ONLINE APPENDICES

Experimentally Estimating Safety in Numbers in a Single-Party Legislature

A.	Randomization	2
B.	Infographics	3
C.	Examples of VNA speeches	5
D.	Randomization inference	7
E.	Summary statistics	7
F.	Tabular results for Figures 1 and 2	9
G.	Graphical presentation of spillover effects	11
H.	Saturation results by original treatment conditions	13
I.	Figure 1 (left panel) without central nominees	14
J.	Figure 2 Chow tests	15

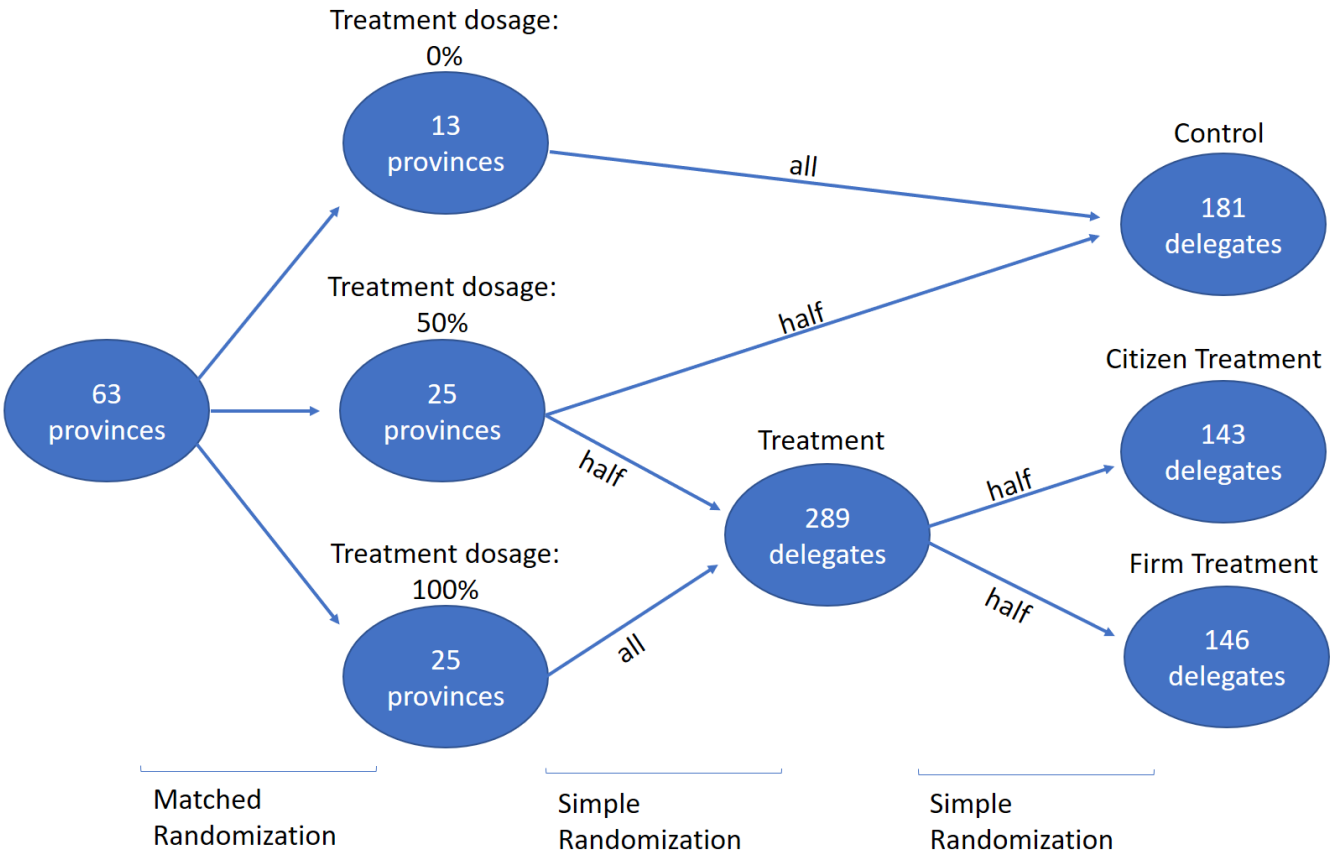

Figure A.1: A three-stage randomization procedure assigned delegates to treatment.

A Randomization

Because all downstream outcomes are potentially contaminated by group caucusing, we followed Baird et al. (2014) in adopting a saturation design with three randomizations across two levels (see Figure A.1). In the first stage, we used a genetic algorithm in the `R` package `Matching` (Sekhon, 2011) to match 63 provinces into three dosages on the basis of 11 covariates: (1) % full-time delegates, (2) % centrally-nominated delegates, (3) % competitively elected delegates, (4) % delegates serving in local assemblies, (5) delegation size, (6) 2016 PAPI score, (7) 2016 PCI score, (8) 2014 GDP per capita, and (9) 2016 fiscal transfers.

All delegates representing provinces assigned to the 100% dosage were assigned to the treatment condition, and all delegates from provinces in the 0% dosage were assigned to the control condition. One simple randomization assigned all delegates from the 50% dosage provinces to one of two conditions, control or treatment. In the third stage, a second simple randomization assigned each delegate in the treatment condition to one of two arms, citizen or firm. After randomization,

approximately 40% of delegates were assigned to the control condition, with around 30% assigned to each treatment arm. The approach generates six different treatment groups: control delegates in 0% and 50% provinces, citizen-treated delegates in 50% and 100% provinces, and firm-treated delegates in 50% and 100% provinces.

