# The Effects of District Magnitude on Voting Behavior[*]

**Simon Hix**

London School of Economics and Political Science

**Rafael Hortala-Vallve**

London School of Economics and Political Science

**Guillem Riambau**

Yale - NUS College

September 8, 2014

**Abstract**. Is there more sincere voting in multi-member districts than in single-member districts? Existing research on this question is inconclusive, at least in part because it is difficult with observational data to isolate the effect of district magnitude on voting behavior independently from voters' preferences or parties' positions. Hence, we investigate this issue in a laboratory experiment, where we vary district magnitude while keeping voters' preferences and parties' positions constant. We find that voting for the preferred party (sincere voting) increases with district magnitude and we are able to explain this in terms of a mechanical effect and a psychological effect. We also find a high incidence of voting for the frontrunner in all our elections, even when there are no incentives for doing so.

# 1. Introduction

The design and reform of electoral systems is a salient policy concern for new democracies as well as many advanced democracies. One key issue in the design of electoral systems is the ideal *district magnitude*: the number of candidates to be elected in each district. Between 2012 and 2014, for example, Romania considered switching from large multi-member districts to single-member districts, Israel considered switching from a single national multi-member district to smaller multi-member districts, the Chilean government proposed to replace two-member districts with large multi-member districts, while Tunisia introduced small multi-member districts for its first democratic elections. What are the consequences of district magnitude in terms of the behavior of voters, the effective number of political parties, and the overall quality of representation and democracy? We still have only partial answers to these questions, and the answers that we do have are mainly based on observational data, where the underlying theoretical mechanisms are difficult to identify clearly. In this paper we analyze the behavior of voters in a controlled laboratory setting, in which we change district magnitude but keep all other relevant political variables constant.

Cox (2007) suggests that as district magnitude increases, the proportion of voters who behave *strategically* decreases, while the proportion who votes *sincerely* for their most preferred party increases. Indeed, Cox (2007, p. 100) claims that: "strategic voting ought to fade out in multi-member districts when the district magnitude gets much above five". This argument is similar to a claim made by Sartori (1968, p. 279) much earlier: "The general rule is that the progression from maximal manipulative impact [via strategic voting] to sheer ineffectiveness follows, more than anything else, the size of the constituency".

These intuitions might explain the patterns observed in elections, where voters in small multi-member districts seem to coordinate around larger parties, which leads to a low number of wasted votes (and a closer relationship between vote-shares and seat-shares) as well as a low number of parties elected to parliaments and fewer parties in government. In contrast, in large multi-member districts, we observe citizens overwhelmingly voting for their most preferred parties (Carey and Hix, 2011).

We still do not know what exactly might be driving these expected empirical regularities. Following Duverger (1954), but with a contemporary twist, we should ask ourselves whether the effect of district magnitude on voting behavior is purely *mechanical* or *psychological*. The former is due to

the fact that as district magnitude increases the proportion of voters whose strategic and sincere motivations coincide increases as the number of viable parties/candidates increases. The latter meanwhile captures changes in voters' strategies possibly due to the increased complexity of the electoral system.

It is impossible to isolate such micro-level effects using actual voting data, which may explain why the presence of strategic (non-sincere) voting does not seem to vary with district magnitude –see Abramson, Aldrich, Blais, Diamond, Diskin, Indridason, Lee, and Levine (2010). In cross-country research, variations in district magnitude covary with a number of other factors which influence how voters behave, like the number of parties, societal cleavage structures, regime type, the level of political and economic development of a country, and so on. In within-country research, district magnitude variations also correlate with other political variations. In Spain, Brazil or Switzerland, for example, where district magnitude varies between larger urban districts and smaller rural districts, elections are held at the same time under the same political institutions and political contexts, but the number and type of candidates and parties competing in each district vary considerably and may be endogenous to expected voting behavior. Similarly, in presidential primary elections in the United States, where there is variation in whether delegates are rewarded in proportion to vote-shares or on a winner-takes-all basis, the effects of the electoral rule covary with the timing of the elections.

It is also difficult to investigate these particular micro-level voting processes with formal models. This is mainly because of the problem of multiple equilibria in multi-candidate and multi-seat elections.[1] Intuitively, it seems that should voters focus their attention in the close races for the last seat, strategic voting should be invariant to district magnitude (see Gerber, Morton, and Rietz, 1998).

We consequently investigate the effect of district magnitude on voting behavior via a laboratory experiment designed to isolate the motivations behind voter choices. Throughout we keep constant the distribution of voters' preferences and the number (and policy location) of parties. This allows us to clearly observe the behavior of voters under different district magnitude treatments. We build our analysis on two stylized types of behavior: (1) *sincere behavior*, where a person votes for the party that yields the highest utility regardless of information about the electoral chances

---

[1] Cox (1994) analyzes voting equilibria under single non transferable vote and shows support for his now classic $M + 1$ rule. We instead analyze voters' behavior in non majoritarian multi member elections.

of the party; and (2) *sophisticated behavior*, where a person takes into account the votes parties have received in previous elections when deciding whom to vote for. Here we define voters being sophisticated when they act "in accordance with both their preferences for the candidates and their perceptions of the relative chances of various pairs of candidates being in contention for victory".[2] Just as pre-election polls serve to inform the electorate about the relative chances of the candidates (Fey, 1997), in our multi-election setting, past voting behavior helps voters form expectations on their chances of influencing the outcome.

In our data we also observe a large proportion of subjects who do not vote sincerely even when their sincere and sophisticated actions coincide – i.e. when there is no tension between the honest expression of their preferences and the consideration of casting a 'useful' vote. We find that subjects not voting for their preferred party (when this vote is also the sophisticated action) vote for the party that obtained most votes in the previous election. This further motivates the characterization of a third type of behavior: *frontrunner behavior*, where a person votes for the party that obtained most votes in the previous election. There is a large literature on *the bandwagon effect*.[3] However in our experimental design this behaviour cannot be rationalised by subjects extracting information from observing a party gaining the largest support. This type of behavioural pattern is analysed in Callander (2008), who suggests that "when voters care about supporting the winning candidate a unique responsive equilibrium exists" , although "the addition of a desire to win creates multiple equilibria".

The paper is organized as follows. We first describe the experimental setup in Section 2. We then summarize the main findings, in Section 3, by showing the proportion of sincere, sophisticated and frontrunner votes in our different experimental treatments. We find that sincere voting increases with district magnitude, sophisticated voting (weakly) decreases, and frontrunner voting can never be discarded in any of our treatments. Later, in Section 4, we analyze individual voting behavior and classify each subject as one of our three types. We show that the increase in sincere voting with district magnitude is due to both mechanical and psychological factors.

---

[2]Myerson and Weber (1993, p. 135)

[3]See Simon (1954) and the large literature that followed.

