Recruiting for Laboratory Voting Experiments: Exploring the (Potential) Sampling Bias

Damien Bol, Simon Labbé St-Vincent, and Jean-Michel Lavoie

1 Introduction

Laboratory experiments are becoming more and more popular in behavioral sciences in general and in research on voting in particular (Druckman et al. 2006). They offer multiple advantages to researchers interested in this topic, including the possibility of reproducing a group dynamic similar to a real-life election or repeating an electoral process several times to investigate voters' learning curve. At the same time, because laboratory experiments allow the researcher to control for a wide range of conditions, they constitute strong instruments to study the (bounded) rationality of electoral behaviors.

However, behavioral laboratory experiments are often criticized for their lack of external validity (Lupia 2002; McDermott 2002). The critiques are particularly strong when it comes to evaluating the tendency of experimenters to rely on student samples instead of proportionality samples representative of the population under study (Druckman and Kam 2011).

In this chapter, we give new insights to this issue by comparing the results of four voting experiments replicated in several laboratories in Europe and Canada. The laboratories use different methods of recruitment, and as a consequence, rely on very different samples. Some of them are almost exclusively composed of students; others also contain highly educated professionals (in the text, we talk about 'heterogeneous' samples to denote this type of samples). In the next section, we discuss the issue of sampling bias in behavioral laboratory experiments; second, we describe the protocol

D. Bol (✉)
Department of Political Economy, King's College London, Strand Campus, Strand Building, WC2R 2LS London, UK
e-mail: damien.bol@kcl.ac.uk

S. Labbé St-Vincent • J.-M. Lavoie
University of Montreal, Montreal, QC, Canada

A. Blais et al. (eds.), *Voting Experiments*, DOI 10.1007/978-3-319-40573-5_14

of the four experiments analyzed here; third, we examine the four laboratories' samples on which these experiments were conducted; and finally, we examine, by sample type, the voting patterns of the subjects engaged in our experiments.

We find very few differences regarding the behavior of subjects in our student-dominated and heterogeneous samples (at the notable exception that free-riding is more common in student-dominated than in heterogeneous samples). We do not claim that our analyses offer a definitive answer to the question of whether there is bias arising from the laboratories' recruitment techniques. We claim that the debate should be somehow reconsidered. Instead of focusing on the differences between students and non-students, we recommend researchers to consider the potential behavioral differences in terms of reasoning abilities.

2 Sampling Bias in Behavioral Laboratory Experiments

The issue of external validity of laboratory experiments has been extensively discussed in the behavioral sciences literature (Lupia 2002; McDermott 2002). Opponents of the method usually claim that the behaviors observed in a laboratory are not the same as those observed in reality. One argument that they make is that the situations created in laboratories are artificial, and that as a result subjects do not react to them in the same way as they would in reality. For example, the use of monetary incentives to induce non monetary-motivated preferences such as a particular policy preference in voting experiments is often criticized for not capturing the reality of political attitudes (Villeval 2007). This has led some researchers to question the findings of behavioral laboratory experiments and to ask what they really tell us about real-world behaviors (Campbell and Stanley 1963; Levitt and List 2007).

The proponents of behavioral laboratory experiments usually dismiss this type of external validity critique. They argue that the generalization of experimental findings to real-world social interactions is not the goal of their research *per se*. Rather, they aim first and foremost to test theories and to isolate the impact of a single factor on a particular behavior as well as to study the relationship between the two. In that sense, they value internal more than external validity (Morton and Williams 2008; Plott 1991).

Another strong argument made by opponents of the method is based on the type of subjects that are recruited to participate in experiments. Researchers, largely for practical reasons, often rely on student samples. Some review pieces note that the majority of political science articles (Kam et al. 2007) as well as economics articles (Danielson and Holm 2007) reporting the results of laboratory experiments used student samples. This tendency has met with general skepticism among opponents of laboratory experiments (Benz and Meier 2008; Chou et al. 2009; Sears 1986). Their main argument is that students are better at abstract thinking than average citizens. In their classes, they exercise this type of cognitive ability constantly. They are thus probably more equipped to identify a gain-maximizing behavior from a set of experimental rules and at adopting it.

In opposition to this argument about the existence of a sampling bias in behavioral laboratory experiments is the fact that some researchers have shown that

students do not score significantly differently than non-students on many attitudinal indicators, including comprehension and cognitive skills. For example, Kam (2005) and Kam et al. (2007) argue that students and non-student adopt similar behaviors in a laboratory. Reflecting on this argument, some authors have replicated simple behavioral experiments on students and unemployed/workers (such as the famous 'beauty contest' and various 'information framing' tests), and found no difference between the two types of subjects (Bosch-Domènech et al. 2002; Kühberger 1998).