B Infographics

Figure B.1 displays English translations of the infographics delivered to treated delegates in Nam Dinh province. In recognition of the widely varying educational backgrounds of the delegates themselves, the infographics were kept simple: a title, five percentages with accompanying illustrations and textual explanations, and footnoted source information. Some items reference specific articles scheduled for debate at the VNA session. Citizen and firm Bullet 1 reference Articles 27–29 on the goals and quality of general and primary education. Firm Bullets 2–3 provide information regarding Article 31 on vocational education. Citizen Bullet 2 provides information on Article 67 concerning school infrastructure funding. Citizen Bullets 3–4 reference Article 80 on fostering professional instructors and Article 70 on teacher morality. Citizen Bullet 4 also connects to Article 105 on adequate compensation for teachers as a way to stave off informal charges to students. Other items, including citizen Bullet 5 and firm Bullets 4–5, describe more general perceptions of education in the province and its economic impact, both of which were related to the debate. All of these cards convey preferences for change, which might also be interpreted as grievances over current outcomes.

Printed infographics were delivered in sealed envelopes containing a short, explanatory note on letterhead from IPPM. To be clear, the infographics differ by province, as well as treatment arm. Data and analyses in online Appendix 6 of Todd et al. (2021), however, examine the relative strengths of the citizen and firm treatments across all 63 provinces, finding little heterogeneity across provinces. While the specific numbers for each province differ slightly, overall scores exhibit little variance. The general message of the infographics was designed to be quite clear—large portions of citizens and firms were dissatisfied with the quality of education and personnel in their provinces and were seeking policy improvements to rectify the situation.

Todd et al. (2021) present qualitative evidence that delegates received the cards, interpreted them as requests for policy improvement, and used them in the debate sessions. Although two delegates directly quoted statistics from the cards in floor and caucuses debates, the impact of the

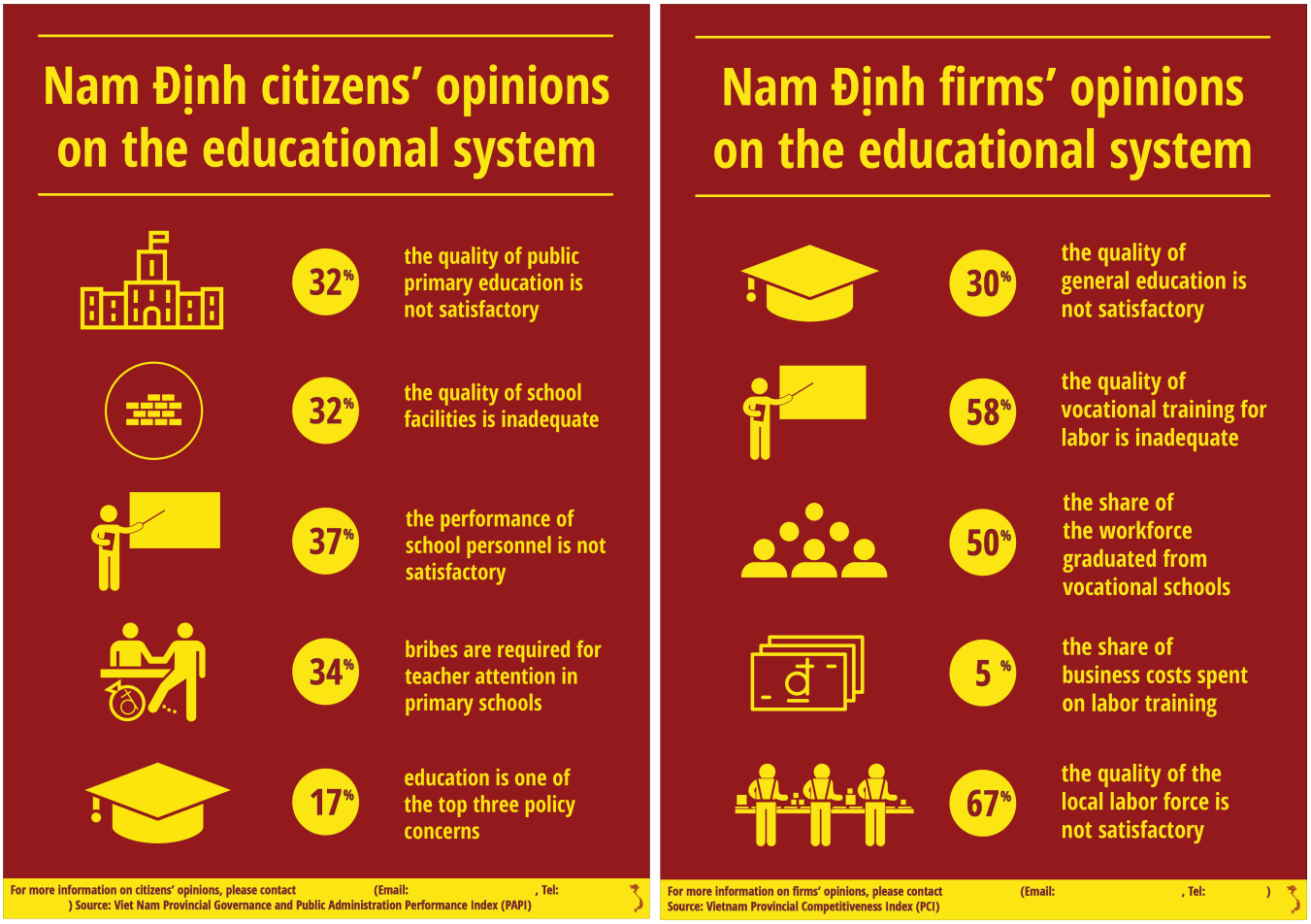

Figure B.1: **Sample infographics for citizens and firms.** Two sets of statistics—one each for citizens and firms—were calculated for each of Vietnam's 63 provinces.

cards was generally more subtle and technocratic as delegates requested amendments to specific words or phrases in the law that reflected the preferences of firms and citizens.[1]

[1]In appendix C, we provide a selection of quotes from delegates in all three treatment groups and legislative fora to illustrate what delegates are saying about the VLOE, paying particular attention to the clauses primed by the cards. The quotes give a sense of the technocratic nature of the debate, illustrating delegates' focus on clause-level details and lack of showboating.

