PSRM

## RESEARCH NOTE

# Incentivized choice in large-scale voting experiments

Tanja Artiga González[1], Georg D. Granic[2], Franziska Heinicke[3], Stephanie Rosenkranz[4] and Utz Weitzel[1,5] (iD)

[1]School of Business and Economics, VU University Amsterdam, 1081 HV Amsterdam, The Netherlands, [2]School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands, [3]Department of Economics, University of Mannheim, Mannheim, Germany, [4]Utrecht University School of Economics, 3584 EC, Utrecht, The Netherlands and [5]Radboud University Nijmegen, Faculty of Management, Elena-Ostrom Building, Nijmegen, The Netherlands
**Corresponding author:** Utz Weitzel; Email: u.weitzel@vu.nl

### Abstract
Survey experiments that investigate how voting procedures affect voting behavior and election outcomes use hypothetical questions and non-representative samples. We present here the results of a novel survey experiment that addresses both concerns. First, the winning party in our experiment receives a donation to its campaign funds inducing real consequences for voting. Second, we run an online experiment with a Dutch national representative sample ($N = 1240$). Our results validate previous findings using a representative sample, in particular that approval voting leads to a higher concentration in votes for smaller parties and strengthens centrist parties in comparison to plurality voting. Importantly, our results suggest that voting behavior is not affected by voting incentives and can be equally reliably elicited with hypothetical questions.

## 1. Introduction

The comparative study of voting procedures is a central area of research within political science (Grofman, 2016). In recent times, there has been a notable surge in utilizing survey experiments to examine the influence of voting procedures on both voting behavior and election outcomes (Dolez *et al.*, 2011; Blais *et al.*, 2016). Whereas novel experimental results have enhanced our understanding of the effects of voting procedure changes, survey experiments can be subject to criticism. The central question of this paper is whether using hypothetical questions and non-representative samples in survey voting experiments is problematic and biases inferences.

Survey experiments on voting procedures are coined *in situ* experiments, henceforth simply referred to as voting experiments or experiments. They often employ an exit-poll design. After casting votes in an official election, voters are asked to vote hypothetically a second time using a different voting procedure (Laslier and Van der Straeten, 2008; Baujard *et al.*, 2014). A variant approach is to conduct online experiments and to ask the same voter to cast votes using one or several different voting procedures hypothetically (Laslier *et al.*, 2015; Bol *et al.*, 2016).

The strengths of such experiments are clear; they facilitate establishing causal claims and exhibit a high external validity as they are conducted against a real political backdrop on a real voting population. However, given the hypothetical question format and non-compulsory participation, earlier experiments face two potential drawbacks. For one, earlier work contains self-selecting samples, i.e., only contain individuals with sufficient motivation to participate in a

study. This could bias inferences as participating individuals may not be representative of the general population. Second, voting behavior in earlier experiments is hypothetical; it bears no consequences on the official election outcome. Biases arising from hypothetical questions are well documented in social sciences (Hertwig and Ortmann, 2001; Baumeister *et al.*, 2007) and include, e.g., a greater tendency to give socially desirable answers, which can affect voter behavior (Hanmer *et al.*, 2014; Morin-Chassé *et al.*, 2017).

In this paper, we present the results of a large-scale online experiment that addresses the problems of non-representativeness and hypothetical questions for voting experiments. We conducted our experiment using a Dutch national representative sample ($N = 1240$) and we incentivize voting behavior by coupling experimental election results to real party donations. That is, the winning party in the experiment receives a donation to its campaign funds inducing real consequences for voting behavior.

To estimate the degree of bias present due to hypothetical questions in voting experiments, we employed a randomized control trial in which we compare hypothetical and incentivized answers. We hereby followed the existing literature in two important ways. First, we conducted the experiment against a real political backdrop, the then upcoming 2021 Dutch general election. Second, we used the two most widely studied voting procedures: first-past-the-post plurality voting and approval voting. Under the latter, voters can approve as many parties as wished and the party with the most approvals wins the election (Brams and Fishburn, 1978; Alós-Ferrer, 2006).

One important message emerges from our experiment: incentivized choice has little impact on voting behavior and on election outcomes in our voting experiment. Differences due to incentives tend to be non-significant and are generally of an order of magnitude that is likely too small to merit sparking a major reevaluation of the conclusions drawn from existing experiments. This result is good news for researchers and reassuring.

