

Designing preference voting

Philipp Harfst^{a,*}, Damien Bol^b, Jean-François Laslier^c

^a Technische Universität Dresden, Germany

^b King's College London, United Kingdom

^c Paris School of Economics (CNRS), France

ARTICLE INFO

Keywords:

Electoral system
Proportional representation
Preference voting
Approval voting
Experiment
Austria

ABSTRACT

Electoral systems in which voters can cast preference votes for individual candidates within a party list are increasingly popular. To the best of our knowledge, there is no research on whether and how the scale used to evaluate candidates can affect electoral behavior and results. In this paper, we analyze data from an original voting experiment leveraging real-life political preferences and embedded in a nationally representative online survey in Austria. We show that the scale used by voters to evaluate candidates makes differences. For example, the possibility to give up to two points advantages male candidates because male voters are more likely to give 'zero points' to female candidates. Yet this pattern does not exist in the system in which voters can give positive and negative points because male voters seem reluctant to actively withdraw points from female candidates. We thus encourage constitution makers to think carefully about the design of preference voting.

1. Introduction

A majority of countries in Europe and Latin America, as well as many in other parts of the world, use proportional representation (PR) to elect national representatives (Bormann and Golder 2013). In these systems, individuals can usually cast a vote for a party list. A growing number of countries also gives them the possibility to express a preference for candidates within the list. These preference votes are then used to decide the set of elected candidates, which sometimes differs from the list ordering decided by the party. This option naturally feels more democratic, as voters have larger expressive possibilities, and effectively have a greater impact on the final electoral outcome (Colomer 2011; Karvonen 2010). Facing growing political distrust, many constitution makers throughout the world have thus decided to reform electoral systems to introduce preference voting, or to increase the influence of preference-voting results on who is elected (Bedock 2017; Dalton 2004; Renwick and Pilet 2016). For example, in 2003 Belgium reformed its electoral system to give twice as much weight to preference votes relative to the list ordering.

There is a large literature on the effects of electoral systems on voting behavior and election results (for a compilation of the most influential papers on the topic, see Farrell and Shugart 2012). For a long time, most studies have focused on the inter-party consequences of electoral systems such as how they affect the number of elected parties (Duverger

1951; Lijphart 1994), or polarization (Cox 1990; Matakos et al. 2016). More recently, in parallel to the rapid proliferation of PR systems with preference voting throughout the world, scholars in the field have turned their attention to the intra-party consequences of election systems such as how they affect the type of candidates recruited by parties (André et al. 2014; Shugart et al., 2005), the nature of their work (André et al. 2015; Crisp et al., 2004), or the focus of the campaigns (Bowler and Farrell 2011; Sudulich and Trumm 2019). However, this strand of literature usually focuses on the effect of the very possibility for voters to cast a preference vote compared to closed-list PR systems, i.e., those in which voters can only vote for a party list, without entering into the details of the preference-voting system.

To the best of our knowledge, there are very few studies that focus on the precise method used to cast preference votes, such as the scale of points that voters can use to support candidates. This question is crucial, and not only for intellectual curiosity: although this is rarely discussed, democracies do use different scales. For example, in Belgium, voters can decide to approve as many candidates as they want; in Switzerland and Luxembourg, they can give up to two points; and in Latvia and in the Soviet Union (in the 1980s), they can give them a positive or negative point. This last scale is particularly interesting, as it could be one that naturally fits the evolution of public opinion in certain countries that is heading to an increase in skepticism towards politicians (Norris 2011) and in affective polarization between ideological camps (Boxell et al.,

* Corresponding author.

E-mail address: philipp.harfst@tu-dresden.de (P. Harfst).

2020). Therefore, given that different scales co-exist in reality and that they are under-researched, it is of utmost importance to evaluate the consequences of preference-vote designs for electoral behavior and results, as well as their respective advantages, as for constitution makers to design electoral institutions in a way that is informed by scientific evidence.

To achieve this goal, we conducted an original voting experiment embedded in a nationally representative online survey in Austria, a country that uses PR with preference voting. A few months after the 2017 national election, we reproduced the ballot of the election in terms of party list and candidates, and asked participants to vote as if they were in the ballot box again. The experiment thus relied on real-life political preferences, which ensures its realism. We are interested in rules where voters evaluate each candidate on their list independently. The treatment randomly assigned to participants was the scale used to evaluate the candidates within the chosen party list.¹ We considered three such scales:

- The one with possibility to approve or not each of the candidate [0,1], that we label Approval Voting AV [0,1]²
- The one with the possibility to give negative on top of positive points [-1,0,1], that we label Evaluative Voting EV [-1,0,1];
- The one with the possibility to give up to two points [0,1,2], that we label Evaluative Voting EV [0,1,2].

With the data generated by this experiment, we can assess how these three preference-voting scales affect the fate of candidates. Although the scales can be considered as mere linear transformations of each other (although there are two categories in AV [0,1] and three in the two others), we find that some candidates receive more preference votes under some scales. This suggests that the numbers associated to categories matter in the eyes of voters and can distort preference votes. We investigate the sources of these differences in two ways. First, we examine whether some of the 120 candidates included in the experiment are (dis)advantaged by some scales depending on their personal characteristics appearing on the ballot or not (gender, age, and academic title). We show that male candidates are better off under the one with the possibility to give two points because some voters seem to be reluctant to give the maximum number of points to female candidates.

Second, we analyze whether these effects are driven by certain subsets of the electorate. We follow the literature on affinity voting and

¹ In our experiment, we gave the possibility to subjects to give a preference vote to each of the candidate of the party list. Note that in reality in most preference-voting systems, the number preference vote is limited to one (e.g., Netherlands or Brazil). We decided not to put any limit for a pragmatic reason related to statistical power. Letting subjects cast a single vote would have given us only limited information about their preference since we were not in the position to decide on the number of subjects (the experiment was embedded in a survey conducted by a separate organization, see below). Yet, we believe our experiment can still inform the reality of elections held systems in which preference voting is limited in the sense that we can think our experimental design as a series of elections in which the subjects choose to support or not each of the candidates appearing on the list. Moreover, a recent experiment with Dutch subjects shows that most voters do not substantially alter their vote when they are granted with the possibility to cast multiple preference votes (Nagtzaam and van Erkel (2017)).

² Strictly speaking, approval voting is an electoral system in which voters decide to approve or not each of the competing candidates (not only those from their favorite party list), and in which a single candidate, the one with most approvals, is elected. Approval voting has attracted a lot of scholarly attention because it is often considered as producing outcomes closer to voters' preferences, although it is almost never used in reality (Laslier and Sanver 2010). The electoral system used in Austria (and many other countries using PR with preference voting) can be seen as a derivative of approval voting in the sense that voters decides to approve one candidate within their favorite party list, and those with the most approvals are elected.

hypothesize that voters are more likely to favor candidates with whom they share characteristics like gender (Sanbonmatsu 2002) or other socio-demographics (Cutler 2002). We find some evidence for affinity voting, and more importantly for the purpose of this paper, some evidence for interaction between affinity voting and the scale. For example, we find that evaluate voting scales gave advantages to male candidates. The reason lies in affinity voting, and the propensity of male voters to give the minimum number of votes to female candidates and the maximum number to male candidates. Yet this effect is weaker under the evaluative voting system in which voters can give negative points. Male voters seemed reluctant to give 'minus one' to female candidates, which attenuates their disadvantage. Taken together, our findings have important implications for constitution makers: the details in the design of preference voting can give an unfair advantage to candidates based on their gender, or other personal characteristics.

