



HAL
open science

Comparing Voting Methods: 2016 US Presidential Election

Herrade Igersheim, François Durand, Aaron Hamlin, Jean-François Laslier

► **To cite this version:**

Herrade Igersheim, François Durand, Aaron Hamlin, Jean-François Laslier. Comparing Voting Methods: 2016 US Presidential Election. *European Journal of Political Economy*, 2022, 71, pp.102057. 10.1016/j.ejpoleco.2021.102057 . halshs-03926997v2

HAL Id: halshs-03926997

<https://shs.hal.science/halshs-03926997v2>

Submitted on 6 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Comparing Voting Methods: 2016 US Presidential Election

Herrade Igersheim ‡

François Durand*

Aaron Hamlin †

Jean-François Laslier§

(Revision: February, 2021)

Abstract. This paper presents data from a survey leading up to the 2016 US presidential elections. Participants were asked their opinions about the candidates and were also asked to vote according to three alternative voting rules, in addition to plurality: approval voting, range voting, and instant runoff voting. The participants were split into two groups, one facing a set of four candidates (Clinton, Trump, Johnson, and Stein) and the other a set of nine candidates (the previous four plus Sanders, Cruz, McMullin, Bloomberg, and Castle). The paper studies three issues: (1) How do US voters use these alternative rules? (2) What kinds of candidates, in terms of individual preferences, are favored by which rule? (3) Which rules empirically satisfy the independence of eliminated alternatives? Our results provide evidence that, according to all standard criterion computed on individual preferences, be there utilitarian or of the Condorcet type, the same candidate (Sanders) wins, and that evaluative voting rules such as approval voting and range voting might lead to this outcome, contrary to direct plurality and instant runoff voting (that elects Clinton) and to the official voting rule (that elected Trump).

Classification JEL: D72, C93.

Keywords: Approval voting, range voting, instant runoff, strategic voting, US Presidential election.

‡ CNRS, BETA and University of Strasbourg, France

* Nokia Bell Labs, France

† The Center for Election Science, Redding, CA, USA

§ CNRS, Paris School of Economics, France

Acknowledgments: We would like to thank Robin Quirke and Tom Bowerman at Policy Interactive in Portland, Oregon for testing question formatting.

1 Introduction

Since November 2016, both the US and worldwide media have emphasized that the sitting Republican president Donald Trump did not get a majority of the votes cast. Indeed, he lost the “popular vote” by almost three million votes to Hillary Clinton, the Democratic Party nominee. The popular vote is defined by the nationwide sum of all ballots in the official election. The Electoral College elects the US president via a different route. The latter is an indirect winner-take-all system that adds up votes state by state. Those state-level results are used to determine electoral voters (based on Senate and House seats within each state) who are then expected to vote for their respective candidates. These two distinct approaches explain why the outcomes might differ. And 2016 was not the first time this discrepancy has occurred. Neubauer et al. (2012) stressed that, since 1780, at least four elections of this kind have occurred—with two cases in the last five elections: this issue has been widely discussed by political scientists and social choice theorists (among others, see Abbott and Levine 1991, Miller 2014, Barthélémy et al. 2014, Kurrild-Klitgaard 2018).

While the official US voting method is often criticized for its two-tier aspect, questions regarding its “*mono-nominal*” character—that a voter can vote for only one candidate—are less often raised, despite the current movement towards the use of different voting rules. For instance, the states of Maine in 2017 and Alaska in 2020 adopted variants of ranked-choice rules and the city of Fargo adopted Approval Voting in 2018, followed by the city of Saint Louis in 2020 (Vox 2018, Alaska Division of elections 2020, AP News 2020a, 2020b). In this paper we will be chiefly interested in evaluative voting, that is “*multi-nominal*” rules where the voter evaluates the candidates one by one, independently. Such processes are commonly used for decision making in practical settings like schools, sports clubs, online applications, et cetera. It has been argued that these kinds of rules would enable voters to be more flexible in casting their ballots and to vote more sincerely if they so wished, thus increasing their overall satisfaction. But only a few voting theorists have analyzed the possible effects of these evaluative processes in general (see Hillinger 2004ab, Gaertner and Xu 2012, Pivato 2013, Macé 2018). Among the various evaluative rules, relatively more attention has been paid to *approval voting*, a particular case where the voter grades each candidate on a two-level scale (0 or 1). The winner of an approval voting election is the candidate who gets more 1-grades than any other candidate (see notably Brams and Fishburn 1978, 1983, Laslier and Sanver 2010).

