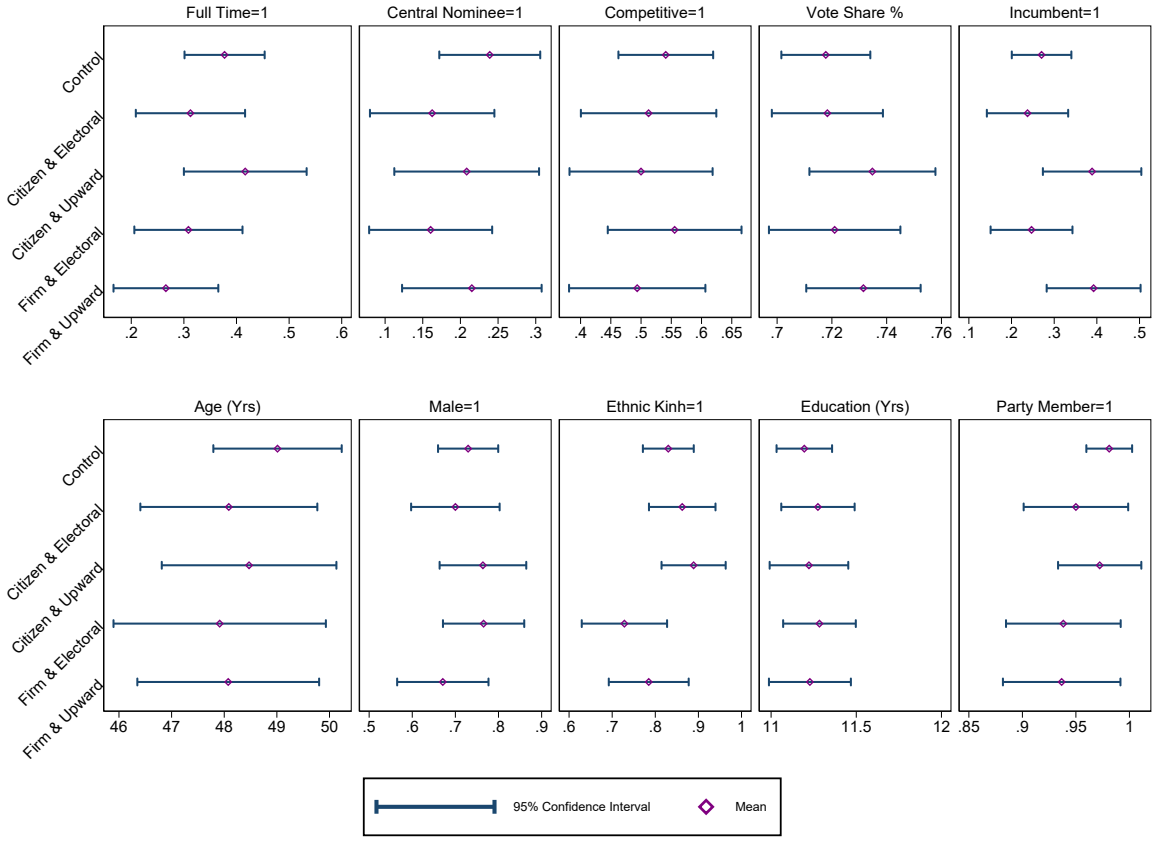
ONLINE APPENDICES

Can Elections Motivate Responsiveness in a Single-Party Regime? Experimental Evidence from Vietnam

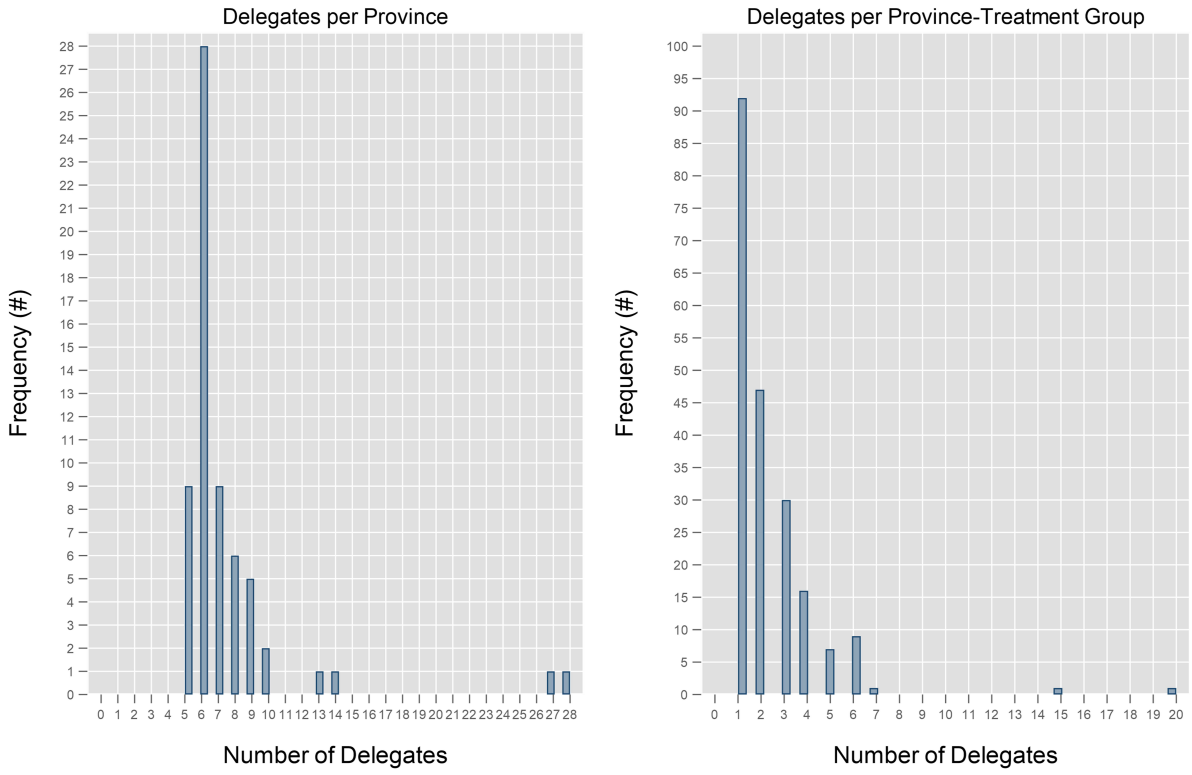
Figure A.1: **Randomization achieved balance.** Figure displays treatment group means and confidence intervals across 10 political and demographic covariates.



A Assessing balance

Because we have four treatment arms distributed across approximately 312 delegates, each treatment cell contains a small number ($\sim$78) of delegates. For this reason, imbalance is a potential concern. Figure A.1 presents confidence interval-based balance tests for several potential confounding variables. Reassuringly, the results clearly show that there are no significant demographic or political differences across treatment groups. On average, delegates display similar VNA, party, and educational histories, and are of comparable age, gender, and ethnicity.

Figure B.1: **Small provincial delegations necessitate randomization inference.** Figure displays delegate counts across provinces and province-treatment groups.



B Descriptive statistics and delegate counts

Table B.1 displays descriptive statistics for all variables used in the primary analyses. The first and second panels present behavioral outcomes; the third and fourth panels present individual and provincial treatment shares; and the final panel presents pre-treatment control variables as well as raw vote shares, unused in the analyses. Figure B.1 summarizes the distribution of provincial delegation sizes (left panel) and province-treatment sizes (right panel). As the left panel attests, most provincial delegations contain between five and ten delegates. Spreading these delegates across four treatment groups and a control condition, we are frequently left with only one or two delegates per province-treatment group. Because of these small group sizes, there is insufficient variation to justify employing fixed effects or province-clustered standard errors. Instead, we opt for a randomization inference approach which naturally accounts for provincial clustering.

Table B.1: Summary statistics for variables used in analyses.

Variable	N	Mean	SD	Min	Max
Speaking behavior					
Spoke in any forum=1	470	0.438	0.497	0.000	1.000
Spoke in caucus=1	470	0.372	0.484	0.000	1.000
Spoke in floor debate=1	470	0.055	0.229	0.000	1.000
Spoke in query session=1	470	0.096	0.294	0.000	1.000
Quality of speech					
Critical of amendment=1	470	0.274	0.447	0.000	1.000
Pro-labor speech=1	470	0.194	0.396	0.000	1.000
Pro-business speech=1	470	0.055	0.229	0.000	1.000
Speech reflected in law=1	470	0.132	0.306	0.000	1.000
`Wordscores` (Pro-labor)	178	38.167	55.654	−48.641	189.489
`Wordscores` (Mean imputed)	470	45.519	34.670	−48.641	189.489
Treatments					
Citizen-Electoral=1	470	0.170	0.376	0.000	1.000
Central-Upward=1	470	0.153	0.361	0.000	1.000
Firm-Electoral=1	470	0.172	0.378	0.000	1.000
Firm-Upward=1	470	0.168	0.374	0.000	1.000
Saturation levels					
Citizen-Electoral share	470	0.170	0.157	0.000	0.666
Citizen-Upward share	470	0.172	0.161	0.000	0.666
Firm-Electoral share	470	0.153	0.153	0.000	0.666
Firm-Upward share	470	0.168	0.163	0.000	0.571
Pre-treatment controls					
Full-time delegate=1	470	0.340	0.474	0.000	1.000
Central nominee=1	470	0.202	0.402	0.000	1.000
Vote share below median=1	470	0.526	0.500	0.000	1.000
Raw vote share (not used)	470	0.723	0.100	0.471	0.954

C Heterogeneous treatment effects: central nominees

Another way to help adjudicate between the cadre advancement and electoral competitiveness explanations for the effects of the CE treatment is to focus upon central nominees. These delegates are slated for advancement in the Vietnamese party-state and boast institutional electoral advantages to aid their election. We therefore adapt our baseline specification (manuscript Equation 1) to include an interaction between the CE treatment and an indicator for central nominees. As with the manuscript's analysis of heterogeneous treatment effects (HTEs) among the competitively elected, here we only interact the central nominee indicator with the CE treatment and simply control for the other treatment groups.

$$\begin{aligned}\Pr(Y_i = 1) = \beta_0 + \beta_1 \mathsf{CE}_i + \beta_2 \mathsf{CentNom}_i + \beta_3 \mathsf{CE}_i \times \mathsf{CentNom}_i \\ + \beta_4 \mathsf{CU}_i + \beta_5 \mathsf{FE}_i + \beta_6 \mathsf{FU}_i + \boldsymbol{\delta} \mathbf{X}_i + \phi \mathsf{S}_p\end{aligned} \tag{1}$$

Surprisingly, Table C.1 reveals that the effect of the CE treatment is especially strong among central nominees. Among central nominees who did not receive the CE treatment, only 41%

Table C.1: **Citizen-electoral effect larger for central nominees.** Central nominees exhibit large effects of CE treatment relative to local nominees.