2. Preference voting in the literature

PR systems in which voters can cast preference votes to individual candidates within party lists are increasingly common in democracies throughout the world (Renwick and Pilet 2016). These preference-voting systems naturally feel more democratic than closed-list systems because voters have more influence over the electoral results and the identity of the candidates seating in parliament (Colomer 2011; Karvonen 2010). People's self-reported satisfaction level, as well as actual voter turnout, is substantially higher in countries using these systems (Farrell and McAllister 2006; Sanz 2017). Also, when they are asked which system they like the most, about two thirds of individuals say that they prefer preference voting to a closed-list system (Laslier et al., 2015). A reason is probably that preference-voting results are quite influential, even beyond the election of parliamentary representatives. In many countries, a candidate who receives a high preference score is also likely to become minister or be appointed to a leading position within the party (André et al., 2017; Bol et al., 2018; Folke et al. 2016; Meriläinen and Tukiainen 2018). Preference voting can thus serve to hold individual politicians' accountable (Folke and Rickne 2020).

However, preference-voting systems can also have undesired effects. Since voters can vote for individual candidates and the election of these candidates depends at least partially on these votes, there is some level of intra-party electoral competition (Carey and Shugart 1995, for a counter argument see Cheibub and Sin 2020). Candidates, even those from the same party, are effectively in competition with one another for the popular votes. This means that they have incentives to engage into corruption and patronage to differentiate themselves from other co-partisan candidates or to finance their individual electoral campaigns (Chang and Golden, 2007; Ames 1995). Studying the functioning of preference voting is thus of tremendous importance.

How do individuals decide for whom to cast a preference vote? Vote choice in general is a function of partisan preferences, in the sense that individuals tend to vote for the candidates of the party they like the most, to which they feel the closest ideologically, or the one that they believe is best for some important issues like the economy (Campbell et al., 1980). Partisan preferences are deeply interiorized by individuals, so that it is impermeable to external stimuli and remains stable throughout a lifetime (Bartels 2002; Schickler and Green 1997).

Yet with preference voting, all the candidates for whom one can potentially vote are from the same party.³ In a context of limited partisanship, voters tend to rely on the personal characteristics of

³ We do not consider here the possibility of panachage. This possibility is effectively rarely used (but examples include Luxembourg and Switzerland) and raises further questions that are out of the scope of this paper.

candidates to make up their mind (McDermott 1997).⁴ These include their gender (Golder et al., 2017; McElroy and Marsh 2010), physical appearance (Berggren et al. 2017; Rosar et al. 2008), or geographical roots (Bol et al., 2016; Jankowski 2016). Evidence suggests the existence of a phenomenon called ‘affinity voting’, in which voters favor candidates with whom they share personal characteristics, supposing that these candidates would best serve their interests. These characteristics include gender (Dolan 2008; Sanbonmatsu 2002), ethnic background (McDermott 1998), socio-demographics at large (Cutler 2002), and pure facial resemblance (Bailenson et al., 2008). In turn, these shortcuts may or may not serve as effective heuristics in the perspective of choosing the candidates that is best for them (Lupia 1994; Popkin 1991).

Nevertheless, there is very little research on the effect of the method or scale used for preference voting in PR systems on voting behavior, and particularly how it interacts with the personal characteristics of the candidates. A reason could be that cross-national observational studies on the topic are made very difficult due to the variety of systems used in reality. Most countries seem to have a unique preference-voting method. For example, in local elections in Italy, voters can cast up to two votes, but only if they vote for candidates of different genders. This is a very specific system that is not used anywhere else in the world (Baltruonaite et al., 2019). In Belgium and the Netherlands, two countries that share a remarkably similar political system, the preference-voting method is substantially different. In Belgium, voters can approve any of the competing candidates, but can also decide to approve the entire party list (which is different from approving each candidate of the list). In the Netherlands, this last possibility does not exist, and voters can only approve one single candidate in the entire party list (Nagtzaam and van Erkel 2017). Because most countries differ on multiple dimensions of the preference-voting method, cross-national observational studies can hardly disentangle the effects of each of these dimensions.

To our knowledge, the very few studies on the topic that use an experimental design ask individuals to vote for real candidates or parties under various methods or scales, and then evaluate how electoral behavior and results change with these methods/scales. With this design, they leverage real-life political preferences, which increases the external validity of the results. Note however that none of them did so in the context of PR with preference voting. Baujard et al. (2017) for example asked a sample of French voters to evaluate the candidates to the 2012 presidential election using various scales including those covered in the present study such as the one in which voters can give a negative point. They show that although these scales affect the way subjects evaluate some of the candidates, the aggregate result remains unchanged. This is because (1) there is only one candidate elected in a presidential election, and (2) in this highly salient election subjects have strong preferences in favor or against a few candidates, and these preferences are not altered by differences in preference scales. Darmann, Grundner, and Klamler (2017) conducted a similar voting experiment in the state of Styria in Austria in 2015. They asked subjects to evaluate the main Austrian political parties using several preferential-voting scales. They also find that, although some scales advantage some of the parties, they do not change the aggregate result.⁵

The voting experiment conducted for the present paper builds on the

⁴ Note that a context in which there is a particularly salient issue that divides parties, in the sense that some candidates are in favor of it whereas some are not, like Brexit in the United Kingdom, voters can use preference votes to favor the candidates who are on the same side of the issue than them (Blumenau et al., 2017). Yet, in more normal contexts, defecting from the official party line does not seem to increase the probability for a candidate to receive more preference votes (Folke and Rickne 2020).

⁵ Other similar voting experiments focusing on the comparison between one preference-voting method and another one like a closed-list system include Alòs-Ferrer and Granić (2012); Baujard et al. (2014); Bol et al. (2016); Blumenau et al. (2017); Nagtzaam and van Erkel (2017).

experiments presented above but brings three key contributions to the study of preference voting. First, the context is truly one of PR with preference voting. We asked subjects to vote for individual candidates within a party list using various preference-voting scales. This is important because just like in real life, the candidates among whom they had to choose shared the same partisan affiliation, which forced them to rely on other heuristics than partisanship such as candidates’ personal characteristics. It is reasonable to think that these preferences are more likely to be affected by differences in scales than deeply interiorized partisan preferences. Also, PR with preference voting implies that several candidates are elected to form the parliament. The scales are thus more likely to make differences in the final electoral result, not in terms of party composition but in terms of who is seating. Even if they do not change whether the most popular candidates are elected, they can change the fate of less-known candidates.

Second, by virtue of the design, the experiment includes 120 real candidates. Contrary to other experimental studies that only analyze attitudes towards a few candidates, we are able to evaluate patterns in votes depending candidates’ personal characteristics like their age, gender, and academic title. Third, our voting experiment is conducted on a representative sample of the voting-age population in Austria, instead of an *ad hoc* or self-selected sample of subjects. This is important because it means that our treatment effects can be generalized beyond the group of individuals willing to volunteer to participate in these types of studies, who are probably those who are the most interested in politics or scientific experimentation.

3. Experimental design

We conducted a voting experiment with a nationally representative sample of Austrian adults in August 2018. It was embedded in a web-based survey organized by *Statistik Austria*, and part of the PUMA scheme.⁶ For the experiment, we created a ballot similar to the one of the legislative election that took place in October 2017.⁷ In this election, Austria uses a PR system with preference voting: voters first choose one party, and can then approve one of the candidates listed in the corresponding party list. Parties receive a number of seats that is proportional to their vote share, and the candidates with most preference votes receive these seats.⁸ Thus, in our experiment, we first asked subjects to vote for one of the main Austrian parties. We included the eight parties with more than 0.5% of the votes in the 2017 election (ÖVP, SPÖ, FPÖ, NEOS, PILZ, GRÜNE, GILT and KPÖ). In a second stage, we showed on the screen the first 15 candidates who appeared on the 2017 national-constituency list of the chosen party, and asked subjects to cast a preference vote for each of them. Note that the total number of candidates included in the experiment is 120 (15 candidates x 8 parties), but each

⁶ See <https://www.puma-plattform.at/en/home/> for details about the scheme. A detailed description of the experiment that follows APSA’s standards is provided in A1.