We further replicate empirical regularities of earlier literature. In particular, we observe an increase in the number of effective parties and a strengthening of centrist candidates under approval voting. This is encouraging as it demonstrates that earlier results generalize to a representative sample.

## 2. Main hypothesis

In this section, we present our main research hypotheses concerning the effect of incentivized choice on behavior in voting experiments. We will focus on first-past-the-post plurality voting and approval voting. Our exposition here focuses on political parties, as we conducted our experiment using Dutch parties, but all our arguments hold equally true for the case of political candidates. Furthermore, we assume that previously reported empirical regularities stemming from self-selected samples translate into representative samples.

Using hypothetical choice in voting experiments on plurality voting could be problematic. According to Duverger's law (Duverger, 1954), plurality voting creates strong incentives for voters to vote tactically, deserting smaller parties in favor of larger parties who stand a better chance of winning the election. Indeed, empirical evidence documents a significant proportion of tactical voting (Eggers and Vivyan, 2020). However, if an election is hypothetical, voters can safely ignore trade-offs and do not need to engage in tactical considerations. In turn, the vote share of smaller parties in a hypothetical election might be artificially higher than in a real election. Incentivized choice can help close this gap as votes in an incentivized experiment bear consequence so that trade-offs underlying tactical voting become salient. Our first research hypothesis can therefore be formulated as follows.

*H1: Due to tactical voting, incentivized choice under plurality voting leads to a higher concentration of votes on larger parties in comparison to hypothetical choice.*

We next turn to approval voting, which generally reduces tactical considerations, because votes for multiple parties are possible. As approvals on a given ballot are non-competing, approving of a smaller party does not come at the expenses of winning for an approved larger party and *vice versa* (Brams and Fishburn, 1978). In a hypothetical experiment, approval voting might therefore invite voters to over-approve less well-known parties that, for example, use attention-grabbing names. Voters in Germany, for example, reported to have looked at the complete list of all available parties for the first time while casting their votes in an approval voting experiment (Alós-Ferrer and Granic, 2012). This can lead to an artificial over-approval of smaller, more issue-focused parties, particularly when voting is without real consequence. In a direct comparison, we therefore expect:

*H2: Incentivized choice under approval voting leads to a higher concentration of votes on larger parties in comparison to hypothetical choice.*

Research hypotheses 1 and 2 concern the effect of incentivized choice on behavior in voting experiments. We also aim to validate our experiment by establishing comparability of our results with findings from the extant literature. The two most widely observed experimental regularities when comparing approval voting and plurality voting are as follows. In comparison to plurality voting, approval voting leads to a strengthening of smaller parties and of centrist parties (Laslier, 2006; Alós-Ferrer and Granic, 2015). Both regularities can be explained by the non-competitiveness of approvals on a given ballot. For the former, voters no longer need to fear wasting their votes on smaller parties and can express their support for them. Approval voting also offers voters more options to express their preferences by casting multiple approvals (Brams and Fishburn, 1978). The latter regularity can be explained by a larger overlap in the voter base of centrist parties in comparison to more extreme parties. Consequently, by a simple political proximity argument, we can expect that approval voting strengthens centrist parties. This leads to the following two research hypotheses.

*H3: Approval voting leads to a higher concentration in votes for smaller parties in comparison to plurality voting.*

*H4: Approval voting strengthens centrist parties and leads to a higher concentration of votes on centrist parties in comparison to plurality voting.*

Taken together, our experiment will allow us to investigate if the findings from the existing literature on *in situ* experiments are robust with respect to using incentivized choice and generalize to a representative sample.

## 3. Experimental design

The experiment was conducted online with the help of the pollster Dynata using a Dutch representative sample in the last two weeks of September 2020 ($N = 1240$).[1] During this period, political parties were in an early stage of campaigning for the 2021 Dutch National Election to be held on 17 March 2021. In the experiment, voters were asked to submit their vote for one of the 24 parties that had indicated their intention to participate in the upcoming election at that time.[2] To test our research hypotheses, we implemented a $2 \times 2$ between-participant experimental design, summarized in Table 1 below.

---

[1]We obtained IRB approval from our home universities (number SBE9/14/2022gwl260).
[2]Detailed sample characteristics and the experimental instructions are included in the online appendix.

