CALTECH/MIT
VOTING TECHNOLOGY PROJECT

A multi-disciplinary, collaborative project of
the California Institute of Technology – Pasadena, California 91125 and
the Massachusetts Institute of Technology – Cambridge, Massachusetts 02139

Assessing the impact of voting technologies on multi-party electoral outcomes: the case of Buenos Aires' 2005 Congressional Election

GABRIEL KATZ
R. MICHAEL ALVAREZ
CALTECH

ERNESTO CALVO
UNIVERSITY OF HOUSTON

MARCELO ESCOLAR
UNIVERSIDAD DE BUENOS AIRES

JULIA POMARES
LONDON SCHOOL OF ECONOMICS

Key words: ***Buenos Aires elections, voting technology, congressional elections***

VTP WORKING PAPER #64
Apr 2008

Assessing the impact of voting technologies on multi-party electoral outcomes: the case of Buenos Aires' 2005 Congressional Election

Gabriel Katz, R. Michael Alvarez, Ernesto Calvo, Marcelo Escolar, and Julia Pomares*

April, 2008

Abstract:

This paper presents the first study on the impact of different voting technologies on election outcomes in multi-party elections, analyzing data from a large-scale voting experiment conducted in the 2005 congressional election in Buenos Aires, Argentina. Combining different regression models and matching methods, we estimate the effect of alternative voting technologies on the probability of support for the competing parties in the elections for congress and state legislature. The results of the different statistical techniques indicate that voters are extremely receptive to the information cues provided by the different voting technologies and associated ballot designs, and that particular voting devices have a significant impact on voter choice, systematically favoring some parties to the detriment of others. We conclude that the choice of alternative electronic voting devices might have considerable effect on electoral outcomes in multi-party electoral systems.

*Gabriel Katz, Graduate student, California Institute of Technology, gabriel@hss.caltech.edu, corresponding author; R. Michael Alvarez, Professor of Political Science, California Institute of Technology; Ernesto Calvo, Associate Professor, University of Houston; Marcelo Escolar, Professor, Universidad de Buenos Aires; Graduate student, London School of Economics.

1. Introduction

An increasing number of countries around the world have adopted electronic voting systems in national and local elections since the 1990s, and many others are conducting pilot projects (Milkota, 2002; Alvarez and Hall, 2004, 2005, 2008). While the academic literature has focused mainly on the reliability and accuracy of different electronic voting technologies (Alvarez et al., 2001; Alvarez and Hall, 2004, 2008; Stewart, 2004; Ansolabehere and Stewart, 2005) only a few empirical studies have directly examined the effect of different voting technologies on election outcomes (Wand, 2004; Card and Moretti, 2007; Herron and Wand, 2007; Herron, Mebane and Wand, 2008). Empirical studies have even been fewer in multiparty electoral systems, where with a larger number of parties and candidates on a ballot, voters might be more responsive to readily available information and thus may resort to different cues in order to identify and distinguish the various electoral options and to select their preferred choice (Conover and Feldman, 1989; Reynolds and Steenbergen, 2006).

In this paper, we analyze how different voting technologies influence voters' choice and election outcomes in multiparty races, examining evidence from a voting pilot conducted in the 2005 congressional election in Buenos Aires, Argentina, in which four e-vote prototypes were tested. Combining alternative regression models and non-parametric matching methods in order to assess the robustness of our results and strengthen the validity of our conclusions, we show that voters alter their electoral behavior and their vote choice in response to different e-vote technologies, and that this might translate into different electoral outcomes across voting devices. Our main findings are in line with the results of Calvo, Escolar and Pomares (2007), in the sense that 'technology matters', and that different voting technologies and associated ballot designs might have substantive effects on election results in multi-party electoral systems.

The remainder of the paper is organized as follows. The next section briefly reviews the literature on the impact of voting technologies on electoral outcomes. Section 3 describes the Buenos Aires pilot project. Section 4 presents preliminary descriptive evidence on the effect of the different voting technologies considered on voters' electoral behavior. In Section 5, we describe the data and the different methodologies used to assess the effect of the voting technologies on voter choice and electoral outcomes. Section 6 presents and comments the main empirical results, and Section 7 concludes.

2. The effect of electronic voting technologies on electoral outcomes

Previous research has underscored several ways in which different electronic voting technologies could favor some parties or candidates over others, thus influencing election results. One possibility that has received much public attention and media coverage, particularly in the aftermath of the 2000 U.S. presidential election, is whether certain devices are more vulnerable to illegal manipulation and vote tampering (Stewart, 2004; Card and Moretti, 2007; Herron and Wand, 2007), raising the possibility of electoral fraud. Also, there has been some recent work, especially that focused on Florida's 13th Congressional district race in 2006, that studies how ballots are formatted and displayed for voters, and whether this influences how ballots are cast (e.g., Frisina et al. 2008). Finally, the adoption of electronic voting devices might have a differential effect on turnout rates among different socio-demographic segments of the population, because certain groups might find computers more intimidating or confusing than others (Tomz and Van Houweling, 2003; Moretti and Card, 2007).

Thus, previous work indicates that characteristics and features of voting devices may influence voters' electoral choices. Voting systems differ in many ways, for example their ballot design, in the information and contents presented to the voters, and in the possibility of changing or correcting a vote (Herrnson et al., 2008). If the characteristics of specific voting technologies disproportionately affect the likelihood of making mistakes and casting spoiled ballots among groups of voters who share a partisan propensity, then the potential bias in recorded votes relative to intended votes exhibited by different voting technologies might substantially affecting election outcomes (Sinclair and Alvarez, 2004; Herrnson et al., 2006; Herrnson et al., 2008). A related concern regarding the design of voting devices is that differences in the amount and the form in which information is presented to voters might affect the cues they use to identify and select the candidates (Herron, Mebane and Wand, 2008). Voting devices that require well-informed voters or that provide different information shortcuts for decision-making might favor some parties to the detriment of others, potentially affecting election results. This is particularly relevant in multiparty elections that impose higher information demands on voters and increase the potential influence of design effects on voters' electoral choice (Reynolds and Steenbergen, 2006).

Empirical evidence regarding the question of whether and to what extent different voting technologies actually affect election results is far from conclusive (Card and Moretti, 2007; Herron and Wand, 2007; Frisina et al., 2008; Herron, Mebane and Wand, 2008). All these studies, however, have analyzed U.S. elections, where the number of competing candidates is limited compared to multi-party electoral systems, and relied on observational data.[1] In this paper, we use data from a field experiment conducted in the City of Buenos Aires in order to test for the impact of ballot design and informational effects on candidate choice and election outcomes in multi-party elections. Our experimental setting mitigates concerns related to vote tampering, differential turnout rates, endogenous adoption of voting technologies (Knack and Kropf, 2002; Saltman, 2006; Herron, Mebane and Wand, 2008) and other challenges posed by observational data (Herron and Wand, 2007).

3. The Buenos Aires' 2005 Pilot Project

Voters in the congressional election held in Buenos Aires in October 2005 elected national representatives and state legislators using a party-list paper ballot system that included candidates for all offices.[2] Seats were allocated using a PR-D'Hont formula with closed party lists of magnitude 13 for representatives and 30 for legislators. Thirty parties presented candidate lists for national representatives, while forty one parties presented lists for the state legislature. Three parties captured approximately 66% of the valid votes in the election of national representatives and 64% in the election of state legislators: President Kirchner's *Frente para la Victoria (FPV)*, the center-left opposition party *Alianza para una*

[1] This is the case of most research on the effect of voting technologies. Important exceptions are Herrnson et al. (2006; 2008) and Sinclair et al. (2000); none of these studies, however, are centered on the impact of design effects on the support for different candidates or parties.

[2] The description of the e-vote pilot borrows from Calvo, Escolar and Pomares (2007). See also Alvarez (2005), and the reference materials at http://www.vote.caltech.edu/Elections2005/05.htm

Republica de Iguales (ARI), and the center-right *Propuesta Republicana (PRO)*.[3] None of the remaining parties obtained more than 5% of the vote in either of the two congressional races (Ministry of Interior of Argentina, 2008). The campaign for national representatives was very intense, with high spending in support of the candidacies of Rafael Bielsa (FPV), Elisa Carrio (ARI), and Mauricio Macri (PRO). By contrast, candidates to the local legislature spent almost no money during the campaign (Calvo, Escolar and Pomares, 2007).

The e-pilot was conducted in 41 precincts randomly distributed throughout the city and included 14,800 participants. After voting in the official election, participants in each precinct were asked to participate in a non-binding election in which they were randomly assigned to one of four possible voting devices and were asked to vote a list of national deputies and a list of local legislators. Because the experiment was carried out in a single electoral district, with participants in each precinct being randomly assigned to the different voting devices and facing similar menus of party choices, we expect no correlation between the characteristics of the district or the election and voters' behavior.[4]

After the vote, participants in the experiment were asked to complete two surveys. The first survey was a short self-administered survey (six questions) conducted with 13,830 respondents. Half of the questions were identical across prototypes, dealing with general

---

[3] If blank ballots are excluded, the vote share of these three parties comes close to 70%.

[4] Organizational problems prevented the testing of all the protoypes in all the precincts, as originally planned (Alvarez, 2005). While *Prototypes 1* and *2* were tested in all the precincts, *Prototype 3* was tested in 40 precincts, and *Prototype 4* in only 17 precincts. Even though we do not expect this to have resulted in serious imbalance between groups of participants assigned to the different prototypes, we take this potential problem into account in the empirical analysis below.

perceptions about their e-vote experience. The remaining questions tested usability issues specific to each device. A fourth of the participants also answered a longer exit poll. This survey further investigated the opinions and attitudes towards electronic voting and its alternative, hand-counted paper ballots. The survey also provided information about the voters' political sophistication, their familiarity with technology, and their patterns of political participation. It is worth mentioning that, while the vote in the experiment was non-binding and the election results did not count as such, survey responses showed that voters were concerned that the results would be "used" as an exit poll, suggesting that there were incentives to choose the same option in the experiment as in the real election.

The four voting devices tested in the pilot were developed with the existing institutional process of Argentina in mind.[5] *Prototype 1* was a direct recording electronic (DRE) design with two separate modules. A screen in the first module allowed voters to review the lists of candidates, and a numerical keypad was used to register the vote. *Prototype 2* was a touch-screen *DRE* machine with a voter verifiable paper trail. Voters could scroll and select party lists directly by tapping onto the screen, and vote information was digitally stored in the machine, which produced a paper trial to comply with Argentine electoral laws. *Prototype 3* was an optical scan (OS) prototype located inside a voting booth, providing a higher degree of privacy. The voter introduced the paper ballot into a rolling scanner that displayed the selected party on the prototype's screen, and would then proceed to confirm her selection.

[5] Several provinces in Argentina have tested e-voting systems and adapted their legislation to allow for its potential introduction. In Tucuman, Argentina's smallest province, electronic voting was introduced in 2005, and is currently the only voting system used for its provincial elections.

This prototype required separate ballots for each race, allowing direct comparison of the marks that identify a party across races. Finally, *Prototype 4* was an optical scan device with a single ballot listing all parties' names and their numbers. The voter marked her preferences for each race with a pencil and then introduced the ballot into a scanner located next to the election desk. All four prototypes asked the voter to confirm her choices at the end of the process, preventing over and under-counts. Also, in all prototypes, participants voted for National Representatives first and State Legislators second.

An important difference between the DRE and OS prototypes was the way in which voters were required to search for their preferred candidates. In the DRE prototypes, party labels were randomly rotated on the screen and, because of space restrictions, a limited number of labels were displayed on each screen. Two and three screens were required to display party labels for national representatives and state legislators in *Prototype 1*, while three and four screens were required in *Prototype 2*. The placement of the party labels rotated randomly for each voter, preventing order effect biases from favoring the same party. In the case of *Prototype 3*, poll workers sorted the paper ballots numerically.[6] According to the information obtained from the polling place workers, however, ballots rapidly mixed in the voting booth, complicating the search for the voters' preferred ballots. Finally, in *Prototype 4*, party names where listed by their official list number in increasing order. The

[6] When registering the candidates running for a specific election, each party is assigned a different list number. Candidates and Parties advertise this number during the campaign, together with the party and candidate's name.

non-random ordering of parties may have increased the likelihood of order effects but it also facilitated the recognition of the same party across races.

A second relevant difference among the prototypes was how voters accessed information about candidates and parties. The first prototype displayed 15 party names on each screen, including the list number and party logo information. In order to view the list of candidates, however, the voter needed to enter the three-digit party number. If the voter did not know the name of the party, she would need to access each list until finding a recognizable candidate name. *Prototype 2*, on the other hand, displayed the name of the first candidate under the party label, together with the number and logo information. The complete list of candidates was then displayed on a second navigation level. Parties with prominent first candidates (such as the pro-Kirchner Rafael Bielsa from the FPV or Mauricio Macri of the center-right PRO) were readily identified by voters.[7] However, given that voters could recognize without any effort the name of their preferred congressional candidate, very little information about the party name or number was recalled when casting their legislative vote. Hence, while voters faced fewer problems in recognizing their preferred choice for national representative, they could not use such information when choosing state legislators.