C Examples of VNA Speeches

Phan Thai Binh, Quang Nam Province, Caucus #15, Citizen Treatment

[Article 77, 91] Through meetings with teacher constituents in Quang Nam province, they expressed objection that teachers may possess bachelor degree in relevant disciplines with a pedagogic training certificate. I also agreed to review the standardized educational qualifications for teachers as well as the reality and unemployment. I am of the view that pedagogical schools would best train teachers.

Qua tiếp xúc cử tri với đội ngũ giáo viên tỉnh Quảng Nam, tất cả nhà giáo đều đề nghị đối với giáo viên sư phạm trong luật có nêu là bỏ đi việc tốt nghiệp ngành khác nhưng được bồi dưỡng chuyên môn sư phạm làm giáo viên. Tôi cũng cho rằng cái này cũng cần đánh giá, khảo sát lại thực trạng và tình trạng thất nghiệp. Theo tôi để chuyên sâu về giáo dục và chất lượng giáo dục thì nên chỉ có các trường sư phạm đào tạo giáo viên là tốt nhất.

Tran Kim Yen, Ho Chi Minh City, Caucus #2, Citizen Treatment

[Article 31] **Voters** are of the view that the whole learning process be assessed to decide if a pupil should get an upper secondary school diploma. It is expensive to organize an exam to determine high school graduation qualifications. I propose that upper secondary school students who meet the requirement would receive the diplomas without having to take graduation exams.

[Đ31] Ý kiến **cử tri** mong muốn phải xem xét quá trình học tập của các em để đánh giá việc có được cấp bằng tốt nghiệp trung học phổ thông không. Việc tổ chức một kỳ thi để xác định tốt nghiệp trung học phổ thông hiện nay rất tốn kém. Tôi đề xuất "có đủ điều kiện", còn điều kiện như thế nào thì có quy định cụ thể thì được cấp bằng, không nhất thiết phải được dự thi.

Nguyen Phi Long, Binh Dinh Province, Caucus #6, Citizen Treatment

[Article 101] Public pre-schools are overly crowded, ..., non-public schools are massively open without assessment and evaluation, many are not unregistered, teachers are not trained and do not meet standards. Thus, **Article 101** should specify the percentage of the budget for pre-school education, especially in remote and border areas, islands of export processing zones. I request the state to prioritize this education level, allocate budget and make it a policy in order to set the ground for investing budget as well as land allocation ... Secondly, there must also be investment in teachers, preschool teachers must have certain status in the education system, in the system of civil servants, officials and managers of the education sector ... It is also necessary to identify critical areas that the State must invest such as industrial parks and export processing zones where a huge number of young workers at the age of reproduction are working, ... as well as remote areas, where socialization of education is not a practical option.

[Đ101] Các trường [mầm non] công lập quá tải,… trường ngoài công lập mở ồ ạt, không có kiểm định đánh giá, nhiều trường tự lập, giáo viên không được đào tạo, không đạt chuẩn. Theo đó, điều 101 cần ghi rõ phần trăm trong ngân sách đầu tư cho giáo dục mầm non đặc biệt ở vùng sâu, vùng xa, biên giới, hải đảo các khu công nghiệp khu chế xuất. Tôi đề nghị Nhà nước ưu tiên đầu tư cho cấp bậc này, giành một nguồn kinh phí nhất định, vì nếu có chính sách thì ngoài điều đầu tư về tiền bạc thì cũng còn phải quy hoạch đất đai... Thứ hai là cũng phải đầu tư về giáo viên, giáo viên mầm non phải có vị trí nhất định trong hệ thống giáo dục, trong hệ thống cán bộ công chức, viên chức của ngành giáo dục, của cán bộ quản lý giáo dục... Tôi đề nghị Nhà nước phải dành kinh phí thích đáng để đầu tư cấp bậc học này. Phải quy định những chỗ trọng điểm là nhà nước phải đầu tư, các khu công nghiệp, khu chế xuất hiện nay, cơ bản là các thanh niên công nhân đều trong độ tuổi xây dựng gia đình... Hay xã hội hóa nhà trẻ mẫu giáo ở vùng sâu vùng xa thì làm sao xã hội hóa được, cái này nhà nước phải dành ngân sách để đầu tư.

Chau Quang Dao, Kien Giang, Caucus #19, Firm Treatment

We are worried about this number. 67% of firms say that vocational training is low quality and 73% of businesses said that the quality of local labor was not satisfactory. Thus, we have made great efforts in vocational training of rural labor and vocational training. But the businesses cost of labor training accounts for 5% of total costs. As such, we still have many reasons to believe the quality of occupational labor that is not satisfactory.

Con số này chúng ta rất lo. 67% doanh nghiệp cho rằng chất lượng đào tạo nghề còn kém, 73% doanh nghiệp cho rằng chất lượng lao động địa phương chưa đạt yêu cầu. Như vậy chúng ta đã nỗ lực rất nhiều trong việc đào tạo nghề lao động nông thôn rồi đào tạo các ngành nghề. Nhưng mà việc chi phí kinh doanh dành cho đào tạo lao động chiếm 5%. Như vậy, chúng ta còn rất nhiều nguyên nhân dẫn đến chất lượng lao động nghề nghiệp chúng ta chưa đạt yêu cầu.

Le Quang Huy, Nghe An Province, Caucus #14, Firm Treatment

In practice, companies have broadly applied advance technologies, requiring workers to obtain new skills. For instance, pie manufacturers in Soc Trang, including small businesses, and textile enterprises have replaced manual labor with automatic factory.

một thực tế là gần đây doanh nghiệp đã ứng dụng tiến bộ kỹ thuật rất mạnh mẽ, đòi hỏi người lao động phải có rất nhiều kỹ năng mới như bánh pía Sóc Trăng kể cả doanh nghiệp nhỏ đã vào làm, một loạt doanh nghiệp dệt may đã đưa thiết bị vào và toàn bộ lao động thủ công bị sa thải, phải chuyển đổi sang kỹ năng mới.