## 2. The experiment

Our experiment consists of four treatments, each corresponding to a different district magnitude: a single-member district (M=1); a two-member district (M=2); a three-member district (M=3); and pure proportional representation (M=PR). Subjects belong to a group of 25 subjects[4] and participate in 60 elections by casting a single vote for one of five parties.[5] In the M=1 treatment, a candidate from the party that receives the most votes is elected, and each subject receives a payoff from the election equivalent to his or her utility for that party. In the M=2 and M=3 treatments, we apply a form of closed-list proportional representation, where seats are allocated to the parties in proportion to their vote-shares (using the Sainte-Laguë divisor method), and each subject receives a pay-off from the election equivalent to his or her utility for the party of each candidate that is elected. Finally, in the PR treatment, each subject receives a pay-off in direct proportion to the share of votes each party receives. This specification of payoffs allows us to identify the strategic incentives of voters as we modify district magnitude. Needless to say it abstracts from the coalition formation game that would most likely arise in non-majoritarian electoral systems.[6]

Table 1 shows how we allocated 212 subjects to our four treatments.[7] Given that each subject participated in 60 elections we have 12,720 observations.

| treatment | M=1 | M=2 | M=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 1. Participants and Treatments

In all treatments, the utility that subjects derived for each of the parties was privately announced. Every five periods, subjects' preferences were redrawn. We leave the specific details of the overall

---

[4]Due to some participants not showing up on time some groups had slightly less participants (see Table 1)–obviously, instructions were modified accordingly.

[5]Casting a vote for a single party is the most common ballot-structure in single-member as well as multi-member districts in national parliamentary elections in democracies (Reynolds and Steenbergen, 2006).

[6]Strategic voting when taking into account post elections coalitions has received growing interest Austen-Smith and Banks, 1988; Blais, Aldrich, Indridason, and Levine, 2006; Bargstad and Kedar, 2009; Duch, May, and Armstrong II, 2010; McCuen and Morton, 2006 but is out of the scope of this paper.

[7]No subject participated in more than one session. Students were recruited through the online recruitment system ORSEE (Greiner, 2004) and the experiment took place on networked personal computers in Centre for Experimental Social Sciences at Nuffield College, Oxford in November 2011. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). The data and program code for the experiment are available upon request.

distribution of preferences to the Appendix. For our purposes what really matters is that subjects never observe other voters' preferences; they can only infer them by observing past voting behavior. Implicit in this particular feature of our experiment is that we want subjects to condition their voting decisions solely on their own preferences and aggregate past voting behaviour.

The same procedure was used in all sessions. Instructions[8] were read aloud and questions answered in private. Students were asked to answer a questionnaire to check that they fully understood the experimental design, the seat-allocation method, and the payoff structure for their particular treatment group. If any of their answers were wrong, we referred the participant to the section of the instructions where the correct answer was provided. Students were isolated and could not communicate with each other.

In the first election each participant was shown a screen with their utility from each of the five parties and was asked to cast a single vote for one of the parties. Abstention was not allowed. The participants were then informed of the outcome of the first election: the number of votes each party received; which candidate(s) was (were) elected; and the payoff they received from the election. The participants were then asked to vote again for one of the parties. This procedure - in which we counted the votes for each party, we assigned seats, and we informed participants about the outcome of the election and their payoff - was repeated for five elections. Then, after five elections, the participants' preferences were redrawn and the participants interacted for a further five elections, after which the preferences were redrawn again. In other words the experiment was organized as 12 sets of five rounds (60 elections in total) and for each set of elections, participants' preferences and party labels were redrawn.

At the end of the last election, the computer randomly selected four elections and subjects were paid the profits they obtained in those four elections (in Pence Sterling). In addition, subjects received a show-up fee of 3 GBP for taking part in the experiment. At the end of each session, participants were asked to fill in a questionnaire on the computer and were given their final payment in private. Session length, including waiting time and payment, was around 90 minutes. The average payment was 15.71 GBP (approximately 26 USD).

---

[8]See the Appendix for the instructions for the M=2 groups and various screenshots of our program.

## 3. Aggregate Results

As an illustration of our results, Table 2 shows the outcomes of elections 11 to 15 for one of the groups in each treatment (we report the votes received by each of the five parties and the candidates assigned to each party). The results with M=1 shows voters *coordinating* around the first two parties A and B, with support for the other three parties declining over time. This suggests a high proportion of sophisticated behavior, with voters whose preferred party was C, D or E realizing that their most preferred party had no chance of winning. When M=2, voters appear to *coordinate* around three parties (A, B and C) as Cox (2007) would have predicted. In contrast, when M=3 party C ran away with the election after a few rounds. In these 3 cases, we always observe some voters supporting uncompetitive parties (D and E). Finally, in the fully proportional (control) treatment group, there were considerable shifts in voting patterns, despite the fact that the optimal behavior for each participant in this treatment is to vote sincerely. We also observe a tendency to vote for the frontrunner candidate – we will leave this aspect aside for now and will revisit it at the end of this Section.

| election | M=1 | M=2 | M=3 | PR | 11 |
|----------|-----|-----|-----|-----|-----|
| 11 | (6,7*,4,4,3) | (7*,6,6*,3,3) | (5*,5*,6*,4,4) | (5,6,3,6,4) | |
| 12 | (7,12*,4,1,0) | (8*,7,8*,1,1) | (4*,7*,10*,1,2) | (6,7,3,5,3) | |
| 13 | (8,12,*3,1,0) | (7,8*,7*,1,2) | (4*,3,16**,0,1) | (9,7,2,4,2) | |
| 14 | (8,14,*2,0,0) | (7,8*,8*,0,2) | (6*,3,13**,1,1) | (11,6,1,5,1) | |
| 15 | (9,14*,1,0,0) | (7,8*,8*,1,1) | (3,4*,14**,2,1) | (13,5,2,3,1) | |

TABLE 2. Sample of election results for each treatment.

In each cell we indicate the votes received by parties A, B, C, D, and E (resp.) and we identify with one or two stars (* or **) the parties that obtained 1 or 2 candidates, respectively.

In what follows we classify a vote as *sincere* when the subject votes for his/her most preferred party, the one that yields maximum payment.[9] A *sophisticated* vote is instead a vote in which the subject not only considers his/her preferences for all parties but also the likelihood that his or her vote will be pivotal – we compute expected utilities and assume that a sophisticated subject votes for the party that maximises expected utility. In the Appendix we offer a detailed explanation of the computation of expected utilities when voting for each party. For this purpose we build on Myerson

---

[9]This kind of behavior is often also referred as expressive, honest or straightforward vote. See Feddersen and Sandroni (2006) or Fischer (1996).

and Weber (1993), and assume that our subjects best respond to previous period play by assuming that the probability that any player votes for each of the parties coincides with the previous period frequency of votes.[10] Our definitions allow a voter to be *simultaneously* sophisticated and sincere: the optimal action may consist of voting towards the subjects' preferred party.[11] The frequency of sincere and/or sophisticated voting behavior in our four treatments is shown in Table 3. The long-standing hypothesis in the electoral studies literature that sincere voting should increase with district magnitude seems to find little support in our aggregated data when comparing our three non-fully proportional treatments.

|  | M=1 | M=2 | M=3 | PR |
|---|---|---|---|---|
| **% sincere** | *70.5* | *72.4* | *72.7* | *89.8* |
| **% sophisticated** | *84.2* | *74.1* | *70.5* | *89.8* |
| **% observed both** | *64.5* | *67.0* | *66.6* | *89.8* |
| **% predicted both** | *72.5* | *85.5* | *91.2* | *100* |

TABLE 3. Frequency of Types of Behavior by Treatment.
"Observed Both": a subject is *both* sincere and sophisticated.
"Predicted Both": the sincere vote and the sophisticated vote coincide

Most surprisingly, we see a decrease in sophisticated voting as we increase district magnitude from 1 to 3. We expected the opposite because, as district magnitude increases, the likelihood that a sophisticated vote coincides with a sincere vote increases. This increased coincidence in both types of voting is captured in the last row of the table, where the 'predicted' values are the percentage of observations where voting sincerely for the most preferred party in an election can also be classified as a sophisticated vote for the party that maximizes the expected payoff.