**Table 1.** Experimental treatments and allocation of voters to treatments

|  | Hypothetical choice | Incentivized choice |
|---|---|---|
| Plurality voting | Treatment *P*, *N* = 320 | Treatment *PI*, *N* = 320 |
| Approval voting | Treatment *A*, *N* = 314 | Treatment *AI*, *N* = 306 |

The experimental treatments differed in the voting procedure used. We studied voting under plurality voting and under approval voting, using these two procedures to hypothetically determine the next Dutch Prime Minister. *De facto*, we implemented hypothetical single-winner elections, whereas the Netherlands uses a proportional representation system. We focus on single winner elections for three important reasons. First, existing work usually compares an official voting method familiar to the voters with an unofficial voting method, unfamiliar to the voters. Existing evidence suggests that unfamiliarity leads to greater differences between hypothetical and real answers (Schläpfer and Fischhoff, 2012). Observed differences between voting methods in earlier work may, hence, be confounded by familiarity. Second, the extant literature focuses on approval voting and similar evaluative voting procedures which are single-winner procedures, so our design allows us to relate our results to earlier work. Third, the proportional system used in the Netherlands has little room for strategic consideration as there is no barring clause to enter parliament. However, the effect of incentives could be strongest when making such trade-offs, as incentivized choice saliently highlights the need to consider voting decisions more carefully.

In the plurality voting treatments, indicated with *P*, participants were informed that they would have to submit one vote. In the approval voting treatments, indicated with *A*, participants could approve of as many parties as they wished. Abstaining was not allowed, and we asked participants to imagine that the leading candidate of the party with the most votes/approvals would automatically become Dutch Prime Minister.

The experimental treatments also differed regarding incentives. In the unincentivized treatments, *P* and *A*, participants stated their vote (hypothetically) without further consequences. In treatments with incentives, *PI* and *AI*, participants were informed that the winning party would receive a donation of EURO 500. This information was presented prominently. It was further stressed that considering the upcoming election, donations presented an important income source for campaigning. We also included a link to a website where proof of the donation would be published shortly after the experiment. We deliberately set the height of the incentives to balance two competing constraints. On the one hand, incentives should be high enough to be meaningful to the participants. On the other hand, incentives may not be too high as to have significant influence on the election outcome for ethical concerns. As we aimed for *N* = 300 participants in each treatment, our average voter donations amount to EURO 1.67, which is close to the annual average per voter party donation of EURO 1.64 observed in the Netherlands.[3]

Participants made their choice from the list of 24 parties. The order of parties was fixed and motivated by the order in which parties occur on the ballot in Dutch elections. All parties that participated in the last election are ordered according to their vote share in the last election; remaining parties are ordered alphabetically. The vote choice was followed by a survey on general demographics, questions about vote intentions and political involvement. We also asked participants to place themselves on economic policy and the GAL/TAN dimension.

The sample was representative regarding gender, age, region, and education. In total, 1240 participants completed the experiment with the (random) allocation across the four treatments (*P*, *A*,

---

[3]In 2018, individual citizens contributed EURO 20.5 million to political parties (Kakebeeke and Knoop, 2019) and the number of eligible voters was 12.5 million in 2018 (https://www.cbs.nl/nl-nl/maatwerk/2018/08/maatwerk-kiesgerechtigde-bevolking).

PI, AI) as indicated in Table 1. On average, participants were between 45 and 49 years old and 50.9 percent were female. Demographics were balanced across experimental treatments.

## 4. Results

Our two main hypotheses H1 and H2 stipulate that incentivized choice shifts the concentration of votes from smaller to larger parties in comparison to hypothetical choice for both plurality voting and approval voting. A straightforward way to measure vote concentration is to compute the number of effective parties EP according to Laakso and Taagepera (1979). The measure counts parties weighted by their relative strength.[4] A lower EP signifies a higher concentration of votes on fewer parties. Hence, we expected that EP would be lower under incentivized choice.

Figure 1 below plots the effective number of parties EP for all four treatments in the experiment with 95 percent confidence intervals. With an EP of 9.08 and 10.03, treatments P and PI respectively come close to the EP of 9.26 realized in the 2021 official Dutch National Election.[5] Looking at the approval voting treatments, we observe an EP of 10.94 and 11.09 in A and AI, respectively.