That being said, a difference does seem to emerge when it comes to more complicated behavioral laboratory experiments. In a recent study, Belot et al. (2010) found that students tend to behave differently than average citizens in experiments where the subjects could anticipate the behavior of others and adopt a behavior that goes against the collective benefits of the group (for example as a result of having monetary incentives to do so). In these experiments where subjects have 'other-regarding' preferences, students seem to be more likely to free-ride, even after controlling for their level of comprehension of the experiment's rules. In the same vein, other researchers find that, on average, students (and especially economics students) demonstrate less altruistic and generous behaviors than non-students (Carpenter et al. 2005, 2008; Engel 2011; Frank et al. 1993). These findings call into question the validity of behavioral laboratory experiments based on student samples since they make clear that the results might be a poor reflection of the reality of social interactions.

To give new insights to this issue, we analyze the data generated by four voting laboratory experiments. Voting experiments are group experiments where subjects play the role of voters in several consecutive elections. Each individual subject is asked to take a decision concerning their vote by anticipating the voting behavior of others. The goal is to examine to what extent the subjects manage to take a decision that maximizes their gain (which sometimes implies free-riding). There are thus typical examples of experiments with other-regarding preferences where we would expect students to free-ride more often and to adopt more rational behaviors than other subjects.

3 Four Voting Laboratory Experiments

Our analyses are based on four voting experiments on turnout and strategic desertion. These experiments were designed within the framework of other projects. Our analyses are essentially re-analyses of already published works with a different focus (i.e. the comparison of results by sample type). For this reason, we sometimes miss some pertinent data we would ideally need to conduct our tests. Also, some details of the payoff structure vary slightly from experiments to experiments.

The four voting experiments were conducted in computer laboratories with relatively large groups of subjects (20 or 21 per session) who played the role of voters. Subjects interacted through repetitive elections, but were not allowed to communicate by other means. All the elections were held under plurality rules and

the winner was the candidate with the highest number of votes.[1] Before each election, the subjects were informed about the structure of preferences of other subjects. After each election, they were presented with the full results. At the end of the experiment, they received a certain amount of money depending on the number of points obtained during the experiment (in addition to a fixed amount of money). The subjects also had to answer a short questionnaire at the end of the session.

3.1 The Distance-to-Winner Experiment

A series of ten elections was conducted on 21 subjects.[2] Subjects were randomly assigned to a different position on a 21-point scale ranging from 0 to 20 (there was one subject per point, with each individual's position randomly changed after each election). Subjects were told that there are candidates located at positions 5 and 15 on the scale. For each election, subjects had to decide to vote for one of these two candidates or to abstain from voting. Voting costs one point. However, regardless of whether they voted or not, the subjects gained a number of points after each election equivalent to 16 minus the distance between themselves and the winning candidate on the 21-point scale.

3.2 The Winner-Takes-All Experiment

A series of six elections was conducted on 20 subjects.[3] Subjects were told that there are two candidates and that they are randomly assigned to one of these two candidates (thus there are ten subjects assigned to each candidate). These assignments were reshuffled after each election. For each election, subjects had to decide to vote for the candidate to whom they were assigned or to abstain from voting. Voting cost two points. However, regardless of whether they voted or not, subjects gained eight extra points if their candidate won the election.

[1] In some experiments, several elections were held under different electoral rules. To facilitate comparison, we only consider elections held under plurality rule.

[2] Subsequent series were organized under different electoral rules. For the reason stated above in footnote 1, we do not analyze these extra series here. For more information, see Blais et al. (2014).

[3] Here again, subsequent series of elections were organized under different rules. However, since these series entailed other treatments designed to test the impact of social norms that exist outside the laboratory such as civic duty on the decision to vote, we are only studying the first series of elections in this chapter. For more information, see Gallego et al. (2016).

3.3 The Unequal-Distribution Experiment

Four series of two elections were conducted on 21 subjects.[4] Subjects are randomly assigned a position on a ten-point scale ranging from 1 to 10 (this position was reshuffled after each series). For the first series, the distribution of subjects on this scale was normal (with a peak of four and five subjects on positions 5 and 6, respectively); for the second series, multimodal (with peaks of three subjects on positions 2, 5 and 6, and of four subjects on position 9); for the third series, bimodal (with peaks of three and four subjects on positions 2 and 9, respectively); and for the fourth series, asymmetric (with peaks of three subjects on positions 5, 6 and 8, and of four subjects on position 7). Subjects were also told that there are four candidates on positions 2, 5, 6 and 9. For each election, subjects had to decide to vote for one of the four candidates or to abstain from voting. Voting cost one point. However, regardless of whether they voted or not, the subjects gained a number of points equivalent to ten minus the distance between themselves and the winning candidate on the ten-point scale after each election.