⁷ Note that the online format of the experiment has a clear advantage here: we were able to reproduce a ballot similar to a real-election ballot on the subjects’ screen, while randomly manipulating some of the aspects of this ballot. This would not have been possible with a telephone survey for instance.

⁸ Like in most PR with preference voting systems, the election of candidates does not only depend on the results of the preference vote, it also depends on the list ordering of the candidates as decided by the party. Yet, from the perspective of a voter, the existence of list ordering does not change that a preference vote increases the probability of a candidate to be elected. Note also that a specificity of the system used in legislative elections in Austria is that voters can express a preference vote in three party lists, one for each constituency level (national, state, and sub-state). For practical reasons, we focus in the experiment on the national constituency.

subject only saw the 15 candidates of the party they picked in the first stage.⁹ In order to account for any list-ordering effect, we randomized the position of the candidates on the list for each subject.

The experimental treatment is the preference-voting method. One third of the subjects (randomly selected) had to decide whether to approve each of the candidates.¹⁰ In other words, they could give them either 0 or 1 point. This scale is common in PR system with preference voting. For example, it is the one used in Belgium and Brazil. We call it AV [0,1]. Another third of the respondents (randomly selected) could give 0, 1, or 2 points to the candidates. This is a scale that is used in Switzerland and Luxembourg for example. We call it EV [0,1,2]. Finally, the last third of the respondents could give a positive, a negative, or no point to the candidates. This is a preference-voting method used in Latvia, and in the 1980s in the Soviet Union. We call it EV [-1,0,1]. In each treatment, voters could only proceed to the next page when they manually filled in evaluations for each of the party's 15 candidates. In order to ensure the simplicity of the experiment and thus subjects' understanding, we decided not to give a detailed explanation of how preference votes would be counted to elect representatives in these three scales. We only informed voters that the total sum of points decides on the election of individual candidates. We can reasonably expect that Austrian voters are familiar with the concept of preference voting, and that they thought that just like in the 2017 election, the more points they gave to a candidate, the more likely this candidate would have been elected. A2 shows the results of a balance test. The assignment of treatments is neither substantially nor statistically related to observable characteristics of experimental subjects (age, gender, education, political knowledge, interest in politics, left-right self-placement, and party choice).

The sample is composed of 617 eligible Austrian voters that are representative of the adult population of the country in terms of age, gender, and education.¹¹ Among them, 456 declared having voted in the 2017 election, are thus included in the analysis. A3 shows the descriptive statistics, and A4 a comparison with census data from Austria on key socio-demographics. The differences between the sample and the population of interest are small, although we find that young and highly-educated voters are slightly over-represented. A5 reports the distribution of parties' vote share in the sample and in the 2017 election. We observe that the largest difference between the two is for the ÖVP (-10 points) and the FPÖ (4 points). The scores of other parties are fairly close to the 2017's benchmark. The discrepancies we observe are probably due to the timing of the survey that was conducted eight months after the election. It is reasonable to think that some voters changed their partisan preferences after the formation of a politically controversial ÖVP-FPÖ coalition and/or some of the decisions that this government made in the meantime. Given that the paper focuses on preference votes for candidates, and not on party votes (actually the treatment was assigned after subjects selected a party), it is not useful to weight the

⁹ The key advantage of relying on real instead of hypothetical candidates is that we leverage real political preferences, which increases the external validity of our results. The disadvantage is that we cannot experimentally manipulate the personal characteristics of the candidates and disentangle the effect of each of them. To address this issue, we include all of the observable personal characteristics of the candidates in the same regression, as to estimate the unique effect of each of them.

¹⁰ Note that unlike other voting experiments of the same kind, each subject had to vote according to one of the three scales, instead of all of them one after the other (Baujard et al., 2017; Darmann et al. 2017). We believe our design is superior as it shuts down possible contamination effects between treatments. Asking subjects to vote according to different methods one after the other could create biases in results, for example because some could be willing to report different votes with different methods to please the experimenter (or the way round, trying to be consistent across methods).

¹¹ The sample is only composed of 617 subjects because *Statistik Austria* only sent the voting experiment to a random subsample of the overall survey sample.

data according to the 2017 election. It would bias the results without increasing the external validity.

4. Results

In this section, we present the results of the voting experiment in three steps. First, we describe how experimental subjects used the three scales to cast their preference vote in order to evaluate whether voters use them as if they were mere linear transformations of each other or not. Second, we check how candidates' personal characteristics affect the number of votes received on each of the three scales, as well as their propensity to receive the minimal and maximum number of points available. To do so, we focus on three characteristics, i.e. gender, age and academic title (BA, Master, PhD, or Professor), that are among the most visible ones, as they either directly appear on the ballot (academic title, age¹²), or can be easily inferred from names on the ballot paper (gender). A6 reports the descriptive statistics of the 120 candidates included in the experiment on these three characteristics. Note that we cannot assess the effect of (perceived) ethnic background since only seven of the 120 candidates have a name that suggests a non-Austrian one. Finally, in the third and last step, we explore whether the differences observed in the second step are driven by some groups of the electorate. In particular, we follow the literature on affinity voting, and check whether voters are favoring the candidates with whom they share some characteristics, and how this interacts with the preference-voting scale. In particular, we check (1) whether male (female) voters are likely to give more points to male (female) candidates, (2) whether old (young) voters are likely to give more points to old (young) candidates, and (3) whether voters with (without) a university degree are likely to give more points to candidates with (without) an academic title.

4.1. How voters used the preference-voting scales

Table 1 presents the proportion of times subjects used each point of the scales, as well as the average number of points received by candidates. As the scales are different, we can expect to find large discrepancies. Yet, it shows that the use of the 'one point' and 'zero point' options is relatively similar across experimental conditions. First, voters overwhelmingly used these two options even in experimental conditions in which they could use others (more than 80% of the time). Second, about the same proportion of voters gave one point in AV [0,1] and EV [0,1,2] (33 and 30% respectively). Similarly, voters used 'zero points' at about the same rate in AV [0,1] and EV [-1,0,1] (67% and 65%

Table 1
Use of the scales.

	AV [0,1]	EV [0,1,2]	EV [-1,0,1]
Points			
-1	.	.	9.54
0	67.11	53.55	65.30
1	32.89	29.72	25.17
2	.	16.73	.
Mean	0.33	0.63	0.15
Std. Dev.	(0.47)	(0.75)	(0.57)
N	2280	2295	2265

Note: Entries are percentages or averages. Standard deviations are in parentheses. The N corresponds to the number of subjects in each experimental condition times the number candidates appearing on the party list (=15).

¹² The original and experimental ballot papers included candidates' year of birth.

Table 2
Number of points received by candidates (bivariate analysis).

	AV [0,1]	EV [0,1,2]	EV [-1,0,1]
Female – Male	.29–.36 = –.07 (.00)	.55–.70 = –.15 (.00)	.10–.20 = –.10 (.00)
Old – Other	.26–.34 = –.08 (.00)	.61–.64 = –.03 (.52)	.12–.17 = –.05 (.07)
Young – Other	.30–.33 = –.03 (.32)	.66–.63 = .03 (.45)	.18–.15 = .03 (.44)
Academic title – No academic title	.32–.33 = –.01 (.65)	.62–.65 = –.03 (.31)	.13–.19 = –.06 (.02)

Note: Entries are the average number of points received by candidates with a certain characteristic (example: female) compared to candidates with another characteristic (example: male). P-values of difference in means t-tests are in parentheses (2-tailed), $N = 2280$ (AV[0,1]), 2295 (EV[0,1,2]), and 2265 (EV [-1,0,1]). Old candidates are those older than 55 years old (3rd quartile), young candidates are those younger than 33.5 years old (1st quartile).

respectively).