Different information was available to voters using the optical scan systems. Ballot papers for *Prototype 3* included all the relevant information, such as party name, party logo, identification number, and the complete list of candidates for each race. The only difficulty in identifying the preferred choice, therefore, was in finding the correct paper ballot. In

[7] Bielsa was President Kirchner's Foreign Relations Minister at that time, while Macri is a famous businessman and was the president of one of the most famous soccer teams in Argentina.

Prototype 4, a booklet provided voters with all the party information; the ballot introduced in the rolling scanner listed only the party name, number and logo. It is also worth mentioning that some of the political parties favored the OS prototypes because of their closer resemblance to the actual voting mechanism used in Buenos Aires congressional election. The main characteristics of the four prototypes tested in the experiment are summarized in this paper's supplementary materials (Appendix I).

4. A first look at the effect of different voting technologies on voter choice

The survey data lets us examine how voters interacted with each prototype and how the different voting technologies and the associated ballot designs affected voters' electoral choice. Table 1 presents data about which ballot features participants used to identify their preferred candidates. Nearly half of the voters cast their ballot based using the party name, followed by the name of the first candidate. The party's name was particularly important for those participants using *Prototype 4*: 53.4% of those using the single-ballot OS device stated that they relied on the party name when casting their vote. In contrast, only 44.3% of respondents in *Prototype 3* used the party name as an information cue. Also, the name of the first candidate was more relevant for those assigned to *Prototype 2*, while participants using *Prototype 1* were less likely to use it as a voting cue, using more frequently the party number instead. This is consistent with the characteristics of the ballot designs associated with the different prototypes: recall that the name of the first candidate figured prominently on the second prototype's screen, while voters using *Prototype 1* could access the candidates' names only after entering each party's number in the keypad. The p-value for the modified Pearson's chi-square test proposed by Loughin and Scherer (1998) with 5,000 bootstrap resampled data sets is 0.07, indicating a significant relationship between the information used by respondents as voting cues and the voting technologies.[8]

Table 1

How voters found their preferred candidates[*]

Information used as voting cue	Prototype 1 (%)	Prototype 2 (%)	Prototype 3 (%)	Prototype 4 (%)	All prototypes (%)
Party name	51.4	51.0	44.3	53.4	49.4
First candidate's name	33.3	50.1	47.1	45.0	44.2
Party Logo	27.3	30.3	22.4	7.4	25.8
Party number	35.4	21.0	19.9	28.6	25.3
Other features	4.1	2.7	7.5	6.4	4.6
N	879	1,158	858	189	3,084

[*] Table entries are the percentage of respondents in each prototype that used each of the ballot features to identify their preferred candidates. Since participants could use several of the ballot features as voting cues, percentages do not necessarily sum to 100 across rows.

[8] Given that respondents could use several of the ballot features in order to identify their preferred choice, the assumption of independence among units required by the standard test of independence is violated. We implemented Loughin and Scherer's (1998) bootstrap resampling method based on a modified Pearson chi-square statistic in order to test for association between voting cue and prototype.

Table 2, in turn, reports the percentage of participants who were not able to vote for their preferred candidate for each of the prototypes, sorted by education and political information levels.[9] The survey data indicates that education levels affected the ability of the participants to vote for their preferred party: while only 3.8% of voters with college education were unable to cast a vote for their preferred option, this figure was almost 2.6 times higher for those with high school education or lower. The difference in the proportions between the two groups is statistically significant, with a 95% confidence interval of [0.04, 0.08] (Newcombe, 1998). Although less educated voters experienced more difficulties using all of the prototypes tested, the gap between participants with college education and the rest was much smaller for *Prototype 2*: the sample odds ratios (Agresti, 2002) for the probability of not being able to vote for the preferred candidate range from 1.7 for this prototype to 4.7 for the Protoype 1, suggesting that the touch-screen DRE device imposed substantially lower barriers on less educated voters than the other voting devices. The p-value of Woolf's (1955) test for homogeneity across prototypes is 0.001, indicating that there are considerable differences across voting technologies regarding the difficulties experienced by less educated participants.

[9] Political information was computed as the average of respondents' number of correct answers to three questions asking them he name of the minister of economy, the minister of education and the minister of health.

Table 2

Percentage of voters who could not vote for their preferred candidate by education and political information level*

Variable	Prototype 1 (%)	Prototype 2 (%)	Prototype 3 (%)	Prototype 4 (%)	All prototypes (%)
Education					
College	3.0	2.7	6.5	3.6	3.8
Secondary or lower	12.6	4.5	13.6	12.9	9.8
N	3,175	3,873	2,743	887	10,678
Non-response rates	21.4	18.4	28.2	27.5	22.8
Political information					
Null	9.9	3.4	11.4	0.0	7.3
Low	7.3	4.1	11.7	2.4	6.9
Medium	1.7	4.3	11.5	7.3	5.7
High	3.0	3.8	10.5	3.8	5.4
N	835	1,108	823	185	2,951
Non-response rates	5.0	4.3	4.1	2.1	4.3

* Table entries are the percentage of respondents in each prototype that were not able to cast a vote for their preferred candidate, among all respondents belonging to each row-category that were assigned to that prototype. The data on education levels was taken from the short self-administered survey, while the data on political information was obtained from the longer exit poll.

When we examine the data by political information levels, again *Prototype 2* seems to have allowed voters with no or low political information to vote for their preferred choice. *Prototype 3,* in contrast, exhibits the highest rates of reported voting problems for all levels of political information. The Cochran-Armitage Trend Test (Agresti, 2002) provides evidence of a modestly negative linear relationship between political information and reported voting problems (two-sided p-value = 0.1), but this is only statistically significant (at the 0.01 level) for *Prototype 1*. Overall, almost 90% of the voters were able to vote for their preferred party, with a rate of success ranging from 93.9% for *Prototype 2* to 82.6% for *Prototype 3;* the hypothesis of independence between voting device and reported voting difficulties is rejected at the usual confidence levels.[10]

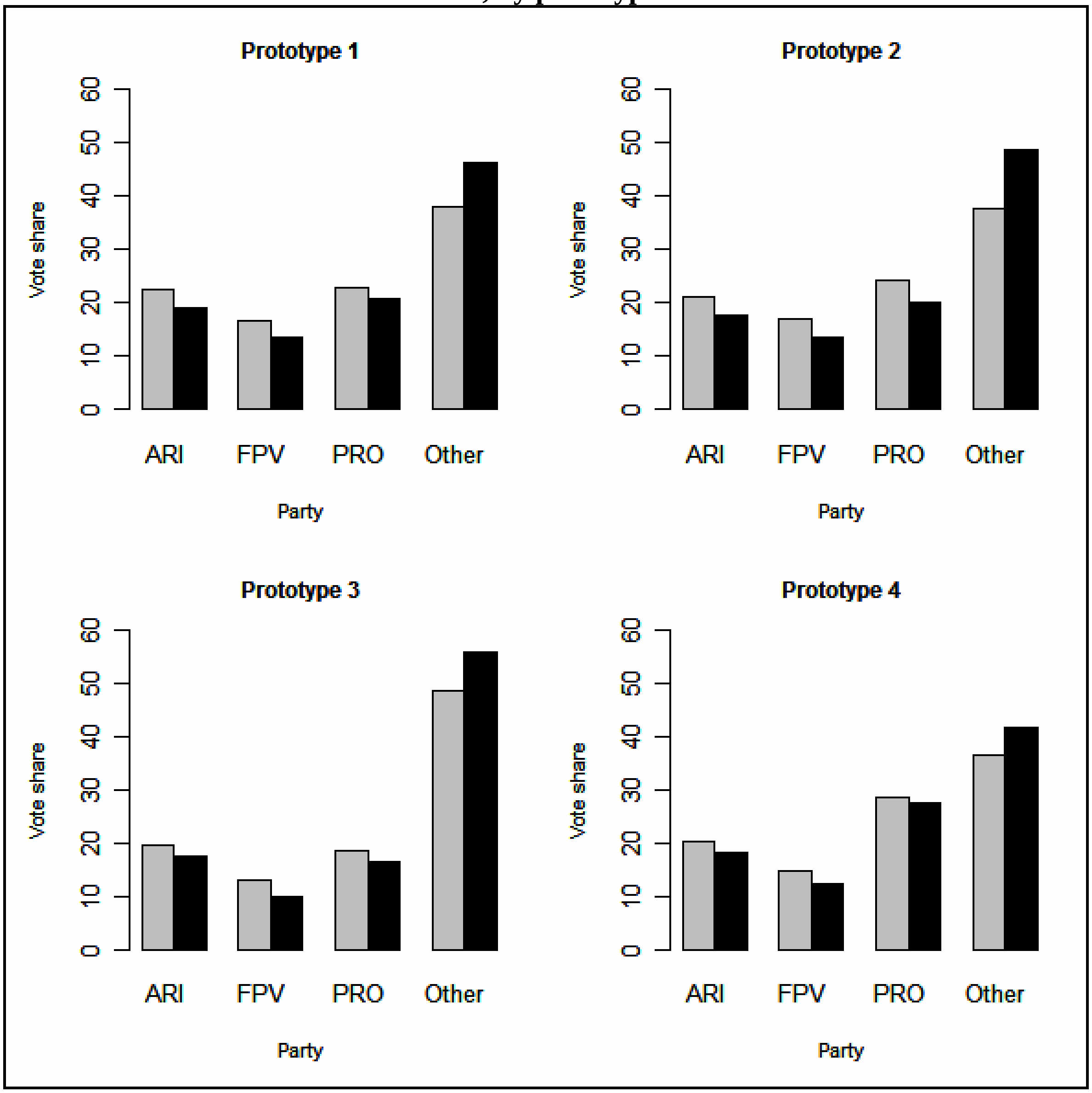
The fact that the four prototypes impose different information demands on voters and seem to have influenced the cues they used in their decision-making process suggests that the voting devices could have had systematic effects on electoral outcomes. For instance, parties with more visible candidates may have fared relatively better among voters using *Prototype 2*, and those with more recognizable names/logos might have benefited from the ballot design and screen display in the DRE devices. Figure 1 explores this issue further, reporting the actual vote-share of ARI, FPV, PRO and Other parties in the election of national representatives and state legislators for each prototype.[11]

---

[10] Pearsons' chi-squared test yields $\chi^2 = 50.3$, with 3 degrees of freedom.

[11] Vote-shares are expressed as percentages of the total number of votes cast for the competing parties in both races, excluding blank and null votes.

Figure 1
Parties' vote share, by prototype and election



Note: The gray bar indicates parties' vote-shares in the election of National Representatives; the black bar corresponds to to the election of State Legislators.

For all the prototypes tested, each of the three largest parties, *Alianza para una Republica de Iguales (ARI)*, *Frente para la Victoria (FPV)* and *Propuesta Republicana (PRO)*, exhibited higher vote-shares in the first election, jointly obtaining 65% of the total vote cast for the parties competing in the election of national representatives. In contrast, the smallest parties gathered almost 50% of the vote in the less visible state legislative race. However, there are remarkable variations in the vote-share of the different parties across prototypes. First, the vote-share of the smaller parties included in the category "Other" is substantially higher in *Prototype 3*, reaching 48.7% in the national representative election and 55.7% in the state legislative election. In contrast, the support for minor parties was the lowest among voters using *Prototype 4*, with 36.4% and 41.6% respectively. Also, the relative support for the three largest parties varied across prototypes. The total vote-share of ARI, FPV and PRO in the National (Local) election was 21.0% (18.2%), 15.6% (12.6%) and 22.6% (19.9%), respectively. However, for both races, ARI fared relatively better among voters using *Prototype 1*, while FPV maximized its vote-share among those using *Prototype 4*; the support for PRO, in contrast, did not vary substantially across voting devices.

As reported in Table 3, based on Pearson's chi-squared test statistic, the hypothesis that the average proportion of votes obtained by each of the parties under the different prototypes is the same is rejected at the 0.9 confidence level in both congressional races. In order to analyze the discrepancy in the mean support of the parties across prototypes, we used bootstrapped Kolmogorov-Smirnov tests (Abadie, 2002; Diamond and Sekhon, 2005; Mebane and Sekhon, 2008) to examine the statistical significance of differences in each party's vote-share between pairs of voting devices. The p-values of these pairwise tests, based on 10,000 replicates, are also reported in Table 3. The results point to statistically significant differences between *Prototype* 3 and the two DRE devices regarding the support for the three largest parties: the Kolmogorov-Smirnov p values are highly significant for the comparisons of FPV and PRO's vote-shares under *Prototypes 1* and *2* and their support under *Prototype 3* in both races, as well as for the comparison of ARI's vote-share under *Prototypes 1* and *3* in the national representative election. In the case of the smaller parties, there are statistically significant differences between their support under *Prototype 3* and each of the remaining prototypes in the two elections analyzed.