	Predicted probability of speaking				Effect of CE treatment	
Group	A	B	C	D	B-A	D-C
CE-treated	No	Yes	No	Yes	Local	Central
Central nominee	No	No	Yes	Yes	nominee ATE	nominee ATE
Spoke	0.420	0.503	0.406	0.718	0.084	0.312
	[0.891]	[0.110]	[0.973]	[0.012]	[0.106]	[0.016]
Critical	0.253	0.301	0.276	0.621	0.048	0.345
	[0.770]	[0.240]	[0.983]	[0.004]	[0.237]	[0.005]
Pro-labor	0.185	0.240	0.183	0.219	0.054	0.036
	[0.843]	[0.173]	[0.652]	[0.362]	[0.173]	[0.359]
Pro-firm	0.059	0.057	0.048	−0.002	−0.002	−0.050
	[0.494]	[0.504]	[0.318]	[0.780]	[0.462]	[0.742]
In final LC	0.119	0.165	0.106	0.441	0.046	0.335
	[0.856]	[0.146]	[0.999]	[0.000]	[0.148]	[0.000]
`Wordscores`	33.28	60.01	63.71	66.98	26.73	3.27
	[0.980]	[0.048]	[0.532]	[0.452]	[0.044]	[0.454]
N	308	67	82	13	375	95

Note: OLS coefficients [RI p-values] , controlling for competitiveness, full-time, and saturation levels. Last row weighted by share of scored words. For calculation of predicted probabilities, central nominee and full-time dummies held at their modal value and saturation level held at the sample mean. Full regression results available in *APSR* Dataverse files for Appendix C.

spoke during the caucuses or floor debates, compared to 72% of CE-treated central nominees. This represents a 31 percentage point marginal effect ($p = 0.016$). Contrast this with the eight percentage point marginal effect of the CE treatment among local nominees ($p = 0.106$). Central nominees were particularly energized after learning about citizen views and electoral competitiveness. The strong effect of the CE treatment on central nominees points toward the cadre advancement explanation, as these are the delegates who already have the highest vote shares and probability of victory, but whose careers are most likely to be derailed by a poor electoral performance.

Taken together, the HTE analysis of electoral competitiveness presented in the manuscript and the HTE analysis of central nominees presented here offer conflicting evidence about the story behind the effects of the CE treatment. On the one hand, finding larger marginal effects among safe seat delegates and central nominees is consistent with the cadre advancement and promotion explanation. On the other hand, the independent positive effects of competitiveness indicate that competitively-elected delegates are already highly attuned toward responsiveness, thereby reducing the effectiveness of reminding them about the upcoming elections. This offers partial support to the electoral explanation for responsiveness.

Table D.1: **Additional subgroup analyses fail to reveal HTEs.** Table presents analyses of full-time delegates, incumbents, party members, and those approaching retirement age. Note that we elect not to use RI for statistical inference.

	Spoke					Critical				
CE	0.075	0.113	0.190	0.137	0.127	0.077	0.073	0.130	0.440	0.101
	(0.076)	(0.073)	(0.275)	(0.113)	(0.065)	(0.069)	(0.067)	(0.103)	(0.249)	(0.059)
	[0.170]	[0.072]	[0.228]	[0.093]	[0.032]	[0.139]	[0.143]	[0.260]	[0.000]	[0.051]
Full-time	0.354					0.282				
	(0.069)					(0.063)				
	[0.916]					[0.759]				
CE×Full-time	0.151					0.072				
	(0.127)					(0.116)				
	[0.120]					[0.281]				
Incumbent		0.016					0.027			
		(0.056)					(0.050)			
		[0.500]					[0.822]			
CE×Incumbent		0.044					0.120			
		(0.137)					(0.124)			
		[0.370]					[0.162]			
Party			0.024					0.145		
			(0.128)					(0.116)		
			[0.341]					[0.043]		
CE×Party			−0.071					−0.355		
			(0.279)					(0.253)		
			[0.617]					[0.910]		
Freshman				−0.011					−0.038	
				(0.054)					(0.049)	
				[0.542]					[0.336]	
CE×Freshman				−0.019					−0.039	
				(0.131)					(0.119)	
				[0.553]					[0.650]	
Retirement					−0.009					−0.053
					(0.201)					(0.104)
					[0.194]					[0.130]
CE×Retirement					−0.472					−0.294
					(0.497)					(0.452)
					[0.973]					[0.984]
Constant	0.309	0.301	0.279	0.312	0.301	0.114	0.107	0.141	−0.028	0.113
	(0.059)	(0.060)	(0.137)	(0.072)	(0.060)	(0.054)	(0.054)	(0.066)	(0.124)	(0.054)
Block FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Saturations	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	470	470	470	470	470	470	470	470	470	470
R^2	0.093	0.091	0.090	0.090	0.092	0.072	0.075	0.074	0.076	0.073
RMSE	0.478	0.479	0.479	0.479	0.478	0.434	0.434	0.435	0.434	0.435

Note: OLS coefficients (standard errors) [RI p-values] . Full regression results available in *APSR* Dataverse files for Appendix D.

D Heterogeneous treatment effects: other subgroups

Readers and reviewers have suggested analysis of heterogeneous treatment effects (HTEs) for several other subgroups, including full-time delegates, incumbents, party members, freshmen, and those approaching retirement age. Almost none of these analyses turned up significant findings, but there was one interesting effect that is very supportive of our theory. Officials past the official retirement age have significantly negative responses to the CE treatment. One interpretation is that, recognizing that they will not run again, these delegates do not alter their behavior in response to the informational treatments. Table D.1 presents all HTE results.

E Saturation design to model potential SUTVA violations

In this appendix, we describe and analyze three potential forms of contamination in our experiment. We begin first by describing the institutional attributes of the VNA that pose potential threats to the Stable Unit Treatment Value Assumption (SUTVA). Next, we describe how our research design aimed to address and model these effects, and the complications this design poses for statistical analysis. Fourth, we discuss robustness analyses to the direct effects presented in manuscript Table 3 which account for two effects of our research design: delegates had unequal probabilities of assignment to treatment and the highest level of dosage lacks a pure control. Fifth, we propose and implement four different approaches to modeling potential contamination that account for different challenges posed by the setting and our research design.

We conclude that our estimates in Table 3 of the main paper are robust to a range of alternative specification choices. In the manuscript's primary analyses, it is reasonable to assume homogeneity of treatment effects across the three sampling bins and to exclude the interaction terms for saturation. We reach this conclusion because all robustness analyses illustrate that the conditional average treatment effects (CATEs) at different levels of treatment dosage are statistically indistinguishable from one another. That is, we uncover no evidence of any of the presumed forms of SUTVA violations. To the extent that omitting the interaction effects might increase bias, such bias would actually militate against finding a significant ATE for the citizen-electoral treatment. Increased saturation appears to be insignificantly but negatively correlated with reduced speaking behavior. In a world without contamination, the effect of the citizen-electoral treatment would have had a greater effect on speaking, because treated delegates would be less likely to steal each other's thunder by using the same material in caucus and floor debates.

E.1 Is contamination a concern in the VNA?

The chronology of debates at the VNA necessitates that provincial delegates caucus together about draft laws after receiving the treatment yet before entering formal debates on the VNA floor. This real-world institutional feature generates the possibility for violation of SUTVA for randomized experiments. Three potential forms of the violation are possible in this particular setting. First, we might observe *spillover effects* where control delegates learn from delegates from treatment groups and change their behavior. This violation would lead to increased activity of the control group, making it more difficult to detect significant experimental effects. Second, in a previous analysis of the VNA using a similar design, Malesky and Todd (2022) found evidence of *reinforcement effects*, especially in floor debates. A reinforcement effect occurs when treated delegates feel empowered to speak in televised sessions as the share of similarly treated delegates in the same province increases, raising their confidence in the credibility of the information. In this case, increasing the share of delegates in a province receiving the same infographic positively increases the effect of speaking, leading to higher probabilities of speaking among delegates receiving the same card. Our pre-analysis plan hypothesized we would observe this effect for the Labor Code as well. Third, *crowding out* effects are also possible, whereby so many delegates have the same information that there is very little new content for the next treated delegate to contribute to the debate, rendering her less likely to speak. In a crowding out setting, increasing the share of delegates from a province with the same information reduces the likelihood of speaking of the marginal delegate.

E.2 A research design to detect potential contamination

There was nothing that we as researchers could do to eliminate these three potential SUTVA violations—they are imposed by the institutional setting. Other scholars have shown, however, that it is possible to develop a research design that allows for modeling the impact of these violations and consequently to account for them in our analysis (Gerber and Green, 2012; Sinclair, McConnell and Green, 2012; Baird et al., 2014). As we showed in Figure 3 in the main paper (reproduced in Appendix Figure E.1 below), we followed the Sinclair, McConnell and Green (2012) strategy of randomizing the density of treatment across geographical units. Thus, in the first stage of randomization, Vietnam's 63 provinces were assigned to one of three dosage bins, where 25%, 50%, or 100% of provincial delegates were treated. The dosage shares are often referred to as "saturation" levels in a saturation model. Non-treated delegates were assigned to the control condition. The approach allows us to directly test for each of the three SUTVA violations by interacting an individual delegate's treatment status with the share of delegates treated in their province (their dosage). In the three subsequent stages treated delegates within each province were then: 2) randomly assigned to treatment or control; 3) to firm or citizen infographic; and 4) to either the electoral or upward motivation treatments.

The initial assignment of provincial dosages is a block randomization, whereas the latter three randomizations are delegate-level Bernoulli trials. The first set of Bernoulli trials is conducted for every delegate according to provincial dosages, while the latter two are equiprobable coin flips conducted only among the treated. In this design, the province a delegate represents is a confounder in that it is associated with both a delegate's propensity score (provincial dosage) and her potential outcomes. Such a design could lead to contamination if delegates within the same province shared information or supported each other in group caucuses.

