The Effects of District Magnitude on Voting Behavior

SUPPLEMENTARY MATERIAL

Appendix

A. Instructions. (treatment DM=2 with 25 subjects)
Thank you for agreeing to participate in our voting experiment. The sum of money you will earn during the session will be given privately to you at the end of the experiment. From now on (and until the end of the experiment) you cannot talk to any other participant. If you have a question, please raise your hand and one of the instructors will answer your questions privately. Please do not ask anything aloud! You belong to a group of 25 participants with whom you will interact for 60 elections. The rules are the same for all participants and for all elections. In each election the group will vote to elect two candidates. The winning candidates will be selected by a form of proportional representation, where each party will win seats in proportion to their share of the vote. After each election you will be announced the outcome and your "profit" in such election. At the end of the experiment you'll be asked to answer a questionnaire.

Voting procedure. The party with the most votes wins the first seat, and its vote-total is then divided by 3. The party with the highest remaining votes wins the second seat. In the case of a tie, the winner is determined randomly. As an illustration consider the following example:

Party	A	B	C	D	E
Votes	2	**9**	2	**7**	5
Votes÷		3			

As a result, parties B and D each obtain a candidate because 9 and 7 are the highest numbers. Now consider a different example where parties obtain the following number of votes:

Party	A	B	C	D	E
Votes	**15**	1	4	2	3
Votes÷	5				

In this second example, Party A obtains 2 candidates because 15 and 5 are the highest numbers.

Profits in each election. The profits you receive in each election depend on the candidates elected by the group regardless of whether you voted for any of them. Your profit will be equal to the sum of your valuation of the party of each elected candidate. The table below shows five hypothetical valuations for each of the five parties:

Party	A	B	C	D	E
Your valuations	500	1200	100	1800	500

So, if the 2 candidates from party A are elected, you obtain a profit of 1000 (= 500 + 500). Alternatively, if one candidate from party C is elected, and one candidate from party D is elected you obtain a profit of 1900 (= 100 + 1800). It is important to note that (a) your valuations are different from the valuations of all other voters; and (b) that no other voter knows the valuations of any other voter.

Final Payment. At the end of the last election, the computer will randomly select 4 elections and you will earn the sum of the profits on those elections in pennies. Additionally you will be paid three pounds for taking part in the experiment.

Questionnaire. (prior to the beginning of the session)
1. When the winners of an election are known, do you know your profit in such election? YES/NO
2. When the winners of an election are known, do you know the profit of any other participant? YES/NO
3. Imagine that party A obtains more votes than party B. Could it ever be the case that party B obtains more candidates than party A? YES/NO
4. Imagine a situation where the votes obtained by each party are given by the table below. What would the outcome of the election be?

Party	A	B	C	D	E
Votes	15	2	1	7	0

- Two candidates from party A
- One candidate from party D and one from party E
- One candidate from party A and one from party D
- Two candidates from party B

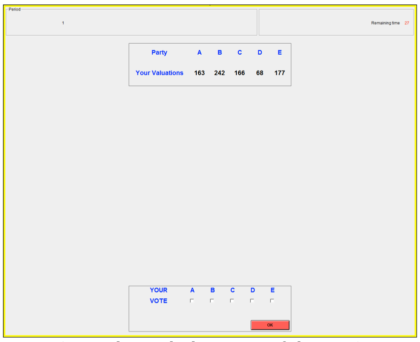
5. Consider a situation where your valuations and the votes obtained by each party are given by the table below. Imagine you voted for party B, what would your profit be?