Table 3

Comparison of parties' vote-shares across prototypes - p-values of the tests

	Comparisons across prototypes	**ARI**	**FPV**	**PRO**	**Other parties**
Election of National Representatives	**Equal support across prototypes[a]**	0.03	0.00	0.00	0.00
	Pairwise comparisons[b]				
	Prototypes 1-2	0.13	0.87	0.33	0.70
	Prototypes 1-3	0.06	0.02	0.06	0.00
	Prototypes 1-4	0.55	0.69	0.32	0.67
	Prototypes 2-3	0.20	0.00	0.00	0.00
	Prototypes 2-4	0.99	0.51	0.54	0.56
	Prototypes 3-4	0.42	0.36	0.20	0.01
Election of State Legislators	**Equal support across prototypes[a]**	0.09	0.00	0.00	0.00
	Pairwise comparisons[b]				
	Prototypes 1-2	0.04	0.51	0.71	0.13
	Prototypes 1-3	0.12	0.02	0.03	0.01
	Prototypes 1-4	0.84	0.85	0.12	0.13
	Prototypes 2-3	0.57	0.01	0.06	0.00
	Prototypes 2-4	0.13	0.84	0.25	0.00
	Prototypes 3-4	0.11	0.18	0.10	0.00

[a] p-values based on Pearson's chi-squared test statistic.

[b] p-values of bootstrapped Kolmogorov-Smirnov tests based on 10,000 replicates.

5. Estimating the effect of different e-vote technologies on election outcomes

While the data presented in the previous section reveals some interesting differences in voters' electoral behavior across voting devices, it does not allow us to assess the relative impact of the different technologies and ballot designs on the voter choice after accounting for the effect of socio-demographic and attitudinal variables. Controlling for these predictors might be relevant in order to estimate the causal effect of the prototypes on voters' choice and election outcomes (Gelman and Hill, 2007), given that not all of the four prototypes were used in all the districts analyzed.[12]

As our data includes the individual level votes for all the participants in the e-vote pilot who were randomly assigned to the different prototypes, we can analyze the aggregate electoral and survey data from voting stations defined by crossing each of the precincts with the four voting devices.[13] The models we estimate below use data from 128 voting stations for the national representative and state legislative election. Our dependent variable is the vote-share of ARI, FPV, PRO and Other parties in the election for national representative and state legislatures in each of the voting stations, where the category "Other parties" comprises all the remaining parties in both races.[14] The independent variables included in

[12] See footnote 4.

[13] Although the individual vote variable can be retrieved from each prototype's logs, privacy considerations prevented us from linking the individual vote with the individual survey data. Also, we dropped 924 observations with missing values from our analysis. Combining the information from the logs and the surveys, we have data from 128 out of the 139 total voting stations.

[14] "Other parties" includes 26 smaller parties in the election for National Representatives and 37 parties in the election for the State Legislature.

the model are: *Education*, measured as the average years of schooling; the mean level of *Political Information,* summarized by the voters' responses to three political knowledge questions; *Interest in politics,* coded on a three-point scale ranging from "not interested" (1) to "very interested" (3); the mean level of participants' *Use of Technology*, estimated using factor analysis on a series of questions asking respondents about their use of cellular phones, personal computers and the internet; *Evaluation of E-voting*, a measure of voters' assessment of the difficulty of electronic voting, coded on a five point scale ranging from 'very difficult' (1) to 'very easy' (5); and four variables measuring the percentage of participants who found their preferred party searching by *Party Name*, by *Party Logo*, by *Party Number*, or by *Candidate Name*. All the independent variables are defined at the voting-station level; descriptive statistics for these variables are provided in Appendix II of this paper's supplementary materials.

In order to estimate the causal effect of different voting technologies on the expected support for the parties competing in the 2005 election, we used both regression models and matching methods. Specifically, we implemented two alternative hierarchical regression models for aggregate polytomous data that allow for extra variation relative to the baseline multinomial model and take into account the pilot project's experimental design. In addition, we conducted a complementary non-parametric analysis, applying matching methods to estimate the causal effects of using alternative voting devices on the probability of support for each of the parties in the two elections under study. Although our analysis is based on experimental data, previous research shows that statistical corrections might be needed even in an experimental setup in order to achieve balance across the units of analysis and improve inferences regarding causal effects (Barnard et al., 2003; Imai and van Dyk, 2004; Imai, 2005). As mentioned before, this might be relevant for the Buenos Aires pilot project. In addition, this non-parametric approach imposes fewer assumptions, allowing us to assess the robustness of the main conclusions drawn from the regression models, which are heavily dependent on functional form assumptions such as linearity and additivity (Herron and Wand, 2007).

The number of votes for the different parties in each voting station forms a vector of counts that can be analyzed using multinomial regression models of count data (Cameron and Trivedi, 1998; McCullagh and Nelder, 1989). As it is known, however, aggregate vote data generally exhibits higher variability than the basic multinomial model can account for, and thus several alternative specifications have been proposed to deal with overdispersion (McCullagh and Nelder, 1989, p. 74), allowing for heterogeneity over units in the allocation of probabilities across categories and seeking to robustify inferences in the presence of outlying observations (Mebane and Sekhon, 2004; Congdon, 2005).

In this paper, we fit and contrast two different hierarchical regression models: a multiple-logit model with a logistic normal distribution for the multinomial probabilities (Leonard and Hsu, 1994; Congdon, 2005) and a modified version of Katz and King's (1999) additive logistic model for vote proportions. The multiple-logit model allows for a more flexible covariance structure than the conjugate multinomial-Dirichlet model for count data, it is more computationally feasible and it provides a more general hierarchical structure by using a multivariate normal distribution for the logits (Leonard and Hsu, 1994; Agresti and Hitchcock, 2005). The additive logistic model, on the other hand, transforms the parties' vote-shares in each voting station into multivariate logits that are assumed to follow a multivariate Student-t distribution, attenuating the influence of outliers and also allowing the covariances between votes for the parties to differ *vis a vis* the basic multinomial model; Katz and King's (1999) approach has been shown to outperform similar models for vote proportions, such as Aitchison's additive logistic normal model (1986), Tomz, Tucker and Wittenberg's (2002) SUR model and Jackson's (2002) heteroskedastic SUR model (Katz and King, 1999; Mebane and Sekhon, 2004). In both specifications, the probabilities of support for the parties are modeled as functions of the voting-station covariates described above. In addition, in order to account for the cluster sampling scheme used in the Buenos Aires experiment and to allow for unobserved heterogeneity across voting stations and for potential correlation in the election results across prototypes and precincts, we include zero-mean random effects for the two non-nested factors (Congdon, 2005; Gelman and Hill, 2007). In both models, we use "Other parties" as the baseline category.

Letting $V_i = \left(V_i^{ARI}, V_i^{FPV}, V_i^{PRO}, V_i^{OTHER}\right)'$ and $VS_i = \left(VS_i^{ARI}, VS_i^{FPV}, VS_i^{PRO}, VS_i^{OTHER}\right)'$ denote the vector of votes for ARI, FPV, PRO and Other parties and the corresponding vector of vote-shares in voting station $i,\ i = 1,...n,$ defined by the intersection of precinct $h,\ h = 1,...,H$ and prototype $p,\ p = 1,...,4,$ the two models considered are:

a) Multiple-logit model:

$$V_i \sim Multinomial\left(N_i, \pi_i\right) \tag{1}$$

$$\pi_i^j = \frac{\exp(\mu_i^j)}{1+\sum_j \exp(\mu_i^j)}, \quad j\text{=ARI, FPR, PRO} \tag{2}$$

$$\pi_i^{Other} = \frac{1}{1+\sum_j \exp(\mu_i^j)} \tag{3}$$

$$\mu_i^j = X_i \beta^j + \lambda_p^j + \eta_h^j, \qquad j = ARI, FPV, PRO \tag{4}$$

where $N_i = \sum_k V_i^k$, k=ARI,FPV,PRO,Other, $\pi_i = \left(\pi_i^{ARI}, \pi_i^{FPV}, \pi_i^{PRO}, \pi_i^{Other}\right)$, X_i is a vector of predictors (including a constant term) defined at the voting station level, β^j is a vector of fixed-effects, and λ_p, η_h, are independent prototype- and precinct- random effects following multivariate normal distributions:

$$\lambda_p = \left[\lambda_p^{ARI}, \lambda_p^{FPV}, \lambda_p^{PRO}\right]' \sim MVN\left(0, \Sigma_\lambda\right), \qquad p = 1, \ldots, 4, \tag{5}$$

$$\eta_h = \left[\eta_h^{ARI}, \eta_h^{FPV}, \eta_h^{PRO}\right] \sim MVN\left(0, \Sigma_\eta\right), \qquad h = 1, \ldots, H \tag{6}.$$

b) Additive logistic Student-t model:

$$Y_i \sim MVT\left(\mu_i, \Sigma_\varepsilon, \upsilon\right) \tag{7}$$

with

$$Y_i^j = \log\left(\frac{VS_i^j}{VS_i^{Other}}\right), \; j = ARI, FPV, PRO \tag{8}$$

$$\mu_i^j = X_i \beta^j + \lambda_p^j + \eta_h^j, \quad j\text{=ARI,FPV,PRO} \tag{9}$$

and X_i β^j, λ_p, η_h are defined as above.[15] Each party's vote is then obtained using the additive-logistic transformations:

$$V_i^j = \left(\frac{\exp\left[Y_i^j\right]}{1 + \sum_j \exp\left[Y_i^j\right]} \right) \times N_i, \quad j = ARI, FPV, PRO; \quad V_i^{OTHER} = \left(\frac{1}{1 + \sum_j \exp\left[Y_i^j\right]} \right) \times N_i \quad (10).$$

Both models were fit by MCMC Gibbs sampling methods (Gelfland and Smith, 1990; Casella and George, 1992). The main advantage of using fully Bayesian estimation is that it allows obtaining arbitrarily precise approximations to the posterior densities, without relying on large-sample theory (Fahrmeier and Knorr-Held, 2000; Jackman, 2004). Additional details of the estimation are provided in Appendix III (Supplementary Materials).

We used posterior predictive simulations (Iyengar and Dey, 2004; Gelman and Hill, 2007) to compare the fit of both models. Specifically, we used two posterior predictive checks. First, we contrasted the ability of each of the models to replicate the overdispersion present in the data by computing

[15] For those voting stations in which the vote-share of at least one of the four categories is zero, we use the "modified Aitchison technique" proposed by Fry, Fry and McLaren (1996). The main advantage of the "modified Aitchison technique" is that, unlike the Box -Cox transformation, it is invariant to the category used as baseline (Fry, Fry and McLaren, 1996).

$$P\left(\chi^2_{\text{Rep}} > \chi^2_{Obs}\right) = \frac{1}{R}\sum_{r=1}^{R}\left(\sum_{i=1}^{I}\sum_{k}\frac{VAR\left(V_i^{\text{k.Rep}(r)}\right)}{E\left(V_i^{\text{k.Rep}(r)}\right)} - \frac{VAR\left(V_i^{\text{k.Obs}}\right)}{E\left(V_i^{\text{k.Obs}}\right)}\right) \tag{11}$$

where, for the voting station defined by crossing precinct i and prototype p and for each party k, $k = ARI, FPV, PRO, Other$, $V_i^{\text{k.Rep}(r)}$ is number of votes sampled from the predictive distribution $f\left(V_i^{\text{Rep}}, \theta \middle| V_i^{\text{Obs}}\right)$ of the model being considered, $V_i^{\text{k.Obs}}$ is the observed number of votes, and R is the number of convergent Gibbs samples of the model's parameters, θ. A satisfactory model will have $P\left(\chi^2_{\text{Rep}} > \chi^2_{Obs}\right)$ around 0.5 (Congdon, 2005).

Also, following Iyengar and Dey (2004), a complementary comparison criteria based on the discrepancy between observed and simulated data would favor the model that minimizes the predictive loss $d\left(VS^{\text{Rep}}, VS^{Obs}\right) = E\left(\left\|VS^{\text{Rep}} - VS^{Obs}\right\|^2 \middle| VS^{Obs}\right)$, where again VS^{Rep} and VS^{Obs} are vectors of replicated and observed vote-shares. Using the Gibbs samples of each model's parameters, d can be estimated as:

$$\hat{d} = \frac{1}{R}\sum_{r=1}^{R}\left(\sum_{i=1}^{I}\left\|VS_i^{Obs} - VS_i^{\text{Rep}(r)}\right\|^2\right) \tag{12}$$

for the two models under consideration.

In addition, we performed a complementary analysis using Genetic Matching (Diamond and Sekhon; 2005; Sekhon 2007), which is a nonparametric method for performing multivariate matching based on an algorithm proposed by Mebane and Sekhon (1998). Genetic Matching has been shown to have better properties than the usual alternative matching methods, such as Mahalanobis distance and propensity score matching (Diamond and Sekhon, 2005; Sekhon, 2007).