We identify the effects of district magnitude on voting behavior via the classic distinction between mechanical and psychological effects of electoral rules. We define *the mechanical effect of district*

---

[10]Forsythe, Myerson, Rietz, and Weber (1993) is the only experimental work we are aware of that uses past election information as the cue from which voters form expectations about their probability of being pivotal. From a game theoretical perspective, we are assuming that subjects engage in a type of fictitious play, by myopically best responding to previous period play.

[11]There is not an agreement in the literature on what is meant by a voter acting *strategically*, hence our use of the word *sophisticated*. For instance, similar to us, Kawai and Watanabe (2013) and Abramson, Aldrich, Blais, Diamond, Diskin, Indridason, Lee, and Levine (2010) consider a voter to be strategic when his/her action maximizes expected utility (regardless of this action being sincere or not); instead, the political behaviour literature only classifies a voter to be strategic when his/her vote maximises expected utility and is *not* sincere.

*magnitude on voting behavior* as the change in sincere behavior not caused by a voter changing their *strategy* but due to the increased likelihood that a sophisticated vote is sincere. Instead, the changes in *strategy* are coined *the psychological effect of district magnitude on voting behavior*. Consider, for instance, a sophisticated voter whose preferred party is ranked third: when district magnitude is 1 his/her vote is less likely to coincide with his/her sincere vote than when district magnitude is 2. Note that the *strategy* of the voter is not changing with district magnitude but the way we classify his/her actions is.

Following this distinction, the results in Table 3 suggest that the small increase in sincere voting (as we increase district magnitude) is not driven by a mechanical effect but by a psychological one: the increase in district magnitude makes more parties viable so subjects are more likely to vote for their preferred party, even when taking into account considerations of pivotability.

Something that seems puzzling in Table 3, though, is the huge difference between the percentage of observations that are both sincere and sophisticated and the situations that are predicted to be so. It seems that in situations in which both sincere and sophisticated actions coincide, the voter should have no conflict about supporting his preferred party. However, we observe that around a 20 percent of subjects fail to choose this action when it is optimal to do so. For whom are they voting? To our surprise we see that 50 percent of the subjects who did not vote for their most preferred party (when sincere and sophisticated actions coincide) voted instead for the party that obtained the most votes in the previous election round: the *frontrunner*. Frontrunner voting is only defined for election rounds 2,3,4 and 5 given that in the first election (round 1) preferences have been redrawn and there is no previous period of play with the same preferences. In common value situations, voting for the winner is usually referred to as *bandwagon* voting and can be understood in terms of herding (Nageeb and Kartik, 2012), information aggregation (Feddersen and Pesendorfer, 1997), or favoring a stable governing party (Riambau, 2013). However, there is no room for such rationalizations in our private values setup. Together with this third classification, our three types of behavior describe more than 90 percent of all vote choices.

Further evidence towards frontrunner behavior is found in our control treatment, with a fully proportional electoral system, where voting sincerely is the dominant strategy (i.e. sincere and sophisticated votes always coincide). In Figure 1 we depict the data of our control treatment aggregated by each set of five elections whenever the frontrunner action does not coincide with the

sincere one. That is, we drop observations in which the subjects' preferred party was the most voted party in the previous election and we see that most vote choices are sincere, yet a large proportion (around 11 percent) of votes are for the frontrunner of the previous election.
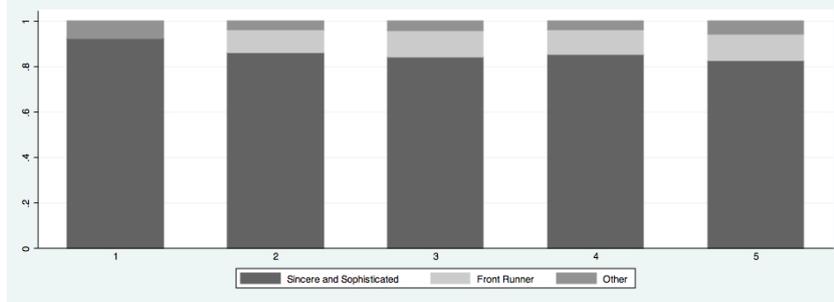


FIGURE 1. Frequencies of Votes by Election Round in the PR treatment when Voting Sincerely is not Voting for the Frontrunner

A key problem when analyzing our data (as with actual voting data) is that many observations can be simultaneously classified as more than one type. Consider for instance a subject whose preferred party is also the one that obtained most votes in the previous period when district magnitude is 1. In this situation, when the subject votes expressively, s/he is also voting for the frontrunner and most likely her/his vote also coincides with her/his sophisticated action. Table 4 shows the subsample of observations in which the three types of actions do not coincide.[12]

|                 | M=1  | M=2  | M=3  |
|-----------------|------|------|------|
| % sincere       | 14.4 | 12.4 | 22.5 |
| % sophisticated | 70.5 | 53.7 | 42.8 |
| % frontrunner   | 7.4  | 24.8 | 22.5 |
| % other         | 7.7  | 9.2  | 12.3 |
| observations    | 312  | 218  | 138  |

TABLE 4. Frequency of Types of Behavior by Treatment when Sincere, Sophisticated and Frontrunner do not coincide.

The last row in the table shows yet another manifestation of the mechanical effect of district magnitude: as district magnitude increases it is more likely that the sincere and sophisticated actions coincide, thus our sample becomes thinner. Even when we are now only considering less than 7 percent of our observations, the patterns we observed earlier are somehow preserved in this

---

[12]Most studies in the literature restrict their analysis to the set of observations in which the type of behaviour is disjoint (e.g. Duch, May, and Armstrong II, 2010)

subsample: sincere voting is greater when district magnitude is 3 rather than 1, and sophisticated voting decreases. Possibly due to the increased complexity of the voting rule we see frontrunner voting and other behavior increasing with district magnitude.

By construction, the disjoint set never includes the first round of elections when subjects vote just after preferences have been redrawn. In these cases there is no previous information, so sincere and sophisticated voting coincide –it is a dominant to vote sincerely for the preferred party. Indeed, we observe a high incidence of such behavior in our data from the very first election – 82 percent of our first election observations are sincere – and such behavior increases as the experiment unfolds reaching 90 percent of sincere observations in the first election of the last round of five elections (election 56).

The patterns we observe are clearly suggestive of both mechanical and psychological effects on voting behavior in our experiment. The increase in sincere voting is (at least partly) driven by the mechanical effect of district magnitude. However, if only mechanical effects are present, we should also be observing an increase in the number of votes that are qualified as sophisticated. Moreover, frontrunner voting should not increase with district magnitude. Overall, there seems to be a tendency to move away from sophisticated actions, towards non-rational behavior (as expressed by sincere and frontrunner voting) as we increase district magnitude.

Tables 3 and 4 both indicate the heterogeneous effects of district magnitude in our population. If all subjects were sincere we should observe 100 percent of observations as sincere, while sophisticated voting should increase with district magnitude due to the mechanical effect, and frontrunner voting should remain unchanged. Instead, if all subjects were sophisticated, sincere voting should increase with district magnitude, sophisticated voting should always be at 100 percent, and frontrunner voting should decrease with district magnitude, because more parties become viable so less voters end up favoring the frontrunner candidate. In the next section we analyze in detail individual voting decisions, to understand whether district magnitude has a systematic effect on the individual behavior of subjects. Our goal is to measure the relative power of the three types of behavior for a representative voter and to see how district magnitude influences the relative weight of the different motivations.