Party	A	B	C	D	E
Votes	15	2	1	7	0
Your Valuations	500	1200	100	1800	500

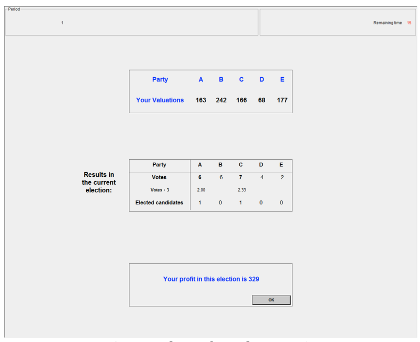
- 1000 or 1200 or 2400 or 0 or 2300

6. Who is going to be paid at the end of the experiment?

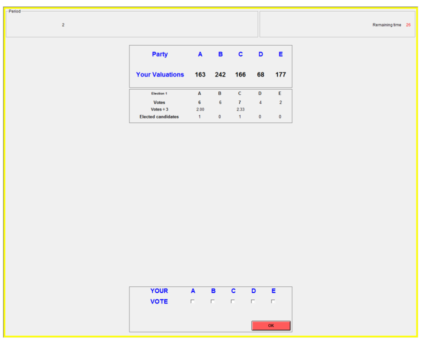
- No one
- 1 person according to his/her profit in various elections
- 2 people according to their profit in various election
- Everyone according to his or her profit in four elections
- Everyone according to his or her average profit throughout the experiment



Screenshot at the beginning of election 1



Screenshot after election 1



Screenshot at the beginning of election 2

Figure A1. Screenshots of the Ztree program for the treatment DM=2.

B. Experimental design.

treatment	*DM=1*	*DM=2*	*DM=3*	*PR*	total
number of groups	*2*	*2*	*3*	*2*	*9*
participants per group	*24,24*	*25,25*	*22,24,24*	*20,24*	*212*

Table B1. Participants and Treatments

C. Location of Political Parties. Subjects were randomly assigned a preferred policy so that the overall distribution in a bounded two dimensional policy space was uniform. Given political parties' position in this policy space, each subject's utility for each of the five parties is computed assuming a quadratic loss function (single-peaked and symmetric preferences). Voters were only told the utility they derived for each party and did not know their policy location relative to the locations of the other voters, nor the relative positions of the parties in the policy space. In all our sessions, the location of the parties and the voters were re-drawn after each set of five elections. We alternated between two types of party locations, shown in the Figure below. So, elections 1 to 5, 11 to 15, 21 to 25, 31 to 35, 41 to 45, and 51 to 55 where held with the Type A locations of the parties. And, elections 6 to 10, 16 to 20, 26 to 30, 36 to 40, 46 to 50, and 56 to 60 where held with the Type B locations of the parties. Note that in the Type B elections we assumed radial symmetry so that there were many more players that were indifferent between various parties. The labelling of parties (from A to E) was randomly allocated in each set of five elections. Given that the participants were not aware of the two-dimensional location of their preferences and the relative location of the parties, when we switched between the two types of party locations we simply announced that the preferences had been redrawn. We clearly stated that the next five elections were independent from the previous five.

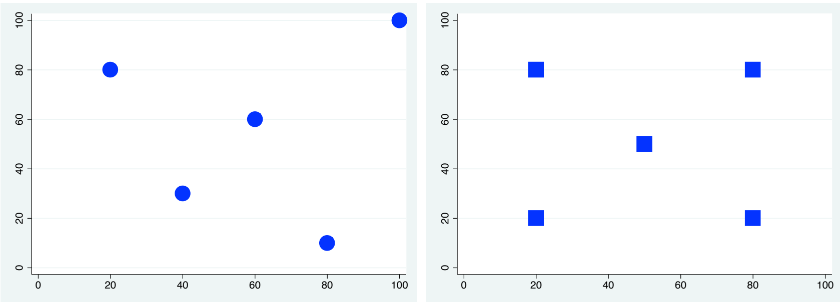


Figure C1. Location of political parties. In all treatments, subjects' preferred policies are uniformly distributed in the two dimensional policy space $[0, 100] \times [0, 100]$.

D. Strategic voting. A subject's *strategic* vote takes into account not only his or her utility but also the probability his or her vote will be pivotal. The latter component depends on the subject's beliefs about the distribution of votes. Using the *theory of voting equilibria* (Myerson and Weber, 1993) we assume that any subject believes that the probability that any other subject votes for party k is equal to the vote share obtained by each party in the previous round of election. That is, $p_k = \frac{v_k}{v_A+v_B+v_C+v_D+v_E}$, where $(v_A, v_B, v_C, v_D, v_E)$ are the votes received by each of the five parties in the previous round of elections.[1] Knowing these probabilities, we can compute the probability that party A gets a votes, candidate B gets b votes, and so on, out of n-1 voters using the following multinomial probability:

$$f(a, b, c, d, e) = \frac{(n-1)!}{a!b!c!d!e!} \cdot p_A^a \cdot p_B^b \cdot p_C^c \cdot p_D^d \cdot p_E^e$$