The other points on the scales reveal largest discrepancies. First, few voters used the extreme category in EV [0,1,2] and even fewer did so in EV [-1,0,1]. 17% of candidates received ‘two points’ in the former, and only 10% received ‘minus one point’ in the latter. Yet, we believe that the number of negative evaluations is substantial given that we are dealing with preferences for candidates on the party list initially chosen by the voter. Second, although voters seem to use the extreme category at the expense of the ‘zero point’ one in EV [0,1,2] (only 54%), they seem to use it at the expense of the ‘one point’ one in EV [-1,0,1] (only 25%). In particular the comparison between EV [-1,0,1] and EV [0,1,2] suggests that voters did not simply use the scales as mere linear transformations of each other. There seems to be some psychological bias depending on the number that is associated to the category.

Finally, it is important to note the presence of these extreme categories is naturally reflected in the average of number of points received by candidates. The average is the largest in EV [0,1,2] (0.63, $sd = 0.75$) and the smallest in EV [-1,0,1] (0.15, $sd = 0.57$). It is important to keep these averages in mind while we compare the effect of personal characteristics on the number of points received by candidates. The effects can only be compared across experimental conditions with regard to the averages and standard deviations.

4.2. Which candidates were advantaged by which scales?

To answer this question, we group candidates by three personal characteristics: gender, age and academic title. We measure gender based on the first name of the candidate, and academic title based on the information put on the ballot. In Austria, candidates add their academic title on the ballot when they have a bachelor’s, master’s or doctoral degree or a professorial title. We measure age based on the date of birth of candidate. Age is a continuous variable but in order to make it comparable to the other two, we split it up in three groups: those who are younger than the 1st quartile of the age distribution of candidates (33.5 years old), those who are older than the 3rd quartile (55 years old), and the rest.

Table 2 presents some bivariate analysis: the average number of points received by each group of candidates by experimental condition. It shows some interesting patterns regarding the popularity of some candidates, as well as how this interacts with the preference-voting scale. First, female candidates received less points than their male counterparts under all three scales. The difference is the largest for EV [0,1,2] and EV [-1,0,1], -0.15 and -0.10 respectively, which are the equivalent of around a fifth of the standard deviation of the distribution of points received by the candidates under these two scales (see Table 1). The difference is smaller for AV [0,1], but it is still -0.07 , i.e. 15% of the

standard deviation.

Second, old candidates received slightly less points than other candidates under both AV [0,1] and EV [-1,0,1], -0.8 and -0.05 . Respectively, which are 17% and 9% of the standard deviation. By contrast, young candidates did not receive less or more points than other candidates. Similarly, the candidate with an academic title only receive slightly less points than candidates without under EV [-1,0,1], -0.06 (11% of the standard deviation).

The result of the bivariate analysis needs to be considered with caution. Because we rely on real candidates, we did not randomize their personal characteristics. There might thus be some underlying patterns that are confounding the results. For example, male candidates might have received more points than female candidates because they are higher in the parties’ internal hierarchy. To have a better idea of the unique effect of each characteristic, we conduct a multivariate analysis, for which we estimate an OLS regression predicting the number of points received by a candidates based on their personal characteristics (age, gender, academic title), the experimental condition, and an interaction between the two. Because the number of points is not perfectly comparable across experimental conditions, we reproduce the analysis with two alternative dependent variables: receiving the maximum number of points on the scale (0–1) and receiving the minimum number of points on the scale (0–1). For these two variables, we use a logit regression. In all three models, we cluster standard errors by candidates and subjects.

We also add some control variables. At the level of the candidate, we control for the incumbency status (member of the national government or parliament or not), and the original position on the party-list in the 2017 Austrian election (as a measure of their popularity).¹³ For each of these variables, we also add an interaction with the treatment. At the level of the voter, we control for age, gender, education, party choice,¹⁴ self-declared political interest, self-declared left-right positioning, and political knowledge (based on the number of correct answers to seven quiz questions about politics in Austria). The descriptive statistics of all these variables can be found in A3 and A6.

Figs. 1–3 present the quantities of interest, whereas the full regression results are in A7.¹⁵ Fig. 1 confirms the results of the bivariate analysis presented above: (1) female candidates received less points than their male counterpart, and this difference is particularly large in EV [0,1,2] and EV [-1,0,1], and (2) old candidates and those with an academic title received less points than others under EV [-1,0,1]. What is interesting is that these effects hold even after controlling for other characteristics of candidates.¹⁶

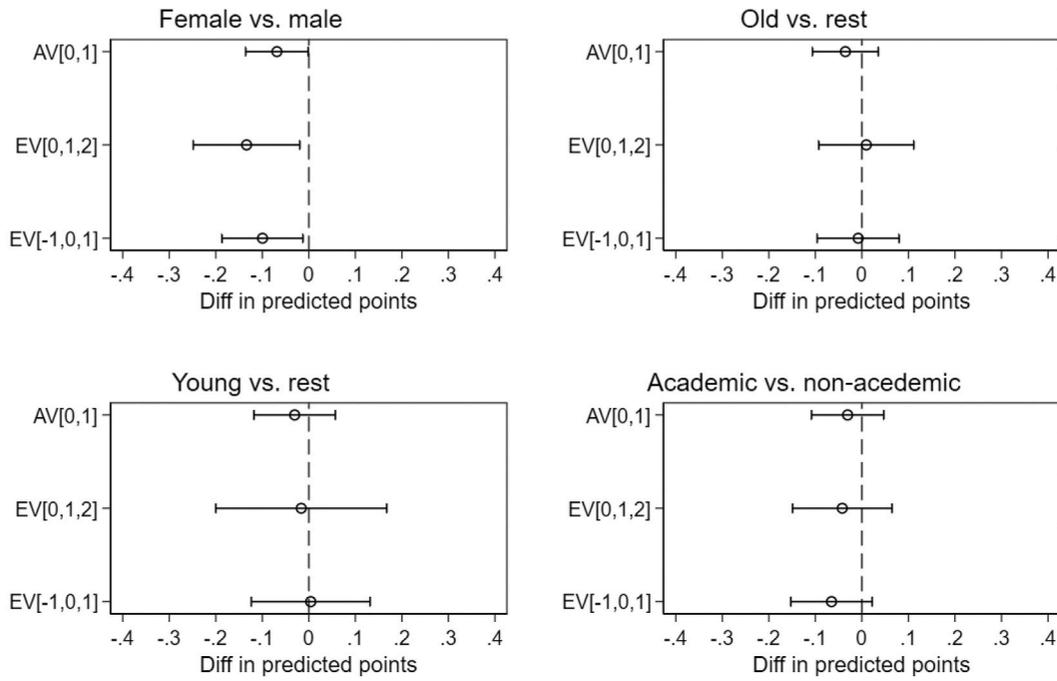
Figs. 2 and 3 presents the results from a different angle. They show that female candidates were less likely to receive the maximum number of points (‘two points’ or ‘one point’) than their male counterparts under both EV [0,1,2] and EV [-1,0,1], by 6 and 8% points respectively. Yet, they were only more likely to receive the minimum number of points under EV [0,1,2] (by 6% points). Under EV [-1,0,1], they were as likely

¹³ We acknowledge is not optimal as the effect of the rank on the candidate’s probability of receiving a preference vote is not linear (Devroey and Wauters 2020). Yet we do not have enough candidates to include a set of dummy variables capturing each rank separately (#1, #2, #3 ...).