E.3 Empirically modeling the potential impact of contamination

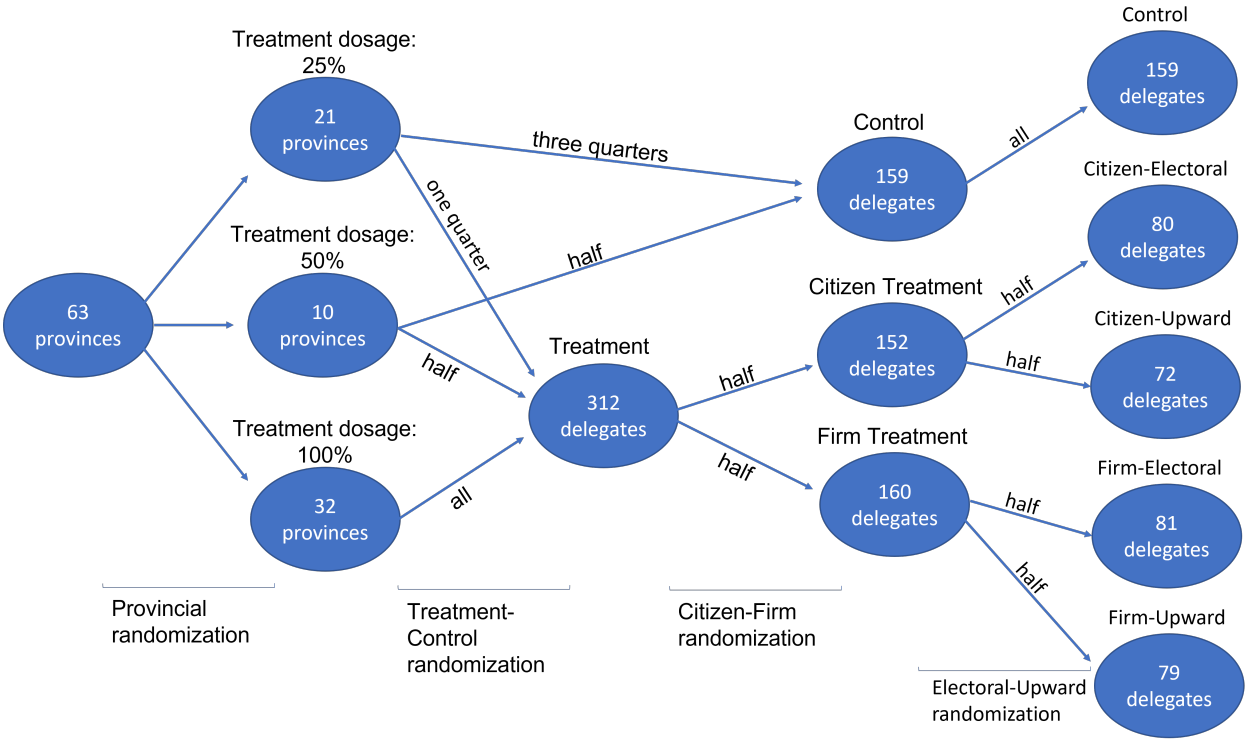
Our research design allows for two possible ways to directly model the saturation interaction effect. The simplest design-based approach, recommended by an anonymous reviewer, is to use a vector of fixed effects ($\mathbf{S}$) reflecting assignment to one of the three dosage bins as the moderator in the interaction. The logic is analogous to treating the assignment probability as a propensity score. Focusing only on the citizen-electoral (CE) condition, because of its positive and sizable independent effect, this generates Equation 2. To capture the effect of the saturation on the other treatment conditions (CU, FE, and FU), further interaction terms can be added (see Equation 3) with some important modeling assumptions that we discuss in more detail below.

$$\Pr(Y_i = 1) = \beta_0 + \beta_1 \text{CE}_i + \boldsymbol{\phi}\mathbf{S}_p + \boldsymbol{\xi}\text{CE}_i \times \mathbf{S}_p + \boldsymbol{\delta}\mathbf{X}_i \tag{2}$$

$$\begin{aligned}\Pr(Y_i = 1) = \beta_0 + \beta_1 \text{CE}_i + \boldsymbol{\phi}\mathbf{S}_p + \boldsymbol{\xi}\text{CE}_i \times \mathbf{S}_p + \beta_2 \text{CU}_i + \boldsymbol{\zeta}\text{CU}_i \times \mathbf{S}_p \\ + \beta_3 \text{FE}_i + \boldsymbol{\eta}\text{FE}_i \times \mathbf{S}_p + \beta_4 \text{FU}_i + \boldsymbol{\theta}\text{FU}_i \times \mathbf{S}_p + \boldsymbol{\delta}\mathbf{X}_i\end{aligned} \tag{3}$$

The impact of increasing delegate dosage, can be easily read off the regression coefficients. β_1 provides the ATE of the CE treatment in the 25% bin, when SUTVA violations have their lowest probability, as treated delegates are least likely to interact with control or other treated delegates. $\boldsymbol{\phi}$ test the possibility of spillover violations, by reporting the impact on the control group when the dosage level is increased. A positive sign on one of $\boldsymbol{\phi}$ indicates *spillover* from treatment to

Figure E.1: Randomization scheme.

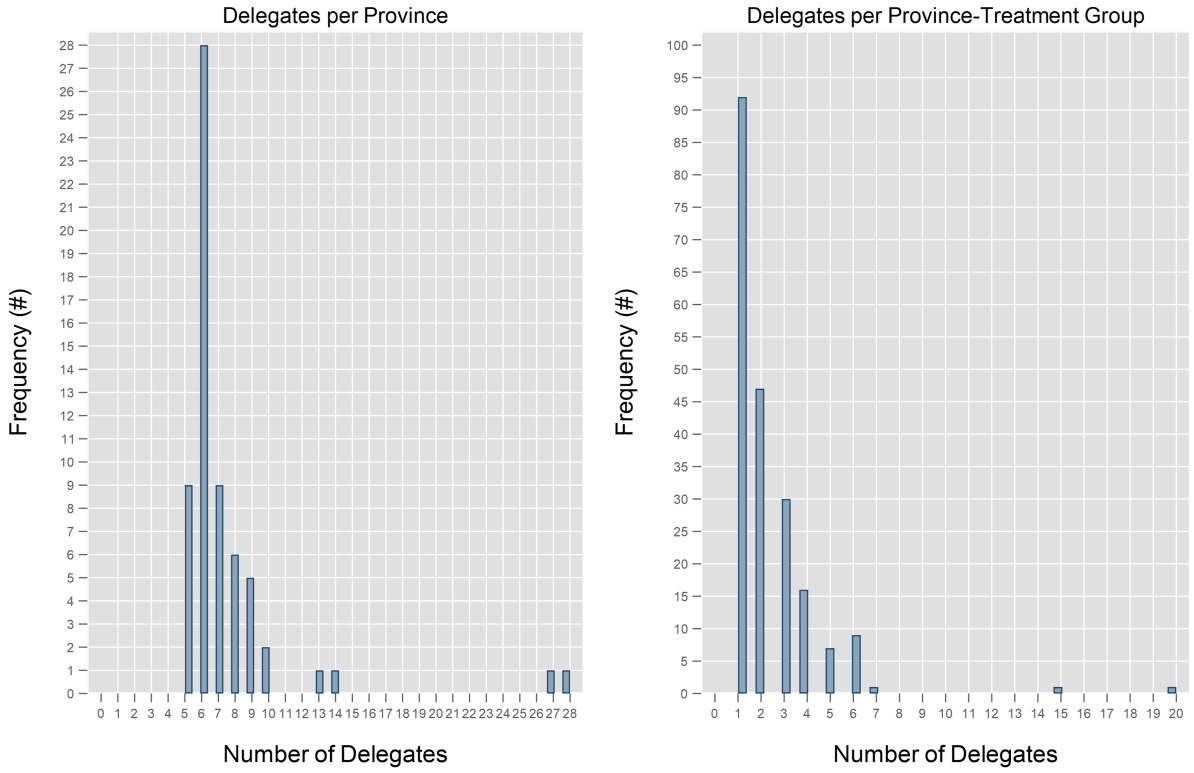


control, illustrating that the higher the probability of interactions between treatment and control, the greater the probability that a control delegate would speak. ξ test the reinforcement or crowding out effects. If one of ξ is positive, it indicates *reinforcement*, because as saturation increases, delegates receiving the CE treatment are more likely to speak, biasing the ATE on CE upward. A negative ξ implies *crowding out*, as increasing saturation reduces the probability of speaking, biasing the ATE on CE downward toward zero.

One concern with using the dosage bins (S) as the moderator in the interaction is that it conflates the four different treatment conditions into a single measure of total treatment. After the provincial dosage assignment, treated delegates within each province were then randomly assigned to an infographic (citizen versus firm) and priming letter (electoral versus upward). Appendix Figure E.2 below depicts the number of delegates assigned to each of the four treatment groups and control within each of the three sampling bins. This figure illustrates that Equation 2, while empirically justified, is theoretically unsatisfactory. Because reinforcement is most likely to occur when delegates possess the same information and are therefore more likely to use such information publicly (Malesky and Todd, 2022), we expect the threat to SUTVA assumptions to be greater when delegates interact with those receiving identical treatments. Crowding out effects are also most likely to occur when delegates receive identical information, such as when a previous speaker has already used the same statistics that a fellow delegate intended to mention.

Because of this theoretical concern, our research design followed Malesky and Todd (2022) in modeling the realized saturation (R) of each treatment. Because delegates in each province were randomly assigned to the four treatment groups, R varies continuously within each province.

Figure E.2: Observations by dosage sampling bin and treatment.