[1] In the first round of elections, just after preferences have been redrawn, there is no previous round on which to condition the voting decision so the strategic vote coincides with the sincere one: voting for most preferred party.

where $a+b+c+d+e=n-1$. We denote the seats assigned to party k given a particular vote distribution as $\Omega_k(a,b,c,d,e)$. The expected utility of a subject that votes for party A is:

$$E\left(u_\theta(\text{vote for party A}\right) = \sum_{(a,b,c,d,e)\in V} \left(\sum_{k=A,B,C,D,E} \Omega_k(a+1,b,c,d,e)\cdot\theta_k\cdot f(a,b,c,d,e) \right)$$

where θ_k ($k\in\{1,2,3,4,5\}$) is the utility the voter he assigns to each party (this is randomly assigned every 5 elections) and V is any possible combination of votes for the five parties so that they add up to (n-1).

Our experimental setting, draws attention to situations in which being strategic is influenced by the probability of being pivotal. There are, however, many motivations that may lead voters to vote non sincerely in real elections. Voters may use their votes to communicate their preferences (Piketty, 2000). They may anticipate the best candidate to be elected given other citizens' votes (Austen-Smith and Banks, 1996; Feddersen and Pesendorfer, 1996, 1997). Or, they may try to influence the coalition formation in the legislature (Austen-Smith and Banks, 1988; Blais, Aldrich, Indridason, and Levine, 2006; Bargstad and Kedar, 2009; Kedar, 2012; Duch, May, and Armstrong II, 2010; McCuen and Morton, 2006; Indridason, 2011). There is also an interesting literature analysing the impact of electoral rules on the strategic behaviour parties which is obviously out of the scope of the current paper (political parties have fixed policy positions in our experimental design); (Calvo and Hellwig, 2011) show that smaller district magnitudes push bigger parties towards the median's preferred policy yet push smaller parties towards more extreme positions.

E. Alternative definition of strategic voting. We now consider the alternative definition of strategic voting by which a subject votes for a party other than her/his most preferred in order to maximize expected utility. Table E1 presents the results using such definition. We can see that patterns across district magnitude are consistent with the patterns we have highlighted throughout the paper: strategic voting sharply decreases as district magnitude increases. To be more precise, at DM=1 around a quarter of the voters act strategically (using the alternative definition). This reduces to around 10% at DM=2 and further to 5% at DM=3.

	DM=1	**DM=2**	**DM=3**
% sincere	*65.32*	*68.63*	*69.02*
% strategic-non-sincere	*24.70*	*8.96*	*4.85*
% frontrunner	*5.25*	*9.67*	*13.54*
% other	*4.73*	*12.75*	*12.59*
observations	*2304*	*2400*	*3360*

Table E1. Frequency of types of behavior when strategic-non-sincere behavior is defined as voting for party that maximizes expected utility *when it is not the most preferred party.*

F. Robustness checks. Some participants did not vote for their most preferred party on the first election round when there was no information on aggregate turnout and the only optimal action is voting for preferred party. This might indicate that some subjects are not understand the basic functioning of the experiment and this could potentially bias our results. We check the robustness of our results, by considering the subsample of subjects that act optimally in most first rounds.

We compute the percentage of times each individual acts non-optimally in each of the twelve 'round one' they face. Then we drop those who act irrationally in the first round two or more times; we drop around half of our observations.[2] Results using this subsample are reported in Table F1. When comparing these results with the ones in Table 4, we see that our findings are now stronger: strategic voting is much more prevalent at $DM=1$, whereas the portion of sincere voting nearly equates that of strategic voting in $DM=3$. Note also that much of the frontrunner effect is driven by those who act unexpectedly in round one.