In this application, we conducted a series of pairwise matching exercises (Ho et al., 2007; Herron and Wand, 2007), where each exercise considered the average causal effect on each party's vote-share of using *Prototype* r versus *Prototype* s, with $r, s = 1,..,4,\ \mathrm{r} < s,$ and voting stations using *Prototype* r taken to be the "treated" group. Using one-to-one matching with replacement, we found sets of matched pairs based on the socio-demographic covariates included in X, where each pair contains one voting station using *Prototype* r and one using *Prototype* s.[16] The average causal effects for each party $k,\ k = ARI, FPV, PRO, OTHER,$ were estimated as the average treatment effect on the treated (ATT):

$$E\left\{E\left(VS_i^k \mid X_i, \text{Prototype}_i = r\right) - E\left(VS_i^k \mid X_i, \text{Prototype}_i = s\right) \mid \text{Prototype}_i = r\right\}, \tag{13}$$

with confidence intervals were computed based on Abadie-Imbens' (2006) standard errors, which account for the asymptotic variance induced by the matching procedure itself.[17] Appendix IV (Supplementary Materials) discusses additional details of the matching

[16] See Mebane and Sekhon (1998), Diamond and Sekhon (2005) and Sekhon (2007) for details on the distance measure and the evolutionary algorithm used to minimize the maximum discrepancy between the matched treated and control covariates in the context of Genetic Matching.

[17] This matching analysis was implemented using "Matching", a package for R developed by Mebane and Sekhon (2008).

analysis and reports the p-values of the two-sample and paired t-tests before and after matching.

6. Results

Tables 4 and 5 report the posterior means and standard deviations of the fixed effects parameters in the multiple-logit and additive logistic Student-t models for both races. The multiple-logit model exhibits a better fit: the posterior predictive loss $\hat{d}$ is smaller than for the compositional model, and it satisfactorily replicates the overdispersion present in the data, with values of $P\left(\chi^2_{\text{Rep}} > \chi^2_{Obs}\right)$ close to 0.5 for the two elections under study. The additive logistic Student-t model, on the other hand, generates predictions that are overdispersed relative to the observations.

Table 4

Estimated posterior means and standard deviations for the fixed effects*

Election of National Representatives

Parameter	Multiple–logit model			Additive logistic Student-t model		
	ARI	*FPV*	*PRO*	*ARI*	*FPV*	*PRO*
Education	0.10 (0.14)	-0.23*** (0.09)	0.29** (0.12)	0.11 (0.11)	-0.21* (0.11)	0.37*** (0.13)
Political information	0.54* (0.32)	0.27 (0.33)	-0.36 (0.34)	0.48 (0.35)	-0.21 (0.39)	-0.06 (0.42)
Interest in Politics	-0.15 (0.19)	0.41* (0.21)	0.24 (0.20)	-0.29 (0.22)	0.39* (0.23)	0.14 (0.25)
Use of Technology	0.05 (0.16)	0.10 (0.17)	0.25 (0.17)	-0.16 (0.19)	0.04 (0.20)	-0.13 (0.19)
Assessment of E-voting	0.19 (0.43)	0.34 (0.35)	0.19 (0.36)	0.23 (0.40)	0.08 (0.43)	0.34 (0.43)
Search by Party Name	-0.54** (0.26)	-0.18 (0.28)	-0.44* (0.26)	-0.81** (0.29)	-0.10 (0.33)	-0.63** (0.33)
Search by Party Logo	0.01 (0.31)	0.02 (0.34)	0.24 (0.33)	-0.31 (0.40)	0.08 (0.42)	-0.30 (0.44)
Search by Party Number	-0.06 (0.32)	0.77** (0.35)	0.43 (0.34)	-0.44 (0.39)	0.87** (0.44)	0.06 (0.42)
Search by Candidate Name	-0.39 (0.25)	-0.06 (0.25)	-0.73*** (0.27)	0.09 (0.30)	-0.20 (0.33)	0.26 (0.31)
Intercept	-1.13	-1.03	-2.73**	-0.78	-0.44	-3.29***

	(1.44)	(0.68)	(1.09)	(0.97)	(1.01)	(1.22)
$P\left(\chi^2_{\mathrm{Rep}} > \chi^2_{Obs}\right)$	0.42			1.00		
$\hat{d}$	2.84			5.69		

* Significance levels: *** 0.01, ** 0.05, *0.1.

Table 5

Estimated posterior means and standard deviations for the fixed effects*

Election of State Legislators

Parameter	**Multiple–logit model**			**Additive logistic Student-t model**		
	ARI	*FPV*	*PRO*	*ARI*	*FPV*	*PRO*
Education	0.14 (0.10)	-0.23** (0.11)	0.29* (0.15)	0.18* (0.10)	-0.29** (0.12)	0.37*** (0.13)
Political information	0.70** (0.30)	-0.01 (0.33)	-0.09 (0.33)	0.49 (0.32)	0.13 (0.43)	0.07 (0.39)
Interest in Politics	-0.09 (0.19)	0.44* (0.22)	0.51*** (0.19)	-0.22 (0.24)	0.49 (0.29)	0.50* (0.28)
Use of Technology	0.01 (0.16)	0.33* (0.18)	0.22 (0.16)	-0.10 (0.17)	0.28 (0.24)	0.06 (0.20)
Assessment of E-voting	0.36 (0.40)	0.05 (0.50)	-0.16 (0.37)	0.27 (0.39)	0.01 (0.50)	0.27 (0.41)
Search by Party Name	-0.11 (0.27)	-0.59** (0.31)	-0.29 (0.27)	-0.56 (0.30)	-0.41 (0.36)	-0.51 (0.35)
Search by Party	-0.05	0.18	0.45	-0.45	0.29	-0.10

Logo	(0.32)	(0.35)	(0.34)	(0.35)	(0.47)	(0.41)
Search by Party Number	-0.21 (0.33)	0.52 (0.39)	0.12 (0.33)	-0.26 (0.39)	0.81 (0.48)	-0.08 (0.41)
Search by Candidate Name	-0.07 (0.24)	0.05 (0.28)	-0.47* (0.27)	-0.01 (0.26)	-0.27 (0.32)	-0.44 (0.31)
Intercept	-2.48** (1.05)	-0.77 (1.15)	-3.44** (1.31)	-1.97* (1.04)	-0.54 (1.16)	-4.35*** (1.15)
$P\left(\chi^2_{\text{Rep}} > \chi^2_{Obs}\right)$	0.57			1.00		
$\hat{d}$	3.02			6.78		

* Significance levels: *** 0.01, ** 0.05, *0.1.

For both model specifications, the results in Tables 4 and 5 reveal some interesting differences regarding the effect of several covariates on the support for the three largest parties. For instance, in the two elections considered, the votes for *Propuesta Republicana (*PRO) increased in voting stations with higher average levels of education, while they decreased for *Frente para la Victoria (*FPV). In contrast, higher average levels of political interest were associated with higher support for FPV. This result is consistent with prior research that emphasizes class and education effects among non-Peronist voters (Mora y Araujo & Llorente, 1980; Calvo & Murillo, 2004). Interestingly, the percentage of respondents interested in politics was also related to the support for PRO in the election for state representatives under both models. Also, the multiple-logit model indicates that the average level of political information was positively related to the vote support for *Alianza Republicana Independiente* (ARI), but not for the other two large parties. Participants' assessments regarding the degree of difficulty posed by the electronic voting procedure had no systematic effect in the linear predictors μ^j, $j = ARI, FPV, PRO$.

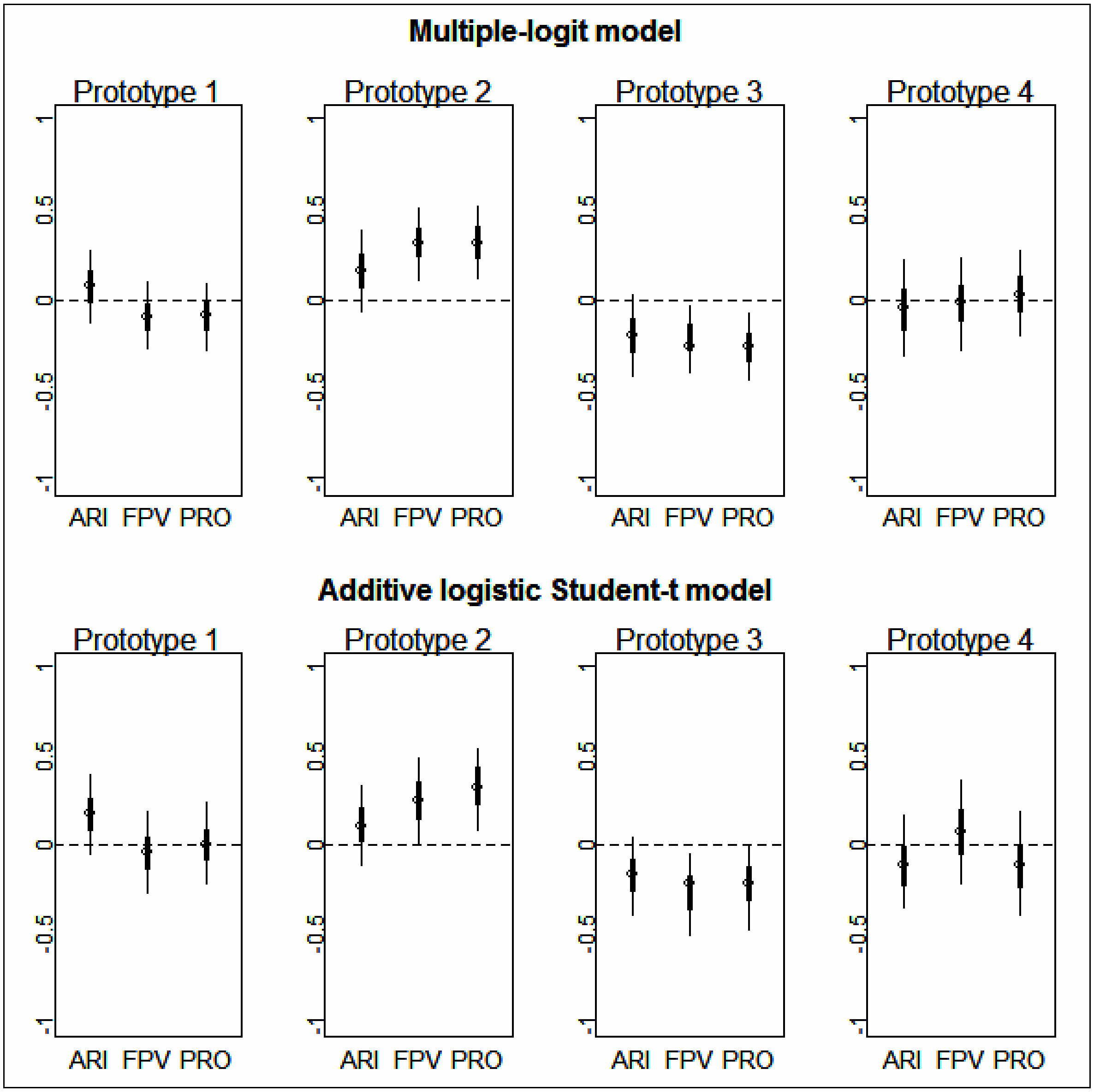
Regarding the effect of the different information cues used by participants when casting their vote, both models show that the support for FPV in the more visible race increased with the percentage of voters relying on the official party number. The votes for ARI and PRO, on the other hand, were negatively related to the percentage of participants using the name of the party in the election for national representatives, while in the multiple-logit specification there is also a negative relationship between *Search by Party Name* and μ^{FPV} in the less visible election. In the case of the multiple-logit model, μ^{PRO} was also negatively associated to the percentage of voters basing their choice on the first candidate's name in both congressional elections; this relationship is not statistically significant at the usual confidence levels for the additive logistic specification. Remarkably, although more than a quarter of the participants in the experiment reported having used the logo information as a voting cue (Table 2), neither of the two models indicates a systematic effect of *Search by Party Logo* on the support for the competing parties in either of the two elections.

The main focus of our analysis, however, lies in the effect of the different voting technologies on the support for the competing parties across elections. Figures 2 and 3 present the estimates and confidence intervals of the centered prototype random-effects for the two hierarchical models, computed as $\tilde{\lambda}_p^j = \lambda_p^j - \frac{1}{4}\sum_{p=1}^{4}\lambda_p^j, \quad j = ARI, FPV, PRO$. This gives us more precise inferences about the relative values of the prototype coefficients and the impact on the support for the three largest parties (Gelman and Hill, 2007).