To test hypotheses 1 and 2, we bootstrapped differences in EP between treatments. Figure 2 below plots the corresponding results. Using one-sample, one-sided t-tests on the bootstrapped samples, we cannot reject the null hypotheses that the EP was smaller or equal in hypothetical choice than in incentivized choice for both plurality voting and approval voting (p-values are 0.182 and 0.807, respectively).[6]

We also investigated hypothesis 2 by analyzing the difference in the average number of approvals cast. If incentives reduce over-approving, voters on average should approve of fewer parties with incentives than with hypothetical choice. We observe that the average number of approvals was 1.90 in treatment A and 1.74 in treatment AI. Using a one-sided, two-sample t-test, the average number of approvals is significantly smaller in AI than in A (p-value is 0.038). This supports our theorizing that hypothetical questions may induce over-approving. The over-approving effect, however, seems to be too small to cause major shifts in party approvals in the aggregate. Overall, all our analyses point to the conclusion that voting behavior in our experiment is statistically not different between incentivized choice and hypothetical choice.

Next, we analyzed tactical voting as a function of proximity. We first obtain party placements in a two-dimensional policy space from "Kieskompas," a popular Dutch voting advice application. For each election, Kieskompas scores all parties along the conservative/progressive dimension (GAL/TAN dimension) and along the economic left/right dimension (Krouwel et al., 2012). Next, for each voter, we calculate the Euclidean distance between the party s/he voted for and the voter's self-placement on these two dimensions. For A and AI, we calculate the averaged Euclidean distance over all approved parties by a voter. Tactical voting should lead voters to vote for parties that are further away in the political space. We found no significant impact of incentives on the distances between voters and the parties they voted for (see online appendix Figure A1 for more details). However, we did find that approval voting reduced tactical voting in comparison to plurality voting in the sense that voters on average approved parties that were closer to their self-placements.

As the last step in our analysis, we aim to validate our experiment by showing that we can replicate two central empirical regularities found in the existing literature. Voting experiments

---

[4] Let $p_i$ denote the vote share of party $i$. $EP = \frac{1}{\sum_{i=1}^{n} p_i^2}$, where $n$ denotes the number of parties. For approval voting, party shares were calculated by (number of approvals received)/(total number of approvals in treatment).

[5] We retrieved the EP for the Dutch National Election in 2021 from Michael Gallagher's database: https://www.tcd.ie/Political_Science/people/michael_gallagher/ElSystems/Docts/ElectionIndices.pdf.

[6] We also run additional tests to assess the effects of incentives on voting behavior. Online appendix Table A.2 shows the distribution of votes across all parties for each treatment. Vote shares were not significantly affected by incentives according to $\chi^2$-tests.
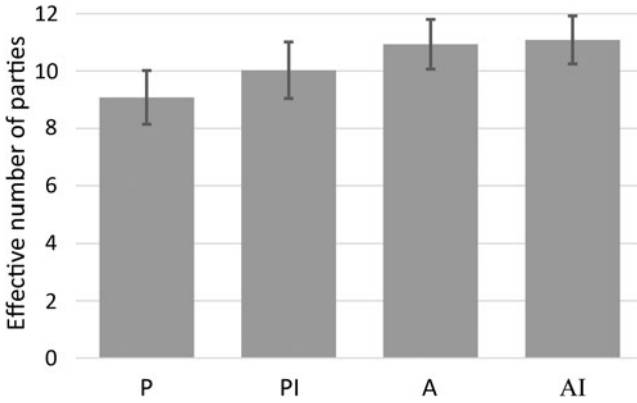
**Figure 1.** The effective number of parties EP for each treatment. Error bars represent 95 percent confidence intervals and were obtained via bootstrapping based on 1000 repetitions assuming an approximately normal distribution.

consistently show that (a) smaller parties and (b) inclusive, more centrist parties benefit the most in terms of vote share when switching from plurality voting to approval voting. The first regularity, our hypothesis 3, implies that the *EP* should be higher under approval voting than under plurality. For example, using Table 3 in Laslier and Van der Straeten (2008), we calculate an *EP* of 8.7 for plurality voting and an *EP* of 13.2 under approval voting for the 2002 French presidential election. Similarly, we obtain an *EP* of 4.9 and 6.8 under plurality voting and approval voting, respectively, using the data in Table 3 of Alós-Ferer and Granic (2012) for the 2009 German Federal Election. Figures 1 and 2 clearly confirm this empirical regularity. One-sided, one-sample *t*-tests on the bootstrapped differences in *EP* between treatments further corroborate our observations showing that the *EP* under approval voting is significantly higher than under plurality voting (p-value incentivized choice was 0.049 and p-value hypothetical choice was 0.002).