[2]We check the robustness of our results to a less demanding condition regarding the percentage of times subjects favoured their preferred party in round 1: (i) not voting for preferred party three or more times

	DM=1	DM=2	DM=3	DM=PR
% sincere	*16.47*	*26.89*	*40.37*	*90.77*
% strategic	*80.70*	*65.4*	*48.42*	
% frontrunner	*2.83*	*7.72*	*11.4*	*9.23*
% other	*4.69*	*10.42*	*10.10*	*1.23*
observations	*1152*	*1104*	*1584*	*1056*

Table F1. Frequency of Types of Behavior, by treatment when participants vote for most preferred party in round 1 at least 10 out of 12 'first rounds'.

Finally note that the percentage of 'other' votes reduces a little bit but remains essentially the same when we compare Table 4 with Table 8. This suggests that the subjects we are dropping are definitely not the one that drive the 'other' type of behavior.

G. Estimation Procedure: Expectation-Maximization algorithm. The EM-algorithm estimates the probability that each observation in a sample comes from a particular distribution.[3] Suppose we have $N = \{n_1, n_2, n_3, ..., n_N\}$ observations coming from two different Normal $N(\mu, \sigma^2)$ distributions: $N(0,1)$ and $N(2,1)$, and we want to find the proportion of observations that come from each distribution. Given that we know the exact p.d.f's of each distribution, we can compute the probability each observation was generated by either. For instance, if observation $n_1 = 1$, we find that its probability of being drawn from each distribution is 0.5. If $n_2 = 0$, then the probability that it is drawn from $N(0,1)$ will be (approximately) 0.98 and 0.02 that it comes from $N(2,1)$. The average of the N probabilities tells us the proportion of observations that come from each distribution.

Now suppose we know that all observations come from two Normal distributions, but that we ignore the means of both Normal distributions (this resembles more our framework). Hence, to start with, we first need to guess two means (μ_1 and μ_2). Once we have both means, we compute the probability that each observation comes from each of both Normals $N(\mu_1, 1)$ and $N(\mu_2, 1)$. We then take the average of the probabilities and using this information we take a new educated guess for the value of the means[4] For instance, if our first guess is $\mu_1 = 0.1$ and $\mu_2 = 40$, we fill find that all observations seemingly belong to distribution 1, so we will take an updated guess that μ_2 is, say, 4 (while keeping our guess for μ_1 close to 0). We will keep slightly modifying our guesses for μ_1 and μ_2 until we find μ_1 and μ_2 that maximize the likelihood that all observations come from such two distributions. Note that the EM algorithm does not give us the true values μ_1 and μ_2, but gives us the best possible estimate of both means given the sample that we have.
The application of the algorithm for our case works in the exact same fashion. The difference is that we call 'types' the different distributions where the observations come from, and that these distributions are not Normal but a conditional logit.

To be precise, we assume that the utility an individual i derives from voting for party j in round t is given by:

$$U_{ijt} = z_i^{str} \frac{e^{\alpha_j^{str} + \beta_j^{str} v_{ijt}}}{\sum_k e^{\alpha_k^{str} + \beta_k^{str} v_{ikt}}} + z_i^{sin} \frac{e^{\alpha_j^{sin} + \beta_j^{sin} u_{ijt}}}{\sum_k e^{\alpha_k^{sin} + \beta_k^{sin} u_{ikt}}} + z_i^{fr} \frac{e^{\alpha_j^{fr} + \beta_j^{fr} FR_{ijt}}}{\sum_k e^{\alpha_k^{fr} + \beta_k^{fr} FR_{ikt}}}, \quad (1)$$

$j, k = \{a, b, c, d, e\}$, where z_i^{τ} is an indicator variable that takes value 1 if individual i is of type τ, $\tau = \{$sincere, strategic, frontrunner$\}$. Furthermore, v_{ikt} is the expected utility i gets from voting for k in round t, as described above in Appendix D; u_{ikt} is the utility i gets from party k in round t; and FR_{ikt} is a dummy

(25% of the subjects are dropped dropped) or (*ii*) six or more times (7.55% dropped). Results do not depend on this.

[3]See Frühwirth-Schnatter (2006) for a detailed description of the algorithm.