¹⁴ Note that the variable party choice can be seen as a candidate-level as well as a voter-level variable because each candidate belongs to only one party, while each voter only votes for one party.

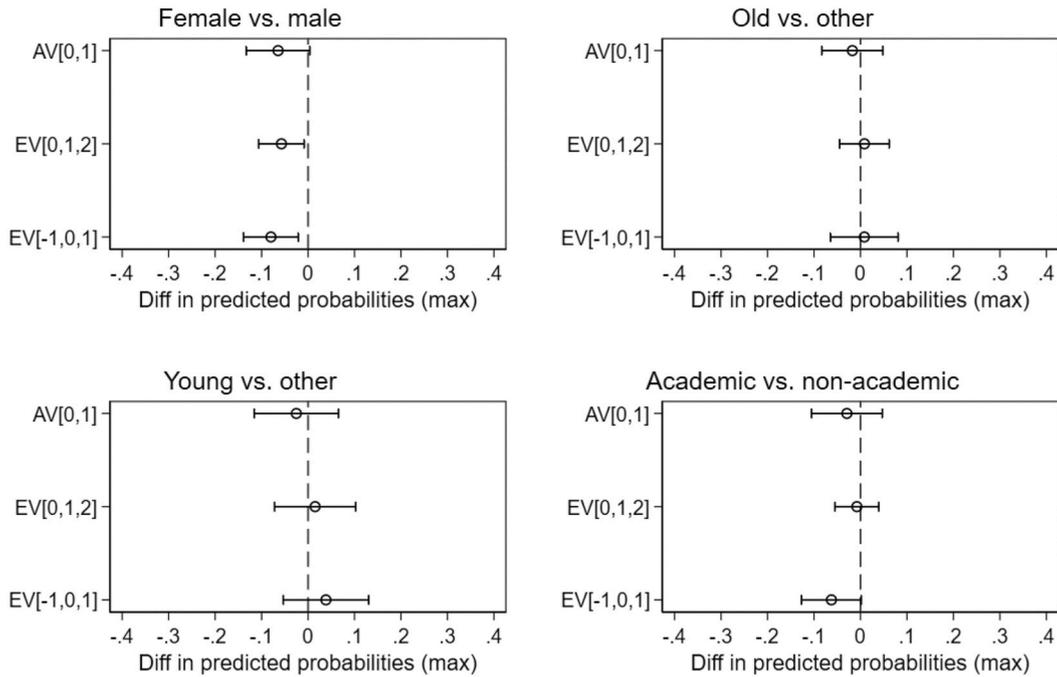
¹⁵ A8 presents the results of the same regressions in removing the candidates who can be considered as Spitzencandidates in the 2017 election (that is those at the top of the list). A9a and A9b replicate it using standard errors clustered by subjects (instead of subjects and candidates). In both instances, results are similar.

¹⁶ Note that the full regression results of A7 reveal that the differences between scales are not statistically significant (see the interactions between the experimental conditions and the personal characteristics of candidates). Although small, we believe the differences between scales are far from being trivial given the design of the study.



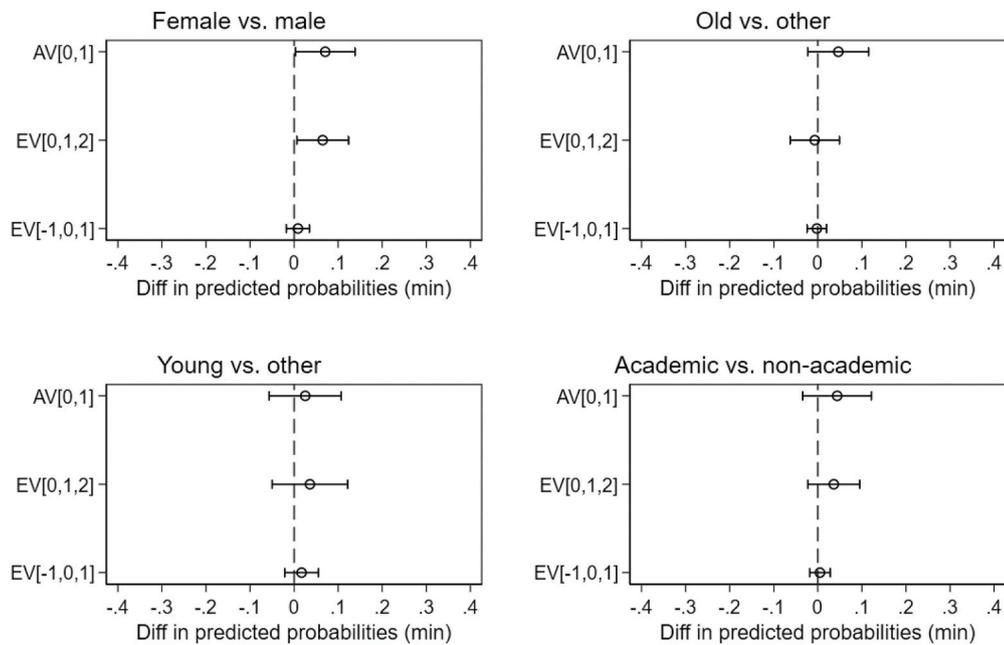
Note: Entries are differences in predicted number of points received between candidates with a certain characteristic (example: female) and candidates with another characteristic (example: male), net of other characteristics. The results are derived from the OLS regression presented in A7. Horizontal bars are 95%-confidence intervals.

Fig. 1. Number of points received by candidates (multivariate analysis).



Note: Entries are differences in predicted probabilities of receiving the maximum number of points between candidates with a certain characteristic (example: female) and candidates with another characteristic (example: male), net of other characteristics. The results are derived from the logit regression presented in A7. Horizontal bars are 95%-confidence intervals.

Fig. 2. Propensity to receive the maximum number of points (multivariate analysis).

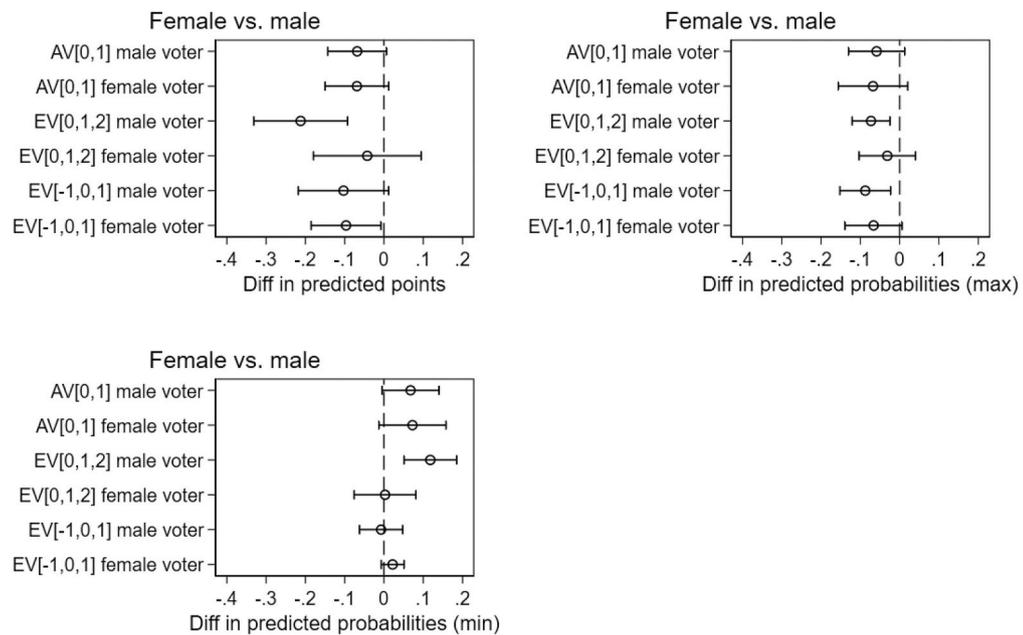


Note: Entries are differences in predicted probabilities of receiving the minimum number of points between candidates with a certain characteristic (example: female) and candidates with another characteristic (example: male), net of other characteristics. The results are derived from the logit regression presented in A7. Horizontal bars are 95%-confidence intervals.