The game-theoretical literature has devoted a lot of attention to the “divided majority problem” that is the coordination game among voters of a same camp that will lose the election if they do not coordinate their (plurality) votes on the same candidate. This is a typical pathology of plurality rule that approval voting might be immune to (see Brams and Fishburn 1983, Forsythe et al. 1993, Bouton et al. 2016). In fact, it is very easy to find such examples, but counter-examples have also been found: see De Sinopoli et al. (2014). In any case, since Fiorina (1976) the empirical literature has established clearly that voters behavior in practice do not conform the purely rational nor the purely expressive model. See on this point Spenkuch (2018) for Germany or the collection Stephenson et al. (2018) for various countries.

A stream of research has conducted experimental tests of approval voting and other multi-nominal rules alongside large-scale official elections. Such studies have been conducted in France since 2002 in parallel with presidential elections (Baujard and Igersheim 2010, Grofman et al. 2011, Baujard et al. 2014), and similar protocols have been used in Germany (Alos-Ferrer and Granic 2010). Several large-scale voting experiments have also been

organized online, in France, in Canada, and throughout Europe; see Laslier et al. 2015 and the website vote.imag.fr. The lessons drawn from these studies are as follows: (i) The principles of multi-nominal voting are easily understood by voters, who show a slight preference for range voting over approval voting. (ii) Such voting methods can indeed modify the aggregate ranking of candidates and can lead to different winners.

The U.S. is one of a rather small number of countries with a presidential regime to elect its president under universal¹ (albeit indirect) suffrage. Another such country is France, where, as noted above, many relevant studies have been conducted. But, to the best of our knowledge, there is no research that aims to test multi-nominal voting rules empirically in the context of a US national election. Compared to countries in Europe and elsewhere, the U.S. political system is rather unique, so that existing results from elsewhere might apply to the U.S. This study thus fills this gap by analyzing how US voters react to multi-nominal voting rules and assessing what the impact of these rules might be on election results. To do so, we use data from a survey conducted in November 2016 in which over 2,000 participants were asked to vote according to the three following alternative voting rules in addition to plurality voting:

- Approval voting (AV): Voters can approve of any number of candidates they want; the winning candidate is the one with the largest number of approvals.
- Range voting (RV): Voters score each candidate on a six-level scale (from 0 to 5 points, inclusive); the winning candidate is the one who receives the largest total number of points.²
- Instant-runoff voting (IRV): Voters rank their three favorite candidates; the candidates with the fewest first-place rankings are sequentially eliminated with their next-preference votes being transferred until a single candidate remains.³

Additionally, participants in the survey were asked to honestly assess the candidates on the same six-level scale. We refer to this evaluation as HA for “honest assessment”.

The participants of the survey were randomly split into two groups. The first group was presented with a “short set” of four candidates (Clinton, Trump, Johnson, and Stein), and the second group was presented with a “long set” of nine candidates (the previous ones plus Sanders, Cruz, McMullin, Bloomberg, and Castle). These two sets were designed to mimic the actual 2016 election, and that same election within a more competitive context.⁴ Our

¹Universal suffrage is for those aged 18 years or older and not currently in prison. Only prisoners in Maine and Vermont may vote while in prison. Most states require former felons to wait an additional waiting period such as parole, probation, or following the subjective decision of a designated official before voting rights can be restored. This point is especially relevant given the US has the highest per capita incarceration rate in the world.

² The term Range voting is often met in the literature but in the survey we used the phrase “score voting” that, we believe, better conveys the definition.

³ We here use the term “instant-runoff”, that is often met in the US literature, but the survey used “ranked choice voting” that might be more descriptive for a US audience. The term “alternative vote”, used in Ireland, is poorly descriptive.