3.4 The Preference-Ranking Experiment

Eight series of three elections were conducted on 21 subjects.[5] Subjects were told that there are four candidates and that they are randomly assigned a specific order of preference for the candidates: a preferred candidate, a second preferred candidate, a third preferred candidate, and a least preferred candidate (these preferences were randomly reshuffled after a series of three elections). For the first four series, one candidate had one more subject that preferred them than the other candidates (that is, six subjects had that candidate as their first preference against five for the other candidates). For the second four series, the distribution of preferences ranking among subjects was such that one candidate had seven first preferences, a second had six first preferences, and the two others had four first preferences. The candidate assigned to these skewed distributions varied across series. For each election, subjects had to vote for one of the four candidates. If their preferred candidate won the election, they received 30 points; if their second preferred candidate won, they gained 20 points; if their third preferred won, they gained ten points; and if

[4]Subsequent sessions were organized under different electoral rules. For the reason stated above, we do not analyze these sessions here. Also, each series of two elections was preceded by three elections where the subjects had no choice but to vote for one of the candidates. As there was no turnout decision to be made for these three elections, they are disregarded. For more information, see Labbé St-Vincent (2013).

[5]This experiment was replicated with different electoral rules. Here again, we do not analyze these extra experimental sessions. For more information, see Labbé St-Vincent et al. (2013).

their last preferred candidate won, they did not gain anything. Voting did not cost any point in this experiment.

4 Four Different Laboratories

We conducted several sessions of the four voting experiments in different behavioral experimental laboratories in Canada and Europe: the Cirano in Montreal (www.cirano.qc.ca), the Leep in Paris (http://leep.univ-paris1.fr/accueil.htm), the Leex in Barcelona (www.upf.edu/leex/), and the Cevipol in Brussels (http://dev.ulb.ac.be/cevipol).[6] Table 1 shows the distribution of the experimental sessions by laboratory for each of the four voting experiments. Most of our sessions were conducted in the Cirano (Montreal) and the Cevipol (Brussels). Importantly, all the experiments were conducted in at least two different laboratories.

In theory, the four behavioral experimental laboratories use a similar strategy to recruit experimental subjects. They circulate calls on the Internet, on local Universities' mailing-lists (where they indicated that they frequently recruit for experiments for which subjects receive money), and they allow adults to sign up as volunteers on the research center's website (or by email). When a laboratory experiment is planned, subjects are randomly chosen from among a full list of volunteers.

Yet in practice, the samples varied quite a lot. After each of the four experiments we asked subjects to answer a short questionnaire, which gave us insights into some of their characteristics. Table 2 shows some of the socio-demographic information gathered, as well as attitudes towards risk and politics, and psychological traits of the subjects, by laboratory. We directly asked subjects in the Cirano (Montreal) and in the Cevipol (Brussels) their occupation (i.e. student or non-student). While about 76 % of the Cevipol sample consisted of students, they made up only about 38 % in the Cirano sample. The Cevipol sample was thus mainly composed of students while the Cirano sample was more heterogeneous.

Furthermore, we can infer from other socio-demographic characteristics that the Leex (Barcelona) and Leep (Paris) samples were also mostly composed of students. Subjects in these two samples were on average younger than those in the Cirano sample (23 years of age compared to 30 years of age) and the standard deviation of age was smaller (3 years compared to 9 years). The Leex and the Leep samples were thus much more similar to the Cevipol sample than to the Cirano sample.

We also see that only 24 % of subjects in the Leex sample had already completed a University degree, as compared to 79 % in the Cirano sample. This suggests that

[6]It is worth noting that we do not address the question of whether cross-national cultural differences influence laboratory experiment results. As pointed by some researchers, the country does not seem to have any influence on the behavior of laboratory subjects (at least in voting experiments, see Van Der Straeten et al. 2010).