## 4. Individual behavior

In this Section we follow the same strategy that Duch, May, and Armstrong II (2010) use to account for coalitional voters in 86 parliamentary democracies.

Our initial specification assumes that subjects are either sincere or sophisticated: the (unobserved) indicator function $z_{it}^{soph}$ ($z_{it}^{sin}$) takes value 1 if in round $t$ subject $i$ behaves in a sophisticated (sincere) way, and 0 otherwise. Our goal is to estimate $z_{it} = (z_{it}^{soph}, z_{it}^{sin})$ for all subjects and rounds, i.e. for each of our treatments we want to identify what is the estimated probability that an action is of each type.[13] The average of all $z_{it}$ across types and rounds is the unconditional probability that each action is of each type, which could also be interpreted as the proportion of subjects who are basing their choice solely on their preferences (sincere types), and the proportion who are also taking into into account their probability of being pivotal (sophisticated types). We will use both interpretations indistinctly throughout the text.

Let $u_{ijt}^*$ be the utility subject $i$ derives from party $j$ in election $t$; and $v_{ijt}^*$ be the expected utility $i$ derives from voting for party $j$ in election $t$. $y_{ijt}$ is a dummy variable that takes value 1 when subject $i$ votes for party $j$ in round $t$. Then:

$$y_{ijt} = 1 \text{ if } u_{ijt}^* \geq \max(u_{i1t}^*, u_{i2t}^*, ..., u_{i5t}^*) \text{ and } i \text{ is sincere}$$

$$y_{ijt} = 1 \text{ if } v_{ijt}^* \geq \max(v_{i1t}^*, v_{i2t}^*, ..., v_{i5t}^*) \text{ and } i \text{ is sophisticated}$$

$$y_{ijt} = 0 \text{ otherwise}$$

where $u_{ijt}^* = \bar{u}_j + \alpha u_{ijt} + \varepsilon_{ijt}^u$ when $z_{it}^{sin} = 1$ and $v_{ijt}^* = \bar{v}_j + \beta v_{ijt} + \varepsilon_{ijt}^v$ when $z_{it}^{soph} = 1$. We assume $\varepsilon_{ijt}^u, \varepsilon_{ijt}^v \sim$ type I extreme value. We model the probability that subject $i$ votes for $j$ in period $t$ as a multinomial logit (MNL):

$$p_{ijt}(u_{ijt}, v_{ijt}; z_{it}) = \frac{e^{\bar{u}_j + \alpha u_{ijt}}}{\sum_{k=1}^J e^{\bar{u}_k + \alpha u_{ikt}}} \qquad \text{if } z_{it}^{soph} = 0$$

$$p_{ijt}(u_{ijt}, v_{ijt}; z_{it}) = \frac{e^{\bar{v}_j + \beta v_{ijt}}}{\sum_{k=1}^J e^{\bar{v}_k + \beta v_{ikt}}} \qquad \text{if } z_{it}^{soph} = 1$$

Note that modeling utilities with a MNL allows us to take into account the cardinality of preferences: the likelihood of voting for one party depends not only on the ranking of the party in an ordinal setting, but on the cardinal utility this party yields relative to all other parties.

---

[13]Note that in this first specification we have that $z_{it}^{soph} = 1 - z_{it}^{sin}$.

Given that the type $z = (z^{sin}, z^{soph})$ is unobserved, we can at most infer the probability that each subject is of each type. We estimate $z$'s using the Expectation-Maximization (EM) algorithm.[14] Arcidiacono, Sieg, and Sloan (2007) elegantly summarizes what we do: given initial values of the parameters, we first calculate the conditional probability that an action is a particular type; using these conditional probabilities as weights, we treat types as observed and we maximize the (now) additively separable log-likelihood function; given the new parameter estimates, we then update the conditional probabilities of being each of the types and iterate until convergence (Arcidiacono, Sieg, and Sloan, 2007).[15] Since the likelihood of being a particular type is always strictly positive, we find that even when a vote seems to unambiguously be of type $z$, the algorithm can only assign it a probability of being of type $z$ which is arbitrarily close to 1, but never 1. In the Appendix we specify all computational details of the EM algorithm.

Throughout, we do a double exercise. First, we pool *all* observations for each district magnitude, and run the EM algorithm just once per district magnitude – i.e. we run it four times (M=0,1,2,3). For each action we find $\hat{z}_{it}$ and report the average across all individuals and periods (panels A in tables below). In other words, we assume that all observations are independent. We also run the algorithm separately for each subject (panels B in tables below).[16] In this case we also report the averages across all subjects. This second set of regressions are run as a robustness check and can be interpreted as controlling for individual fixed effects (we are estimating $z$'s for each subject separately). We can advance that results differ slightly among our two types of regressions but the patterns are the same regardless of which specification we use.

Table 5 reports the results for the two types model. For each district magnitude we report the percentage of individuals classified as sincere and those who are classified as sophisticated. We also report the percentage of correct vote predictions using $\hat{z}_{it}$ and the parameters of the model. In both panels A and B the estimated proportion of sincere types sharply increases with district magnitude. This is a strong result, yet we have to remember that for a large proportion of our observations, sincere and sophisticated actions are observationally equivalent. This means that we are forcing our

---

[14]See Dempster, Laird, and Rubin (1977) and Frühwirth-Schnatter (2006) for detailed descriptions of the algorithm.

[15]As mentioned earlier, our approach is analogous to Duch, May, and Armstrong II (2010) with the only difference that these authors use Bayesian methods while we use the computationally simpler EM algorithm.

[16]Given that we do not include first round observations when sophisticated and sincere coincides we have 48 observations per subject.

algorithm to decide between two types when both are potentially correct. Note that there are no results for our fully proportional treatment because in these cases both sincere and sophisticated behavior are equivalent (we also dropped the first rounds when sincere and sophisticated votes coincide). Note also that the percentage of observations we correctly classify is always quite high yet decreases with district magnitude.

|  | M=1 | M=2 | M=3 |
|---|---|---|---|
| **Panel A: Whole sample** | | | |
| **% sincere** | *21.98* | *33.49* | *55.99* |
| **% sophisticated** | *78.02* | *66.51* | *44.01* |
| % correctly predicted votes | *87.24* | *76.79* | *68.60* |
| observations | *2,304* | *2,400* | *3,360* |
| **Panel B: Average across subjects** | | | |
| **% sincere** | *30.56* | *36.24* | *51.50* |
| **% sophisticated** | *69.44* | *63.76* | *48.50* |
| % correctly predicted votes | *92.07* | *81.30* | *76.29* |
| observations | *2,256* | *2,352* | *3,312* |

TABLE 5. Proportion of subjects of each type (2 types)

We check statistical significance on the increase in the proportion of sincere voters by running a Kolmogorov-Smirnov test: for each district magnitude, the null hypothesis is that the cumulative distribution function of the distribution of $z_{ijt}^{sin}$s for $M+1$ is not larger than the same one for $M$. We do not report the results but we find this hypothesis rejected in both cases. That is, the distribution of $z_{it}^{sin}$s increase with district magnitude.[17]

Following our previous discussion, we next introduce a third type that captures whether subjects vote for the frontrunner party in the previous election. We denote it $z^{FR} = 1 - z^{sin} - z^{soph}$. Formally, $y_{ijt}$ is equal to 1 if in the previous round party $j$ obtained the most votes, and is equal to 0 otherwise. The utility of this type of voter is given by:

$$p_{ijt}(u_{ijt}, v_{ijt}, FR_{ijt}; z_{it}) = \frac{e^{\bar{\gamma}_j + \alpha FR_{ijt}}}{\sum_{k=1}^{k=J} e^{\bar{\gamma}_k + \alpha FR_{ikt}}} \qquad \text{if } z_{it}^{FR} = 1$$

---

[17]Note that panel B has slightly less observations than panel A. This is because in each treatment we had to drop one subject to reach convergence in our algorithm.

where $FR_j$ is a dummy that takes value 1 if $j$ was the party with most votes in the previous round and 0 otherwise.