Figure 2

Centered Prototype random coefficients $\tilde{\lambda}_p^j$ for both hierarchical models

Election of National Representatives

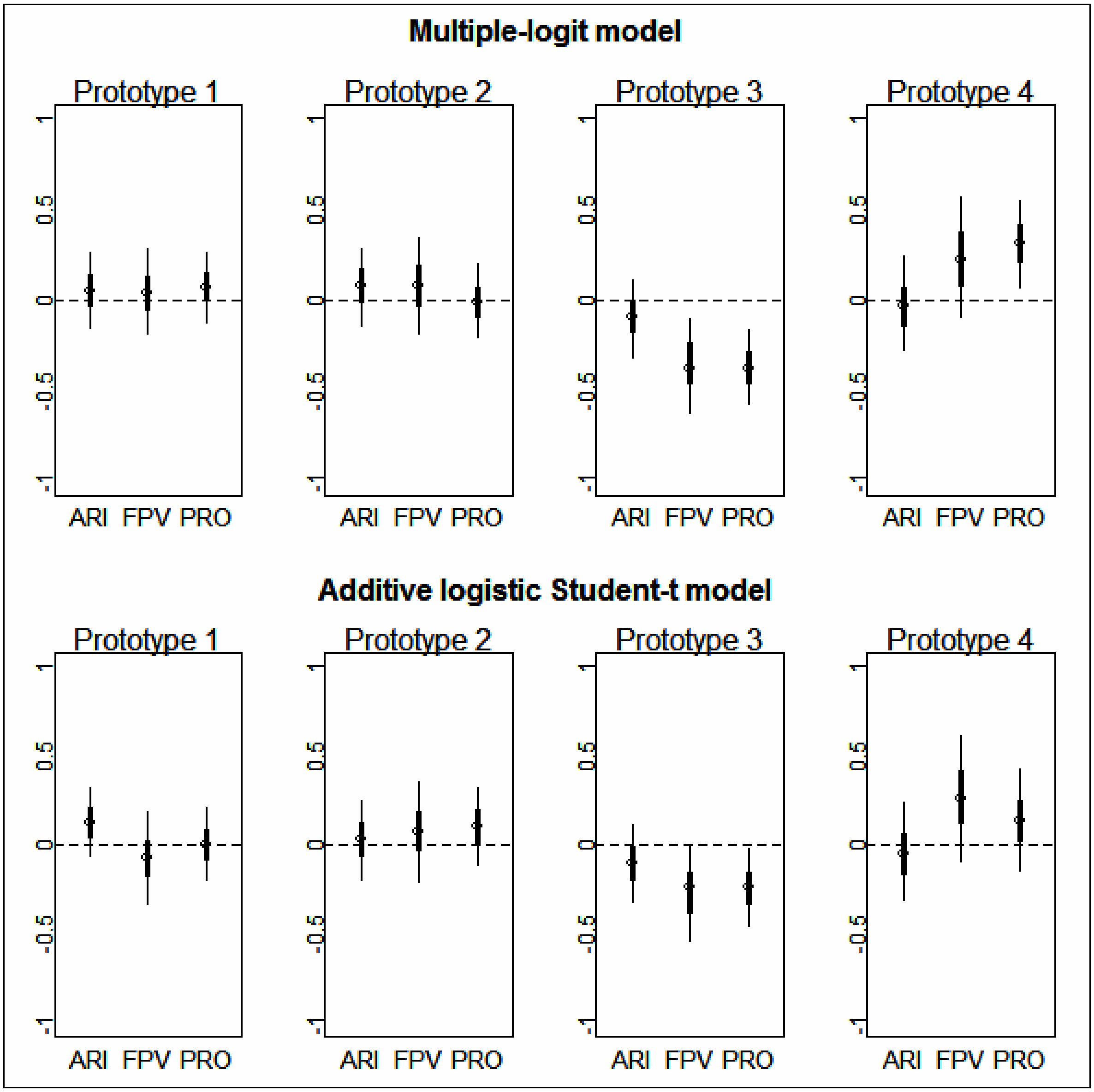


Note: The center dots correspond to the posterior means of $\tilde{\lambda}_p^j$, $j = ARI, FPV, PRO$, the thicker lines to the 50% confidence interval, and the thinner lines to the 90% confidence interval.

Figure 3

Centered Prototype random coefficients $\tilde{\lambda}_p^j$ for both hierarchical models

Election of State Legislators



Note: The center dots correspond to the posterior means of $\tilde{\lambda}_p^j, \quad j = ARI, FPV, PRO$, the thicker lines to the 50% confidence interval, and the thinner lines to the 90% confidence interval.

The evidence presented in these figures indicates that different voting devices have potential influences on electoral outcomes, even after controlling for socio-demographic and behavioral variables. For both model specifications, the effect of the voting technologies and the associated ballot designs vary considerably across parties and, to a certain extent, across races. Moreover, the main substantive conclusions resulting from Figures 2 and 3 are essentially the same for the multiple-logit and additive logistic models, suggesting that the results are robust to alternative specifications and modeling strategies. For instance, while the Optical Scan device with separate ballots (*Prototype 3)* had a significantly negative effect on the votes for FPV and PRO in both congressional elections, the touch-screen DRE device (*Prototype 2*) had the opposite effect, raising the support for FPV and PRO in the election for national representatives, although not in the election for state legislators. As mentioned above, the name of the first candidate of each party figured prominently in the screen of *Prototype 2*, and more than half of the participants using this device cast their vote based on this information. Hence, a possible interpretation of this result is that, while the first candidates of the FPV and PRO, Rafael Bielsa and Mauricio Macri, were renowned figures who were easily identifiable by voters, participants generally did not recognize the candidates running for the local legislature of any of the competing parties (Calvo, Escolar and Pomares, 2007), and thus the relative advantage obtained by the FPV and PRO in the more salient election disappeared in the less visible race. Nonetheless, it is interesting to note that, as seen in Tables 5 and 6, neither of the models indicates a systematic effect of the percentage of respondents using the first candidate’s name on the linear predictor for FPV in either of the races, while the multiple-logit model even points to a negative relationship between *Search by Candidate Name* and μ^{PRO} in the two congressional races. More generally, the effect of *Prototype 2* on the vote support for FPV and PRO is positive and statistically significant even after controlling for the effect of the different voting cues used by participants and their evaluation of the difficulty of electronic voting. This indicates that the prototype-effects might be capturing additional sources of variability in the dependent variables, beyond that explained by the aggregate survey data. On the other hand, there is no evidence of systematic advantages induced by *Prototype 1* in favor of any of the competing parties, while, in the case of *Prototype 4*, $\tilde{\lambda}_4$ is statistically significant only for *Propuesta Republicana* under the multiple-logit model, and just in the election for state legislators. Also, in contrast to our findings for FPV and PRO, the results in Figures 2 and 3 indicate the support for the other large party, ARI, was not significantly related to any of the four prototypes tested in the experiment at the usual confidence levels.

Table 6 complements the information presented in Figures 2 and 3, reporting the mean posterior and 90% confidence intervals of the pairwise differences in the average probabilities π^k, $k = ARI, FPV, PRO, OTHER$, across prototypes. Given the better fit of the hierarchical multiple-logit specification, we use the Gibbs samples of the parameters in this model in order to compute the causal effects of the different voting technologies on π^k *via* average predictive comparisons (Gelman and Hill, 2007). The main substantive findings are similar for the additive logistic model.

Table 6

Pairwise differences in the probability of support for each party across prototypes[a]

(in percentage points)

	Pairwise comparisons	π^{ARI}	π^{FPV}	π^{PRO}	π^{OTHER}
Election of National Representatives	Prototypes 1-2	2.1 (-4.2, 8.7)	-3.6 (-8.4, 1.0)	-5.2 (-10.9, 0.4)	6.6 (0.8, 12.4)
	Prototypes 1-3	3.4 (-3.2, 9.9)	0.2 (-4.0, 4.4)	1.2 (-4.6, 2.1)	-4.8 (-10.8, 1.7)
	Prototypes 1-4	2.9 (-0.4, 6.0)	-1.0 (-4.3, 1.9)	-2.4 (-6.0, 1.0)	0.5 (-3.6, 4.3)
	Prototypes 2-3	1.3 (-0.5, 3.3)	3.8 (2.1, 5.7)	6.3 (4.4, 8.3)	-11.0 (-13.7, -9.2)
	Prototypes 2-4	0.7 (-6.9, 8.0)	2.6 (-3.3, 8.3)	2.7 (-4.2, 9.4)	-6.1 (-13.0, 1.0)
	Prototypes 3-4	-0.5 (-7.7, 6.5)	-1.2 (-6.3, 4.9)	-3.6 (-9.8, 2.4)	5.3 (-2.1, 13.4)
Election of State Legislators	Prototypes 1-2	-0.6 (-5.4, 5.1)	-0.55 (-5.5, 4.6)	1.5 (-4.2, 6.9)	-0.4 (-7.7, 6.5)
	Prototypes 1-3	-0.2 (-5.3, 5.2)	2.7 (-1.5, 7.4)	5.3 (-0.1, 10.4)	-7.8 (-15.1, -0.8)
	Prototypes 1-4	2.5 (-0.49, 5.35)	-1.7 (-5.0, 1.4)	-4.0 (-7.8, -0.3)	3.1 (-0.9, 7.2)
	Prototypes 2-3	0.4 (-1.4, 2.2)	3.3 (1.8, 4.9)	3.8 (1.9, 5.7)	-7.5 (-9.8, -5.1)
	Prototypes 2-4	3.1 (-2.8, 8.7)	-1.1 (-7.8, 5.2)	-5.5 (-12.2, 1.5)	3.5 (-4.6, 12.3)
	Prototypes 3-4	2.7 (-3.0, 8.0)	-4.4 (-10.6, 1.1)	-9.3 (-15.6, -3.1)	11.0 (2.8, 19.5)

[a] 90% confidence intervals are reported in parenthesis

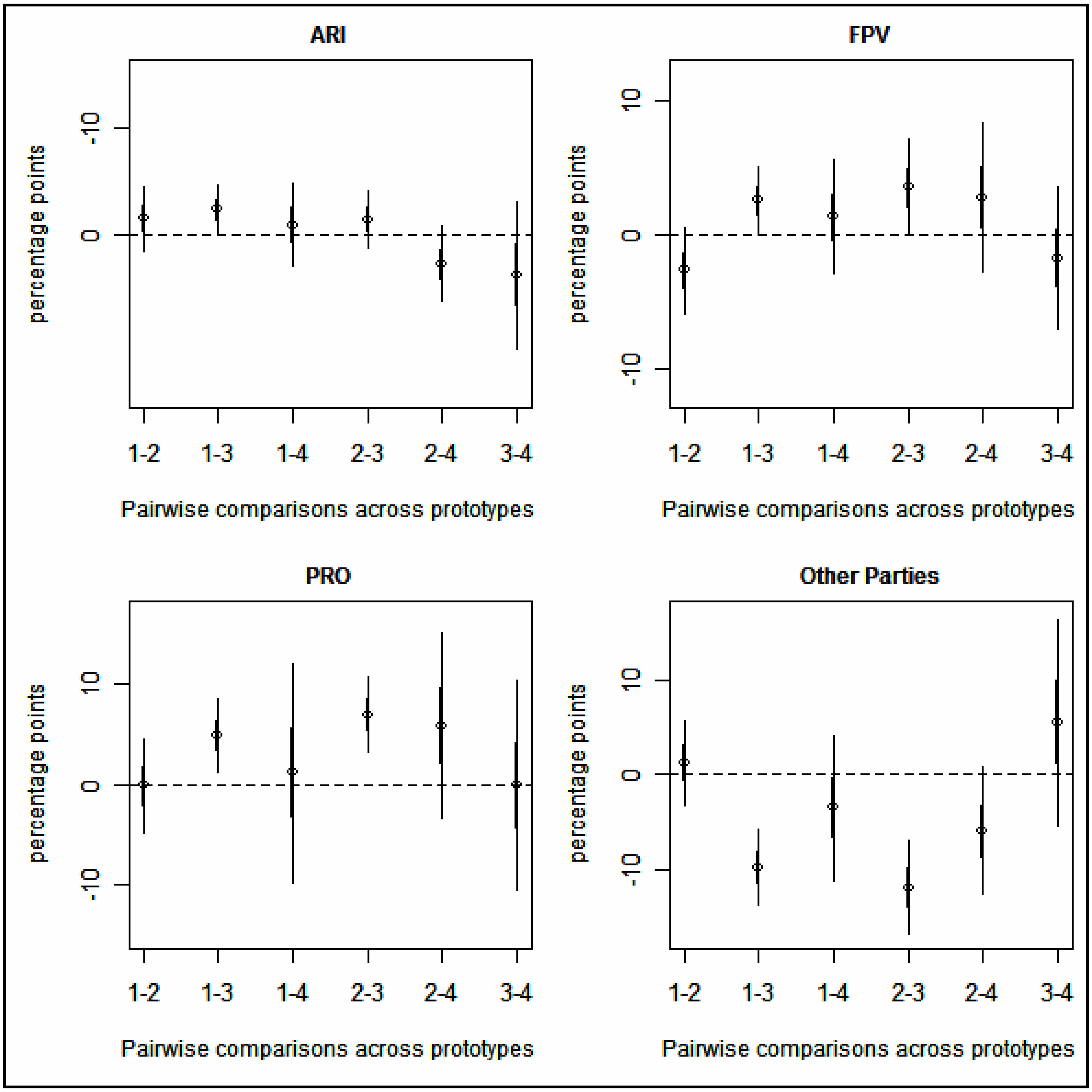
Others things equal, the support for the largest parties tends to be higher for the two DRE devices than for *Prototype 3*, although the differences between *Prototype 1 and 3* are not statistically significant at the usual confidence levels. In contrast, in the cases of FPV and PRO, there are significant differences between their support for *Prototypes 2* and *3*: the touch-screen DRE device leads to an increase of 3.8 and 6.3 percentage points in the their vote-shares, respectively, in the election for national representatives, and of 2.7 and 5.3 percentage points in the election for state legislators; these difference are significant at the 0.01 level. While, as evidenced in Figures 2 and 3, these differences stem both from an increase in the support for FPV and PRO induced by *Prototype 2* in the more visible race and a reduction of their support for *Prototype 3*, the results in the election for state legislators are entirely driven by the higher vote-share of the smaller parties included in the category “Other” under the OS device with separate ballots. In fact, the relative support for the smaller parties tends to be consistently higher with *Prototype 3* in both races: in the national representative election, the vote-share of the minor parties is 11 percentage points higher under *Prototype 3 vis a vis Prototype 2,* while in the state legislature election their vote-share with this prototype is systematically higher when compared against all the other voting devices. Also, note that in the national representative election, the relative support for the smaller parties is lower with *Prototype 2* than *Prototype 1*. Hence, in the more visible race, the touch-screen DRE device consistently favors the parties with more renowned candidates, to the detriment of the smaller ones. This result is consistent with Calvo, Escolar and Pomares (2007), in the sense that voters’ electoral choices are affected by the cues and the information demands from the different e-vote technologies.

