *American Economic Journal: Microeconomics* 9(2): 95–138.

- Avdeenko, Alexandra. 2018. "Long-Term Evidence of Retrospective Voting: A Natural Experiment from the German Democratic Republic." *European Economic Review* 103(1): 83–107.
- Barro, Robert J. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14(1): 19–42.
- Bhattacharya, Sourav, John Duffy, and Sun-Tak Kim. 2014. "Compulsory versus Voluntary Voting: An Experimental Study." *Games and Economic Behavior* 84(1): 111–31.
- Berry, Christopher, and William Howell. 2007. "Accountability and Local Elections: Rethinking Retrospective Voting." *Journal of Politics* 69(3): 844–58.
- Blume, Andreas, Ernest Lai, and Wooyoung Lim. 2020. "Strategic Information Transmission: A Survey of Experiments and Theoretical Foundations." In *Handbook of Experimental Game Theory*, ed. Monica Capra, Rachel Croson, Mary Rigdon, and Tanya Rosenblat. Cheltenham, UK: Edward Elgar, 311–47.
- Callander, Steven, and Simon Wilkie. 2007. "Lies, Damned Lies, and Political Campaigns." *Games and Economic Behavior* 60(2): 262–86.
- Camera, Gabriele, and Marco Casari. 2014. "The Coordination Value of Monetary Exchange: Experimental Evidence." *American Economic Journal: Microeconomics* 6(1): 290–314.
- Camerer, Colin F., and Robin M. Hogarth. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework." *Journal of Risk and Uncertainty* 19(1/3): 7–42.
- Clarke, Harold, and Marianne Stewart. 1994. "Prospections, Retrospections, and Rationality: The 'Bankers' Model of Presidential Approval." *American Journal of Political Science* 38(4): 1104–23.
- Corazzini, Luca, Sebastian Kube, Michel A. Maréchal, and Antonio Nicolò. 2014. "Elections and Deceptions: An Experimental Study on the Behavioural Effects of Democracy." *American Journal of Political Science* 58(3): 579–92.
- Downs, Anthony. 1957. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy* 65(2): 135–50.
- Duch, Raymond, and Randolph Stevenson. 2013. "Voter Perceptions of Agenda Power and Attribution of Responsibility for Economic Performance." *Electoral Studies* 32(3): 512–16.
- Eckel, Catherine C., and Philip J. Grossman. 2008. "Forecasting Risk Attitudes: An Experimental Study of Actual and Forecast Risk Attitudes of Women and Men." *Journal of Economic Behavior & Organization* 68(1): 1–17.
- Elinder, Mikael, Henrik Jordahl, and Panu Poutvaara. 2015. "Promises, Policies and Pocketbook Voting." *European Economic Review* 75(1): 177–94.
- Engle-Warnick, James, and Robert L. Slonim. 2006. "Learning to Trust in Indefinitely Repeated Games." *Games and Economic Behavior* 54(1): 95–14.
- Erikson, Robert, and Thomas Palfrey. 1998. "Campaign Spending and Incumbency: An Alternative Simultaneous Equations Approach." *Journal of Politics* 60(2): 355–73.
- Erikson, Robert, Michael MacKuen, and James Stimson. 2000. "Bankers or Peasants Revisited: Economic Expectations and Presidential Approval." *Electoral Studies* 19(2): 295–312.
- Feltovich, Nick. 2005. "Critical Values for the Robust Rank-Order Test." *Communications in Statistics: Simulation and Computation* 34(3): 525–47.
- Feltovich, Nick, and Francesco Giovannoni. 2015. "Selection vs. Accountability: An Experimental Investigation of Campaign Promises in a Moral-Hazard Environment." *Journal of Public Economics* 126(1): 39–51.
- Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50(1/3): 5–25.
- Finkel, Steven, and John Geer. 1998. "A Spot Check: Casting Doubt on the Demobilizing Effect of Attack Advertising." *American Journal of Political Science* 42(2): 573–95.
- Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.
- Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10(2): 171–78.
- Fox, Justin, and Kenneth W. Shotts. 2009. "Delegates or Trustees? A Theory of Political Accountability." *Journal of Politics* 71(4): 1225–37.
- Freedman, Paul, and Ken Goldstein. 1999. "Measuring Media Exposure and the Effects of Negative Campaign Ads." *American Journal of Political Science* 43(4): 1189–1208.
- Frenkel, Sivan. 2014. "Competence and Ambiguity in Electoral Competition." *Public Choice* 159(1/2): 219–34.
- Gordon, Sanford, Gregory Huber, and Dimitri Landa. 2007. "Challenger Entry and Voter Learning." *American Political Science Review* 101(2): 303–20.
- Hale, Jon, Jeffrey Fox, and Rick Farmer. 1996. "Negative Advertisements in U.S. Senate Campaigns: The Influence of Campaign Context." *Social Science Quarterly* 77(2): 329–43.
- Harrington, Joseph E. 1992. "The Revelation of Information through the Electoral Process: An Exploratory Analysis." *Economics and Politics* 4(3): 255–76.
- Healy, Andrew J., Mikael Persson, and Erik Snowberg. 2017. "Digging into the Pocketbook: Evidence on Economic Voting from Income Registry Data Matched to a Voter Survey." *American Political Science Review* 111(4): 771–85.
- Jacobson, Gary. 1990. "The Effects of Campaign Spending in House Elections: New Evidence for Old Arguments." *American Journal of Political Science* 34(2): 334–62.
- Jacobson, Gary. 2013. *The Politics of Congressional Elections*. New York: Pearson.
- Kahn, Kim Friedkin, and Patrick Kenney. 1999. *The Spectacle of U.S. Senate Campaigns*. Princeton, NJ: Princeton University Press.
- Kartik, Navin, Richard Van Weelden, and Stephane Wolton. 2017. "Electoral Ambiguity and Political Representation." *American Journal of Political Science* 61(4): 958–70.
- Konrad, Kai A., and Raisa Sherif. 2019. "Sanctioning, Selection, and Pivotality in Voting: Theory and Experimental Results." *Constitutional Political Economy* 30(3): 330–57.
- Landa, Dimitri. 2010. "Selection Incentives and Accountability Traps: A Laboratory Experiment." Paper presented at the annual meeting of the American Political Science Association, Washington, DC.
- Lau, Richard, and Gerald Pomper. 2002. "Effectiveness of Negative Campaigning in U.S. Senate Elections." *American Journal of Political Science* 46(1): 47–66.