Table 6 shows the results for the three types model. We consistently find the same patterns: in both panels A and B, the proportion of sincere actions increases with district magnitude. This result is statistically significant when using the Kolmogorov-Smirnov test. Interestingly, it is worth noting that around one in every ten votes is classified as frontrunner. This is indeed the proportion we observed in our control treatment with a fully proportional electoral system.

| | M=1 | M=2 | M=3 | PR |
|---|---|---|---|---|
| **Panel A: Whole sample** | | | | **% sincere** |
| **% sincere** | _19.99_ | _20.41_ | _35.86_ | _85.91_ |
| **% sophisticated** | _71.04_ | _63.73_ | _53.89_ | |
| **% frontrunner** | _8.97_ | _15.86_ | _10.25_ | _14.09_ |
| % correctly predicted votes | _91.32_ | _79.58_ | _78.24_ | _91.10_ |
| observations | _2,304_ | _2,400_ | _3,360_ | _2,112_ |
| **Panel B: Average across subjects** | | | | **% sincere** |
| **% sincere** | _27.30_ | _29.41_ | _41.62_ | _86.95_ |
| **% sophisticated** | _63.67_ | _55.82_ | _38.50_ | |
| **% frontrunner** | _9.03_ | _14.77_ | _19.88_ | _13.05_ |
| % correctly predicted votes | _95.30_ | _87.07_ | _85.45_ | _94.41_ |
| observations | _2,256_ | _2,352_ | _3,312_ | _2,112_ |

TABLE 6. Proportion of subjects of each type (3 types)

As was noted in Section 3, one of the main issues with our experimental design is that many actions are observationally equivalent. As a robustness check we now look at the subsample of observations in which the three types of action do not coincide. Note that there is no self-selection into this subsample: utilities are randomly assigned every five elections. Besides, expected utilities are computed given the behavior of *all* subjects in the previous round so there is no way that a subject can choose a particular ordering of parties according to utilities or expected utilities. The median number of observations per individual is 16, 9 and 5 in the treatments with district magnitude 1, 2, and 3, respectively. This pattern in the number of observations captures the mechanical effect of district magnitude: as district magnitude increases, there is a decrease in the proportion of observations where sophisticated and and sincere actions do not coincide. Table 7 reports results

for the subsample of disjoint observations (the PR treatment observations are considered disjoint when sincere does not coincide with frontrunner): once again we observe that the unconditional probability that an action is sincere increases. Even when the average values do not seem to confirm this last statement (for M=2), the Kolmogorov-Smirnov test rejects that the cumulative distribution function for $(\hat{z}^{sin}|M=3)$ is not larger than the c.d.f for $(\hat{z}|M=2)$ – this contradicts the reported average levels because the distribution of $\hat{z}^{sin}$'s is more sparse when $M=2$. A key lesson we take from this last table is the fact that the cases $M=2$ and $M=3$ are almost equivalent and both are distinctly different than $M=1$ and full PR.

|  | M=1 | M=2 | M=3 | PR |
|---|---|---|---|---|
| **Panel A: Whole sample** |  |  |  | **% sincere** |
| **% sincere** | 14.47 | 21.16 | 20.24 | 86.34 |
| **% sophisticated** | 79.39 | 57.17 | 59.63 |  |
| **% frontrunner** | 6.14 | 21.67 | 20.13 | 13.66 |
| **% correctly predicted votes** | 79.10 | 75.69 | 63.77 | 92.17 |
| **observations** | 312 | 218 | 138 | 1405 |
| **Panel B: Weighted average** |  |  |  | **% sincere** |
| **% sincere** | 13.95 | 21.96 | 20.67 | 86.67 |
| **% sophisticated** | 80.14 | 57.24 | 57.75 |  |
| **% frontrunner** | 5.91 | 20.80 | 21.57 | 13.33 |
| **% correctly predicted votes** | 78.71 | 76.46 | 63.55 | 91.87 |
| **observations** | 312 | 218 | 138 | 1405 |

TABLE 7. Proportion of subjects of each type (3 types) when M=1, 2, or 3 and Sincere, Sophisticated and Frontrunner do not coincide or in the case of PR, when Sincere and Frontrunner do not coincide. In Panel B, observations are weighed by the number of times the subject appears in the disjoint sample.

4.1. **Robustness Check.** We finally show that our results are not driven by the particular choice of the number of types or the EM algorithm. We do not run a conditional probit as such a model would estimate the values of the parameters for the average subject and we want to allow for subject heterogeneity – the mixed logit is our model of choice given that we can estimate parameters for each individual and report the distribution of such parameters among our population.

In our approach we have implicitly assumed that the value of each of our parameters is the same within each type and we have then assigned each of our subjects a probability of being each

type ($z$'s). Instead, the mixed logit assumes that all subjects simultaneously take into account utilitarian, sophisticated and frontrunner considerations (and allows subjects to place different weight on each component). Formally we have that:

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

where

$$\begin{pmatrix} \beta^i_{sin} \\ \beta^i_{soph} \\ \beta^i_{FR} \end{pmatrix} \sim \mathbf{N} \begin{pmatrix} \beta_{sin} & \sigma^2_{sin} & \sigma_{sinsoph} & \sigma_{sinFR} \\ \beta_{soph} & , & \sigma_{sinsoph} & \sigma^2_{soph} & \sigma_{sophFR} \\ \beta_{FR} & \sigma_{sinFR} & \sigma_{sophFR} & \sigma^2_{FR} \end{pmatrix}$$

We assume that the distribution of the vector of parameters $\beta$ follows a multivariate normal distribution. Below we report the estimated means of our mixed logit computations:

| | M=1 | M=2 | M=3 | PR |
|---|---|---|---|---|
| **Mean $\beta_{sin}$** | 2.97 | 1.92 | 3.78 | 14.99 |
| | (0.50) | (0.26) | (0.32) | (1.20) |
| **Mean $\beta_{soph}$** | 90.84 | 80.24 | 63.61 | |
| | (9.57) | (6.14) | (6.68) | |
| **Mean $\beta_{FR}$** | 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 8. 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 similar trends to those we observed earlier: sincere considerations (as captured by $\beta_{sin}$) increase with district magnitude in all cases apart from $M = 1$ to $M = 2$; sophisticated considerations (as captured by $\beta_{soph}$) decrease with district magnitude in all cases; and finally, frontrunner considerations are always present. Finally, we would like to highlight the advantage of our types model and EM algorithm by comparing the predictive power (percentage of correct observations) of different approaches. Analogously to Duch, May, and Armstrong II (2010) we define the *Proportional Reduction in Error* from model A to model B (PRE) as:

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

where $N$ is the total number of observations. We compare our approach to the more standard and prevalently used conditional logit model, and also to the mixed logit model. We find that our approach does a much better job: PRE is consistently above 30% for all district magnitudes and both models, and up to 60% when DM=1 (see Table 9, below).

|  | **M=1** | **M=2** | **M=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 9. Proportional Reduction in Error of our 3 types model with respect to three other specifications.