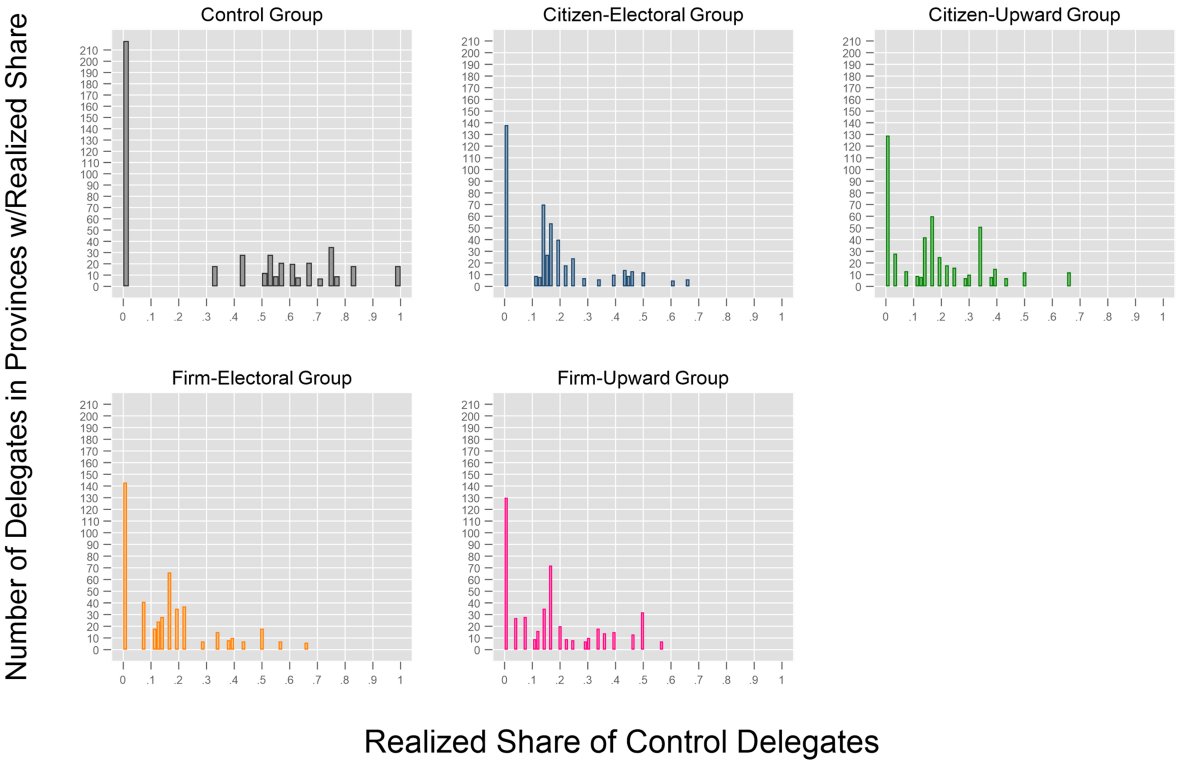
This can be seen in the histograms in Appendix Figure E.3 below. Each panel plots the number of delegates who face a particular realized share of delegates in each treatment condition, which we refer to as realized saturation. This alternative modeling approach can be seen in Equation 4 below. Interpretation of the coefficients is very similar to Equation 2, except that R varies continuously and captures assignment specifically to CE, rather than to one of the four treatments. Again, the model can be extended to control for interactions of all other treatment conditions with their own realized shares (see Equation 5), given important assumptions discussed below.

$$\Pr(Y_i = 1) = \beta_0 + \beta_1 \text{CE}_i + \beta_2 \text{R}_p + \beta_3 \text{CE}_i \times \text{R}_p + \boldsymbol{\delta} \mathbf{X}_i \tag{4}$$

$$\begin{aligned} \Pr(Y_i = 1) = \beta_0 &+ \beta_1 \text{CE}_i + \beta_2 \text{R}_p^{CE} + \beta_3 \text{CE}_i \times \text{R}_p^{CE} \\ &+ \beta_4 \text{CU}_i + \beta_5 \text{R}_p^{CU} + \beta_6 \text{CU}_i \times \text{R}_p^{CU} \\ &+ \beta_7 \text{FE}_i + \beta_8 \text{R}_p^{FE} + \beta_9 \text{FE}_i \times \text{R}_p^{FE} \\ &+ \beta_{10} \text{FU}_i + \beta_{11} \text{R}_p^{FU} + \beta_{12} \text{FU}_i \times \text{R}_p^{FU} + \boldsymbol{\delta} \mathbf{X}_i \end{aligned} \tag{5}$$

An additional concern is that the decision to have a 100% dosage bin, which was made to increase the share of treated delegates and thereby boost statistical power, implies that for delegates in the largest dosage bin, there are no pure control delegates to which the four treatment conditions can be compared. Practically, this means that perfect collinearity prohibits us from

Figure E.3: Histograms of realized treatment shares by treatment group.



estimating Equations 3 and 5 while including all treatment conditions and the 100% dosage bin. Theoretically, the counterfactual for the ATE on the citizen-electoral treatment in the 100% dosage bin is different than the ATE in the other two bins. In the 100% saturation bin, the CE treatment is being compared to the other treatments, while in the other two bins, it is being compared to a pure control. Thus, at the highest levels of dosage/saturation, we cannot differentiate between the effect of increasing the share of treated delegates or the change in the comparison condition. Statistically, this is known as a lack of overlap, which is well known in propensity matching methods (Crump et al., 2009).

Although this problem cannot be solved perfectly, we try two different approaches to estimate the saturation effect. In the first approach (Equations 2 and 4), we collapse all three other treatment conditions into a common control, essentially comparing the CE treatment to an aggregate of the three other treatments and control. While this assumes homogeneity between the treatment conditions, we believe it is justified based on the preponderance of null effects of these treatments in Tables 3, 4, and 5 in the main paper. In other words, the effects of these treatments are not significantly different from the control in their direct effect and therefore can be handled similarly. Of course, this approach assumes, perhaps incorrectly, that spillover did not bias the ATEs in the other treatments to zero. In the second approach (Equations 3 and 5), we run the full set of interactions for all treatments and dosage bins (or realized saturations), but of necessity drop all delegates in the 100% bin and compare the ATE in the 25% bin to that in

the 50% bin. This approach reduces statistical power by reducing the sample size by 46%. This makes finding significant effects less likely, but removes the need to assume homogeneity of the control condition across bins or to assume no spillover from the other treatments.

In sum, this generates four different analyses, which are clarified in Appendix Table E.1 and modeled in Appendix Table E.3. Appendix Table E.1 also notes the related estimations for the direct effects in Appendix Table E.2 that control for saturation levels, but do not use the saturation in interaction terms.

Table E.1: Four analyses of saturation.

	Moderator	
Modeling approach	Assigned dosage bin	Realized treatment saturation
Other treatments as controls	Equation 2 Table E.2, Model 5 Table E.3, Model 1	Equation 4 Table E.2, Model 6 Table E.3, Model 2
Dropping 100% dosage	Equation 3 Table E.2, Model 8 Table E.3, Model 9	Equation 5 Table E.2, Model 9 Table E.3, Model 10

E.4 Addressing saturation and assignment probability in direct effects

As a preliminary step before directly modeling the interaction, Appendix Table E.2 tests the robustness of the direct effects of the results presented in manuscript Table 3 to reasonable adjustments suggested by the research design issues described above. Appendix Table E.2 begins by reproducing the four models presented in the revised main text. Recall that these were (1) an unadjusted model with no controls; (2) a model including a parsimonious set of pre-treatment control variables; (3) a model that controlled for the realized saturation levels directly; and (4) a model that used inverse probability weights to account for the fact that the different sampling bins imposed different probabilities of selection on the delegates, which produces a coefficient that can be considered a national population average.

In Appendix Table E.2, Model 5 replicates Model 2, but adds dosage bin fixed effects (ϕ) for the three different saturation categories (See Equation 6). As described above, because the 100% dosage bin does not have a pure control, one cannot include a model with the other treatment conditions and bin fixed effects in the same estimation due to perfect collinearity. Consequently, we drop all of the other treatment conditions, except CE—the only statistically significant treatment group—treating the delegates in the other treatment conditions as additional controls. Note that this estimation biases the coefficient on CE negatively (toward zero), because in Models 1–4, the other treatment conditions, while insignificant, had effects that were more positive than the control. As a result, the ATE is smaller (9.7 percentage points), but still significant at the 0.1 level in the RI estimation.

$$\Pr(Y_i = 1) = \beta_0 + \beta_1 \text{CE}_i + \boldsymbol{\delta}\mathbf{X}_i + \boldsymbol{\phi}\mathbf{S}_p \tag{6}$$

Table E.2: Manuscript Table 3 robust to saturation design adjustments.

	Table 3 replication				Saturation adjustments					
DV: Spoke	Standard analysis		Saturation adjust.		CU/FE/FU as controls			Dropping 100% bin		
	Unadjusted	Controls	Realized shares	Inverse prob. weights	Assigned bin fixed effects	Realized shares	Inverse prob. weights	Assigned bin fixed effects	Realized shares	Inverse prob. weights
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Citizen-Electoral	0.114	0.120	0.152	0.137	0.097	0.123	0.135	0.136	0.165	0.162
	(0.068)	(0.066)	(0.082)	(0.078)	(0.060)	(0.064)	(0.072)	(0.102)	(0.108)	(0.106)
	[0.041]	[0.033]	[0.048]	[0.038]	[0.057]	[0.035]	[0.030]	[0.126]	[0.110]	[0.098]
Citizen-Upward	0.033	0.013	0.051	0.020				0.054	0.090	0.042
	(0.071)	(0.068)	(0.085)	(0.078)				(0.108)	(0.112)	(0.111)
	[0.288]	[0.392]	[0.273]	[0.370]				[0.318]	[0.254]	[0.354]
Firm-Electoral	0.008	0.011	0.006	-0.006				-0.008	-0.019	-0.027
	(0.068)	(0.066)	(0.085)	(0.075)				(0.112)	(0.122)	(0.116)
	[0.432]	[0.418]	[0.463]	[0.537]				[0.539]	[0.572]	[0.590]
Firm-Upward	0.006	0.048	0.071	0.003				0.041	0.074	-0.044
	(0.069)	(0.066)	(0.086)	(0.070)				(0.117)	(0.122)	(0.105)
	[0.481]	[0.219]	[0.218]	[0.473]				[0.389]	[0.290]	[0.621]
Constant	0.411	0.279	0.302	0.300	0.296	0.308	0.301	0.304	0.405	0.315
	(0.040)	(0.049)	(0.059)	(0.054)	(0.038)	(0.044)	(0.044)	(0.059)	(0.089)	(0.061)
Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dosage bin FEs	No	No	No	No	Yes	No	No	Yes	No	No
Realized saturation	No	No	Yes	No	No	Yes	No	No	Yes	No
IPW	No	No	No	Yes	No	No	Yes	No	No	Yes
N	470	470	470	470	470	470	470	252	252	252
R^2	0.007	0.086	0.092	0.070	0.088	0.086	0.070	0.064	0.075	0.062
RMSE	0.497	0.478	0.479	0.483	0.477	0.477	0.481	0.487	0.487	0.488

Note: OLS coefficients (standard errors) [RI p-values]. Full regression results available in *APSR* Dataverse files for Appendix E.