Table 1 Sessions by laboratory

	Leex	Cevipol	Leep	Cirano
Distance-to-winner experiment		4 sessions		4 sessions
Winner-takes-all experiment	4 sessions			4 sessions
Unequal-distribution experiment		1 session	1 session	2 sessions
The preference-ranking experiment		1 session		1 session

Table 2 Socio-demographics, attitudes, and personality traits, by laboratory

	Leex	Cevipol	Leep	Cirano
Socio-demographics				
Age	21.1 (2.6)	22.8 (3.5)	23.0 (3.12)	29.9 (9.2)
N	80	126	21	227
Gender (0 = Female, 1 = Male)	0.43 (0.50)	0.61 (0.49)	0.62 (0.50)	0.49 (0.50)
N	80	126	21	227
Occupation (0 = other, 1 = student)	[a]	0.76 (0.44)		0.38 (0.50)
N		21		21
Highest degree obtained (0 = other, 1 = Uni.)	0.24 (0.43)			0.79 (0.41)
N	80			80
Attitudes				
Being a risk taker (0–1 scale)		0.50 (0.27)	0.60 (0.28)	0.48 (0.30)
N		126	21	147
Interest in politics (0–1 scale)	0.66 (0.27)	0.81 (0.21)	0.64 (0.27)	0.63 (0.30)
N	80	42	21	143
Duty to vote (0–1)	0.35 (0.48)	0.71 (0.22)	0.74 (0.13)	0.63 (0.38)
N	80	126	21	227
Altruism (0–1)	0.22 (0.23)	0.55 (0.31)	0.28 (0.22)	0.43 (0.32)
N	80	126	21	227
Personality traits (–1 to +1)				
Extroverted	0.15 (.021)	0.02 (0.22)		0.04 (0.28)
Agreeable	0.05 (0.19)	0.15 (0.19)		0.12 (0.27)
Conscientious	0.16 (0.18)	0.17 (0.24)		0.23 (0.22)
Emotionally stable	0.01 (0.23)	0.01 (0.22)		0.07 (0.27)
Open	0.65 (0.23)	0.15 (0.15)		0.67 (0.28)
N	80	21		101
Sample type	Student-dominated	Student-dominated	Student-dominated	Heterogeneous

Note: Entries are the mean of the variable. Standard deviations are in parentheses

[a]The questionnaires asked at the end of the experiments were not always exactly similar. This explains why we have missing values and inconsistent Ns in Table 1

the Leex subjects were still undergraduate students. Thus, the Leex, Cevipol, and Leep samples were mostly composed of students, while the Cirano sample was more heterogeneous (though not representative of the average population since it was mainly composed of young, highly educated individuals). Another indicator of this heterogeneity is the gender balance. While the proportion of males and females was virtually equal in the Cirano sample, there were more female in the Cevipol sample and more men in Leep and Leex samples.

The reason for these differences in makeup of the four samples might be due to several factors. First, there is a practical issue concerning the location of the laboratories. While the Cirano is located in the business district of the city, the other three are located on a University campus. Second, the Cirano makes efforts to circulate its calls for volunteers outside of the Universities, while the others do not. For example, it has conducted recruiting campaigns in metro stations.

Table 2 also includes some information about the subjects' attitudes and personality traits. The questionnaire we circulated among them focused on self-reported attitudes towards risk (i.e. risk aversion) and interest in politics, ranking these on 10-point scales. Results show that samples were quite similar in this respect. The only exception was the Cevipol sample, where subjects were relatively more interested in politics (the average was 0.8 on a 0–1 scale, compared with 0.6 in the other samples).

We also asked them whether they consider voting to be a duty or a choice. Here the Leex sample was the outlier. Subjects in this sample had a much lower sense of civic duty (the average was 0.4 on a 0–1 scale, compared with 0.7 in the other samples).

After our four laboratory voting experiments, we conducted a dictator game as a way to measure the altruism of subjects. This dictator game consisted of offering the possibility to each subject to anonymously give a small fraction of the points that they had accumulated to another subject.[7] The most altruistic subjects were found in the Cevipol and Cirano samples (the average was 0.5, compared with 0.25 in the other samples).

Finally, we asked questions related to the Big Five personality traits in all but the Leep sample. We constructed indicators of extraversion, agreeableness, conscientiousness, emotional stability and openness from the classic set of ten items. There was no clear pattern between the samples.

However, it is worth noting that standard deviations in political attitudes and personality indicators were almost always higher in the Cirano sample. This evidence gives further weight to the suggestion that the Leep, Leex, and Cevipol samples were more homogeneous. We can thus conclude that we had two types of sample: three student-dominated samples (Leep, Leex, and Cevipol) and one heterogeneous sample (Cirano).

[7]These were quasi hypothetical points (only two subjects were chosen as the giver-receiver pair), based on a random draw in all experiments except in the winner-takes-all experiment.

5 Results

To test whether the behavioral patterns of subjects engaged in the four laboratory voting experiments described above were different in our four samples, we constructed several variables capturing their behavior and then systematically compared these across samples: The number of points received, of votes casted, of sincere votes casted, and of decision that maximizes one's payoff. The first variable was the amount of points received. As all experiments relied on a specific structure of payoff (see above) and a neutral framework (the candidates were labeled with letters so as to ensure that the underlying political preference of subjects did not interfere with their decision), we might reasonably assume that subjects made decisions so as to maximize their gains. The gains are directly related to the number of points obtained (for example one point is converted into 0.25$). Therefore, this number is a direct indicator of how good the subjects were at understanding the experimental rules, anticipating the behavior of other subjects, and making consistent decisions.