Fig. 3. Propensity to receive the minimum number of points (multivariate analysis).

as male candidates to receive ‘minus one point’. Similarly, old candidates and those with an academic title were less likely to receive the maximum number of points under this scale, but as likely as others to receive the minimal number of points. Estimates point that the

preference-voting scale EV [-1,0,1] leads to some interesting patterns: voters seemed to be reluctant to use the ‘minus one point’ on this scale, which could lead to less discrimination, especially against female candidates. This suggests that the negative charge associated to the category



Note: Entries are differences in predicted number of points/probabilities of giving the max/min number of points on the scale between male and female candidates, net of other characteristics. The results are derived from the OLS/logit regressions presented in A10. Horizontal bars are 95%-confidence intervals.

Fig. 4. Affinity voting between male/female voters and male/female candidates.

'minus one point' affects patterns of preference votes under this scale. By contrast, the difference in propensity to give the maximum and minimum number of points between male and female candidates is relatively moderate under AV [0,1].

4.3. Which groups in the electorate drive the differences between scales?

As a final step, we replicate the OLS and logit regressions presented above, but in adding triple interactions between candidates' characteristics, experimental conditions, and voters' characteristics. In order to keep the analysis as simple as possible, we restrict ourselves to voters' characteristics that are similar to candidates' characteristics. In doing so, we effectively test the 'affinity voting' hypothesis. To be clear, we reproduce the OLS and logit regressions presented above, in adding a triple interaction with one voter characteristic at a time. Figs. 4–6 show the quantities of interest, whereas the full results are presented in Tables A10 to A12.

Fig. 4 presents the results when we add the triple interaction between candidates' characteristics and the gender of voters. It provides evidence for affinity voting in the behavior of male voters, but it also shows that the pattern is stronger under some preference-voting scales. Male voters gave less points to female candidates compared to their male party fellows under all three scales, but especially under EV [0,1,2]. Although the probability to give the maximum number of points is similar across all three scales (male voters being more likely to give the maximum number of points to male candidates and the minimum number of points to female candidates), there are differences when it comes to the probability to give the minimum number of points: male voters seemed reluctant to give 'minus one point' to female candidates under EV [-1,0,1], but did not seem to have any problem giving them 'zero points' under AV [0,1] and, particularly, EV [0,1,2].

Fig. 5 presents the results when we add a triple interaction with the age of voters. In general, we also find evidence for affinity voting, in the sense that the oldest voters gave more points to old candidates than youngest voters (and vice versa).¹⁷ Yet, we also find differences between scales, with affinity voting being more pronounced under EV [0,1,2]. Again, this effect is driven by the probability to give the minimum number of points: young (old) voters seemed reluctant to give the minimum number of points to old (young) voters under EV [-1,0,1] (and to some extent under AV [0,1], but not under EV [0,1,2]).

Finally, Fig. 6 presents the results when we re-estimate the regression with a triple interaction including voters' education. Again, we find evidence for affinity voting. Voters who have a university degree gave more points to candidates with an academic title (and vice versa for voters without any school degree), especially under EV [0,1,2], which once more seems to be driven by the reluctance of voters to give the minimum number of points under both EV [-1,0,1] and AV [0,1]. Yet, the results are less clear cut than for age and gender.

5. Conclusion

Many countries that use a PR system also use a system of preference voting, in which voters can vote for individual candidates within their favorite party list. Although there is a vast literature on the advantages and drawbacks for the functioning of elections and democracy of preference voting compared to closed list PR systems, we know very little about the effects of the type of preference voting. Several types indeed co-exist in democracies throughout the world, and yet, this institutional specificity has been overlooked in the literature.

¹⁷ In the regression, we use a continuous measure of age. Yet, in Fig. 5, we take the empirical minimum and maximum in the experimental sample to define oldest and youngest voters. Oldest voters are 74 years old (note that there was a single voter with 110 years, but it does not seem realistic), and youngest are 16. In Austria voting age is indeed 16 years old.

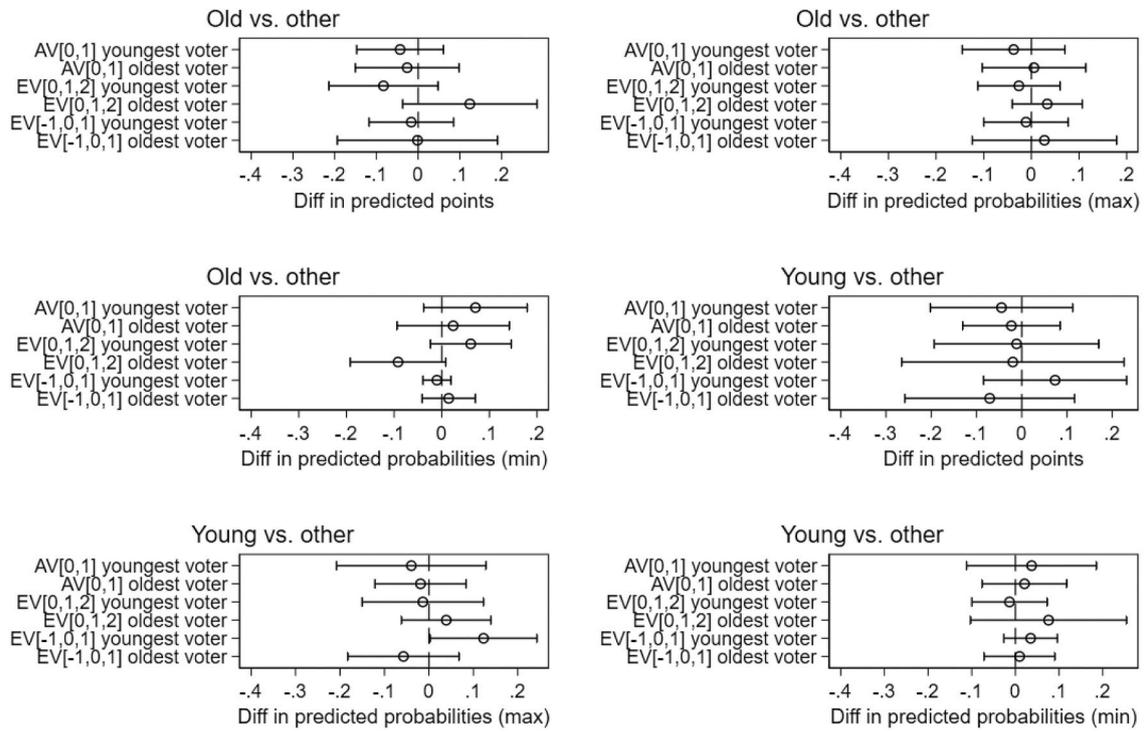
To fill this gap, we conducted a web-based voting experiment with a nationally representative sample from Austria, a country that uses preference voting. We made the subjects cast a preference vote in recreating the ballot of the 2017 national election that occurred just a few months before the experiment. The originality is that we randomized the scale under which they could cast a preference vote: either approval voting (approving candidates with one point or not) or evaluative voting (EV [0,1,2] with the possibility to give up to two points to candidates; or EV [-1,0,1] with the possibility to give negative as well as positive points). On the ballot, we showed them real candidates that genuinely competed in the 2017 election. We thus leveraged real political preferences to increase the realism of the study. In total, 120 candidates were included in the experiment, which allows us to test how personal characteristics (age, gender, and academic title) affect the number of points they received, and how this interacts with the scale under which the subjects casted their vote.