Model 6 replaces the bin fixed effects with a control variable for the realized saturation of CE, denoted by R above. This produces a 12.3pp ATE, which is significant at the 0.05 level (RI p=0.035). Model 7 tests the robustness of using inverse probability weights as in Model 4, uncovering a 13.5pp ATE that is significant at the 0.05 level (RI p=0.030). The ATEs from Model 6 and 7 are substantively similar, although slightly smaller, to the main models found in Table 3.

Models 8, 9, and 10 address the perfect collinearity issue in a different way by dropping delegates in the 100% dosage bin. This approach has the benefit of ensuring a pure control is available as a comparison group, but severely reduces statistical power by removing 118 (46%) of the 470 observations. Estimations follow the same pattern as the previous three models with bin fixed effects (Model 8), controls for realized saturation (Model 9), and inverse probability weights (Model 10). In all three models, the substantive size of the ATE is larger than in equivalent specifications, ranging from 13.6pp with bin fixed effects to 16.5pp when controlling for realized saturations. However, due to the reduction in power, the RI p-values are slightly over conventional thresholds of 0.1 (RI p-values = 0.126, 0.110, and 0.098 respectively).

These results leave us confident that the findings in manuscript Table 3 are robust to model specification choices in dealing with saturation and differences in treatment assignment probability.

E.5 Modeling saturation through interaction effects

Appendix Table E.3 provides the full modeling of the strategies presented in Equations 1–4, which use interaction effects to model the conditional average treatment effects (CATEs) at different levels of saturation. In Model 1 of Appendix Table E.3, we compare the CE treatment to the combined other treatments and control and interact the treatment with the assigned dosage bin. This analysis is repeated for the other treatment conditions in Models 3, 5, and 7. In Model 2, we continue with the combined control, but interact the CE treatment with the realized saturation of CE in each province (p). This analysis is repeated for the other treatments in Models 4, 6, and 8. In Models 9 and 10, we include the full set of interactions in Equations 2 and 4, but drop delegates in provinces assigned to the 100% dosage bin. Model 9 uses the assigned sampling bins as the moderator (Equation 2), while Model 10 uses the four different realized saturations (Equation 4). For presentation purposes, the table only shows the coefficients of the CE interaction for Models 9 and 10, so they can be more easily compared to Models 1 and 2.

We use speaking as our outcome variable, as it is the outcome studied in manuscript Table 3 and the one with the highest amount of variance, where a significant effect is most likely to be uncovered. To remain consistent with results in the main paper, we present p-values from both an asymptotic analysis and from permutation tests for joint-sharp null effects which test the null that both direct and joint effects are zero for all delegates (Athey, Eckles and Imbens, 2018). As the p-values are reasonably similar, we rely on the asymptotically generated confidence intervals to present the marginal and predicted effects of saturation. Figure E.4 depicts the predicted effects of the CE treatment and control treatment groups in four different specifications and Figure E.5 illustrates the conditional average treatment effects (CATEs). To ease interpretation, these are laid out in precise accordance with Appendix Table E.1 above: the outcomes of Model 1 appear in the northwest corner, Model 2 in the northeast corner, Model 9 in the southwest corner, and Model 10 in the southeast corner.

Table E.3: Interaction models reveal no SUTVA violations in key finding.

	One treatment at a time								Dropping 100% bin	
DV: Spoke	Citizen-Electoral		Citizen-Upward		Firm-Electoral		Firm-Upward		All treatments	
	Assigned bin (1)	Realized share (2)	Assigned bin (3)	Realized share (4)	Assigned bin (5)	Realized share (6)	Assigned bin (7)	Realized share (8)	Assigned bin (9)	Realized share (10)
Treatment	0.175	0.231	0.030	-0.117	-0.053	0.027	-0.233	0.094	0.174	0.595
	(0.119)	(0.059)	(0.843)	(0.264)	(0.723)	(0.749)	(0.244)	(0.483)	(0.136)	(0.010)
	[0.153]	[0.050]	[0.417]	[0.897]	[0.622]	[0.396]	[0.920]	[0.245]	[0.163]	[0.037]
50% dosage bin	-0.042		-0.071		-0.073		-0.104		-0.085	
	(0.521)		(0.289)		(0.269)		(0.121)		(0.313)	
	[0.750]		[0.867]		[0.872]		[0.949]		[0.844]	
100% dosage bin	0.011		0.022		0.015		-0.003			
	(0.843)		(0.680)		(0.791)		(0.956)			
	[0.415]		[0.320]		[0.371]		[0.522]			
Treatment×50% bin	-0.247		-0.009		0.047		0.386		-0.191	
	(0.285)		(0.965)		(0.827)		(0.110)		(0.429)	
	[0.847]		[0.506]		[0.417]		[0.055]		[0.778]	
Treatment×100% bin	-0.091		-0.086		0.015		0.241			
	(0.502)		(0.612)		(0.931)		(0.258)			
	[0.682]		[0.694]		[0.452]		[0.100]			
Realized share		-0.006		0.109		-0.065		0.050		0.118
		(0.971)		(0.455)		(0.683)		(0.767)		(0.741)
		[0.529]		[0.215]		[0.668]		[0.383]		[0.371]
Treatment×share		-0.397		0.504		-0.317		-0.261		-2.201
		(0.291)		(0.275)		(0.411)		(0.510)		(0.029)
		[0.831]		[0.106]		[0.786]		[0.743]		[0.944]
Constant	0.299	0.298	0.323	0.297	0.328	0.329	0.333	0.303	0.338	0.378
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dosage bin FEs	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Realized share	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
CU/FE/FU shares	No	No	No	No	No	No	No	No	Yes	Yes
N	470	470	470	470	470	470	470	470	252	252
R^2	0.090	0.088	0.084	0.084	0.083	0.082	0.088	0.080	0.076	0.098
RMSE	0.478	0.477	0.480	0.478	0.480	0.479	0.479	0.480	0.488	0.485

Note: OLS coefficients (standard errors) [RI p-values]. Full regression results available in *APSR* Dataverse files for Appendix E.

The first thing to notice is that there is no effect of spillover or contamination in any of the models. In Appendix Table E.3, the coefficients on the dosage bins (50% and 100%) are never significantly different from zero. Increasing saturation, either through the sampling bin or realized share, does not lead to increases in speaking activity of control delegates. Delegates in low-saturation provinces are equally as likely to speak as delegates in high saturation provinces with about a 40% speaking rate. In Figure K3, the 95% confidence intervals around the control predictions for each sampling bin overlap. This is true whether the control condition is only the pure control in the lower panel or the aggregate control, including other treatments, in the upper panel. It is also true whether or not the moderator is the assigned sampling bin or realized share.

There is also no evidence of reinforcement, contra Malesky and Todd (2022). The marginal effect of increased saturation is negative and insignificant in Table E.3, implying that as saturation increased, the average treatment effect of the marginal delegate declined. This can also be seen in the declining predicted effects in Figure E.4 as saturation levels increase. One way to reconcile these null findings with previous work concerns the venue in which speaking occurred: while recent studies found reinforcement effects in televised floor debates Malesky and Todd (2022), most delegates spoke only in the closed-door caucuses during the Labor Code debates.

Although highly speculative, both because the CATEs are not significantly different at the 95% confidence level and because there are limited observations in very high realized saturation, the direction of the saturation effects is consistent with a crowding out effect. That is, as the share of treated delegates increased, delegates found it more difficult to find something new to say, particularly in the caucuses. In Model 10, when realized saturation of the CE treatment was less than 20%, delegates receiving the CE treatment were significantly more likely than the control to speak. At higher levels of saturation, there is no difference between CE and control delegates. We emphasize that this conclusion should be treated with caution due to the limited statistical power in each group and consequently large confidence intervals.

E.6 Implications for primary findings

What are the implications of the saturation analysis for the interpretation of results in the main body of the paper, particularly the analysis of speaking in manuscript Table 3? First, the findings in Table 3 are quite robust to specification choices. All alternatives yield substantively similar results. The small differences in ATE size and statistical significance are reasonable and often strengthen the conclusions drawn in the main paper. Second, we believe it is reasonable to assume homogeneity of treatment effects across the three dosage bins and to exclude an interaction term for saturation in the estimations of manuscript Table 3, because all analyses indicate that the CATEs in the sampling bins are statistically indistinguishable from one another. Third, should omitting the interaction effects increase bias, such bias would militate against a significant finding as increased saturation most likely reduces speaking behavior. In a perfect world without contamination, the effect of the citizen-electoral treatment would have been higher, because there would have been less of a possibility that another delegate could steal another delegate's thunder in the debate.

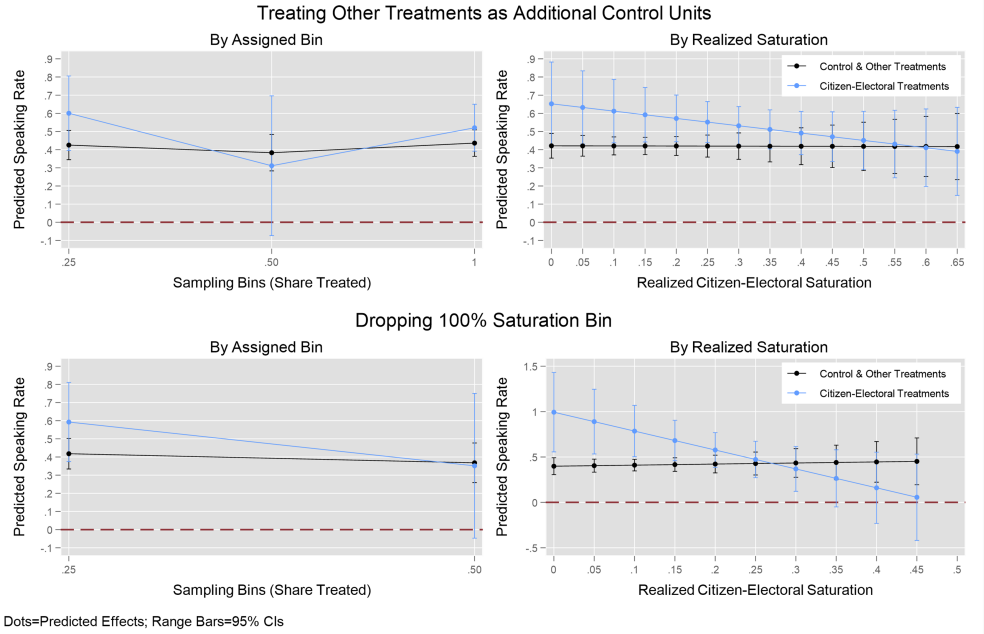


Figure E.4: Predicted saturation effects reveal no spillover or reinforcement, but possible crowding out effects. All continuous (categorical) variables held at their sample means (modes). Full regression results available in *APSR* Dataverse files for Appendix E.