[4]Technically, we maximize the complete log-likelihood function. See the technical explanation below.

variable that takes value 1 if k won the most votes in the previous election, 0 otherwise. Hence, the probability that agent i votes for j in election t is given by (we disregard the use of subscript t henceforth to simplify notation)

$$(P_{ij}|z_i^{sin}=1)=\frac{e^{\alpha_j^{sin}+\beta_j^{sin}u_{ijt}}}{\sum_k e^{\alpha_k^{sin}+\beta_k^{sin}u_{ikt}}};\quad (P_{ij}|z_i^{strat}=1)=\frac{e^{\alpha_j^{str}+\beta_j^{str}v_{ijt}}}{\sum_k e^{\alpha_k^{str}+\beta_k^{str}v_{ikt}}};\quad (P_{ij}|z_i^{fr}=1)=\frac{e^{\alpha_j^{fr}+\beta_j^{fr}FR_{ijt}}}{\sum_k e^{\alpha_k^{fr}+\beta_k^{fr}FR_{ikt}}} \tag{2}$$

Denote the p.d.f of observation i by $p(\mathbf{X}_i|z_i^\tau,\beta^\tau,\alpha^\tau)$, where $\mathbf{X}_i=[u_i,v_{,i}\,,FR_i,y_i]$ and y_i denotes i's vote. Then,

$$f_i^\tau=p(\mathbf{X}_i,\beta^\tau,\alpha^\tau|z_i^\tau=1)=\prod_{j=a}^{e}(P_{ij}|z_i^\tau=1)^{y_{ij}},\quad \text{where } y_{ij}=1 \text{ if } i \text{ votes for } j,\ 0 \text{ otherwise} \tag{3}$$

Let π^τ be our main parameter of interest, i.e., the unconditional probability that an agent is of type τ. To simplify notation, let $\theta=(\alpha,\beta,\pi)$.[5] *If we knew all zs*, the complete data likelihood function would take the following form:

$$p(\mathbf{X},z|\theta)=p(\mathbf{X}|z,\theta)p(z|\theta)=\prod_i^N p(\mathbf{X}_i|z_i,\theta)p(z_i|\theta)=\prod_i^N\left(\pi^{sin}f_i^{sin}\right)^{z_i^{sin}}\left(\pi^{str}f_i^{str}\right)^{z_i^{str}}\left(\pi^{fr}f_i^{fr}\right)^{z_i^{fr}} \tag{4}$$

Taking logs,

$$log\mathscr{L}(\mathbf{X},z;\theta)=\sum_{i=1}^N\sum_\tau\{z_i^\tau log\,(\pi^\tau)+z_i^\tau log\,(f_i^\tau)\} \tag{5}$$

To find $\hat\pi$, it would suffice to maximize (5) with respect to θ. However, we do not observe the individual types (zs). Hence, we need to estimate θ using the Expectation-Maximiztion (EM) algorithm. We first describe how the algorithm works, and then proceed to explain why the estimated values of θ we get using the EM algorithm are arbitrarily close to those we would get were we to observe the individual types.

1. We assume that the conditional expectation of $z|\theta,\mathbf{X}$ is given by

$$\hat z_i\equiv E(z_i^\tau|\theta,\mathbf{X}_i)=\frac{\pi^\tau f_i^\tau(\mathbf{X}_i,\beta,\alpha)}{f(\mathbf{X}_i,\beta,\alpha,\pi)}=\frac{\pi^\tau f_i^\tau(\cdot)}{\pi^{sin}f_i^{sin}(\cdot)+\pi^{str}f_i^{str}(\cdot)+\pi^{fr}f_i^{fr}(\cdot)}\qquad\forall i \tag{6}$$

2. (E-STEP) We take an educated guess on $\{\pi,\alpha,\beta\}$. Then we take the expected value of z conditional on the guess $\{\pi,\alpha,\beta\}$ as specified in the previous step. I.e., $E(z_i^\tau|\theta,\mathbf{X}_i)$ is the expectation conditional on the data and the parameters of the model. In other words, $\hat z_i=E(z_i^\tau|\theta,\mathbf{X}_i)$ is the posterior probability that i belongs to type τ. Once all $\hat z_i$s are computed, we replace the unobserved z_is with $\hat z_i$ in (5).