Our findings confirm our intuition that the scale of preference voting matters for electoral results. We find that some groups of candidates were advantaged by some scales. Female candidates, but also to some extent old candidates and those with an academic title, were hit by evaluative voting systems. These effects were (partly) attenuated in the system in which voters can give positive and negative points. Voters seemed to be reluctant to give 'minus one point' to candidates they only dislike because of socio-demographic characteristics. Actively removing candidates from the list appears to represent a higher hurdle than just not supporting someone.

We believe that the effects are informative of the reality of elections in various contexts. First, although we conducted the experiment in a single country, i.e. Austria, we did it with a representative sample of the population, instead of a convenient sample of say students. Second, Austria is an interesting case-study to examine preference voting: the country uses such a system in real-life, which means that citizens are used to cast preference votes. Then, we can reasonably think that the behavior of subjects is more meaningful than it would have been if the experiment would have been conducted in a country that does not use preference voting. In such a country, the behavior of subjects would have been mostly hypothetical. Yet, we cannot discard the possibility that there is something specific about Austria and its political culture that would harm the generalizability of the findings to other countries. Third, in the experiment, we leveraged real-life political preferences in using actual candidates in the ballot instead of fictitious candidates unknown from the subjects. This is yet another important element that increases the realism of the design. Third, although we let subjects vote for as many candidates as they wanted, which is not a very common preference-voting system (for an exception, see Belgium and Latvia), which could have increased subjects' inclination to base their preference vote on easily accessible characteristics of the candidates like gender, age, and education level, it is important to note that we employed scales that are used in real-life. For these reasons, we do believe that our results are meaningful and informative. Yet, we acknowledge that the extent to which the differences between scales can affect final electoral results depends on other contextual factors like how much the election of individual candidates also hinges on their rank on the party list or solely on the number of preference votes that they obtained. Overall, we believe our paper opens new avenues for research on the type of preference voting and how it can affect elections and democracy.

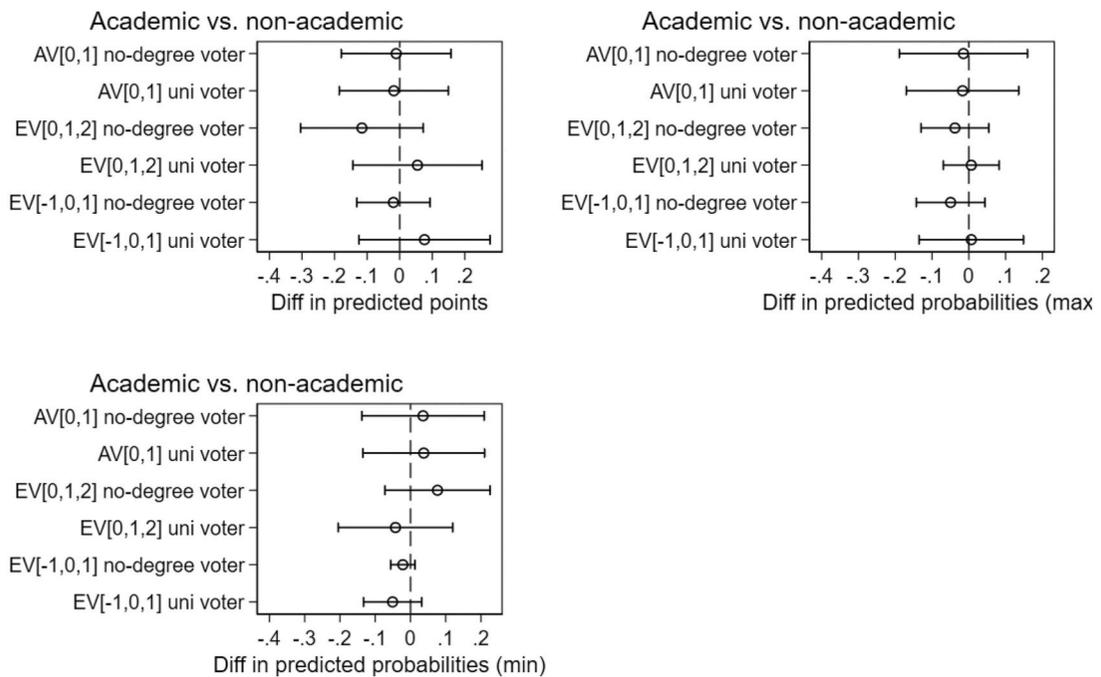
Funding

Data has been collected in the context of a survey administered by PUMA – Platform für Umfragen, Methoden und empirische Analysen (Austria). J.-F. Laslier thanks the support of an Agence Nationale de la Recherche (France) EUR grant (ANR-17-EURE-0001).



Note: Entries are differences in predicted number of points/probabilities of giving the max/min number of points on the scale between old and young candidates, net of other characteristics. The results are derived from the OLS/logit regressions presented in A11. Horizontal bars are 95%-confidence intervals.

Fig. 5. Affinity voting between young/old voters and old/young candidates.



Note: Entries are differences in predicted number of points/probabilities of giving the max/min number of points on the scale between candidate with and without an academic title, net of other characteristics. The results are derived from the OLS/logit regressions presented in A12. Horizontal bars are 95%-confidence intervals.

Fig. 6. Affinity voting between educated/non-educated voters and candidates with/without academic title.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.electstud.2020.102262>.