Ly Tiet Hanh, Binh Dinh Province, Caucus #6, Control Group

[Article 27] From practical experience, we concur that the goals of general education have been reasonably designed; however, it is necessary to ensure the integration and connection between education levels ... **Voters** are also interested in professional orientation for students from general education level ... I suggest that the law emphasizes on and provides specific solutions for this matter.

[Đ27] Thực tế công tác tình hình địa phương, chúng tôi thấy mục tiêu giáo dục phổ thông chúng ta đề ra tương đối sát, tuy nhiên cần bảo đảm có sự liên thông, liên kết giữa cấp học ... **Cử tri** quan tâm vấn đề định hướng nghề nghiệp cho học sinh, trong đây chỉ có đến cấp học trung học phổ thông thì mới đề cập vấn đề kĩ thuật và hướng nghiệp ..., tôi đề nghị trong luật cần nhấn mạnh và có giải pháp cụ thể.

Ngo Thi Kim Yen, Da Nang City, Caucus #4, Control Group

[Article 77] Currently there is a circular regulating maximum number of pupils per class at general education levels; however, the average numbers, especially those of big cities, are in fact much higher. It is therefore recommended that the Drafting Board take this factor into account to train a sufficient number of qualified teachers. Besides, it is suggested to consider whether the standardized educational qualifications for teachers in subjects such as art and music at the elementary level must necessarily be at the university level.

[Đ77] Hiện nay có một thông tư quy định sĩ số học sinh ở các cấp giáo dục phổ thông; song thực tế sĩ số trung bình, đặc biệt ở các đô thị lớn, cao hơn nhiều so với mức quy định. Do đó đề nghị Ban soạn thảo tính đến yếu tố này để có lộ trình đào tạo đủ lượng giáo viên có trình độ chuẩn được đào tạo. Bên cạnh đó, đề nghị cân nhắc trình độ chuẩn được đào tạo của giáo viên các môn như mỹ thuật, âm nhạc cấp tiểu học có nhất thiết phải là trình độ đại học không.

Nguyen Thi Xuan, Dac Lac, Query, Control Group

Many voters were concerned about Official Letter No. 4612 issued by the Ministry on October 3, 2017, which prohibit teaching of contents beyond textbooks. Is this a ban, and whether such ban on teaching non-textbook content contradicts the spirit of educational innovation, when the Ministry is designing open textbooks for general education?

Nhiều cử tri băn khoăn về Công văn số 4612 do bộ ban hành ngày 03/10/2017, trong đó quy định không được dạy những nội dung ngoài sách giáo khoa. Thưa Bộ trưởng, đây có phải là một lệnh cấm và lệnh cấm dạy những nội dung ngoài sách giáo khoa có mâu thuẫn với tinh thần đổi mới giáo dục, khi chính Bộ đang xây dựng một chương trình sách giáo khoa phổ thông theo hướng mở.

D Randomization inference

In lieu of regression tables with asymptotic p-values, we instead adopt a randomization inference-based approach coupled with graphical presentation of all results (Fisher, 1937). Randomization inference (RI) has been increasingly recommended for analyzing experiments in "low information" settings such as those with complex randomization procedures, binary outcomes, clustering of observations, or a small number of observations (Gerber and Green, 2012). Because RI is nonparametric and simply replicates the original randomization procedure, results are not model dependent and are therefore less influenced by the analyst's specification choices (Imbens and Rubin, 2015). An additional feature of RI is that is designed to test the more conservative sharp null hypothesis of no treatment effect for all subjects. Alwyn Young, for instance, shows RI reduced statistically significant results in top economic journals by as much as 22 percent (Young, 2019).

First, we reassigned delegates to treatment and control groups 10,000 times in precise accordance with the three-stage randomization procedure detailed above; covariates and outcomes remained undisturbed. Ideally, all potential randomizations should be realized, but when the combinatorics do not permit complete enumeration a large sample provides a good approximation (Gerber and Green, 2012). In the second step, we conducted all analyses on each of the newly randomized datasets. Finally, by comparing our actual experimental estimates to the distribution of re-randomized estimates, we obtained an answer to the question: Under the sharp null hypothesis of no effect, just how unusual are our experimental results? If, for example, an experimental result is smaller than one (or exceeds 39) out of every 40 re-randomized results, then it is deemed significant at the 0.05 level.

E Summary statistics

Table E.1 presents summary statistics, including mean and standard deviation, for each treatment arm. The top section includes our primary control variables—indicators for full-time, centrally nominated, and competitively elected delegates—as well as the proportion of delegates receiving each treatment. The bottom section displays summary statistics for the two primary outcomes, delegates' survey responses regarding debate preparation and whether delegates spoke in any legislative forum. Also included are the number of missing observations from the VNA-administered survey. For balance and further discussion of missingness, see the previous paper.

	Control (N=181)			Citizen (N=143)			Firm (N=146)		
	Mean	SD	NA	Mean	SD	NA	Mean	SD	NA
Full-time	0.343	0.476		0.343	0.476		0.336	0.474	
Central nominee	0.199	0.400		0.224	0.418		0.185	0.390	
Competitively elected	0.552	0.499		0.510	0.502		0.507	0.502	
% citizen	0.116	0.139		0.451	0.170		0.343	0.179	
% firm	0.112	0.145		0.351	0.189		0.484	0.204	
Prepared for debate	0.481	0.502	73	0.709	0.457	57	0.576	0.497	47
Spoke	0.409	0.493		0.510	0.502		0.459	0.500	

Table E.1: Basic summary statistics.