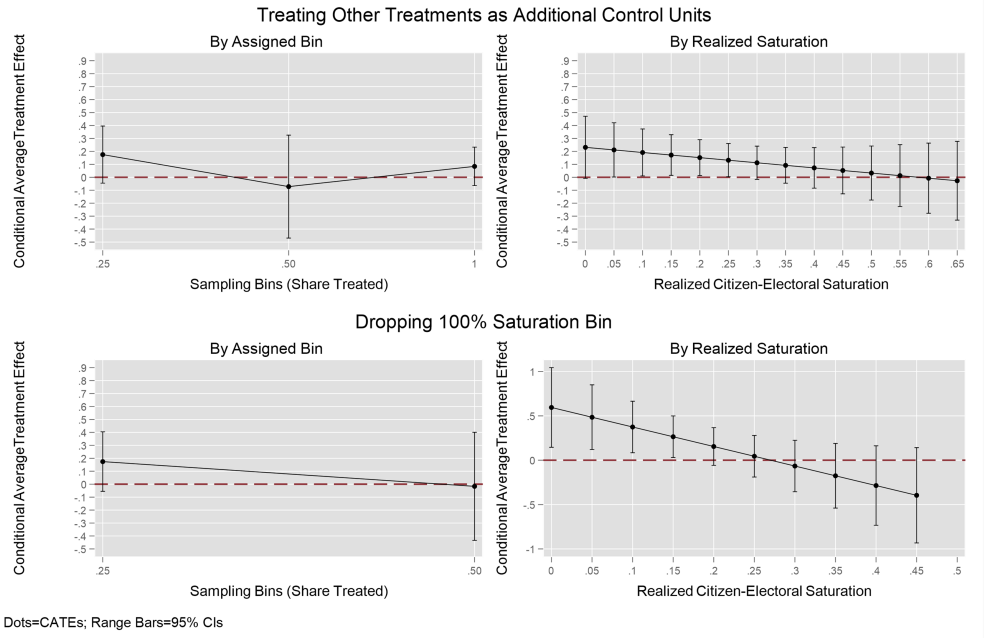


Figure E.5: CATEs reveal no spillover or reinforcement, but possible crowding out effects. All continuous (categorical) variables held at their sample means (modes). Full regression results available in *APSR* Dataverse files for Appendix E.

Table F.1: **Primary results robust to alternative specifications.** Table replicates paper Figure 4 (1) and Table 3 Column 3 (2) and presents alternative specifications with clustered standard errors (3), a probit (4), and a probit with clustered SEs and fixed effects (5). Results remain qualitatively unchanged.

DV: Spoke	Unadjusted	Table 3(3)	Cluster SEs	Probit	All
Citizen-Electoral=1	0.114^{*}	0.152^{*}	0.152^{***}	0.164^{***}	0.130^{***}
	(0.068)	(0.082)	(0.007)	(0.004)	(0.005)
Citizen-Upward=1	0.033	0.051	0.051	0.057^{**}	0.022
	(0.071)	(0.085)	(0.024)	(0.023)	(0.023)
Firm-Electoral=1	0.008	0.006	0.006	0.006	0.019
	(0.068)	(0.085)	(0.031)	(0.035)	(0.026)
Firm-Upward=1	0.006	0.071	0.073	0.077	0.063
	(0.069)	(0.086)	(0.071)	(0.071)	(0.063)
Constant/Probability	0.411^{***}	0.302^{***}	0.302^{**}	0.438	0.438
	(0.040)	(0.059)	(0.034)		
Model	OLS	OLS	OLS	Probit	Probit
Block FEs	No	Yes	Yes	Yes	Yes
Saturation controls	No	Yes	Yes	Yes	No
Cluster SEs	No	No	Yes	Yes	Yes
Dosage bin FEs	No	No	No	No	Yes
N	470	470	470	470	470
Clusters			3	3	3
R^2	0.007	0.092	0.092	0.070	0.067
RMSE	0.497	0.479	0.479		

Note: Standard errors in parentheses (***p<0.01, **p<0.05, *p<0.1). Full regression results available in *APSR* Dataverse files for Appendix F.

F Robustness to alternative specifications

Adopting a saturation design and randomization inference allowed us to replicate prior results (Todd et al., 2021; Malesky and Todd, 2022) and accumulate scientific knowledge, while simultaneously handling our provincially clustered treatments despite a small sample size. Despite these arguments in favor of our empirical strategy, other specification choices (e.g., dosage bin-clustered standard errors, dosage bin fixed effects, or probit) are worth examining. Consequently, Tables F.1, F.2, F.3, F.4, F.5, and F.6 present robustness tests for all of our primary outcome variables. The first five of these tables juxtapose the main paper's unadjusted (1) and adjusted (2) results with five alternative specifications: (3) dosage bin-clustered standard errors, (4) probit, and (5) clustered standard errors, fixed effects, and probit. Table F.6, which examines a non-dichotomous outcome, omits the probit specifications. In each case, our primary results are robust to alternative specifications; generally, our preferred specification is the most conservative of the options.

Table F.2: **Criticism results robust to alternative specifications.** Table replicates paper Figure 5 (1) and Table 4 Column 2 (2) and presents alternative specifications with clustered standard errors (3), a probit (4), and a probit with clustered SEs and fixed effects (5). Results remain qualitatively unchanged.

DV: Critical	Unadjusted	Table 4(2)	Cluster SEs	Probit	All
Citizen-Electoral=1	0.122^{**}	0.127^{*}	0.127	0.140^{**}	0.140^{***}
	(0.061)	(0.074)	(0.052)	(0.058)	(0.046)
Citizen-Upward=1	0.064	0.025	0.025	0.029	0.058
	(0.063)	(0.077)	(0.033)	(0.039)	(0.050)
Firm-Electoral=1	0.068	0.043	0.043	0.049	0.084^{**}
	(0.061)	(0.077)	(0.045)	(0.050)	(0.036)
Firm-Upward=1	0.025	0.046	0.046	0.050	0.067
	(0.062)	(0.078)	(0.080)	(0.091)	(0.071)
Constant/Probability	0.228^{***}	0.111^{**}	0.111^{*}	0.274	0.274
	(0.036)	(0.054)	(0.037)		
Model	OLS	OLS	OLS	Probit	Probit
Block FEs	No	Yes	Yes	Yes	Yes
Saturation controls	No	Yes	Yes	Yes	No
Cluster SEs	No	No	Yes	Yes	Yes
Dosage bin FEs	No	No	No	No	Yes
N	470	470	470	470	470
Clusters			3	3	3
R^2	0.010	0.072	0.072	0.060	0.059
RMSE	0.446	0.435	0.435		

Note: Standard errors in parentheses (***p<0.01, **p<0.05, *p<0.1). Full regression results available in *APSR* Dataverse files for Appendix F.

Table F.3: **Pro-labor results robust to alternative specifications.** Table replicates paper Figure 5 (1) and Table 4 Column 4 (2) and presents alternative specifications with clustered standard errors (3), a probit (4), and a probit with clustered SEs and fixed effects (5). Results remain qualitatively unchanged.

DV: Pro-labor	Unadjusted	Table 4(4)	Cluster SEs	Probit	All
Citizen-Electoral=1	0.093^{*}	0.058	0.058	0.070^{**}	0.084^{***}
	(0.054)	(0.065)	(0.030)	(0.030)	(0.026)
Citizen-Upward=1	0.065	0.014	0.014	0.018	0.028
	(0.056)	(0.067)	(0.047)	(0.053)	(0.041)
Firm-Electoral=1	0.028	-0.002	-0.002	0.001	0.006
	(0.054)	(0.067)	(0.055)	(0.056)	(0.036)
Firm-Upward=1	0.033	0.019	0.019	0.026	0.049
	(0.054)	(0.068)	(0.084)	(0.093)	(0.069)
Constant/Probability	0.157^{***}	0.039	0.039^{*}	0.193	0.193
	(0.031)	(0.047)	(0.013)		
Model	OLS	OLS	OLS	Probit	Probit
Block FEs	No	Yes	Yes	Yes	Yes
Saturation controls	No	Yes	Yes	Yes	No
Cluster SEs	No	No	Yes	Yes	Yes
Dosage bin FEs	No	No	No	No	Yes
N	470	470	470	470	470
Clusters			3	3	3
R^2	0.007	0.091	0.091	0.083	0.082
RMSE	0.396	0.382	0.382		

Note: Standard errors in parentheses (***p<0.01, **p<0.05, *p<0.1). Full regression results available in *APSR* Dataverse files for Appendix F.

Table F.4: **Pro-firm results robust to alternative specifications.** Table replicates paper Figure 5 (1) and Table 4 Column 8 (2) and presents alternative specifications with clustered standard errors (3), a probit (4), and a probit with clustered SEs and fixed effects (5). Results remain qualitatively unchanged.

DV: Pro-firm	Unadjusted	Table 4(8)	Cluster SEs	Probit	All
Citizen-Electoral=1	−0.007	−0.012	−0.012	−0.005	−0.002
	(0.031)	(0.039)	(0.017)	(0.010)	(0.007)
Citizen-Upward=1	0.012	0.005	0.005	0.008	0.015
	(0.033)	(0.040)	(0.015)	(0.017)	(0.015)
Firm-Electoral=1	0.017	0.008	0.008	0.012	0.021
	(0.031)	(0.040)	(0.037)	(0.035)	(0.030)
Firm-Upward=1	−0.032	−0.019	−0.019	−0.018	−0.019
	(0.032)	(0.041)	(0.054)	(0.060)	(0.053)
Constant/Probability	0.057***	0.044	0.044*	0.553	0.553
	(0.018)	(0.028)	(0.011)		
Model	OLS	OLS	OLS	Probit	Probit
Block FEs	No	Yes	Yes	Yes	Yes
Saturation controls	No	Yes	Yes	Yes	No
Cluster SEs	No	No	Yes	Yes	Yes
Dosage bin FEs	No	No	No	No	Yes
N	470	470	470	470	470
Clusters			3	3	3
R^2	0.005	0.026	0.026	0.050	0.048
RMSE	0.229	0.229	0.229		

Note: Standard errors in parentheses (***p<0.01, **p<0.05, *p<0.1). Full regression results available in *APSR* Dataverse files for Appendix F.