Table 3 reports the average number of points by subject (and standard deviation) for each of the four experiments by sample type. We see that in all four laboratory voting experiments, with the exception of the unequal-distribution experiment, subjects from student-dominated samples gained more than those from the heterogeneous sample. However, this difference is small. It represents at most (in the winner-takes-all experiment) 10 % of the average number of points of the subjects from student-dominated samples. None of these differences are statistically significant.

Second, we looked more closely at the type of decision made by subjects. Table 3 reveals that subjects from student-dominated samples abstained more often in the distance-to-winner and winner-takes-all experiments. This difference is especially striking in the winner-takes-all experiment, since subjects in these samples voted on average one time less than subjects in the heterogeneous sample (out of six elections). This result suggests that students free-ride more often than the average population, though the difference is not statistically significant. Since voting was costly in these two laboratory voting experiments, this difference in free-riding practices explains why the average gain of subjects in student-dominated samples was greater than the gain of subjects from the heterogeneous sample.

However, the average number of votes by subject in the unequal distribution experiment somehow complicates the picture. Over eight elections, it seems that subjects from student-dominated samples voted 1.3 times more than those from the heterogeneous sample (this difference is significant at a level of 0.01). This third laboratory voting experiment was different from the first two as the decision subjects had to make was not only about voting or abstaining, but was also about which candidate to vote for (there were four candidates). The calculations each subject needed to make to maximize their gain was thus much more complex. This probably explains why we do not observe that abstaining was more common in student-dominated samples in this experiment.

Table 3 Points, votes, sincere votes, and maximizing decisions by sample type

	Distance-to-winner (10 elections)	Winner-takes-all (6 elections)	Unequal-distribution (8 elections)	Preference-ranking (24 elections)
Points				
Student sample	88.81 (13.47)	19.08 (9.13)	55.55 (6.67)	463.81 (58.09)
Heterogeneous sample	88.14 (13.13)	17.15 (9.13)	57.48 (6.49)	459.51 (52.77)
Difference	0.67	1.93	–1.93	4.29
N	168	160	84	42
Votes				
Student sample	6.90 (2.06)	2.46 (1.83)	4.90 (2.66)	
Heterogeneous sample	7.57 (2.27)	3.42 (2.01)	3.57 (2.82)	
Difference	–0.67	–0.96	1.33*	
N	168	160	84	
Sincere votes				
Student sample			2.76 (2.10)	2.53 (2.09)
Heterogeneous sample			2.38 (2.26)	3.43 (2.27)
Difference			0.38	–0.90
N			84	42
Maximizing decisions				
Student sample	3.82 (1.82)	2.92 (1.17)	3.24 (2.39)	23.29 (0.85)
Heterogeneous sample	3.97 (1.81)	3.11 (1.19)	3.71 (2.68)	23.52 (0.60)
Difference	–0.15	–0.19	–0.48	–0.23
N	168	160	84	42

Note: Entries are the mean by individual. Standard deviations are in parentheses. *$p < 0.05$, **$p < 0.01$ (from two-tailed t-tests)

Table 3 also reports the number of sincere votes (i.e. votes for a candidate associated with the subject's highest payoff) in laboratory voting experiments in which there were more than two candidates (that is, the unequal-distribution experiment and the preference-ranking experiment). No clear patterns emerge from these indicators. Subjects from student-dominated samples voted as sincerely as those from the heterogeneous sample.

Finally, we also looked at the number of maximizing decisions for all elections. The decision of a subject is considered to be maximizing when, considering the decision made by all other subjects, his or her gain could not have been higher.[8] It is thus the best indicator of how good the subjects are at understanding the

[8]For the unequal-distribution experiment, the number of maximizing decisions is calculated from the subjects' perceptions. Before each decision, subjects were asked whether their decision will change the result of the election, and so is a proxy for a maximizing decision.

experimental game, anticipating the behavior of other subjects, and making consistent decisions.

The results in Table 3 reveal that the differences were once again very minimal (and not statistically significant). In the three first experiments (that is, the distance-to-winner experiment, the winner-takes-all experiment, and the unequal-distribution experiment), subjects made 40–50 % of maximizing decisions regardless of whether they were in student-dominated samples or in the heterogeneous sample. Even in the distance-to-winner experiment and the winner-takes-all experiment where subjects from student-dominated samples free-rode more than those in the heterogeneous sample, the difference was only 3 % of all decisions. The proportion of maximizing decisions was much higher in the preference-ranking experiment (more than 90 %). However, the absence of cost of voting made virtually all decisions a maximizing decision.