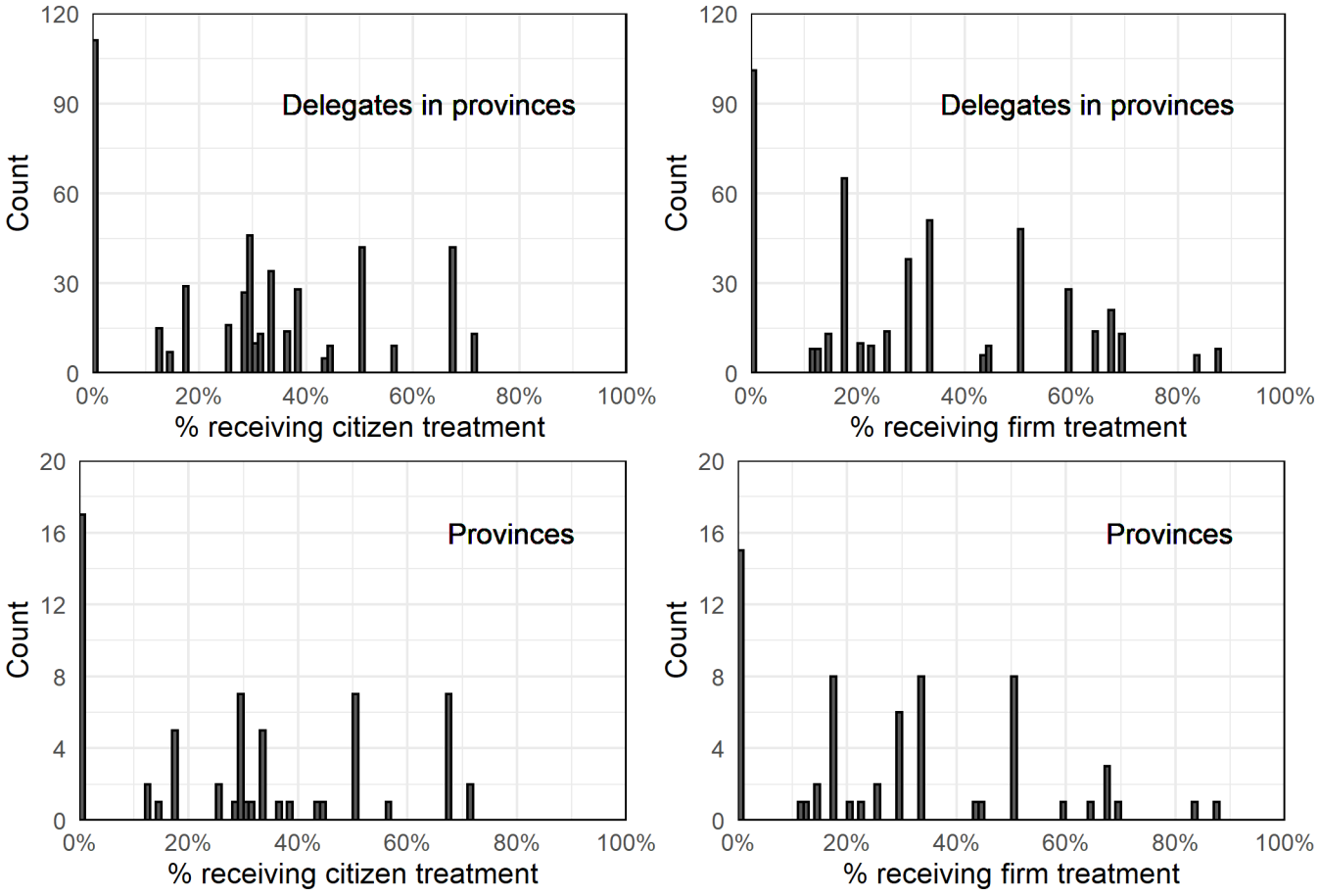

Figure E.1: Histograms of realized treatment shares at individual (top row) and provincial (bottom row) levels. Left column displays citizen treatment shares and right column the firm equivalent.

Figure E.1 depicts the realized provincial treatment shares for each treatment. The top row counts *delegates* by provincial treatment shares (i.e., 45 delegates represent provinces in which

50% of the delegation was citizen-treated). The bottom row counts *provinces* by treatment shares (i.e., seven provinces were 50% citizen-treated). Citizen treatment shares are presented in the left column; firm treatment shares to the right. Due to the presence of 0% dosage provinces, there is an abundance of 0% treated delegates. Otherwise, treatment shares are reasonably continuous and well distributed from 0% to around 75%.

F Tabular results for Figures 1 and 2

Table F.1 contains the primary regression results used in the paper, with statistical significance assessed by randomization inference. The parenthetical quantity underneath each regression coefficient represents that coefficient's quantile as compared to 10,000 coefficients estimated from repeated randomizations. For example, a parenthetical value of 0.999 means that our experimental estimate exceeds 99.9% of the coefficients resulting from re-randomized treatment assignments and would be deemed statistically significant at conventional levels. The treatment share coefficients (% Citizen; % Firm) correspond to spillover effects upon the untreated. The coefficients on the interactions between treatment dummies and treatment shares (Citizen $\times$ % Citizen; Firm $\times$ % Firm) point to reinforcement—or strength-in-numbers—effects.

Table F.1: Tabular results for Paper Figures 1 and 2.