- Lau, Richard, and Ivy Brown Rovner. 2009. "Negative Campaigning." *Annual Review of Political Science* 12(1): 285–306.
- Lau, Richard, Lee Sigelman, and Ivy Brown Rovner. 2007. "The Effects of Negative Political Campaigns: A Meta-Analytic Reassessment." *Journal of Politics* 69(4): 1176–1209.
- Li, Hao, and Wei Li. 2013. "Misinformation." *International Economic Review* 54(1): 253–77.
- Lindbeck, Assar, and Joergen W. Weibull. 1987. "Balanced-Budget Redistribution as the Outcome of Political Competition." *Public Choice* 52(3): 273–97.
- Mailath, George J., and Larry Samuelson. 2006. *Repeated Games and Reputations: Long-Run Relationships*. Oxford: Oxford University Press.
- Mansergh, Lucy, and Robert Thomson. 2007. "Election Pledges, Party Competition, and Policy-Making." *Comparative Politics* 39(3): 311–29.
- Matthies, Theres. 2020. "Retrospective Pledge Voting: A Comparative Study of the Electoral Consequences of Government Parties' Pledge Fulfilment." *European Journal of Political Research* 59(4): 774–96.
- Naurin, Elin. 2011. *Election Promises, Party Behaviour and Voter Perceptions*. New York: Palgrave MacMillan.
- Naurin, Elin, Terry Royed, and Robert Thomson, eds. 2019. *Party Mandates and Democracy: Making, Breaking, and Keeping Election Pledges in Twelve Countries*. Ann Arbor: University of Michigan Press.
- Persson, Torsten, Gerard Roland, and Guido Tabellini. 1997. "Separation of Powers and Political Accountability." *Quarterly Journal of Economics* 112(4): 1163–1202.
- Pétry, François, and Benoît Collette. 2009. "Measuring How Political Parties Keep Their Promises: A Positive Perspective from Political Science." In *Do They Walk Like They Talk? Studies in Public Choice*, ed. Louis Imbeau. New York: Springer, 65–80.
- Polborn, Matthias, and David Yi. 2006. "Informative Positive and Negative Campaigning." *Quarterly Journal of Political Science* 1(4): 351–71.
- Przeworski, Adam, Susan C. Stokes, and Bernard Manin. 1999. *Democracy, Accountability, and Representation*. Cambridge: Cambridge University Press.
- Royed, Terry 1996. "Testing the Mandate Model in Britain and the United States: Evidence from the Reagan and Thatcher Eras." *British Journal of Political Science* 26(1): 45–80.
- Schnakenberg, Keith. 2016. "Directional Cheap Talk in Electoral Campaigns." *Journal of Politics* 78(2): 527–41.
- Siegel, Sidney, and N. John Castellan Jr. 1988. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
- Skaperdas, Stergios, and Bernard Grofman. 1995. "Modeling Negative Campaigning." *American Political Science Review* 89(1): 49–61.
- Snowberg, Erik, and Leeat Yariv. 2021. "Testing the Waters: Behavior across Participant Pools." *American Economic Review* 111(2): 687–719.
- Snyder, James, and David Strömberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy* 118(2): 355–408.
- Snyder, James, and Michael Ting. 2002. "An Informational Rationale for Political Parties." *American Journal of Political Science* 46(1): 90–110.
- Squire, Peverill. 1995. "Candidates, Money, and Voters: Assessing the State of Congressional Elections Research." *Political Research Quarterly* 48(4): 891–97.
- Theilmann, John, and Allen Wilhite. 1998. "Campaign Tactics and the Decision to Attack." *Journal of Politics* 60(4): 1050–62.
- Tomz, Michael, and Robert P. Van Houweling. 2009. "The Electoral Implications of Candidate Ambiguity." *American Political Science Review* 103(1): 83–98.
- Woon, Jonathan. 2012. "Democratic Accountability and Retrospective Voting: A Laboratory Experiment." *American Journal of Political Science* 56(4): 913–30.

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Appendix A:** Additional analysis

**Appendix B:** Sample instructions