In order to assess the robustness of these results, Figures 4 and 5 plot the average treatment effects on the vote-share of each of the parties and the 50% and 90% confidence intervals obtained from pairwise comparisons across prototypes using genetic matching (Diamond and Sekhon, 2005; Sekhon, 2007). While the point estimates obtained from this non-parametric analysis do not necessarily coincide with those reported in Table 6, the general conclusions are in line with those drawn from the two regression models: in both elections, the vote-share of the smaller parties increases under *Prototype 3* when compared against the two DRE devices, and the average effect is larger in the more visible race. The opposite occurs for the three more established parties, those with more campaign spending and higher name recognition; again, *Prototype 2* tends to increase the vote-share of the two parties with more visible candidates, FPV and PRO, especially *vis a vis Prototype* 3 in the national representative election.

Figure 4

Average causal effect on each party's vote-share across prototypes

Election of National Representatives

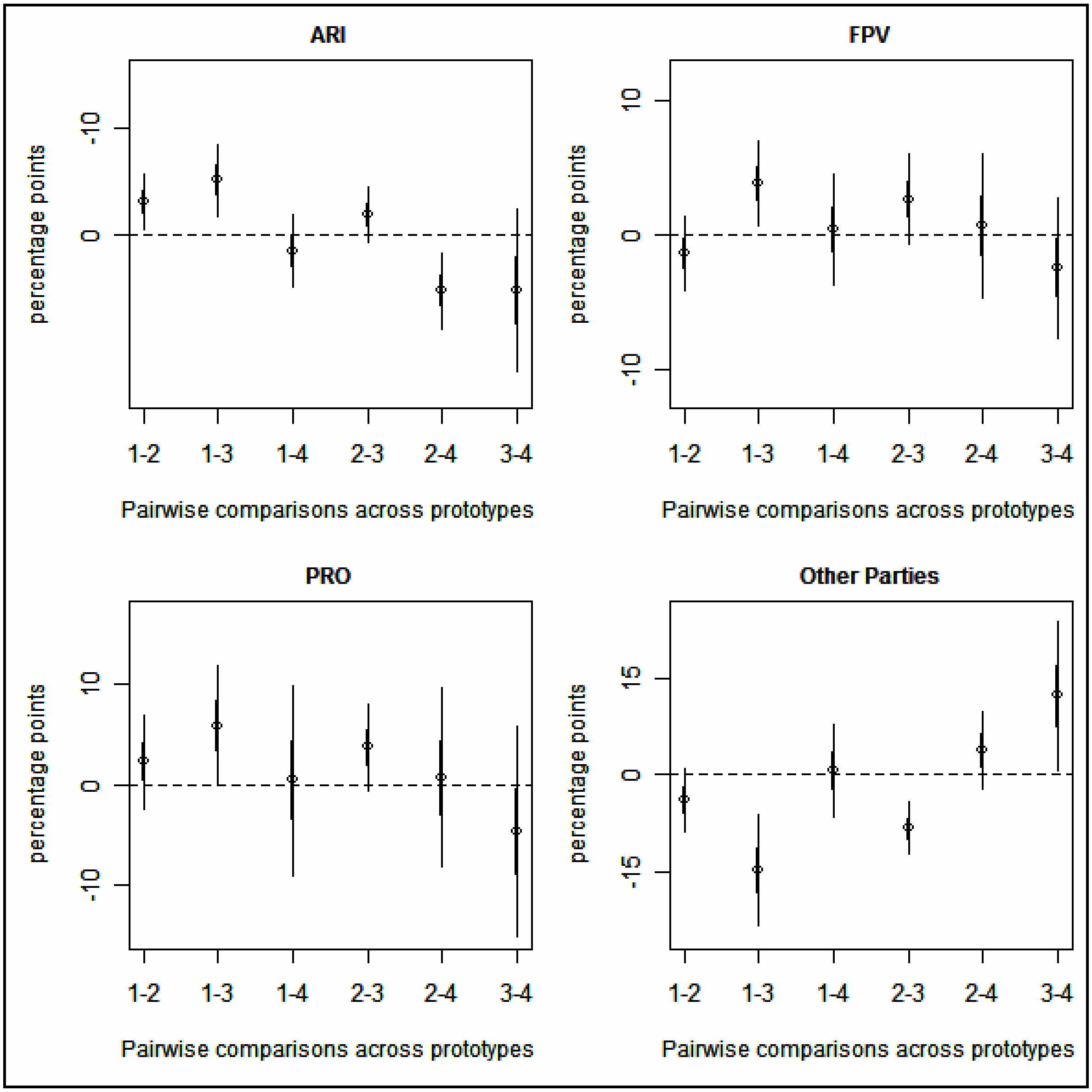


Note: The center dots correspond to the estimated average causal effect on the probability of supporting each of the parties comparing prototypes r and s, $r, s = 1,..,4,\ \ r < s$. The thinner lines correspond to the 90% confidence intervals, and the thicker lines to the 50% confidence intervals.

Figure 5

Average causal effect on each party's vote-share across prototypes

Election of State Legislators



Note: The center dots correspond to the estimated average causal effect on the probability of supporting each of the parties comparing prototypes r and s, $r, s = 1,..,4$, $r < s$. The thinner lines correspond to the 90% confidence intervals, and the thicker lines to the 50% confidence intervals.

The results presented above provide strong evidence in support of the hypothesis that alternative voting technologies may have substantive influence on the electoral support for the different parties competing in future elections using the types of voting technologies tested in the Buenos Aires pilot. The relevant question thus becomes: how would the election outcomes vary under different voting technologies? In order to answer this question, we estimate the expected electoral outcome assuming only one prototype had been used in each voting-station, while holding all the remaining variables constant. Based on the posterior Gibbs samples from the hierarchical multiple-logit model, Table 7 reports the aggregate election outcomes in both races for each of the four prototypes and compares them to the actual results in all the voting-stations used in the experiment.

The results indicate that different voting technologies would in fact have led to quite different election outcomes. For instance, if *Prototype 1* had been used in all voting-stations, *Alianza para una Republica de Iguales (ARI)* would have had the highest expected number of votes in the race for the election for national representatives, rather than the actual winner, *Propuesta Republicana (*PRO). ARI would have had the highest expected vote-share in the election for state legislators under *Prototype 3*. In contrast, the vote-shares of PRO and FPV in the national election would have been maximized under *Prototype 2*, increasing their support at the expense of ARI and, especially, of the smallest parties; in the less visible election, however, in which voters could generally not recognize the candidates' names, the advantage enjoyed by PRO and FPV under the touch-screen DRE device would have virtually vanished. Finally, the expected vote-share of the minor parties in both races would have increased substantially under *Prototype 3*, obtaining almost 46% of the support in the national representative election and 55.0% in the state legislative contest, against 40.2% and

49.2% in the actual experiment, respectively. Hence, the choice among different voting technologies could have had substantive implications in terms of the election results in both congressional races.

Table 7

Expected versus actual election outcomes for all the voting-stations

Vote-shares, in percentage points

	ARI	**FPV**	**PRO**	**Other Parties**
Election of N. Representatives				
Prototype 1	22.77	14.52	21.59	41.12
Prototype 2	20.64	18.13	26.74	34.49
Prototype 3	19.36	14.33	20.40	45.91
Prototype 4	19.89	15.52	23.99	40.60
Actual results	**21.03**	**15.58**	**23.16**	**40.24**
Election of S. Legislators				
Prototype 1	18.00	12.97	21.87	47.16
Prototype 2	18.57	13.52	20.38	47.53
Prototype 3	18.16	10.25	16.59	55.00
Prototype 4	15.47	14.64	25.84	44.05
Actual results	**18.04**	**12.31**	**20.43**	**49.22**

7. Conclusion

Multi-party races impose substantial demands on voters, who have to gather enough information to be able to distinguish between the positions of the different parties before the elections and to identify their preferred choice at the polls. In this paper, we present the first study on the potential impact of different voting technologies on election outcomes in multi-party races, analyzing data from a large-scale pilot in Buenos Aires using a combination of parametric and non-parametric methods.

Our findings indicate that different voting devices could have considerable influence on the relative support for different parties across races, even after controlling for relevant socio-demographic and behavioral predictors. In particular, we show that amount and the form in which information is presented to voters might influence their propensity to choose some parties over other, and this effect may be large enough to actually affect the election results. These substantive results are similar for the different empirical methods used in the analysis, strengthening the validity of our conclusions and marking an important difference with relevant studies on this topic examining U.S. elections, most of which have found that of impact of alternative voting technologies on election outcomes are quite small (Card and Moretti, 2007; Herron and Wand, 2007; Herron, Mebane and Wand, 2008). In this sense, our results are in line with the findings of Reynolds and Steenbergen (2006), who concluded that some aspects of the ballot design, such as symbols, photographs, layout, and color, play a crucial role as political cues and may have a considerable influence on voting behavior, particularly in multi-party elections.

The evidence presented in this paper is particularly significant in view of the increasing trend towards electronic voting and the growing number of countries moving from traditional paper ballots to electronic voting systems (Milkota, 2002, E-Voting.CC, 2008). In many of these countries, political parties have repeatedly expressed concerns about the possibility of being systematically disadvantaged by the new voting technologies.[18] Our results suggest that this might actually be the case, rather than just a myth fuelled by politicians, and raise the possibility that some voting technologies may in fact shape the electoral outcomes, rather than merely recording voters' preferred choices. In addition, the evidence presented here underscores the importance of comparing not only electronic voting *vis a vis* hand-counted ballots, but also different types of voting technologies. While this might be quite apparent in the U.S., given the wide variety of voting systems used, it is definitely not the case in other regions (Europe, Latin America), where the debate in many countries switching to electronic voting systems has focused on the differences between the traditional paper ballots and the "voting machines".

[18] For instance, several French political parties expressed such concerns during the 2007 Presidential election, the first time electronic voting machines were used for a presidential election in the country (*Le Figaro*, 04/18/2007).

References

Abadie, Alberto. 2002. "Bootstrap Test for Distributional Treatment Effect in Instrumental Variable Models". *Journal of the American Statistical Association*, 97(457), 284-292.

Abbadie, Alberto, and Guido Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects". *Econometrica*, 74, 235-267.

Agresti, Alan. 2002. *Categorical Data Analysis*. New Jersey: John Wiley and Sons.

Agresti, Alan, and David B. Hitchcock. 2005. "Bayesian inference for categorical data analysis." *Statistical Methods and Applications*, 14(3), 297-330.

Alvarez, R. Michael, Stephen Ansolabehre, Erik Antonsson, Jehoshua Bruck, Stephen Graves, Nicolas Negroponte, Thomas Palfrey, Ron Rivest, and Charles Stewart. 2001. "Residual Votes Attributable to Technology: An Assessment of Reliability of Existing Voting Equipment". Caltech/MIT Voting Technology Project, http://www.hss.caltech.edu/~voting/CalTech MIT Report Version2.pdf.

Alvarez, R. Michael and Thad E. Hall. 2004. *Point, Click and Vote: The Future of Internet Voting.* Washington, D.C.: The Brookings Institution Press.

Alvarez, R. Michael and Thad E. Hall. 2005. "Lessons and Trends in e-voting: Initiatives in the US and abroad." Caltech/MIT Voting Technology Project, Working Paper 38, http://www.vote.caltech.edu/media/documents/wps/vtp_wp38.pdf.

Alvarez, R. Michael and Thad E. Hall. 2008. *Electronic Elections: The Perils and Promises of Digital Democracy*. Princeton, NJ: Princeton University Press.

Alvarez, R. Michael. 2005. "Qualitative Evaluation: 'VotoElectronico: Prueba Piloto 2005, Ciudad De Buenos Aires'". Working Paper, Caltech/MIT Voting Technology Project, http://www.vote.caltech.edu/media/documents/wps/vtp_wp43.pdf.