Table F.5: **Results on congruence between speeches and resulting law robust to alternative specifications.** Table replicates paper Figure 5 (1) and Table 4 Column 6 (2) and presents alternative specifications with clustered standard errors (3), a probit (4), and a probit with clustered SEs and fixed effects (5). Results remain qualitatively unchanged.

DV: Reflected	Unadjusted	Table 4(6)	Cluster SEs	Probit	All
Citizen-Electoral=1	0.072^{*}	0.082	0.082	0.097^{**}	0.081^{*}
	(0.042)	(0.050)	(0.032)	(0.047)	(0.047)
Citizen-Upward=1	-0.015	-0.061	-0.061	-0.105^{***}	-0.091^{***}
	(0.044)	(0.052)	(0.033)	(0.032)	(0.027)
Firm-Electoral=1	0.016	-0.006	-0.006	-0.006	-0.002
	(0.042)	(0.052)	(0.020)	(0.006)	(0.025)
Firm-Upward=1	0.001	0.000	0.000	-0.018	-0.010
	(0.042)	(0.053)	(0.023)	(0.017)	(0.018)
Constant/Probability	0.119^{***}	0.050	0.050	0.174	0.174
	(0.024)	(0.036)	(0.030)		
Model	OLS	OLS	OLS	Probit	Probit
Block FEs	No	Yes	Yes	Yes	Yes
Saturation controls	No	Yes	Yes	Yes	No
Cluster SEs	No	No	Yes	Yes	Yes
Dosage bin FEs	No	No	No	No	Yes
N	470	470	470	470	470
Clusters			3	3	3
R^2	0.009	0.096	0.096	0.098	0.094
RMSE	0.306	0.294	0.294		

Note: Standard errors in parentheses (***p<0.01, **p<0.05, *p<0.1). Full regression results available in *APSR* Dataverse files for Appendix F.

Table F.6: `Wordscores` **results robust to alternative specifications.** Table presents unadjusted results (1), Table 5 Column 2 (2), and an alternative specification with clustered standard errors (3). Results remain qualitatively similar.

DV: Reflected	Unadjusted	Table 5(2)	Cluster SEs
Citizen-Electoral=1	14.092^{*}	25.595^{***}	25.595^{***}
	(7.356)	(9.060)	(1.378)
Citizen-Upward=1	15.469^{**}	25.329^{***}	25.329
	(7.861)	(9.630)	(13.587)
Firm-Electoral=1	-17.637^{**}	-6.606	-6.606
	(7.743)	(9.901)	(3.330)
Firm-Upward=1	-16.146^{**}	-3.876	-3.876
	(7.833)	(10.105)	(6.619)
Constant	44.129^{***}	48.864^{***}	48.864^{***}
	(4.579)	(6.955)	(0.820)
Model	OLS	OLS	OLS
Block FEs	No	Yes	Yes
Saturation controls	No	Yes	Yes
Cluster SEs	No	No	Yes
Dosage bin FEs	No	No	No
N	470	470	470
Clusters			3
R^2	0.053	0.108	0.108
RMSE	55.47	54.26	54.26

Note: Standard errors in parentheses (***p<0.01, **p<0.05, *p<0.1).
Full regression results available in *APSR* Dataverse files for Appendix F.

G Multiple outcomes adjustments

In this section, we address the heightened possibility of Type I errors given repeated hypothesis testing. Because it assumes independence of outcomes, a simple Bonferroni correction for Family Wise Error Rates (FWER) is inappropriate here. Our outcomes are linked theoretically: a delegate must speak to be critical, and subjectively coded pro-labor speeches are likely correlated with their pro-labor Wordscores.

To account for the correlation structure in our outcomes and in recognition that our p-values were constructed using randomization inference, we utilize our re-randomized treatment assignments to calculate the probability of obtaining a given number of significant effects under the null hypothesis of no effects for all delegates and all outcomes.[1] Specifically, we analyzed five outcomes—delegate spoke, delegate's speech was critical, delegate's speech was pro-labor, delegate's speech was reflected in law, and delegate's speech's pro-labor Wordscore—across two specifications—with and without the realized saturation shares (e.g., manuscript Table 3, Models 2 and 3). Setting alpha to 0.05 and examining our RI-based p-values, six of the 10 results are significant: spoke, critical, pro-labor, reflected, spoke-saturation, and Wordscores-saturation. We then set out to determine how likely we were to find six significant effects.

For each of our 1,000 re-randomized treatment assignments, we then counted how many of these 10 ATEs were significant at the 0.05 level. From this, we calculated the probability of finding z or more significant effects and identified the smallest z such that Pr(significant effects $\geq z|z, \alpha = 0.05) < 0.05$. In our data, $z = 4$. In other words, at that alpha level, we were statistically unlikely (p<0.05) to encounter four or more significant effects under the null hypothesis. At the same time, that means it was plausible (p≥0.05) to find a maximum of three significant effects, all of which were false positives by virtue of the re-randomization. Subtracting the three false positives from the six significant ATEs yields a minimum bound on true positives; the three effects with the lowest RI p-values—whether delegates spoke, spoke critically, and saw speech reflected in law—are indeed significant at the 0.05 level.

Relaxing our alpha level to 0.10 and reexamining our RI p-values, another two results attain significance, for a total of eight out of 10. Taking the same approach, we find that $z = 4$, four or more significant results would be implausible, and a maximum of three significant effects would be plausible. Subtracting these three false positives from the eight significant ATEs yields five as the minimum bound on true positives. Sorting the results by p-value, we identify critical, reflected, spoke, Wordscores-saturation, and pro-labor as true positives that survive our multiple outcomes adjustment.

By repeating this procedure under a fine grid of alphas from 0.005 to 0.99, we are able to revise the p-values upward for each of the 10 results in a way that accounts for the correlation structure across outcomes. These adjusted p-values are presented in Table G.1. The distribution of true positives across alpha levels is presented (along with *unadjusted* RI p-values) in Figure G.1.

As an additional test, we subject the ten Citizen-Electoral ATEs for delegate speaking from Appendix Table E.2 to the same multiple outcomes adjustment procedure. The results are presented in Table G.2 and G.2.

[1] Our approach is similar in spirit to Westfall and Young (1993).

Table G.1: Primary results are robust to multiple outcomes adjustment accounting for the correlation structure across outcomes.

DV	Table (model)	Realized saturation	ATE	asymptotic SE	unadjusted RI p-value	adjusted RI p-value
Spoke	3(2)	No	0.120	(0.066)	0.033**	0.042**
Critical	4(1)	No	0.131	(0.060)	0.011**	0.033**
Pro-labor	4(3)	No	0.095	(0.052)	0.042**	0.064*
Reflected	4(5)	No	0.076	(0.040)	0.027**	0.037**
Wordscores	5(3)	No	14.973	(12.299)	0.102	1.000
Spoke	3(3)	Yes	0.152	(0.082)	0.048**	0.065*
Critical	4(2)	Yes	0.127	(0.074)	0.064*	1.000
Pro-labor	4(4)	Yes	0.060	(0.065)	0.176	1.000
Reflected	4(6)	Yes	0.082	(0.050)	0.060*	0.176
Wordscores	5(4)	Yes	27.140	(15.486)	0.037**	0.054*

Note: Full results available in *APSR* Dataverse files for Appendix G.

Figure G.1: **True positives as a function of alpha level.** Full results available in *APSR* Dataverse files for Appendix G.

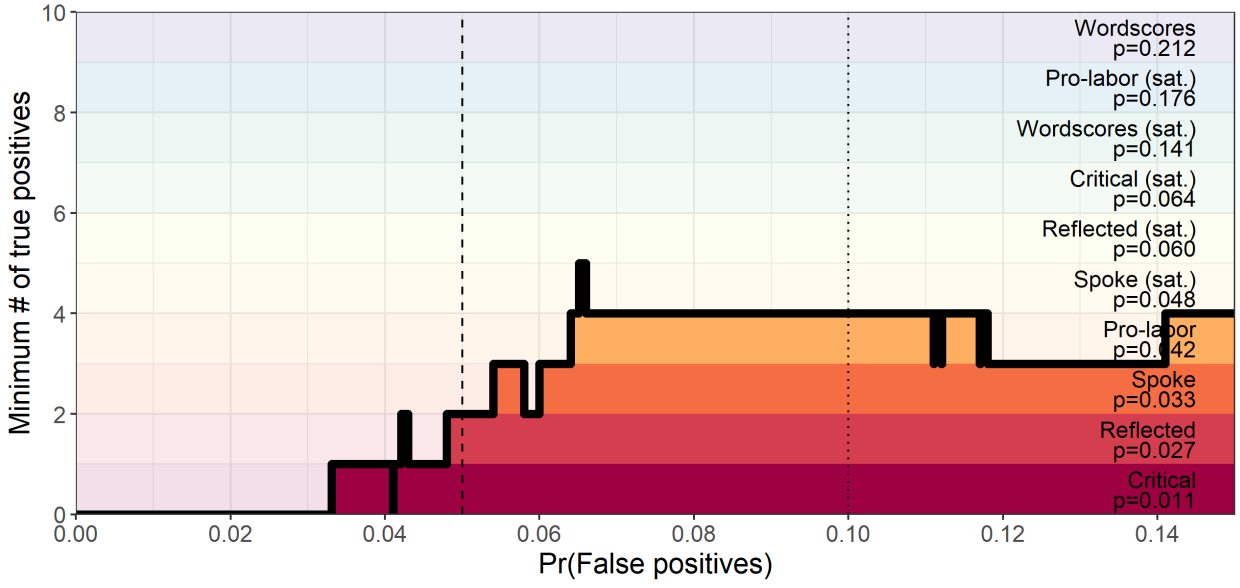
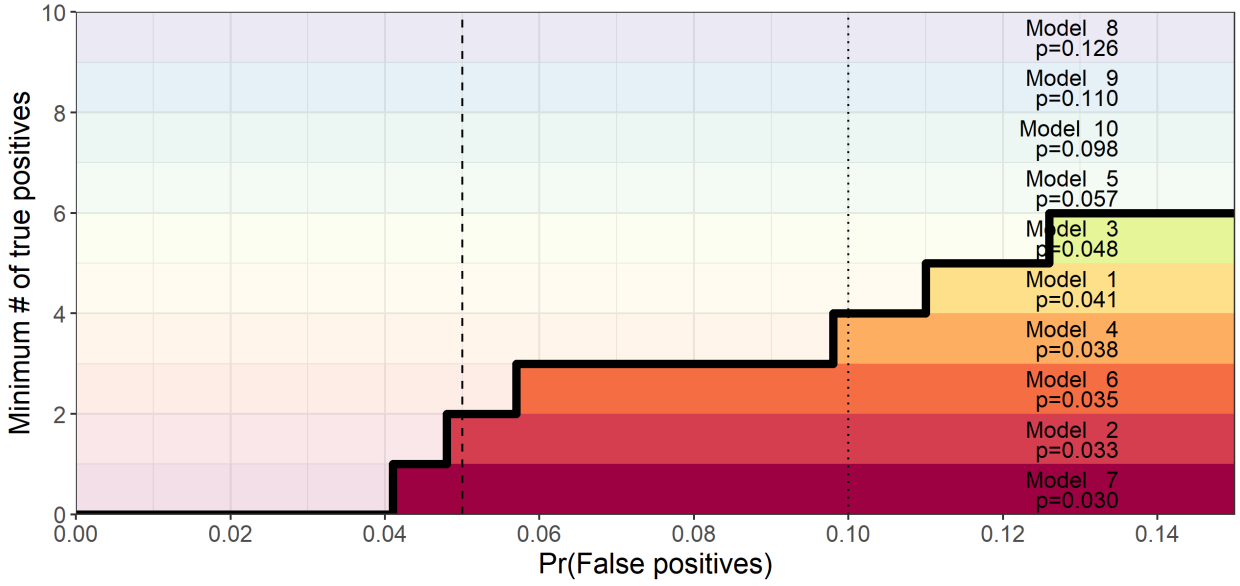