Delegate...	Prepared	Spoke		
	Survey	Pooled	Floor	Caucus
Citizen	0.148	0.084	-0.105	0.073
	(0.770)	(0.743)	(0.056)	(0.713)
% Citizen	-0.343	-0.268	-0.164	-0.192
	(0.090)	(0.042)	(0.020)	(0.094)
Citizen × % Citizen	0.441	0.231	0.392	0.196
	(0.868)	(0.774)	(0.999)	(0.747)
Firm	0.113	0.245	0.072	0.235
	(0.709)	(0.966)	(0.877)	(0.961)
% Firm	0.001	0.023	-0.053	-0.064
	(0.508)	(0.531)	(0.294)	(0.308)
Firm × % Firm	0.115	-0.283	-0.053	-0.163
	(0.629)	(0.131)	(0.288)	(0.254)
Full-time	-0.12	0.129	0.007	0.140
	(0.020)	(0.528)	(0.004)	(0.298)
Central nominee	0.074	0.032	-0.016	0.016
	(0.725)	(0.366)	(0.993)	(0.721)
Competitive	-0.093	0.085	0.050	0.034
	(0.062)	(0.506)	(0.658)	(0.664)
Constant	0.598	0.340	0.059	0.291
	(0.219)	(0.373)	(0.949)	(0.326)
Observations	293	470	470	470

Randomization inference based on 10,000 randomizations.

Quantile of experimental estimate in parentheses.

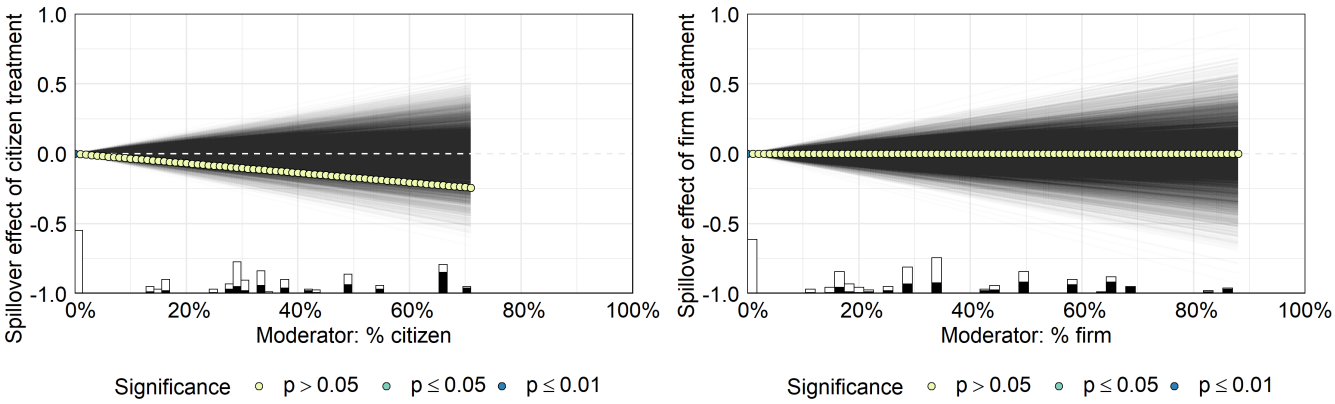

Figure G.1: Insignificant negative spillover from the citizen treatment on debate preparation.

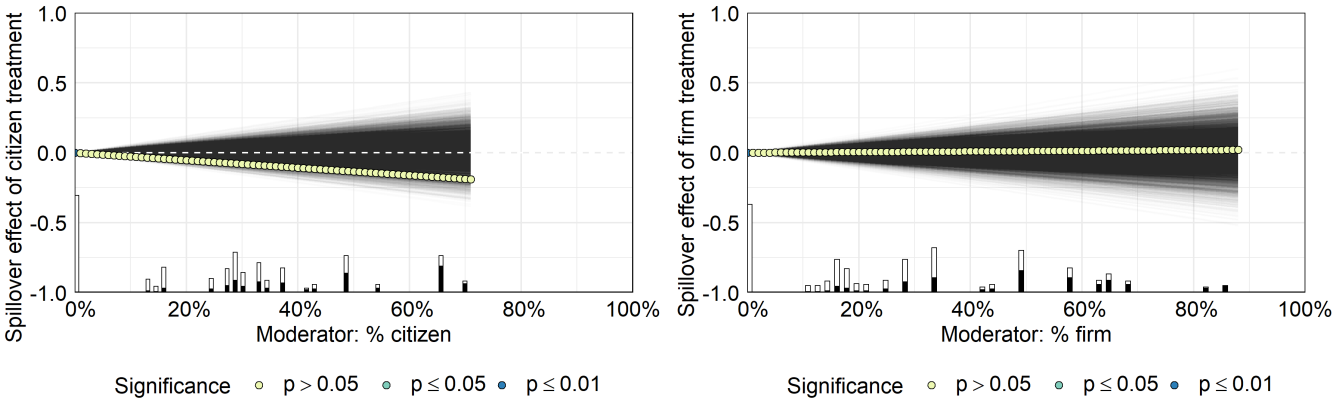

Figure G.2: Insignificant negative spillover from the citizen treatment on delegate speaking.

G Graphical presentation of spillover effects

Figures G.1, G.2, and G.3 present spillover results from the same specification as the reinforcement results presented in paper Figures 1 and 2. With the exception of Figure G.3 Panel B, all results are similar: weak and statistically insignificant spillover from the citizen treatment and no spillover effects from the firm treatment. In Figure G.3 Panel B, however, the weak negative spillover of the citizen treatment attains statistical significance.