Ansolabehre, Stephen, and Charles Stewart III. 2005. "Residual Votes Attributable to Technology". Journal of Politics, 67(2), 365-389.

Barnard, John, Constantine E. Frangkasis, Jennifer L. Hill and Donald B. Rubin. 2003. "Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City". *Journal of the American Statistical Association*, 98(462), 299-323.

Calvo, Ernesto, and Maria Victoria Murillo. 2004. "Who Delivers? Partisan Clients in the Argentine Electoral Market.". *American Journal of Political Science*, 48(4), 742-757.

Calvo, Ernesto, Marcelo Escolar, and Julia S. Pomares. 2007. "Ballot Design and Split Ticket Voting in Multiparty Systems: experimental evidence on information effects and vote choice." Unpublished manuscript.

Cameron, A. Colin, and Pravin K. Trivedi. 1998. *Regression analysis of Count Data*. New York: Cambridge University Press.

Card, David, and Enricco Moretti. 2007. "Does Voting Technology Affect Election Outcomes? Touch-screen Voting and the 2004 Presidential Election". *Review of Economics and Statistics,* 89 (4), 660-673.

Casella, George, and Edward L. George. 1992. "Explaining the Gibbs Sampler". *The American Statistician*, 46(3), 167-174.

Congdon, Peter. 2005. *Bayesian Models for Categorical Data*. New York: New York: John Wiley & Sons.

Conover, Pamela J., and Stanley Feldman. 1989. “Candidate Perception in an Ambiguous World: Campaigns, Cues, and Inference Processes.” *American Journal of Political Science*, 33(4), 912-940.

Diamond, Alexis, and Jasjeet S. Sekhon. 2005. “Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies”. Mimeo.

E-Voting.CC. The E-Voting Database. On-line resource, http://db.e-voting.cc/ (accessed February 2008).

Fahrmeier, Ludwig, and Leonhard Knorr-Held. 2000. “Dynamic and semiparametric models.” In *Smoothing and Regression: Approaches, Computation and Application*, M. Schimek (ed.), 513-544. New York: John Wiley & Sons.

Frisina, Laurina, Michael C. Herron, James Honaker, and Jeffrey B. Lewis. 2008. “Ballot Formats, Touchscreens and Undervotes: A Study of the 2006 Midterm Elections in Florida”. *Election Law Journal* ,7(1), 25-47.

Fry, Jane M., Tim R. Fry, and Keith R. McLaren. 1996. “Compositional Data Analysis and Zeros in Micro Data”. General Paper No. G-120, Centre of Policy Studies, Monash University.

Gelfland, Alan E., and Adrian F. Smith. 1990. “Sampling-Based Approaches to Calculating Marginal Densities”. *Journal of the American Statistical Association*, 85(410), 398 – 409.

Gelman, Andrew, and Jennifer L. Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Herron, Michael C., and Jonathan Wand. 2007. “Assessing artisan bias in voting technology: The case of the 2004 New Hampshire recount”. *Electoral Studies*, 26, 247-261.

Herron, Michael C., Walter R. Mebane and Jonathan Wand. 2008. “Voting Technology and the 2008 New Hampshire Primary”. Working paper.

Herrnson, Paul S., Richard G. Niemi, Michael J. Hanmer, and Michael Traugott. 2006. “The Not so Simple Act of Voting: An Examination of Voter Errors with Electronic Voting." Paper presented at the annual meeting of the The Midwest Political Science Association, Palmer House Hilton, Chicago, Illinois, April, 2006.

Herrnson, Paul S., Richard G. Niemi, Michael J. Hanmer, Benjamin B. Bederson, and Frederick C. Conrad. 2008. *Voting Technology: The Not-So-Simple Act of Casting a Ballot*. Brookings Institution Press.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis*, 15, 199-236.

Imai, Kosuke. 2005. “Do Get-Out-The-Vote Calls Reduce Turnout? The importance of Statistical Methods for Field Experiments”. *American Political Science Review*, 99(2), 283-300.

Imai, Kosuke, and David A. van Dyk. 2004. “Causal Inference With General Treatment Regimes: Generalizing the Propensity Score.” Journal of the American Statistical Association, 99 (467), 854 – 866.

Iyengar, Malini and Dipak K. Dey. 2004. “Bayesian analysis of compositional data”. In *Generalized linear Models: A Bayesian Perspective*, Dipak K. Dey, Sujit K. Ghosh and Bani K. Mallick (eds.), 349-364. New York: Marcel Dekker.

Jackman, Simon. 2004. “Bayesian Analysis for Political Research”. *Annual Review of Political Science*, 7, 483-505.

Jackson, John E. 2002. “A Seemingly Unrelated Regression Model for Analyzing Multiparty elections.” *Political Analysis*, 10(10, 49-65.

Katz, Jonathan, and Gary King. 1999. “A Statistical Model for Multiparty Electoral Data.” *American Political Science Review*, 93, 15 – 32.

Knack, Stephen, and Martha Kropf. 2002. “Who uses inferior voting technology?” *Political Science and Politics*, 35(3), 541-548.

Leonard, Tom, and John S. Hsu. 1994. “The Bayesian analysis of categorical data. A selective review”. *In Aspects of Uncertainty: A Tribute to DV Lindley*, P. Freeman and A. Smith, (eds.). New York: John Wiley and Sons.

Loughin, Thomas M. and Peter N. Scherer. 1998. “Testing for Association in Contingency Tables with Multiple Column Responses”. *Biometrics*, 54(2), 630-637.

McCullagh, Peter, and John A. Nelder. 1989. *Generalized Linear Models*. New York: Chapman & Hall.

Mebane, Walter R., and Jasjeet S. Sekhon. 1998. “Genetic Optimization Using Derivatives: The rgenoud package for R”. *Journal of Statistical Software*. http://sekhon.berkeley.edu/papers/rgenoudJSS.pdf.

Mebane, Walter R., and Jasjeet S. Sekhon. 2004. “Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data”. *American Journal of Political Science*, 48(2), 391-410.

Mebane, Walter R., and Jasjeet S. Sekhon. 2008. The Matching Package. R library. http://cran.r-project.org/web/packages/Matching/Matching.pdf.

Milkota, Kristina. 2002. “E-democracy conference : a summary of findings”. EVE conference "E-democracy: scenarios for 2010”, Paris, France 15-16 Oct 20022002.

Ministry of Interior of Argentina. National Elections, 2005. On-line resource, http://www.mininterior.gov.ar/elecciones/2005/inicio.asp (accessed January 2008).

Mora y Araujo, Manuel, and Ignacio Llorente. 1980. *El voto peronista: ensayos sobre la sociología electoral argentina*. Buenos Aires: Editorial Sudamericana.

Newcombe Robert G. 1998. “Interval Estimation for the Difference Between Independent Proportions: Comparison of Eleven Methods.” *Statistics in Medicine* 17, 873–890.

Reynolds, Andrew, and Marco Steenbergen. 2006. “How the world votes: the political consequences of ballot design, innovation and manipulation.” *Electoral Studies* 25(3): 570-598.

Saltman, Roy G. 2006. *The History and Politics of Voting Technology*. New York: Palgrave MacMillan.

Sekhon, Jasjeet S. 2007. “Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R.” *Journal of Statistical Software.* http://sekhon.berkeley.edu/papers/MatchingJSS.pdf.

Sinclair, Betsy and R. Michael Alvarez. 2004. “Who Overvotes, Who Undervotes, Using Punchcards? Evidence from Los Angeles County.” *Political Research Quarterly*, 57(1), 15-25.

Sinclair, Robert C., Melvin M. Mark, Sean E. Moore, Carrie A. Lavis, and Alexander S. Soldat. 2000. “*An electoral butterfly effect*”. Nature, 408, 665-666.

Stewart, Charles. 2004. “The Reliability of Electronic Voting Machines in Georgia”. Caltech/MIT Voting Technology Project, VTP Working Paper 20, http://www.vote.caltech.edu/media/documents/wps/vtp_wp20.pdf

Tomz, Michael, Joshua A. tucker and Jason Wittenberg. 2002. “An Easy and Accurate Regression Model for Multiparty Electoral Data”. Political Analysis 1(1), 66-83.

Tomz, Michael, and Robert P. Van Houweling. 2003. “How Does Voting Equipment Affect the Racial Gap in Voided Ballots?” *American Journal of Political Science*, 47, 46-60.

Wand, Jonathan N. 2004. Evaluating Voting Technologies: 2004 New Hampshire Democratic Primary. Technical Report, Stanford University,

http://wand.stanford.edu/elections/us/NH/nh2004primary.pdf

Woolf, B. 1955. “On estimating the relation between blood group and disease.” *Ann. Human Genet*. 19, 251-253.

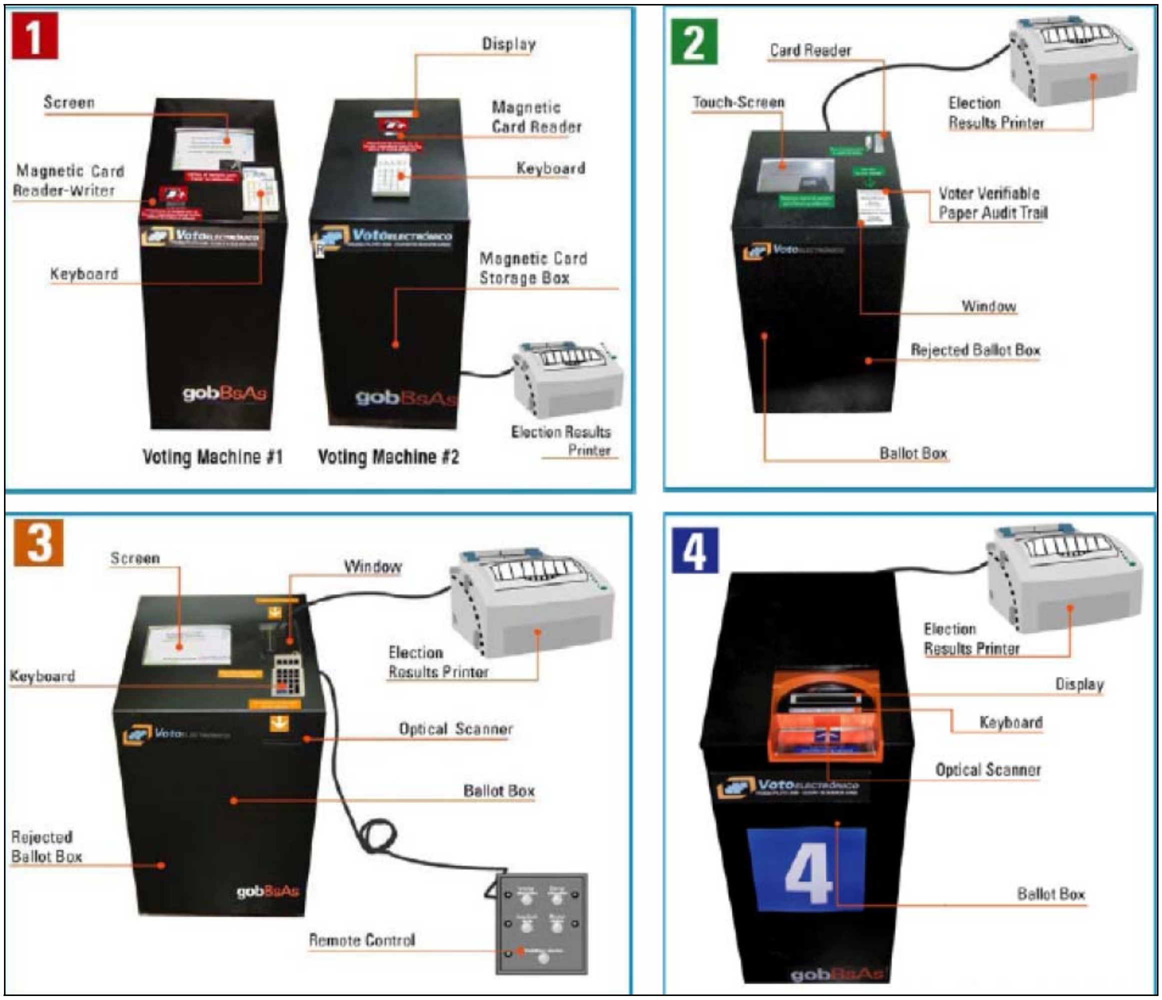
Supplementary Materials - Assessing the impact of voting technologies on multi-party electoral outcomes: the case of Buenos Aires' 2005 Congressional Election

Appendix I - Characteristics of the four e-voting prototypes

Table S. 1

	How to cast a vote	Information on the candidates
Prototype 1: DRE, Keypad	The voter inserts the smart card to initiate the voting process. Using a numeric keypad, she over party labels and selects a party list for each race.	Party name, number and logo displayed on the screen. Entering the party number displays the list of candidates. Labels for National Representatives and State Legislators displayed in two and three screens, respectively.
Prototype 2: DRE, Touch-screen	The voter inserts the smart card to initiate the voting process. Using a touch-screen device, she scrolls over party labels and selects a party list for each race	Party name, number, logo and first candidate of the list displayed on the screen. Selecting the party label displays the remaining candidates. Labels for National Representatives and State Legislators displayed in three and four screens, respectively.
Prototype 3: OS, Two ballots	The voter is asked to insert a paper ballot or press the CONTINUE button to cast a blank vote. Separate paper ballots for each party and for each race.	All information displayed on the ballot: party name, number, logo, and complete list of candidates.
Prototype 4 OS, Single ballot	T he voter browses the party lists in a booklet and marks her preferences on the paper ballot. An optical device scans the ballot. Order of lists based on Party number.	Only the party name and number are displayed on the ballot. Complete information on the party and candidates is in a separate booklet.