To further assess the potential behavioral differences across sample type, we specifically concentrated on the preference-ranking experiment. Subjects who participated in this experiment were asked whether they are full-time students or not. We thus got a direct measure of their occupation. Table 4 reports the overall amount of gains, the number of sincere votes, and the number of maximizing decisions by occupation. In line with what we find above, the results reveal that the differences between students and non-students were very small (and not statistically significant). The largest difference concerned the number of sincere votes. On average, non-students voted for their preferred party about 3 % more than students (this explains the difference in the gains obtained). There was virtually no difference between students and non-students as far as the number of maximizing decisions is concerned.

To make their decision, i.e. to abstain or vote, and if the latter, for which candidate, subjects rely on various information. Given the nature of the experimental games, subjects engaged in laboratory voting experiments typically use the results of previous elections. This is because they do not have other information available to them about the other participants' behavior to make their decision.

To examine whether the heuristics on which voters rely to make their decision differ depending on sample type, we estimated probit models.[9] For the two first experiments on turnout (that is, the distance-to-winner experiment and the winner-takes-all experiment), we predicted the probability of each subject voting based on whether they had voted in the previous election, and based on the absolute difference in votes obtained by the two competing candidates in the previous election. We expected this difference to be negatively associated with the probability of voting. If the absolute difference is low at an election T_{-1}, a subject might think that it will also be low at election T_0. Therefore, they should be more inclined to cast a vote as their vote is more likely to be pivotal (and vice versa if the absolute difference at T_{-1} is high).

[9]The estimated probit and logit models give similar results and are available upon request to the authors.

Table 4 Points, sincere votes, and maximizing decisions by status (preference-ranking experiment, 24 elections)

	Students	Non-students	Difference
Points	432.92	437.78	4.86
Sincere votes	2.71	3.33	0.63
Maximizing decisions	23.33	23.50	0.17
N	24	18	44

Note: Entries are the mean by individual. Standard deviations are in parentheses. *$p < 0.05$, **$p < 0.01$ (from two-tailed t-tests)

Table 5 shows the results of the probit models for the two turnout experiments by sample type. As expected, we find that the absolute difference in votes obtained by the two competing candidates is negatively associated with the probability of voting (statistically significant at a level of 0.05 or 0.01 for the two experiments). This supports the idea that subjects use this information to estimate their probability of being pivotal.

When we differentiate these analyses by sample type, we see, once again, very few differences between student-dominated and heterogeneous samples. The coefficient estimate associated with the absolute difference of votes obtained by the two competing candidates is very similar (although sometimes not statistically significant).

Also, we find in Table 5 that voting at T_{-1} is a little bit more associated with the propensity to vote at T_0 in heterogeneous samples than in student-dominated samples (in both experiments). It suggests that subjects in heterogeneous samples were less reactive to the previous election's result than subjects in student samples. However, the difference is small.

To evaluate whether the heuristics used by subjects from student-dominated samples are similar to those used by subjects from heterogeneous samples in strategic desertion experiments (i.e., the unequal-distribution experiment and the preference-ranking experiment), we estimated probit models predicting the probability of each subject casting a sincere vote at T_0 based on whether this subject cast a sincere vote at T_{-1}, as well as the difference between the number of votes obtained by the winning candidate and his or her preferred candidate at T_{-1}. We expected that the higher this difference, the less likely a subject was to cast a sincere vote. If the preferred candidate lost by a large number of votes in the previous election, the subjects might expect they will lose again in the next election. They should thus desert their preferred candidate and opt for their second most viable option in the next election as a way to maximize their gain.

Table 6 reports the results of the probit models. As expected, the difference between the number of votes obtained by the winning candidate and the preferred candidate at T_{-1} is negatively associated with the probability of casting a sincere vote at T_0 in both experiments. This association is however small, and not statistically significant for the unequal-representation experiment and only significant at a level of 0.05 for the preference-ranking experiment. However, the subjects seem