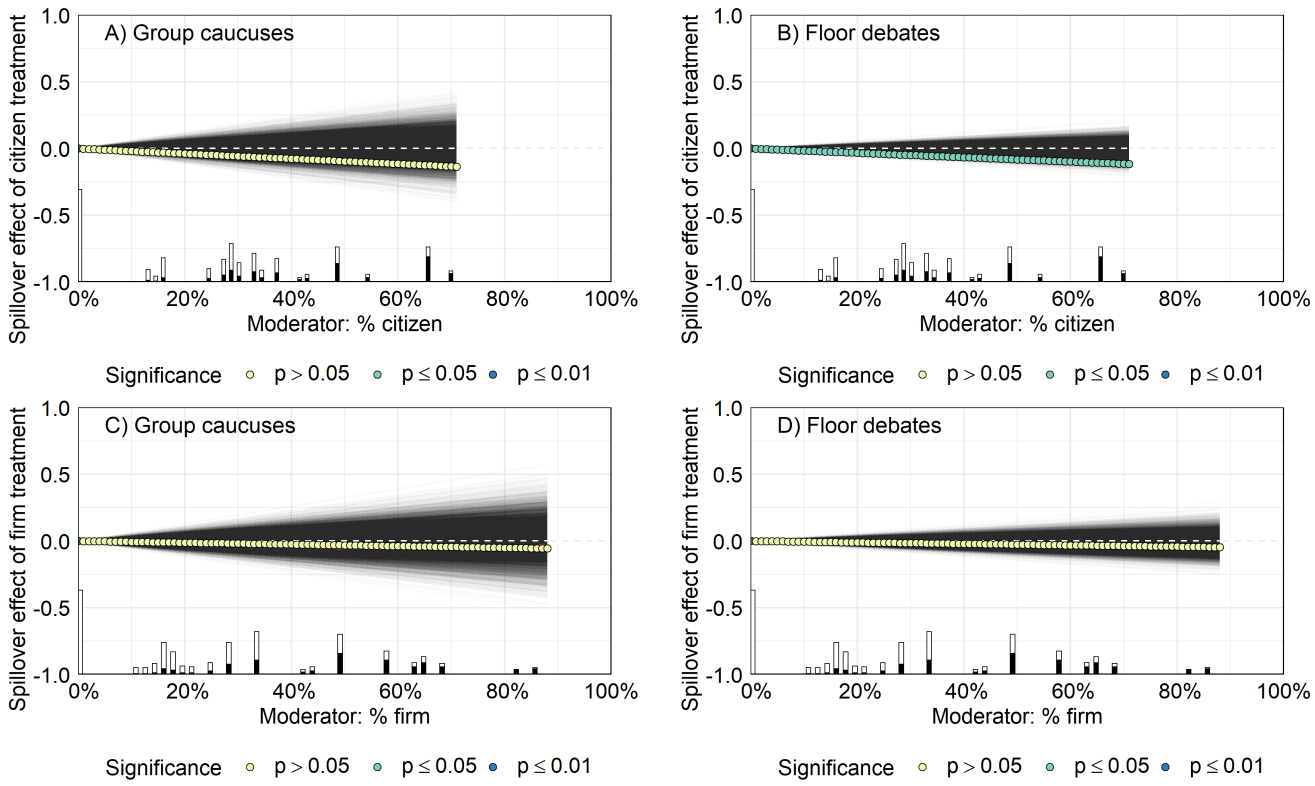

Figure G.3: Weak yet significant negative spillover from the citizen treatment in floor debates.

H Saturation results by original treatment conditions

Here we examine treatment effects by provincial-level saturation and individual treatment assignment. We do so by examining raw outcomes across six bins: (1) untreated delegates in untreated provinces, (2) untreated in 50% treated provinces, (3) citizen-treated delegates in 50% treated provinces, (4) citizen-treated delegates in fully treated provinces, (5) firm-treated delegates in 50% treated provinces, and (6) firm-treated delegates in fully treated provinces. Table H.1 displays the results when we regress the debate preparation and speaking outcomes on these simple dummy variables with intercepts suppressed.

Table H.1: Examining results by provincial saturation.

	Prepared	Spoke
Control (0% Saturation)	0.460	0.425
	(0.324, 0.596)	(0.320, 0.530)
Control (50% Saturation)	0.500	0.394
	(0.374, 0.626)	(0.292, 0.495)
Citizen (50% Saturation)	0.694	0.536
	(0.557, 0.831)	(0.429, 0.643)
Citizen (100% Saturation)	0.730	0.475
	(0.572, 0.888)	(0.347, 0.602)
Firm (50% Saturation)	0.636	0.461
	(0.507, 0.766)	(0.348, 0.573)
Firm (100% Saturation)	0.500	0.457
	(0.355, 0.645)	(0.340, 0.574)
N	293	470
R^2	0.598	0.460
RMSE	0.488	0.499
F-Statistic	71.133	65.880

Note: Linear regressions with suppressed intercepts. Coefficient 95% confidence intervals displayed in parentheses.

Consistent with our primary results using continuous treatment shares, we find that the citizen treatment increases responsiveness relative to untreated delegates in 50% treated provinces, regardless of how responsiveness is measured. Also consistent with earlier results, as we increase a province's treatment saturation from 50% to 100%, we find that citizen-treated delegates are more

prepared for debate. We do not, however, observe a similarly-signed difference on the speaking outcome. Because provincial treatment saturations include both the citizen and firm treatments simultaneously, increasing the saturation does not necessarily correspond to increasing a province's share of citizen-treated delegates. We find no evidence of reinforcement effects of the firm treatment across either outcome. Because control delegates do not exhibit increased responsiveness in the 50% treated provinces relative to the untreated provinces, we also find no evidence of spillover effects.

We also conducted formal F-tests for the equality of coefficients to assess direct (Citizen-50% vs. Control-50%; Firm-50% vs. Control-50%), reinforcement (Citizen-100% vs. Citizen-50%; Firm-100% vs. Firm-50%), and spillover (Control-50% vs. Control-0%) effects. As Table H.2 shows, the only statistically significant differences were those between citizen-treated and untreated delegates in 50% treated provinces.

Table H.2: Spillover and reinforcement effects not reflected in saturation analysis.

	Prepared		Spoke	
	F-stat	p-value	F-stat	p-value
Direct (50% Citizen vs. 50% Control)	4.192	0.042	3.597	0.059
Direct (50% Firm vs. 50% Control)	2.204	0.139	0.755	0.385
Spillover (100% Control vs. 50% Control)	0.180	0.671	0.182	0.670
Reinforcement (100% Citizen vs. 50% Citizen)	0.114	0.736	0.520	0.471
Reinforcement (100% Firm vs. 50% Firm)	1.909	0.168	0.002	0.967

Note: F-tests for equality of coefficients from Table H.1.