References

- Alòs-Ferrer, C., Granić, D.-G., 2012. Two field experiments on approval voting in Germany. *Soc. Choice Welfare* 39 (1), 171–205.
- Ames, B., 1995. Electoral strategy under open-list proportional representation. *Am. J. Polit. Sci.* 39 (2), 406–433.
- André, A., Depauw, S., Deschouwer, K., 2014. Legislators' local roots: disentangling the effect of district magnitude. *Party Polit.* 20 (6), 904–917.
- André, A., Depauw, S., Martin, S., 2015. Electoral systems and legislators' constituency effort: the mediating effect of electoral vulnerability. *Comp. Polit. Stud.* 48 (4), 464–496.
- André, A., Depauw, S., Shugart, M.S., Chytilík, R., 2017. Party nomination strategies in flexible-list systems: do preference votes matter? *Party Polit.* 23 (5), 589–600.
- Bailenson, J.N., Iyengar, S., Yee, N., Collins, N.A., 2008. Facial similarity between voters and candidates causes influence. *Publ. Opin. Q.* 72 (5), 935–961.
- Bartels, L., 2002. Beyond the running tally: partisan bias in political perceptions. *Polit. Behav.* 24 (2), 117–150.
- Baltrunaite, A., Casarico, A., Profeta, P., Savio, G., 2019. Let the voters choose women. *J. Publ. Econ.* 180, 104085.
- Baujard, A., Gavrel, F., Igersheim, H., Laslier, J.-F., Lebon, I., 2017. How voters use grade scales in evaluative voting. *Eur. J. Polit. Econ.* 55, 14–28.
- Baujard, A., Igersheim, H., Lebon, I., Gavrel, F., Laslier, J.-F., 2014. Who's favored by evaluative voting? An experiment conducted during the 2012 French presidential election. *Elect. Stud.* 34, 131–145.
- Bedock, C., 2017. *Reforming Democracy: Institutional Engineering in Western Europe*. Oxford University Press.
- Berggren, N., Jordahl, H., Poutvaara, P., 2017. The right look: conservative politicians look better and voters reward it. *J. Publ. Econ.* 146, 79–86.
- Blumenau, J., Eggers, A.C., Hangartner, D., Hix, S., 2017. Open/closed list and party choice: experimental evidence from the UK. *Br. J. Polit. Sci.* 47 (4), 809–827.
- Bol, D., Blais, A., Gillard, X., Nunze, L., Pilet, J.-B., 2018. Voting and satisfaction with democracy in flexible-list PR. *Elect. Stud.* 56, 23–34.
- Bol, D., Harfst, P., Blais, A., Golder, S., Laslier, J.-F., Stephenson, L., Van der Straeten, K., 2016. Addressing Europe's democratic deficit: an experimental evaluation of the pan-European district proposal. *Eur. Union Polit.* 17 (4), 525–545.
- Bormann, N.-C., Golder, M., 2013. Democratic electoral systems around the world, 1946–2011. *Elect. Stud.* 32 (2), 360–369.
- Bowler, S., Farrell, D.M., 2011. Electoral institutions and campaigning in comparative perspective: electioneering in European Parliament elections. *Eur. J. Polit. Res.* 50 (5), 668–688.
- Boxell, L., Gentzkow, M., Shapiro, J.M., 2020. *Cross-Country Trends in Affective Polarization*. Stanford Mimeo available at: <http://web.stanford.edu/~gentzkow/research/cross-polar.pdf>.
- Campbell, A., Converse, P.E., Miller, W.E., Stokes, D.E., 1980. *The American Voter*. University of Chicago Press.
- Chang, E.C.C., Golden, M.A., 2007. Electoral systems, district magnitude and corruption. *Br. J. Polit. Sci.* 37 (1), 115–137.
- Carey, J.M., Shugart, M.S., 1995. Incentives to cultivate a personal vote: a rank ordering of electoral formulas. *Elect. Stud.* 14 (4), 417–439.
- Cheibub, J.A., Sin, G., 2020. Preference vote and intra-party competition in open list PR systems. *J. Theor. Polit.* 32 (1), 70–95.
- Colomer, J., 2011. *Personal Representation: the Neglected Dimension of Electoral Systems*. ECPR Press.
- Cox, G.W., 1990. Centripetal and centrifugal incentives in electoral systems. *Am. J. Polit. Sci.* 34 (4), 903–935.
- Crisp, B.F., Escobar-Lemmon, M.C., Jones, B.S., Jones, M.P., Taylor-Robinson, M.M., 2004. Vote-seeking incentives and legislative representation in six presidential democracies. *J. Polit.* 66 (3), 823–846.
- Cutler, F., 2002. The simplest shortcut of all: sociodemographic characteristics and electoral choice. *J. Polit.* 64 (2), 466–490.
- Dalton, R.J., 2004. *Democratic Challenges, Democratic Choices: the Erosion of Political Support in Advanced Industrial Democracies*. Oxford University Press.
- Darmann, A., Grundner, J., Klamer, C., 2017. Election outcomes under different ways to announce preferences: an analysis of the 2015 parliament election in the Austrian federal state of Styria. *Publ. Choice* 173, 201–2016.
- Devroe, R., Wauters, B., 2020. Does high on the ballot means highly competent? Explaining the ballot position effect in list-PR system. *Acta Politic.* 55 (3), 454–471.
- Duverger, M., 1951. *Les Partis Politiques*. Armand Colin.
- Dolan, K., 2008. Is there a 'gender affinity effect' in American politics? *Polit. Res. Q.* 61 (1), 79–89.
- Farrell, D.M., Shugart, M.S., 2012. *Electoral systems: six-volume set*. Sage.
- Farrell, D.M., McAllister, I., 2006. Voter satisfaction and electoral systems: does preferential voting in candidate-centred systems make a difference? *Eur. J. Polit. Res.* 45 (5), 723–749.
- Folke, O., Rickne, J., 2020. Who wins preference votes? An analysis of party loyalty, ideology, and accountability to voters. *J. Theor. Polit.* 32 (1), 11–35.
- Folke, O., Persson, T., Rickne, J., 2016. The primary effect: preference votes and political promotions. *Am. Polit. Sci. Rev.* 110 (3), 559–678.
- Golder, S., Stephenson, L., Van der Straeten, K., Blais, A., Bol, D., Harfst, P., Laslier, J.-F., 2017. Votes for women: electoral systems and support for female candidates. *Polit. Gend.* 13 (1), 107–131.
- Jankowski, M., 2016. Voting for locals: voters' information processing strategies in open-list PR systems. *Elect. Stud.* 43 (3), 72–84.
- Karvonen, L., 2010. *The Personalisation of Politics: A Study of Parliamentary Democracies*. ECPR Press.
- Laslier, J.-F., Blais, A., Bol, D., Golder, S., Harfst, P., Stephenson, L., Van der Straeten, K., 2015. The EuroVotePlus experiment. *Eur. Union Polit.* 16 (4), 601–615.
- Laslier, J.-F., Sanver, R., 2010. *Handbook of Approval Voting*. Springer.
- Lijphart, A., 1994. *Electoral Systems and Party Systems*. Oxford University Press.
- Lupia, A., 1994. Shortcuts versus encyclopedias: information and voting behavior in California insurance reform elections. *Am. Polit. Sci. Rev.* 88 (1), 63–76.
- Matakos, K., Troumpounis, O., Xefteris, D., 2016. Electoral rule disproportionality and platform polarization. *Am. J. Polit. Sci.* 60 (4), 1026–1043.
- McDermott, M.L., 1997. Voting cues in low-information elections: candidate gender as a social information variable in contemporary United States elections. *Am. J. Polit. Sci.* 41 (1), 270–283.
- McDermott, M.L., 1998. Race and gender cues in low-information elections. *Polit. Res. Q.* 51 (4), 895–918.
- McElroy, G., Marsh, M., 2010. Candidate gender and voter choice: analysis from a multitember preferential voting system. *Polit. Res. Q.* 63 (4), 822–833.
- Meriläinen, J., Tukiainen, J., 2018. Rank effects in open-list elections. *Publ. Choice* 177, 87–109.
- Nagtzaam, M.A.M., van Erkel, P.F.A., 2017. Preference votes without preference? Institutional effects on preference voting: an experiment. *J. Elections, Public Opin. Parties* 27 (2), 172–191.
- Norris, P., 2011. *Democratic Deficit: Critical Citizens Revisited*. Harvard University Press.
- Popkin, S.L., 1991. *The Reasoning Voter. Communication and Persuasion in Presidential Campaigns*. The University of Chicago Press.
- Renwick, A., Pilet, J.-B., 2016. *Faces on the Ballot: the Personalization of Electoral Systems in Europe*. Oxford University Press.
- Rosar, U., Klein, M., Beckers, T., 2008. The frog pond beauty contest: physical attractiveness and electoral success of the constituency candidates at the north rhine-westphalia state election of 2005. *Eur. J. Polit. Res.* 47 (1), 64–79.
- Sanbonmatsu, K., 2002. Gender stereotypes and vote choice. *Am. J. Polit. Sci.* 46 (1), 20–34.
- Sanz, C., 2017. The effect of electoral systems on voter turnout: evidence from a natural experiment. *Pol. Sci. Res. Methods* 5 (4), 689–710.
- Schickler, E., Green, D.P., 1997. The stability of party identification in western democracies: results from eight panel studies. *Comp. Polit. Stud.* 30 (4), 450–483.
- Shugart, M.S., Valdini, M.E., Suominen, K., 2005. Looking for locals: voter information demands and personal vote-earning attributes of legislators under proportional representation. *Am. J. Polit. Sci.* 49 (2), 437–449.
- Sudulich, L., Trumm, S., 2019. A comparative study of the effects of electoral institutions on campaigns. *Br. J. Polit. Sci.* 49 (1), 381–399.