3. (M-STEP) Given $\hat z$, we maximize (5) with respect to α, β and π . This yields a set of updated estimates $\hat\alpha,\hat\beta$ and $\hat\pi$.

4. (E-STEP) Given the current set of estimates $\hat\alpha,\hat\beta$ and $\hat\pi$, we estimate the updated expected value of the vector z: $\hat z'=\hat z'(\mathbf{X};\hat\alpha,\hat\beta,\hat\pi)$.

5. (M-STEP) Given $\hat z'$, we maximize (5) with respect to α, β and π and get $\hat\alpha',\hat\beta',\hat\pi'$.

6. We repeat 4 and 5 until convergence.

The estimates $\{\alpha_r,\beta_r,\pi_r\}$ we get after r iterations, once the algorithm has converged, can be shown to be the best possible estimates. Recall that, generally, we can describe our problem as follows: we have an unobserved random variable Z, with a given realization z. We want to find θ which maximizes $P(\mathbf{X}|z,\theta)P(z|\theta)$, $\theta=(\alpha,\beta,\pi)$.

In Borman (2009) it is shown that

$$\max_\theta\ log\left\{\sum_z P(\mathbf{X}|z,\theta)P(z|\theta)\right\}\geq\max_\theta\sum_z P(z|\mathbf{X},\theta_r)\ log\ P(\mathbf{X},z|\theta), \tag{7}$$

where θ_r is the set of estimated parameters at iteration r. That is, given a set of estimates θ_r, $\sum_z P(z|\mathbf{X},\theta_r)\,log\,P(\mathbf{X},z|\theta)$ is bounded above by $log\,\{\sum_z P(\mathbf{X}|z,\theta)P(z|\theta)\}$. Therefore, given an unobserved vector z and an educated initial guess θ_0, maximizing the RHS of (7) will yield us $\hat\theta_r$, the best approximation to the estimated value

[5]That is, $\beta=[\beta^{sin},\beta^{str},\beta^{fr}]$; $\alpha=[\alpha^{sin},\alpha^{str},\alpha^{fr}]$; $\pi=[\pi^{sin},\pi^{str},\pi^{fr}]$.

of θ were we to observe z (i.e., the value we would estimate if we could maximize the LHS of (7)).

In other words, the best approximation to the estimates of θ we would get from (5) is given by

$$\underset{\alpha,\beta,\pi}{\text{argmax}} \sum_{i=1}^{N}\sum_{\tau} \hat{z}_i^{\tau} log\pi^{\tau} + \sum_{i=1}^{N}\sum_{\tau} \hat{z}_i^{\tau} \left\{ \sum_j y_{ij} log(P_{ij}|z_i^{\tau}=1) \right\} \qquad \tau = \{strat, sin, fr\} \tag{8}$$

which is what we actually maximize.

H. Mixed Logit. As a robustness check we modify our specification, allowing for within-subject variability. That is, we assume that all agents place different weights to the different types of action. Hence we can estimate the (weight) parameters for each individual and report the distribution of such parameters among our population. Formally, a mixed logit:

$$p_{ijt}(u_{ijt}, v_{ijt}, FR_{ijt}; \alpha, \beta) = \frac{e^{\alpha_j + \beta^i_{sin} u_{ijt} + \beta^i_{strat} v_{ijt} + \beta^i_{FR} FR_{ijt}}}{\sum_{k=1}^{J} e^{\alpha_k + \beta^i_{sin} u_{ikt} + \beta^i_{strat} v_{ikt} + \beta^i_{FR} FR_{ikt}}}$$

where

$$\begin{pmatrix} \beta^i_{sin} \\ \beta^i_{strat} \\ \beta^i_{FR} \end{pmatrix} \sim \mathbf{N} \begin{pmatrix} \beta_{sin} & & \sigma^2_{sin} & \sigma_{sinstrat} & \sigma_{sinFR} \\ \beta_{strat} & , & \sigma_{sinstrat} & \sigma^2_{strat} & \sigma_{stratFR} \\ \beta_{FR} & & \sigma_{sinFR} & \sigma_{stratFR} & \sigma^2_{FR} \end{pmatrix}$$

$u_{ijt}, v_{ijt}, FR_{ijt}$, are measured as described in Appendix G. We assume that the distribution of the vector of parameters β follows a multivariate normal distribution. Table H1 below reports the estimated means of our mixed logit computations.