The second regularity, our hypothesis 4, postulates that centrist parties benefit the most under approval voting. To study treatment effects for votes for centrist parties, we construct a new variable at the voter level. We again obtain party placements in the two-dimensional policy space
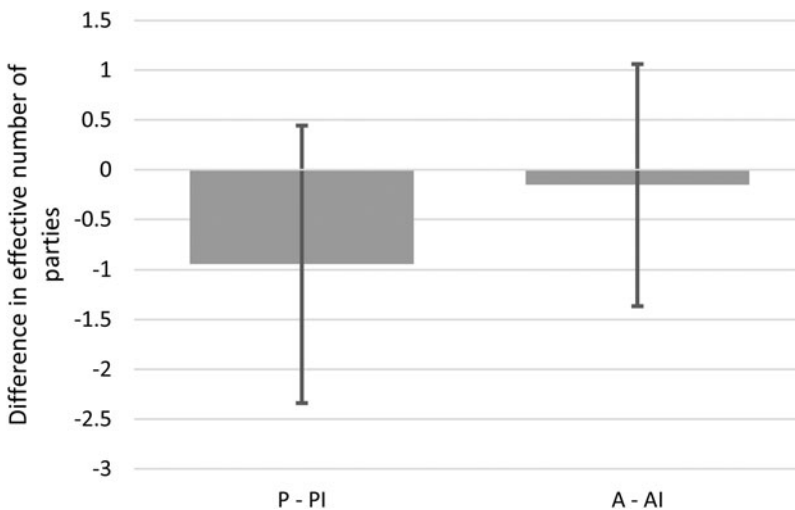


**Figure 2.** Estimated treatment differences in the effective number of parties EP in the experiment. Error bars represent 95 percent confidence intervals and were obtained via bootstrapping based on 1000 repetitions assuming an approximately normal distribution.

from "Kieskompas" on the GAL/TAN and the economic left/right dimension. Next, for each voter, we calculate the Euclidean distance between the party s/he voted for and the center of the two-dimensional policy space. For *A* and *AI*, we calculate the averaged Euclidean distance from the center over all approved parties by a voter. This gives us a measure of how concentrated voting behavior is around the political center. The smaller the distance the more "centrist" the vote. Figure 3 below presents the corresponding treatment averages.

Supporting hypothesis 4, we observe a higher concentration of votes in the political center under approval voting than under plurality voting. The average distance to the political center of parties voted for drops from 1.52 and 1.51 under *P* and *PI* to 1.25 and 1.30 under *A* and *AI*, respectively. These observed differences between approval voting and plurality voting are significant according to one-sided, two-sample *t*-tests (p-value *A* versus *P* was <0.001 and p-value *AI* versus *PI* was <0.001). Again, we do not detect any significant impact of incentives on the Euclidean distance to the political center (both p-values >0.17).

## 5. Conclusions

This paper presented the first evidence on whether using hypothetical questions and self-selected samples in large-scale voting experiments is problematic. To this end, we introduced a novel experimental design that incentivized voting behavior by coupling experimental election results to real party donations on a Dutch representative sample. Whereas we discussed several plausible ways in which hypothetical questions may distort voting behavior, our results did not show any systematic, statistically significant differences between incentivized choice and hypothetical choice. Our results are good news for researchers and reassuring, suggesting that the standard design in large-scale voting experiments seems appropriate.

Our findings contrast with the ones of neighboring social sciences that do document issues in using hypothetical questions. The debate about the usage of hypothetical versus incentivized choice essentially revolves around the question to which extent answers to hypothetical questions translate to relevant real-world settings. Our usage of real Dutch parties and proper explanations of voting procedures, including how the winner is determined, may have offered a high degree of relevance to our representative Dutch sample. It is plausible that these participants brought their home-grown values and convictions to the experiment so that even hypothetical questions may have appeared real enough to them to care about the experiment and to report their voting behavior honestly.

Of course, interpreting null results should always come with a caution. It may be that incentivized questions do matter in voting experiments under circumstances not covered by our
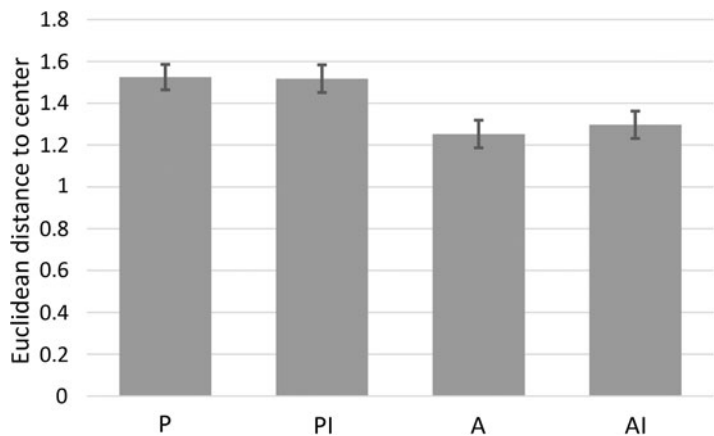