Table G.2: Covariate-adjusted and nationally representative ATEs are robust to multiple outcomes adjustment accounting for the correlation structure across outcomes.

DV: Spoke Specification	App. E (model)	Saturation adjustment	ATE	asymptotic SE	unadjusted RI p-value	adjusted RI p-value
Unadjusted	(1)	None	0.114	(0.068)	0.041**	0.110
Adjusted	(2)	None	0.120	(0.066)	0.033**	0.048**
Realized shares	(3)	Shares	0.152	(0.082)	0.048**	0.126
IPW	(4)	IPW	0.137	(0.078)	0.038**	0.098*
Bin FE	(5)	CU/FE/FU	0.097	(0.060)	0.057*	1.000
Realized shares	(6)	CU/FE/FU	0.123	(0.064)	0.035**	0.057*
IPW	(7)	CU/FE/FU	0.135	(0.072)	0.030**	0.041**
Bin FE	(8)	Drop 100%	0.136	(0.102)	0.126	1.000
Realized shares	(9)	Drop 100%	0.165	(0.108)	0.110	1.000
IPW	(10)	Drop 100%	0.162	(0.106)	0.098*	1.000

Note: Full results available in *APSR* Dataverse files for Appendix G.

Figure G.2: **True positives as a function of alpha level.** Full results available in *APSR* Dataverse files for Appendix G.



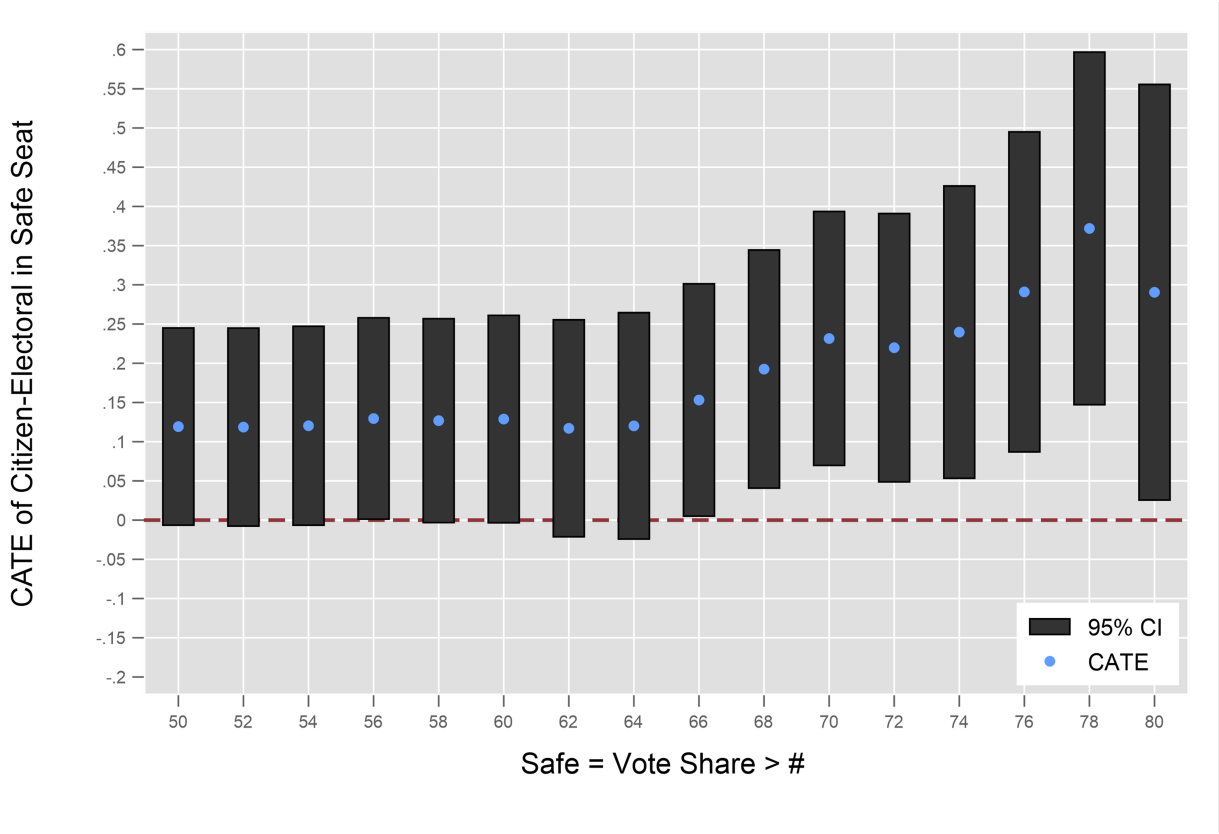


Figure H.1: Safe-seat delegates were more responsive than those in competitive seats, regardless of threshold. Full regression results available in *APSR* Dataverse files for Appendix H.

H Alternative thresholds for competitiveness

To distinguish delegates elected in "competitive" seats from those in "safe seats," we followed the measurement strategy of Todd et al. (2021). In that paper, the competitively elected are defined as those delegates whose vote shares fall below the median vote share for all elected delegates. Indeed, because both papers examine delegates from the 14th VNA (2016–2021), following their measure implies that exactly the same delegates are classified as competitively elected: all those whose vote share falls below 72.4%. By maintaining consistency with existing work on the VNA, we avoid arbitrarily adopting a "convenient" threshold.

Nonetheless, to demonstrate that our results are not overly sensitive to our measure of competitiveness, Figure H.1 presents safe seat CATEs and asymptotic 95% confidence intervals under various measurement thresholds from 50% to 80%. Our original CATE of 0.22 indicates that the speaking rate for safe-seat delegates was 22 percentage points higher than control. The ATE is positive and significant at all thresholds greater than 64%. Although the effect size declines, it remains positive and very close to statistical significance as we reduce the threshold to 50%, at which point the ATE is 12 percentage points and significant at the 0.10 level. In short, our results are relatively robust to alternative thresholds for competitiveness, and safe-seat delegates were more responsive to the treatment than those in competitive seats.

I Background on Resolution 27/2012/QH13

The document referenced in the upward incentive primes, Resolution 27/2012/QH13, is not a legislative document from the VNA. Rather, it was issued by the VNA Standing Committee, a permanent body controlling VNA activities when the latter is not in session. The Chairman of the VNA Standing Committee, who is also the fourth-ranked member of the VCP's Politburo, presides over VNA sessions, authenticates VNA output, and liaises with other party-state organs. That Resolution 27 was signed by VNA Standing Committee Chairman Nguyen Sinh Hung indicates it was approved at the highest levels of the VCP. Evidence of this is plain in Resolution 27's use of language from the VCP Central Committee's 2011 Political Report to the VCP Party Congress. Specifically, the Political Report called for the VNA to

> [r]enew organization and operation of the National Assembly; ensure that the National Assembly is in essence the highest representative body of the people, the highest power body of the State; improve the mechanisms of National Assembly elections so that voters can select and elect to Parliament representatives who really represent them; improve the quality of the National Assembly to increase full-time representatives to a reasonable number; develop a mechanism to support National Assembly members in linking closely to and being responsible to voters...(VCP Central Committee, 2011)

When delegates received the upward incentive prime, they would have immediately recognized Resolution 27 as the means to achieving ends laid out in the quoted Political Report.

J Contents of Dataverse supplemental information

Additional information which could not be included in these online appendices is available in the *APSR* Dataverse files which accompany this paper. The supplemental information, available at https://doi.org/10.7910/DVN/JPVEX5, includes the following:

References

Athey, Susan, Dean Eckles, and Guido W. Imbens. 2018. "Exact p-Values for Network Interference." *Journal of the American Statistical Association* 113 (521): 230–240.

Baird, Sarah, J. Aislinn Bohren, Craig McIntosh, and Berk Özler. 2014. *Designing Experiments to Measure Spillover Effects*. New York: World Bank.

Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik. 2009. "Dealing with Limited Overlap in Estimation of Average Treatment Effects." *Biometrika* 96 (1): 187–199.

Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton & Company.

Malesky, Edmund J., and Jason Douglas Todd. 2022. "Experimentally Estimating Safety in Numbers in a Single-Party Legislature." *The Journal of Politics* p. `https://doi.org/10.1086/716967`.

Sinclair, Betsy, Margaret McConnell, and Donald P. Green. 2012. "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments." *American Journal of Political Science* 56 (4): 1055–1069.

Todd, Jason Douglas, Edmund J. Malesky, Anh Tran, and Anh Le. 2021. "Testing Legislator Responsiveness to Citizens and Firms in Single-Party Regimes: A Field Experiment in the Vietnamese National Assembly." *The Journal of Politics* 83 (4): 1573–1588.

VCP Central Committee. 2011. "Political Report to the 11th National Party Congress." Vietnam Law & Legal Forum. `https://tinyurl.com/VCP-political-report`

Westfall, Peter H., and S. Stanley Young. 1993. *Resampling-based Multiple Testing: Examples and Methods for p-value Adjustment*. New York: John Wiley & Sons.