	DM=1	**DM=2**	**DM=3**	**PR**
Mean β_{sin}	*2.97*	*1.92*	*3.78*	
	(0.50)	*(0.26)*	*(0.32)*	*14.99*
Mean β_{strat}	*90.84*	*80.24*	*63.61*	*(1.20)*
	(9.57)	*(6.14)*	*(6.68)*	
Mean β_{FR}	*1.66*	*0.77*	*1.02*	*1.98*
	(0.15)	*(0.12)*	*(0.11)*	*(0.88)*
% correctly predicted votes	*78.26*	*70.75*	*66.90*	*85.23*
observations	*2,304*	*2,400*	*3,360*	*2,112*

Table H1. Mixed Logit Results
(standard errors in brackets)

Note that the scale of the utility and expected utility values are not the same so parameters in our four different specifications are not directly comparable. However, we can observe the trends across DM. We find that these are similar those we observed earlier: sincere considerations (as captured by β_{sin}) increase with DM in all cases apart from $DM = 1$ to $DM = 2$; strategic considerations (as captured by β_{strat}) decrease with DM in all cases; and finally, frontrunner considerations are always present across DMs.[6]

[6]Callander (2008) is the only formal model we are aware of that introduces an explicit preference for voting for the frontrunner.

I. Who is strategic? In order to identify what triggers strategic voting, and who is more likely to be strategic, we run an OLS regression where the dependent variable is the estimated probability of being strategic for each subject in each election. Table I1 reports the results (which are referred to in Section 4 in the paper). As argued throughout this paper, strategic behavior decreases with district magnitude. We can also observe that agents learn to be strategic as the experiment unfolds: the more periods they play, the more strategic their votes become, which hints at some learning effect. In column (2) we add the *vote margin in the previous round*, which captures how contentious the fight for the last seat is. Results show that it is negatively correlated with being strategic.[7] The smaller the vote margin, the more salient the strategic action and the more likely subjects are to target their vote to the race between those two parties. In column (3) we further add *effective number of parties* to check whether concentration of votes affects behavior. Results suggest it does not (the F–test of joint significance of *margin of votes* and *effective number of parties* yields a p–value of 0.6). Columns (2) and (3) include the socio-demographic characteristics we elicited at the end of the experiment. Most variables we include are self-explanatory: *economics* (whether the students has/is studying economics), *Non–FPP* (whether the electoral system of the subject's country of origin is not first past the post), *time* (how long it took the participant to cast a vote, in seconds) or *British* (54% of our subjects are British). The only one with a significant effect is *time*, suggesting that participants who took longer to decide their vote were more likely to be acting strategically. All regressions included *age*, *male* and *experience* (number of experiments the subject has participated in), but none of these turned out as significant.

	(1)	(2)	(3)
DM=2	0.098***	0.071***	0.082***
	(0.02)	(0.03)	(0.03)
DM=1	0.17***	0.16***	-
	(0.02)	(0.02)	
Round	0.0032	0.0062**	0.0025
	(0.00)	(0.00)	(0.00)
Period	0.0018***	0.0015***	0.0014***
	(0.00)	(0.00)	(0.00)
Margin prev. round		-0.0056**	-0.0031
		(0.00)	(0.00)
Time		0.0024**	0.0020*
		(0.00)	(0.00)
Economics		-0.0076	0.0065
		(0.03)	(0.04)
Non-FPP		0.022	0.013
		(0.02)	(0.02)
British		-0.032	-0.0031
		(0.03)	(0.03)
Effective # of parties			0.010
			(0.02)
Other controls	Y	Y	Y
Observations	8064	8064	5760
R^2	0.080	0.100	0.066

Clustered standard errors in parentheses, ***p<0.01, **p<0.05, *p<0.1

Table I1. Linear regression on the estimated probability of being strategic. The dependent variable is the probability that the vote cast is strategic, as estimated in Table 4 in the paper. *DM=3* is the base category in all cases. *DM=PR* excluded in all regressions. *DM=1* drops in (3) since the effective number of parties is always one. Round= $\{1, 2, 3, 4, 5\}$. Period= $\{1, 2, ..., 60\}$. Other controls used: age, gender, experience. Standard errors clustered at the individual level.