**Figure 3.** Average Euclidean distance between parties voted for and the Dutch political center, by treatment.

experiment. For instance, a different mechanism to incentivize choice might have been able to detect significant difference between hypothetical voting behavior and incentivized voting behavior. However, at least with regard to using party donations, the observed effect of incentives was of an order of magnitude that is likely too small to merit sparking a major reevaluation of the conclusions drawn from existing experiments. We hope that our results will prove useful for future research, and we hope to stimulate further investigations into the robustness of conclusions drawn from the standard *in situ* experimental design.

# References

Alós-Ferrer C (2006) A simple characterization of approval voting. *Social Choice and Welfare* **27**, 621–625.

Alós-Ferrer C and Granić Đ-G (2012) Two field experiments on approval voting in Germany. *Social Choice and Welfare* **39**, 171–205.

Alós-Ferrer C and Granić Đ-G (2015) Political space representations with approval data. *Electoral Studies* **39**, 56–71.

Baujard A, Igersheim H, Lebon I, Gavrel F and Laslier J-F (2014) Who's favored by evaluative voting? An experiment conducted during the 2012 French presidential election. *Electoral Studies* **34**, 131–145.

Baumeister RF, Vohs KD and Funder DC (2007) Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspectives on Psychological Science* **2**, 396–403.

Blais A, Laslier J-F and Van der Straeten K (eds) (2016) *Voting Experiments*. Switzerland: Springer International Publishing.

Bol D, Harfst P, Blais A, Golder SN, Laslier J-F, Stephenson LB and Van der Straeten K (2016) Addressing Europe's democratic deficit: an experimental evaluation of the pan-European district proposal. *European Union Politics* **17**, 525–545.

Brams SJ and Fishburn PC (1978) Approval voting. *American Political Science Review* **72**, 831–847.

Dolez B, Grofman B and Laurent A (eds) (2011) *In Situ and Laboratory Experiments on Electoral Law Reform: French Presidential Elections*. New York: Springer.

Duverger M (1954) *Political Parties*. London: Methuen; New York: J. Wiley.

Eggers AC and Vivyan N (2020) Who votes more strategically? *American Political Science Review* **114**, 470–485.

Grofman B (2016) Perspectives on the comparative study of electoral systems. *Annual Review of Political Science* **19**, 523–540.

Hanmer MJ, Banks AJ and White IK (2014) Experiments to reduce the over-reporting of voting: a pipeline to the truth. *Political Analysis* **22**, 130–141.

Hertwig R and Ortmann A (2001) Experimental practices in economics: a methodological challenge for psychologists? *Behavioral and Brain Sciences* **24**, 383–403.

Kakebeeke P and Knoop B (2019) "De moeizame zoektocht van politieke partijen naar geld (en leden)". Het Financieele Dagblad. 10/03/2019. Available at https://fd.nl/economie-politiek/1317864/de-moeizame-zoektocht-van-politieke-partijen-naar-geld-en-leden

Krouwel A, Vitiello T and Wall M (2012) The practicalities of issuing vote advice: a new methodology for profiling and matching. *International Journal of Electronic Governance* **5**, 223–243.

Laakso M and Taagepera R (1979) "Effective" number of parties: a measure with application to West Europe. *Comparative Political Studies* **12**, 3–27.

Laslier J-F (2006) Spatial approval voting. *Political Analysis* **14**, 160–185.

**Laslier J-F and Van der Straeten K** (2008) A live experiment on approval voting. *Experimental Economics* **11**, 97–105.

**Laslier J-F, Blais A, Bol D, Golder SN, Harfst P, Stephenson LB and Van der Straeten K** (2015) The EuroVotePlus experiment. *European Union Politics* **16**, 601–615.

**Morin-Chassé A, Bol D, Stephenson LB and St-Vincent SL** (2017) How to survey about electoral turnout? The efficacy of the face-saving response items in 19 different contexts. *Political Science Research and Methods* **5**, 575–584.

**Schläpfer F and Fischhoff B** (2012) Task familiarity and contextual cues predict hypothetical bias in a meta-analysis of stated preference studies. *Ecological Economics* **81**, 44–47.