Figure S.1 – The four e-voting devices



Appendix II – Summary descriptive statistics for the independent variables included in the hierarchical models

Table S. 2

Mean value of the regressors, by prototype and for the whole sample

Variable	Prototype 1	Prototype 2	Prototype 3	Prototype 4	Whole sample	Range
Education	6.7	6.7	6.6	7.0	6.7	
Political information	1.96	1.90	1.91	1.94	1.92	0 - 3
Interest in politics	2.12	2.12	2.15	2.04	2.12	1 - 3
Assessment of E-voting	4.15	4.19	4.05	4.07	4.13	1 – 5
Cell-phone owners	75.31	73.14	73.54	74.07	73.93	0 - 100
Personal-computer owners	79.49	76.10	78.25	82.95	78.45	0 - 100
Internet users	80.32	79.97	78.55	82.54	79.83	0 - 100
Use of Technology	0.02	-0.02	-0.00	0.06	0.00	
Search by Party Name	51.4	51.0	44.3	53.4	49.4	0 - 100
Search by Party Logo	51.4	51.0	44.3	53.4	49.4	0 - 100
Search by Party Number	27.3	30.3	22.4	7.4	25.8	0 - 100
Search by Candidate Name	33.3	50.1	47.1	45.0	44.2	0 - 100

Appendix III - Estimation of the hierarchical models used in the analysis

Adopting non-informative conjugate priors for the fixed-effects and precision matrices:

$$\beta_l^j \sim N\left(\delta_\beta, \sigma_\beta^2\right),\ j = ARI, FPV, PRO,\ l = 1, \dots L \tag{S.1}$$

$$\begin{aligned} &\Sigma_\varepsilon^{-1} \sim Wishart(E, \rho_E), \quad \left|E\right| > 0, \rho_E \geq 3 \\ &\Sigma_\lambda^{-1} \sim Wishart(P, \rho_P), \quad \left|P\right| > 0, \rho_P \geq 3 \\ &\Sigma_\eta^{-1} \sim Wishart(Q, \rho_Q), \quad \left|Q\right| > 0, \rho_Q \geq 3 \end{aligned} \tag{S.2}$$

and assuming conditional independence throughout, the joint posterior densities of the unknown parameters in the multiple-logit and additive logistic models presented in Section 5 are given by:

$$\begin{aligned} & f\left(\beta, \Sigma_u, \Sigma_v, \upsilon \middle| V\right) \propto \\ & \prod_{i=1}^{n} \prod_{j} \left[\frac{\exp\left(X_i \beta^j - \lambda_p^j - \eta_h^j\right)}{\sum_k \exp\left(X_i \beta^k - \lambda_p^k - \eta_h^k\right)} \right]^{V_i^j} \times \sigma_\beta^{-L/2} \exp\left\{ -\frac{1}{2} \sigma_\beta^{-2} \left(\beta - \delta_\beta I_L\right)' \left(\beta - \delta_\beta I_L\right) \right\} \\ & \times \left|\Sigma_\lambda\right|^{-p/2} \exp\left\{ -\frac{1}{2} \sum_{p=1}^{4} \lambda_p ' \Sigma_\lambda^{-1} \lambda_p \right\} \times \left|\Sigma_\lambda^{-1}\right|^{\frac{\rho_P - 2 - 1}{2}} \exp\left\{ -\frac{1}{2} tr\left(P \Sigma_\lambda^{-1}\right) \right\} \\ & \times \left|\Sigma_\eta\right|^{-H/2} \exp\left\{ -\frac{1}{2} \sum_{h=1}^{H} \eta_h ' \Sigma_\eta^{-1} \eta_h \right\} \times \left|\Sigma_\eta^{-1}\right|^{\frac{\rho_Q - 2 - 1}{2}} \exp\left\{ -\frac{1}{2} tr\left(Q \Sigma_\eta^{-1}\right) \right\} \end{aligned} \tag{S.3}$$

with $\beta^{OTHER}, \lambda_p^{OTHER}, \eta_h^{OTHER} = \vec{0}$, and

$$f\left(\beta,\Sigma_e,\Sigma_u,\Sigma_v,\upsilon\middle|Y\right)\propto$$

$$\prod_{i=1}^{n}\frac{\Gamma\left(\frac{(\upsilon+3)}{2}\right)}{\upsilon^{3/2}\Gamma(\upsilon/2)\left|\Sigma_\eta\right|^{1/2}}\left[1+\frac{\left(Y_i-X_i\beta-\lambda_p-\eta_h\right)^T\Sigma_\varepsilon^{-1}\left(Y_i-X_i\beta-\lambda_p-\eta_h\right)}{\upsilon}\right]^{-(3+\upsilon)/2}$$

$$\times\sigma_\beta^{-L/2}\exp\left\{-\frac{1}{2}\sigma_\beta^{-2}\left(\beta-\delta_\beta I_L\right)^T\left(\beta-\delta_\beta I_L\right)\right\}\times\left|\Sigma_\lambda\right|^{-p/2}\exp\left\{-\frac{1}{2}\sum_{p=1}^{4}\lambda_p{}'\Sigma_\lambda^{-1}\lambda_p\right\}\times p(\upsilon) \qquad \text{(S.4).}$$

$$\times\left|\Sigma_\lambda^{-1}\right|^{\frac{\rho_P-2-1}{2}}\exp\left\{-\frac{1}{2}tr\left(P\Sigma_\lambda^{-1}\right)\right\}\times\left|\Sigma_\eta\right|^{-H/2}\exp\left\{-\frac{1}{2}\sum_{h=1}^{H}\eta_h{}'\Sigma_\eta^{-1}\eta_h\right\}$$

$$\times\left|\Sigma_\eta^{-1}\right|^{\frac{\rho_Q-2-1}{2}}\exp\left\{-\frac{1}{2}tr\left(Q\Sigma_\eta^{-1}\right)\right\}\times\left|\Sigma_e^{-1}\right|^{\frac{\rho_E-2-1}{2}}\exp\left\{-\frac{1}{2}tr\left(\mathrm{E}\Sigma_\varepsilon^{-1}\right)\right\}$$

Inference on the parameters of interest can be performed by Markov chain Monte Carlo (MCMC) simulations using Gibbs sampling to repeatedly draw samples from each unknown parameter’s full conditional posterior distribution (Gelfland and Smith, 1990; Casella and George, 1992).[1] Under mild regularity conditions (Geman and Geman, 1984; Gilks, Richardson and Spiegelhalter, 1996), for a sufficiently large number of iterations, samples from these complete conditionals approach samples from the marginals used for Bayesian inference. The means and standard deviation of the convergent samples can used to summarize the posterior distributions of the model's coefficients and to compute the marginal effects of the different e-voting devices on each party’s vote-share (using the corresponding transformations in the case of the additive logistic model; Katz and King, 1999; Bhaumik, Dey and Ravishanker, 2003).

---

[15] In the case of (S.4), while the conditional posterior densities have no closed forms, draws of the unknown parameters can be obtained using Adaptative Rejection Sampling (Gilks and Wild, 1992).

The two models were fit in WinBUGS 1.4, as called from R 2.4.1.[2] For both specifications, the fixed effects β_k^j, were assigned $N(0,100^2)$ priors, while a $Wishart\left(I_3,4\right)$ prior was used for the precision matrices of the random effects Σ_u^{-1} and Σ_v^{-1}. In the case of the additive logistic Student-t model, the precision matrix of the voting-station errors was assigned a $Wishart\left(I_3,4\right)$ prior, while a $Uniform\left(2,100\right)$ distribution was used for the degrees of freedom parameter υ. Routine sensitivity analyses were performed in order to examine the effect of the priors on the model fit, and alternative values for the hypeparameters were tried, yielding similar results.

Three parallel chains with dispersed initial values reached approximate convergence after 85,000 iterations, with a burn-in of 3,500 iterations; the results reported in Section 5 are based on 1,000 samples of the pooled chains of deviates.[3] The means and standard deviation of the convergent Gibbs samples were used to summarize the posterior distributions of the model's coefficients and to compute the effects of the different e-voting technologies on each party's support.

[2] The code is available from the authors upon request.

[3] Approximate convergence is achieved for values of Gelman and Rubin's (1992) estimated Potential Scale Reduction factor below 1.1.

Appendix IV - Balance Tests Before and After Matching

As mentioned in the text, we also performed a series of matching exercises using Genetic Matching (Diamond and Sekhon; 2005; Sekhon 2007a), where each exercise considered the average causal effect on each party's vote-share of using *Prototype* r versus *Prototype* s, with $r, s = 1,..,4,\ \mathrm{r} < s,$ and voting stations using *Prototype* r taken to be the "treated" group. The criterion for determining whether sets of voting stations have been matched is based on whether the means of observable pre-treatment variables are indistinguishable between the two groups being tested. Table S.3 presents the set of matching covariates and reports the p-values of two-sample and paired t-tests of the hypothesis of no difference between the means of these variables before and after matching. Including additional socio-demographic covariates or higher-order terms yields essentially identical results.

As seen in the Table, for most of these variables, the difference in means tests were statistically insignificant at the usual confidence levels before matching. Nonetheless, Genetic Matching minimizes the maximum discrepancy between the matched treated and control covariates. In the few cases in the two-sample t-tests were significant, the pairs of matched subsamples are indistinguishable in their means. It is worth mentioning, however, that large values for these p-values do not guarantee that the level of level of balance achieved is sufficient for reliable inferences about the effects of the prototypes on parties' vote-shares (Diamond and Sekhon, 2005; Sekhon, 2007b).

Table S. 3: p-values of the two-sample and paired t-tests

Variable	Two-sample t-tests before matching						Paired t-tests After matching					
	1-2	1-3	1-4	2-3	2-4	3-4	1-2	1-3	1-4	2-3	2-4	3-4
Education	0.86	0.90	0.12	0.94	0.17	0.14	0.60	0.93	0.20	0.83	0.68	0.19
$Education^2$	0.83	0.91	0.22	0.91	0.16	0.14	0.65	0.76	0.30	0.96	0.64	0.17
Political information	0.87	0.68	0.92	0.52	0.84	0.87	0.68	0.82	0.40	0.98	0.96	0.99
Interest in politics	0.48	0.60	0.18	0.34	0.15	0.10	0.27	0.19	0.19	0.33	0.12	0.12
Use of Technology	0.29	0.83	0.38	0.33	0.08	0.27	0.57	0.93	0.82	0.85	0.39	0.97
Education * Political information	0.97	0.65	0.64	0.39	0.62	0.86	0.64	0.73	0.20	0.98	0.74	0.89
Education * Use of Technology	0.51	0.94	0.79	0.53	0.07	0.15	0.68	0.82	0.78	0.99	0.57	0.15
Number of observations												
Treated	38	38	38	38	38	36	38	38	38	38	38	36
Control	38	36	13	36	13	13	38	38	38	38	38	36

References

Bhaumik, Amitabha, Dipak K. Dey, and Nalini Ravishanker. 2003. "A dynamic Linear Model Approach for Compositional Time Series Analysis". Technical Report, University of Connecticut.

Diamond, Alexis, and Jasjeet S. Sekhon. 2005. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies". Mimeo.

Gelfland, Alan E., and Adrian F. Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities". *Journal of the American Statistical Association*, 85(410), 398 – 409.

Gelman, Andrew, and Donald B. Rubin. 1992. "Inference for iterative simulation using multiple sequences". *Statistical Science*, 7, 457-472.

Geman, S., and D. Geman. 1984. "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Gilks, Walter. R., and P. Wild. 1992. "Adaptive rejection sampling for Gibbs sampling." *Applied Statistics*, 41, 337-348.

Gilks, Walter R., Sylvia Richardson, and David J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Katz, Jonathan, and Gary King. 1999. "A Statistical Model for Multiparty Electoral Data." *American Political Science Review*, 93, 15 – 32.

Sekhon, Jasjeet S. 2007a. “Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R.” *Journal of Statistical Software.* http://sekhon.berkeley.edu/papers/MatchingJSS.pdf.

Sekhon, Jasjeet S. 2007b. “Alternative Balance Metrics for Bias Reduction in Matching Methods for Causal Inference”. Working Paper, http://sekhon.berkeley.edu/papers/SekhonBalanceMetrics.pdf