I Figure 1 (left panel) without central nominees

Figure I.1 replicates the analysis of paper Figure 1 (left panel) on a subset of the data, dropping all central nominees as those least likely to be responsive to local constituencies. This model is substantively similar, but demonstrates a much larger reinforcement effect for the citizen treatment—the marginal effect is 0.58 at the maximum observed treatment share. This indicates that the reinforcement effect is much more important for local- and self-nominated delegates. Although not shown

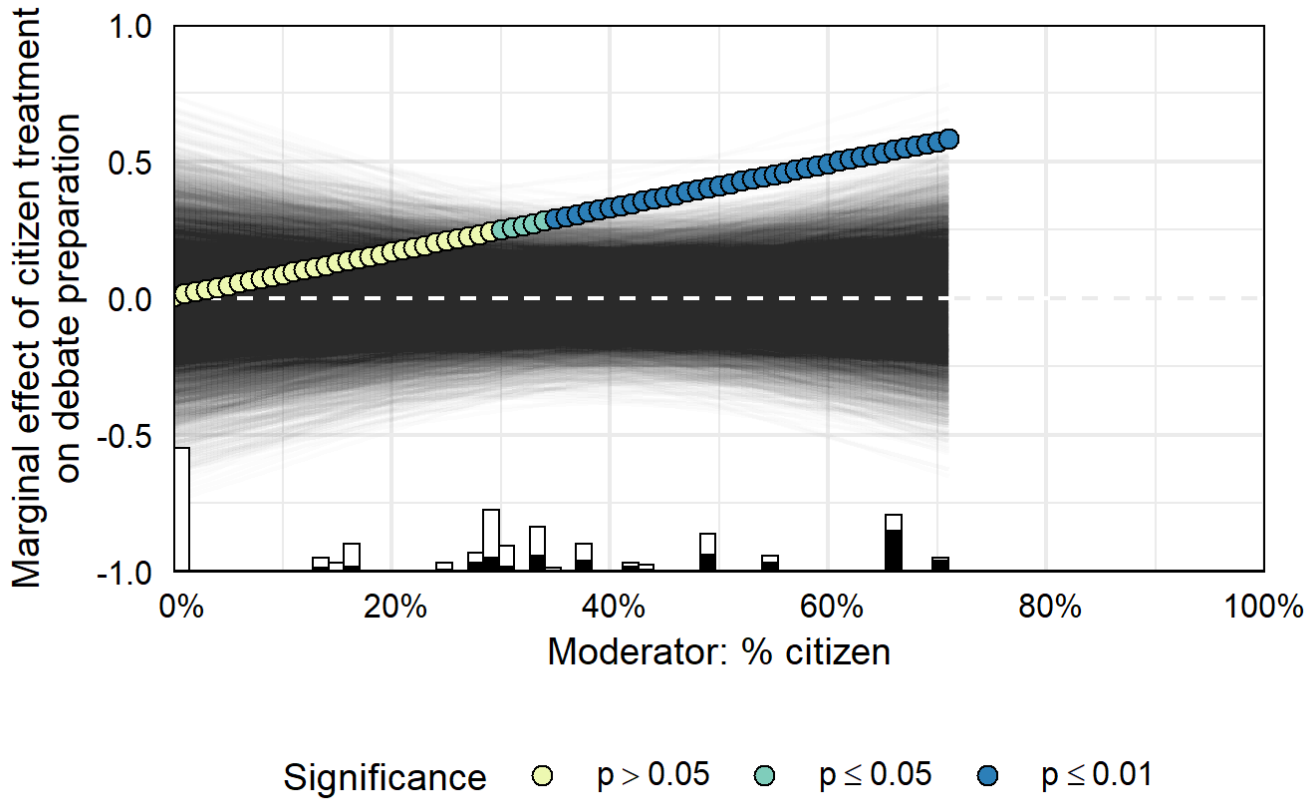

Figure I.1: The safety-in-numbers effect of the citizen treatment on debate preparation is stronger for local nominees.

here, the negative spillover effect of the citizen dosage on those not receiving the citizen treatment remains, and is in fact strengthened. While we blocked on central nominees during the randomization, excluding them from analyses was not part of our pre-analysis plan and such inferences should therefore be treated with caution.

J Figure 2 Chow tests

In the paper we report the results of a Chow test comparing regression coefficients from the fully interactive specification across two datasets—the group caucus transcripts and the floor debate transcripts. The resulting F-statistic, 23.262, is highly significant, indicating that the visible difference between the two legislative fora is statistically meaningful. Table J.1 presents the results of additional Chow tests on three increasingly parsimonious specifications, all of which reveal meaningful differences across legislative fora. Row 2 drops the three control variables, full-time, central

nomination, and competitive elections. Row 3 drops firm treatment dummies, provincial shares of firm treated, and the interaction between them. The final row drops both controls and all firm terms.

Table J.1: All Chow test specifications reveal a meaningful difference between the caucuses and floor debates.

Specification	F-stat.	p-value
Full interactive	23.262	< 0.001
Dropping controls	34.965	< 0.001
Dropping firm terms	30.383	< 0.001
Dropping controls + firm terms	57.070	< 0.001

Full interactive specification includes Cit., %Cit., Firm, %Firm, Fulltime, CentNom, and Competitive.

References

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

Fisher, Ronald Aylmer. 1937. *The Design of Experiments*. London: Oliver and Boyd.

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

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences*. New York: Cambridge University Press.

Sekhon, Jasjeet S. 2011. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The `Matching` Package for `R`." *Journal of Statistical Software* 42 (7): 1–52.

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* `DOI: 10.1086/715169`.

Young, Alwyn. 2019. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *The Quarterly Journal of Economics* 134 (2): 557–598.