Table 5 Explaining the decision to vote in turnout experiments

	All samples	Student-dominated samples	Heterogeneous sample
Distance-to-winner experiment			
Intercept	–0.71** (0.06)	–0.44** (0.12)	–0.46 (0.28)
Voting $_{T-1}$	0.07** (0.01)	0.03 (0.02)	0.06** (0.02)
|Number of votes for Candidate A – Number of votes for Candidate B|$_{T-1}$	–0.05* (0.02)	–0.04** (0.01)	–0.03 (0.06)
χ^2	127.58**	11.66**	10.31**
N	800	400	400
Winner-takes-all experiment			
Intercept	0.36** (0.08)	0.36** (0.12)	0.31* (0.15)
Votes $_{T-1}$	0.40** (0.10)	0.23* (0.10)	0.57** (0.14)
|Number of votes for Candidate A – Number of votes for Candidate B|$_{T-1}$	–0.02** (0.01)	–0.00 (0.02)	–0.01 (0.03)
χ^2	25.75**	8.03*	17.81**
N	1512	756	756

Note: Entries are coefficient estimates of probit models predicting the decision of subject to vote at T_0. Standard errors clustered by session are in parentheses. *$p < 0.05$, **$p < 0.01$ (two-tailed)

to have relied on this information to make their decision to vote sincerely or to desert their preferred candidate at T_0.

When we differentiate the analyses by sample, we find that subjects from student-dominated samples and those from heterogeneous samples essentially used the same heuristics. The coefficient estimates are similar in the unequal-distribution experiment. However, they are different in the preference-ranking experiment. The negative association between the difference in the number of votes obtained by the winning candidate and the preferred candidate at T_{-1} is larger when the sessions were conducted on student-dominated samples (and statistically significant at a level of 0.05). This is further evidence in support of the idea that subjects from student-dominated samples were more reactive to the previous election's result than those from heterogeneous samples.

6 Conclusion

Laboratory voting experiments are more and more popular in the political science and economics literature. However, this method is often criticized for its lack of sample validity. In particular, the practice of focusing on student samples is said to overestimate the rationality of actors. In this chapter, we give new evidence related

Table 6 Explaining voting behavior in strategic desertion experiments

	All samples	Student samples	Heterogeneous sample
Unequal-distribution experiment			
Intercept	–1.03** (0.15)	–1.15** (0.29)	–1.04** (0.23)
Voting for the preferred candidate $_{T-1}$	0.18** (0.02)	0.19** (0.04)	0.19** (0.04)
Number of votes for winning candidate – Number of votes for the preferred candidate at T_{-1}	–0.04 (0.02)	–0.03 (0.03)	–0.03 (0.04)
χ^2	165.11**	113.00**	50.76**
N	672	336	336
Preference-ranking experiment			
Intercept	–0.56* (0.26)	–0.42 (0.39)	–0.72* (0.37)
Voting for the preferred candidate $_{T-1}$	–0.05* (0.03)	–0.08* (0.04)	–0.03 (0.04)
Number of votes for winning candidate – Number of votes for the preferred candidate at T_{-1}	–0.05* (0.02)	–0.07* (0.03)	–0.03 (0.03)
χ^2	5.44	4.77	1.09
N	1008	504	504

Note: Entries are coefficient estimates of autoregressive probit models predicting the decision of subjects to vote sincerely at T_0. Standard errors clustered by session are in parentheses. *$p < 0.05$, **$p < 0.01$ (two-tailed)

to this issue in comparing the results of four laboratory voting experiments on strategic voting and turnout that are replicated on student-dominated and heterogeneous (although highly educated) samples. Our analysis of the behavioral patterns of experimental subjects reveals that the differences between the two types of samples were rather small. Even if subjects of student-dominated samples tended to free-ride more often in simple turnout games, this pattern disappeared once the game got more complex.

We also examined the heuristics used by subjects to make their decision to vote or to abstain in turnout experiments, or to vote sincerely or to desert their preferred candidate in strategic desertion experiments. Here, we found that subjects from student-dominated samples were a little bit more reactive to the previous election's result than subjects from the heterogeneous sample. However, our results also indicate that, regardless of the sample, subjects tended to rely on the same information, and to apply the same strategy (all coefficient estimates have the same sign). It is reasonable to think that if we had organized more sessions on our heterogeneous sample, we would find also similar statistically significant patterns.

In conclusion, we claim that the debate concerning the validity of results derived from voting experiments conducted on students should be somehow reconsidered. The difference in behavioral patterns between student-dominated and heterogeneous samples appears to be minimal. If there is a sampling bias, it is probably not arising from the tendency of laboratories to recruit students *per se*.

However, our heterogeneous samples were mostly composed of highly educated subjects. Highly educated individuals are, by definition, former students. During their study, they have had the chance to develop abstract reasoning capacities. Even though they might not have all had the opportunity to put these capacities into practice recently (if one assumes that their current occupation is less abstract than the subjects taught in colleges and Universities), they are probably as able as students to anticipate the behavior of other subjects and adopt a rational behavior.