## 5. WHO IS SOPHISTICATED?

What drives sophisticated behaviour? Following our previous discussions, it seems reasonable that subjects in the first past the post treatments would be the most likely to be sophisticated given that it is easier to identify where a vote is most *pivotal*. As we increase district magnitude, the number of situations in which a vote can be pivotal increases and with it the complexity of identifying the sophisticated action. Below (see Table 10) we run a simple linear regression (with clustered standard errors by individual) where the dependent variable is the $z_{soph}$ we estimated for each subject in each election.[18] We drop the PR and first round observations where, by construction, sincere and sophisticated actions coincide.

We see that subjects are more likely to be sophisticated as we decrease district magnitude. The positive (and significantly different than 0) coefficient in the variable *period* captures the existence of some learning –along each of the sessions, subjects learn to take into account the pivotability of their vote and act accordingly. Towards the end of the experiment subjects are almost 20% more likely to act sophisticated than in the initial periods. The variable *time* captures the number of seconds subjects take in deciding their vote: subjects that take longer tend to be more sophisticated.

---

[18]We use the same specification as in Table 6 panel A.

| variables | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| DM2 | -0.073*** <br> (0.02) | -0.088*** <br> (0.02) | -0.090*** <br> (0.03) | -0.087*** <br> (0.02) |
| DM3 | -0.171*** <br> (0.02) | -0.166*** <br> (0.02) | -0.167*** <br> (0.020) | -0.155*** <br> (0.02) |
| round | 0.003 <br> (0.003) | 0.003 <br> (0.003) | 0.003 <br> (0.003) | 0.002 <br> (0.003) |
| period | 0.001*** <br> (0.0002) | 0.001*** <br> (0.0002) | 0.001*** <br> (0.0002) | 0.001*** <br> (0.0002) |
| time | | 0.002** <br> (0.001) | 0.002** <br> (0.001) | 0.002** <br> (0.001) |
| herfindhal | | | 0.117 <br> (0.086) | 0.133 <br> (0.085) |
| margin vote | | | -0.008*** <br> (0.002) | -0.007*** <br> (0.002) |
| age | | | | -0.006** <br> (0.002) |
| male | | | | 0.003 <br> (0.018) |
| experience | | | | -0.007 <br> (0.005) |
| economics | | | | -0.012 <br> (0.033) |
| non FPP | | | | 0.002 <br> (0.023) |
| British | | | | -0.029 <br> (0.025) |
| $R^2$ | 0.08 | 0.08 | 0.09 | 0.10 |
| Observations | 8064 | 8064 | 8064 | 8064 |

TABLE 10. Linear regression on the estimated probability of being sophisticated
the dependent variable is $z_{soph}$ as estimated in Table 6 panel A, standard errors (in brackets) are clustered at the individual level, ***p<0.01, **p<0.05, *p<0.1

In column (3) we add two variables that capture the overall vote distribution. On the one hand the coefficient on the *Herfindhal Index*[19] shows that more concentrated votes in the previous election allow subjects to be more sophisticated. This might be due to the increased salience of the sophisticated action. Similarly, the variable *vote margin* captures how contentious the fight for the last seat is and is negatively correlated with being sophisticated.[20] Once again, the smaller the vote margin, the more salient the sophisticated action and the more likely subjects are to target their vote to the race between those two parties. Finally, in the last column we add the socio-demographic characteristics we elicited at the end of the experiment.[21] The only effect we observe is that younger subjects tend to be more sophisticated than the rest of subjects.

---

[19]This index is commonly used to capture the concentration in votes: it is the sum of each party's squared vote share.

[20]*Vote margin* 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.

[21]Most variables we include are self-explanatory: *age*, *male*, *experience* (number of experiments the subject has participated in), *economics* (whether the students has/is studying economics), *non FPP* (whether the

We are also interested in finding out whether acting in a sophisticated manner impacts positively in the payoff that subjects achieve. The fact that subjects are seldom pivotal makes us a little skeptical about the findings of such an analysis, however, in table 11 we find that acting in a sophisticated way has a significantly positive effect on the payoff subjects obtain.

| variables | (1) | (2) |
|---|---|---|
| $z_{soph}$ | 10.578*** | 10.076*** |
| | (2.516) | (2.322) |
| lagged payoff | 0.679*** | 0.698*** |
| | (0.016) | (0.016) |
| round | -0.448 | -1.174** |
| | (0.481) | (0.555) |
| period | -0.115*** | -0.109*** |
| | (0.038) | (0.041) |
| time | 0.285** | 0.275*** |
| | (0.114) | (0.102) |
| DM2 | | 4.378** |
| | | (2.149) |
| DM3 | | 8.188*** |
| | | (2.048) |
| herfindhal | | 18.502 |
| | | (13.256) |
| margin vote | | 0.172 |
| | | (0.513) |
| age | | 0.136 |
| | | (0.156) |
| male | | 0.288 |
| | | (1.434) |
| experience | | 0.362 |
| | | (0.416) |
| economics | | -1.114 |
| | | (2.614) |
| non FPP | | 1.623 |
| | | (2.371) |
| British | | 2.001 |
| | | (1.955) |
| Individual fixed effects | YES | NO |
| $R^2$ | 0.52 | 0.52 |
| Observations | 6864 | 6864 |

TABLE 11. Linear regression on payoff subjects obtain

standard errors (in brackets) are clustered at the individual level, ***p<0.01, **p<0.05, *p<0.1

In the first column we run an individual fixed effects regression (with clustered standard errors at the individual level) and find that acting in a sophisticated manner (as captured by our estimated variable $z_{soph}$) has a positive and significant impact on the payoff obtained in that period. Note that the fixed effects regression implies that the positive coefficient reflects that being *more* sophisticated

---

electoral system of student's country of origin is not first past the post), or *British* (54% of our subjects are British.

in a particular election correlates positively with a higher payoff. We are not saying that more sophisticated individuals make higher profits but rather that per period small variations on the likelihood of being sophisticated leads to variations in that period's payoff.

The positive and significant coefficient on the likelihood of being sophisticated remains unchanged when we look at an alternative specification without fixed effects but with socio-demographic and electoral system controls (column 2). We see that subjects that take more time to vote usually obtain a higher payoff.

## 6. Conclusion

A widely-held assumption in political science is that non-sincere voting should be lower in higher-magnitude districts. Yet, formal work on the probability of being pivotal as well as actual voting data from elections suggests that voters are just as sophisticated in multi-member districts as they are in single-member districts (see for instance Abramson, Aldrich, Blais, Diamond, Diskin, Indridason, Lee, and Levine, 2010). One problem for empirical research, however, is that it is almost impossible to isolate the effect of district magnitude on voting behavior independently of voters' preferences, parties' positions or expectations about post-election coalition formation. We hence designed a lab experiment to isolate this effect, by varying district magnitude while keeping voters' preferences and parties' positions constant and excluding post-election coalition calculations.

We found an increase in sincere voting as district magnitude increased. This was due to a mechanical effect: as district magnitude increased, the proportion of voters who found that their most preferred party yielded the highest expected utility increased. However, not all of this increase in sincere voting was due to a mechanical effect because sophisticated voting did not increase with district magnitude and many voters failed to vote sincerely. In other words, part of the increase in sincere voting was due to psychological effects. The source of these effects remains a fruitful area for future research.