[7] *Vote margin in previous round* is the difference of votes between the party that obtains a seat with least votes and the party that does not obtain a seat with most votes.

J. Model performance. Analogously to Duch, May, and Armstrong II (2010), when N is the total number of observations, we define the *Proportional Reduction in Error* from model A to model B (PRE) as:

$$\mathbf{PRE} = \frac{\#\{\text{correct predictions model } A\} - \#\{\text{correct predictions model } B\}}{N - \#\{\text{correct predictions model } B\}}$$

We compare our approach to (*i*) the more standard and prevalently used conditional logit model, (*ii*) the mixed logit model (where agents put different weights on sincere, strategic and frontrunner considerations – see Appendix H), and (*iii*) a two-types model (where frontrunner type is ignored). We find that our approach does a much better job: PRE is consistently above 30% for all DMs, and up to 60% when *DM=1* (see Table J1, below).

	DM=1	**DM=2**	**DM=3**	**PR**
Conditional Logit	*58.50%*	*30.50%*	*34.22%*	*47.21%*
Mixed Logit	*60.07%*	*30.19%*	*34.26%*	*39.74%*
2 types model	*31.97%*	*12.02%*	*30.70%*	-

Table J1. Proportional Reduction in Error of our 3 types model with respect to three other specifications. Two types model allows for strategic and sincere behavior, and ignores frontrunner considerations.

References

Austen-Smith, D., and J. S. Banks (1988): "Elections, Coalitions and Legislative OutcomeS," *American Political Science Review*, 2(82), 405–422.

Austen-Smith, D., and J. S. Banks (1996): "Information Aggregation, Rationality and the Condorcet Jury Theorem," *American Political Science Review*, 1(90), 35–45.

Bargstad, M. A., and O. Kedar (2009): "Coalition-Targeted Duvergerian Voting: How Expectations Affect Voter Choice under Proportional Representation," *American Journal of Political Science*, 2(53), 307–323.

Blais, A., J. Aldrich, I. Indridason, and R. Levine (2006): "Do Voters Vote for Government Coalitions? Testing Downs' Pessimistic Conclusion," *Party Politics*, 6(12), 691–705.

Borman, S. (2009): "The Expectation Maximization Algorithm A short tutorial," mimeo.

Callander, S. (2008): "Majority Rule when Voters like to Win," *Games and Economic Behaviour*, 2(64), 393–420.

Calvo, E., and T. Hellwig (2011): "Centripetal and Centrifugal Incentives under Different Electoral Systems," *American Journal of Political Science*, 55(1), 27–41.

Duch, R., J. May, and D. A. Armstrong II (2010): "Coalition-directed Voting in Multiparty Democracies," *American Political Science Review*, 104, 698–719.

Feddersen, T. J., and W. Pesendorfer (1996): "The Swing Voter's Curse," *American Economic Review*, 3(86), 408–424.

Feddersen, T. J., and W. Pesendorfer (1997): "Voting Behavior and Information Aggregation in Elections With Private Information," *Econometrica*, 5(65), 1029–1058.

Frühwirth-Schnatter, S. (2006): *Finite Mixture and Markov Switching Models*. Springer Series in Statistics, New York.

Indridason, I. H. (2011): "Proportional Representation, Majoritarian Legislatures, and Coalitional Voting," *American Journal of Political Science*, 4(55), 955–971.

Kedar, O. (2012): "Voter Choice and Parliamentary Politics: An Emerging Research Agenda," *British Journal of Political Science*, 42, 537–553.

McCuen, B., and R. B. Morton (2006): "Tactical Coalition Voting and Information in the Laboratory," *Electoral Studies*, 3(29), 316–328.

Myerson, R. B., and R. J. Weber (1993): "A Theory of Voting Equilibria," *American Political Science Review*, 1(87), 102–114.

Piketty, T. (2000): "Voting as Communicating," *Review of Economic Studies*, 1(67), 169–191.