For this reason, our chapter does not offer a definitive answer to the question of the potential existence of a sampling bias in voting experiments (or all other behavioral group experiments). We recommend researchers interested in the topic to go beyond the rather crude distinction between students and non-students, and to analyze the differences in how experimental subjects behave depending on their reasoning abilities. A good proxy could be, for example, the highest degree obtained (or to targeted for students).

References

Belot, M., Duch, R., & Miller, L. (2010). *Who should be called to the lab? A comprehensive comparison of students and non-students in classic experimental games* (CESS Working Paper). 51p.

Benz, M., & Meier, S. (2008). Do people behave in experiments as in the field?: Evidence from donation. *Experimental Economics, 11*, 268–281.

Blais, A., Pilet, J.-B., Van Der Straeten, K., Laslier, J.-F., & Heroux-Legault, M. (2014). To vote or to abstain? An experimental test of rational calculus in first past the post and PR elections. *Electoral Studies, 36*, 39–50.

Bosch-Domènech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2002). One, two, (three), . . ., infinity, . . .: Newspapers and lab beauty-contest experiments. *American Economic Review, 92*, 1687–1701.

Campbell, D., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Carpenter, J., Burks, S., & Verhoogen, E. (2005). Workers: The effects of social framing on behavior in distribution games. *Research in Experimental Economics, 10*, 261–290.

Carpenter, J., Connolly, C., & Myers, C. K. (2008). Altruistic behavior in a representative dictator experiment. *Experimental Economics, 11*, 282–298.

Chou, E., McConnell, M., Nagel, R., & Plott, C. R. (2009). The control of game form recognition in experiments: Understanding dominant strategy failure in a simple two-person 'guessing game'. *Experimental Economics, 12*, 159–179.

Danielson, A. J., & Holm, H. J. (2007). Do you trust your Bret hen? Eliciting trust attitudes and trust attitudes in a Tanzanian congregation. *Journal of Economic Behavior and Organization, 62*, 255–271.

Druckman, J. N., Green, D. P., Kuklinski, J. H., & Lupia, A. (2006). The growth and development of experimental research in political science. *American Political Science Review, 100*, 627–635.

Druckman, J. H., & Kam, C. D. (2011). Students as experimental participants: A defense of the 'narrow data base'. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *Cambridge handbook of experimental political science* (pp. 41–57). Cambridge: Cambridge University Press.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics, 14*, 583–610.

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). Does studying economics inhibit cooperation? *The Journal of Economic Perspectives, 7*, 159–171.

Gallego, A., Galais, C., Guinjoan, M., Lavoie, J.- M., & Blais, B. (2016). Visibility and sanctions: The social norm of voting in the lab. In A. Blais, J. -F. Laslier, & K. Van der Straeten (Eds.), *Voting experiments*. Cham: Springer, pp. 127–146.

Kam, C. D. (2005). Who toes the party line?: Cues, values, and individual differences. *Political Behavior, 27*, 163–182.

Kam, C. D., Wilking, J. R., & Zechmeister, E. J. (2007). Beyond the 'narrow data base': Another convenience sample for experimental research. *Political Behavior, 29*, 415–440.

Kühberger, A. (1998). The influence of framing on risky decisions. *Organizational Behavior and Human Decision Processes, 75*, 23–55.

Labbé St-Vincent, S. (2013). An experimental test of the pivotal voter model under plurality and PR elections. *Electoral Studies, 32*, 795–806.

Labbé St-Vincent, S., Pilet, J.- B., & Blais, A. (2013). *Where is the electoral sweet spot?* Paper presented at EPSA annual conference, Barcelona.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences tell us about the real world? *Journal of Economic Perspectives, 21*, 153–174.

Lupia, A. (2002). New ideas in experimental political science. *Political Analysis, 10*, 319–324.

McDermott, R. (2002). Experimental methodology in political science. *Political Analysis, 10*, 325–342.

Morton, R. B., & Williams, K. C. (2008). Experimentation in political science. In J. M. Box-Steffensmeier, H. E. Brady, & D. Collier (Eds.), *Handbook of political methodology* (pp. 339–356). Oxford: Oxford University Press.

Plott, C. R. (1991). Will economics become an experimental science? *Southern Economic Journal, 57*, 901–919.

Sears, D. O. (1986). College sophomores in the laboratory: Influence of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51*, 515–530.

Van der Straeten, K., Laslier, J.-F., Sauger, N., & Blais, A. (2010). Sincere, strategic, and heuristic voting under four election rules. *Social Choice and Welfare, 35*, 435–472.

Villeval, M.-C. (2007). Experimental economics: Contributions, recent challenges. *French Politics, 5*, 178–186.