We also found that there was a high incidence (above 10% of our observations) of voting for the winner of the previous election (the frontrunner) in all our treatments. We even found this effect in our control treatment with fully proportional elections. The presence of frontrunner voting has not received much attention in the empirical and theoretical studies of voting. However, as Hinich

(1981) points out: "voting for the winner is no less plausible than the assumption that voters believe they can be pivotal".

## References

ABRAMSON, P. R., J. H. ALDRICH, A. BLAIS, M. DIAMOND, A. DISKIN, I. H. INDRIDASON, D. J. LEE, AND R. LEVINE (2010): "Comparing Strategic Voting Under FPTP and PR," *Comparative Political Studies*, 1(43), 61–90.

ARCIDIACONO, P., H. SIEG, AND F. SLOAN (2007): "Living Rationally Under the Volcano? An Empirical Analysis of Heavy Drinking and Smoking," *International Economic Review*, 48(1), 37–65.

AUSTEN-SMITH, D., AND J. S. BANKS (1988): "Elections, Coalitions and Legislative OutcomeS," *American Political Science Review*, 2(82), 405–422.

——— (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.

CAREY, J., AND S. HIX (2011): "The Electoral Sweet Spot: Low-Magnitude Proportional Electoral Systems," *American Journal of Political Science*, 2(55), 383–339.

COX, G. W. (1994): "Strategic Voting Equilibria Under the Single Nontransferable Vote," *American Political Science Review*, 3(88).

——— (2007): *Making Votes Count: Strategic Coordination in the World's Electoral Systems.* Cambridge University Press, Cambridge.

DEMPSTER, A., N. M. LAIRD, AND D. B. RUBIN (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of Royal Statistical Society, Series B (Methodological)*,

1(39), 1–38.

DUCH, R., J. MAY, AND D. A. ARMSTRONG II (2010): "Coalition-directed Voting in Multiparty Democracies," *American Political Science Review*, 104, 698–719.

DUVERGER, M. (1954): *Political Parties: Their Organization and Activity in the Modern State.* John Wiley, New York, NY.

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

——— (1997): "Voting Behavior and Information Aggregation in Elections With Private Information," *Econometrica*, 5(65), 1029–1058.

FEDDERSEN, T. J., AND A. SANDRONI (2006): "A Theory of Participation in Elections," *American Economic Review*, 4(96), 1271–1282.

FEY, M. (1997): "Stability and Coordination in Duverger s Law: A Formal Model of Preelection Polls and Strategic Voting," *American Political Science Review*, 1(91), 135–147.

FISCHBACHER, U. (2007): "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics*, 2(10), 171–178.

FISCHER, A. J. (1996): "A Further Experimental Study of Expressive Voting," *Public Choice*, 1(88), 171–180.

FORSYTHE, R., R. B. MYERSON, T. A. RIETZ, AND R. J. WEBER (1993): "An Experiment on Coordination in Multi-candidate Elections: The Importance of Polls and Election Histories," *Social Choice and Welfare*, 3(10), 223–247.

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

GERBER, E. R., R. B. MORTON, AND T. A. RIETZ (1998): "Minority Representation in Multi-Member Districts," *American Political Science Review*, 1(92), 127–144.

GREINER, B. (2004): "The Online Recruitment System ORSEE 2.0: A Guide for the Organization of Experiments in Economics," *Working Paper Series in Economics - University of Cologne*, 10.

HINICH, M. J. (1981): "Voting as an Act of Contribution," *International Economic Review*, 36(1), 135–140.

KAWAI, K., AND Y. WATANABE (2013): "Inferring Strategic Voting," *American Economic Review*, 103, 624–62.

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.

NAGEEB, A., AND N. KARTIK (2012): "Herding with Collective Preferences," *Economic Theory*, 51, 601–626.

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

REYNOLDS, A., AND M. STEENBERGEN (2006): "The Political Consequences of Ballot Design, Innovation and Manipulation," *Electoral Studies*, 3(25), 570–598.

RIAMBAU, G. (2013): "Bandwagon in Israel? Note on Bargsted and Kedar (2009)," mimeo.

SARTORI, G. (1968): *Political Development and Political Engineering.* Cambridge University Press, New York, NY.

SIMON, H. A. (1954): "Bandwagon and underdog effects and the possibility of election predictions," *Public Opinion Quarterly*, 18, 245–253.

## 7. Appendix

### 7.1. **Instructions.** (treatment M=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.

#### 7.1.1. *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.

#### 7.1.2. *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.

#### 7.1.3. *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.

#### 7.1.4. *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
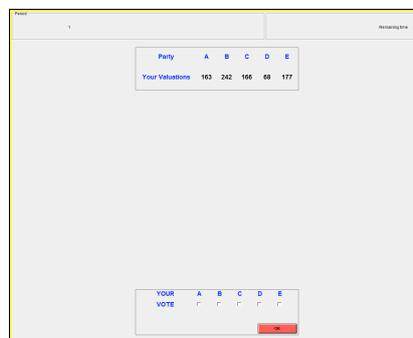

Screenshot at the beginning of election 1


Screenshot after election 1
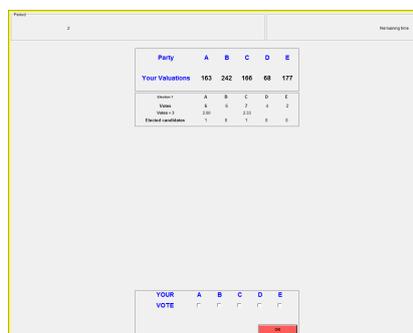

Screenshot at the beginning of election 2

FIGURE 2. Screenshots of the Ztree program for the treatment M=2.

7.2. **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.
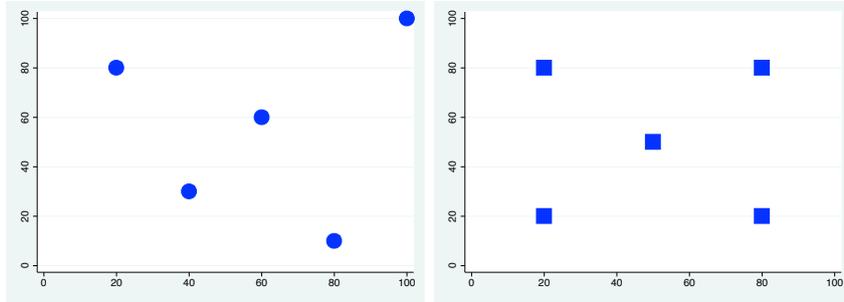


FIGURE 3. Location of political parties. In all treatments, subjects' preferred policies are uniformly distributed in the two dimensional policy space $[0, 100] \times [0, 100]$.

7.3. **Sophisticated voting.** A subject's *sophisticated* 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.[22] 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:

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

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 $\theta$ is the type of the voter (the utility he assigns to each party) and $V$ is any possible combination of votes for the five parties so that they add up to (n-1).

In our experimental setting, we restrict our attention to situations in which being sophisticated is solely driven 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,

---

[22]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 sophisticated vote coincides with the sincere one: voting for most preferred party.

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; Duch, May, and Armstrong II, 2010; McCuen and Morton, 2006).

7.4. **Estimation Procedure.** Suppose we want to find $\theta$ which maximizes $P(\mathbf{X}|\theta)$, $\theta = (\alpha, \beta, \pi)$. Say that we have an unobserved random variable $Z$, with a given realization $z$ (WLOG, for simplicity let us assume that $z = z^{soph}$, and that $\pi$ is the unconditional probability that an action is of type 'sophisticated'. Also for simplicity we will omit subscript 't'). Then

$$\hat{\theta}_0 \equiv \operatorname*{argmax}_{\theta} \; logP(\mathbf{X}|\theta) = \operatorname*{argmax}_{\theta} \; log\left\{\sum_z P(\mathbf{X}|z, \theta)P(z|\theta)\right\} \tag{1}$$

In Borman (2009) it is shown that

$$\max_{\theta} \; log\left\{\sum_z P(\mathbf{X}|z, \theta)P(z|\theta)\right\} \geq \max_{\theta} \sum_z P(z|\mathbf{X}, \theta_\tau) \; log \; P(\mathbf{X}, z|\theta) \tag{2}$$

where $\theta_\tau$ is the set of estimated parameters at iteration $t$. That is, given a set of estimates $\theta_\tau$, $\sum_z P(z|\mathbf{X}, \theta_\tau) \; log \; P(\mathbf{X}, z|\theta)$ is bounded above by $logP(\mathbf{X}|\theta)$. Therefore, given an unobserved vector $z$ and an educated initial guess $\theta_0$, maximizing the RHS of (2) will yield us $\hat{\theta}$, the best approximation to $\hat{\theta}_0$.

Since in our case

$$P_i(\mathbf{X}_i|z_i, \theta) = \left(f_i^{soph}\right)^{z_i} \left(f_i^{pty}\right)^{(1-z_i)}$$
$$P_i(z_i|\theta) = \pi$$
$$P_i(z_i|\mathbf{X}_i, \theta_\tau) = \hat{z}_i(\mathbf{X}_i, \alpha_\tau, \beta_\tau, \pi_\tau) \qquad \text{and}$$
$$P_i(\mathbf{X}_i, z_i|\theta) = \left(\pi f_i^{soph}\right)^{z_i} \left((1-\pi)f_i^{sin}\right)^{(1-z_i)}$$

we have that the best approximation to

$$\operatorname*{argmax}_{\alpha, \beta, \pi} \quad log\mathcal{LIK}(\mathbf{X}|z, \alpha, \beta, \pi) = \sum_i^N log\left\{\pi\left(f_i^{soph}\right)^{z_i} + (1-\pi)\left(f_i^{sin}\right)^{(1-z_i)}\right\}$$

is given by

$$\operatorname*{argmax}_{\alpha, \beta, \pi} \quad \sum_{i=1}^N \hat{z}_i(.)log\left(\pi f_i^{soph}\right) + (1-\hat{z}_i(.))log\left((1-\pi)f_i^{sin}\right) =$$

$$\operatorname*{argmax}_{\alpha, \beta, \pi} \quad \sum_{i=1}^N \hat{z}_i(.)log(\pi) + \hat{z}_i(.)log\left(f_i^{soph}\right) + (1-\hat{z}_i(.))log(1-\pi) + (1-\hat{z}_i(.))log\left(f_i^{sin}\right)$$

which is what one actually maximizes.

7.5. **The M-STEP in detail.** We will omit subscript $\tau$ for ease of notation. At iteration $\tau$, in order to get estimates of $\{\alpha, \beta, \pi\}$, we need to solve

$$\operatorname*{argmax}_{\alpha, \beta, \pi} \sum_{i=1}^N \{\hat{z}_i log\pi + (1-\hat{z}_i)log(1-\pi)\} + \sum_{i=1}^N \hat{z}_i \left\{\sum_j y_{ij} log(P_{ij}|z_i=1)\right\}$$
$$+ \sum_{i=1}^N (1-\hat{z}_i) \left\{\sum_j y_{i1} log(P_{ij}|z_i=0)\right\} \tag{3}$$

Recall that

$$(P_{ij}|z_i=0) = \frac{e^{u_{ij}\alpha + \delta_j}}{\sum_{j=1}^J e^{u_{ij}\alpha + \delta_j}} \qquad (P_{ij}|z_i=1) = \frac{e^{v_{ij}\beta + \delta_j}}{\sum_{j=1}^J e^{v_{ij}\beta + \delta_j}}$$
$$\delta_j = [\delta_1, \delta_2, ..., \delta 5] \text{ with } \delta_1 = [0,0,0,0,0] \text{ for the base outcome}$$
$$\text{and } 1_{y_{ij}} = 1 \text{ if } i \text{ votes for } j, 0 \text{ otherwise.} \tag{4}$$

FOCs yield:

$$[\pi] \Rightarrow \hat{\pi} = \frac{\sum_i^N \hat{z}_i}{N}$$

$$[\alpha] \Rightarrow \frac{\partial log\mathcal{LIK}}{\partial \alpha} = \sum_i^N 1_{y_{ij}} \times \hat{z}_i \times (\Psi_{ij}|z_i = 1) \times v_{ij} +$$

$$\sum_i^N 1_{y_{ij}} \times (1 - \hat{z}_i) \times (\Psi_{ij}|z_i = 0) \times u_{ij} \overset{!}{=} 0$$

$$[\beta_{s,j}] \Rightarrow \frac{\partial log\mathcal{LIK}}{\partial \beta_{s,j}} = -\sum_i^N 1_{y_{ij}} \times \hat{z}_i \times (P_{ij}|z_i = 1) \times x_i^a -$$

$$\sum_i^N 1_{y_{ij}} \times (1 - \hat{z}_i) \times (P_{ij}|z_i = 0) \times x_i^a \overset{!}{=} 0$$

where

$$(\Psi_{ij}|z_i = 1) = \frac{\sum_j e^{v_{ij}\alpha + \delta_j} \times v_{ij}}{\sum_j e^{v_{ij}\alpha + \delta_j}}, \qquad \delta_j = [0, 0, ..., 0] \text{ for the base outcome}$$

$$(\Psi_{ij}|z_i = 0) = \frac{\sum_j e^{u_{ij}\alpha + \delta_j} \times u_{ij}}{\sum_j e^{u_{ij}\alpha + \delta_j}},$$

Note: Since the likelihood conditional on any type is always strictly positive, then we find that even when a vote is obviously sincere, the algorithm can only assign it a probability of being sincere which is arbitrarily close to one, but never one (and, in some cases, given the available information, is not necessarily close to one). Moreover, the probability assigned to each single action depends not only on all the information we have on that action (embedded in the likelihood function), but also on the average of all the probabilities assigned to all actions in the previous round. Therefore, even if in one particular round a voter cast a vote that was unambiguously sincere, it is likely that the probability assigned by the algorithm converges to *but is not* one. for each action, the probability that it is of type $z$ is a weighted average of the likelihoods conditional on that action being of all possible types (sincere, sophisticate, frontrunner). More specifically, these probabilities are defined at each iteration $r$ as follows by the EM algorithm :

$$\Pr(\text{action } a_{it} = \text{type } z) \text{ in iteration } r \equiv$$

$$\hat{z}_{it}(r) = \frac{\Pi_z \cdot \text{Likelihood } a_{it} \mid a_{it} = \text{type } z}{\Pi_{sin} \cdot (\text{Lik. } a_{it} \mid a_{it} = \text{sin}) + \Pi_{sin} \cdot (\text{Lik. } a_{it} \mid a_{it} = \text{soph.}) + \Pi_{FR} \cdot (\text{Lik. } a_{it} \mid a_{it} = \text{FR})}$$

$$z = \{\text{sincere, sophisticated, frontrunner}\}$$

$$\Pi_z = \sum_i^N \sum_t^T \hat{z}_{it}(r - 1)$$

(5)