⁴The idea of varying the number of candidates as an experimental trick to study a theoretical point also appears in Crowder-Meyer et al. (2020) in their study of which cues do voters use to decide their vote when they have relatively few information about the candidates.

analyses focus on three issues. (1) How do US voters use these alternative rules? (2) How do candidate options, including the presence of little-known candidates, influence the election outcome? (3) Do alternative voting methods affect the outcome compared to plurality voting?

According to our data, Bernie Sanders stands out according to voters' true preferences, in the sense that, as will be seen in section 3.3, if one considers the honest assessment of the candidates by the voters (HA), he wins according to several aggregation rules, whether of the majoritarian or the utilitarian tradition.

Given this, the main conclusions of this study are the following: (1) Multi-nominal voting rules such as range voting and potentially approval voting are able to elect such "best" candidates: range voting would clearly elect Sanders, and approval voting would elect either Sanders or Clinton. (2) Such is not the case for direct plurality and instant runoff voting (both would elect Clinton), or the indirect, official, voting rule (that elected Trump).

The next section describes the survey and data. A third section presents the global statistics and results per candidate depending on the voting methods tested. From these, one can draw some lessons regarding the first issue at stake: how do voters use these rules? A fourth section develops further analyses focused on the two remaining issues: the fate of minor candidates and the election outcome. The fifth section concludes. Additional tables and figures as well as the survey questions are to be found in the on-line appendix.

2 Survey and statistical strategy

The Center for Election Science designed and contracted the survey with the international firm Growth from Knowledge (GfK). The survey ran from November 3, 2016, to the day of the US presidential election on November 8, 2016 (noon). GfK used a panel of 4,181 members of which 2,552 responded and 2,367 completed the survey. The survey had six questions defined by the Center plus additional socio-demographic questions (such as age, gender, education, race, etc.). The survey was conducted electronically on a panel of representative American participants, but with no specific target beyond that: general population adults (18+), English- and Spanish-language survey-takers. In the on-line appendix B, we describe the socio-economic profile of the sample and provide evidence that it is indeed representative. Like almost all published surveys at that time, ours over-estimated Clinton's final vote tally: At the national level she was finally ahead of Trump by 2% while this margin was estimated in most surveys between 2 and 6%⁵.

Participants were also separated into two groups. While a first group had to give their vote and opinion on a short set of four candidates, a second group was asked to do the same for a set of nine candidates. The four candidates of the short set were the nominees of their respective party: Hillary Clinton for the Democratic Party, Gary Johnson for the Libertarian Party, Jill Stein for the Green Party, and Donald Trump for the Republican Party. The five remaining candidates of the long set were either popular candidates in the Republican and Democratic primaries (such as Ted Cruz and Bernie Sanders) or independent candidates. Evan McMullin was only on one state ballot and Michael Bloomberg decided not to run because he feared splitting the left-leaning vote with Hillary Clinton. Finally, Darrell Castle was the

⁵ See the list of such surveys on the wikipedia page http://en.wikipedia.org/wiki/Nationwide_opinion_polling_for_the_2016_United_States_presidential_election

nominee of the Constitution Party, although he was not on enough state ballots to theoretically win enough electoral votes.⁶ 1,198 participants were asked to answer the long-set survey and 1,169 were asked to answer the short-set survey. For the two sets, the four different voting rules were presented in random order, as were the candidate names. In all, 2,367 participants filled in the survey. Upon cleaning, a small number of responses (10 for the long set and 10 for the short set) had to be deleted—see appendix B for details—so what follows is thus based on a set of $1,188 + 1,159 = 2,347$ respondents.

To check the statistical validity of our conclusions in a unified manner we report the standard errors or confidence levels obtained from bootstrapping the dataset 10,000 times. But note that the question of extrapolating the observations made with this particular survey to statement about what would have been the 2016 presidential election in other circumstances (under different voting rules) is a different question not of a statistical nature. First, the campaign was not over at the time of the survey and such a survey does not intend to forecast the actual election. Second, and more importantly, the campaign dynamics (and even the set of candidates) might be different depending on the rule.

3 Results and main lessons on voters' behavior

3.1 Overall findings and assessment of the tested voting rules

Overall summary statistics offer us a quick overview of the participants' behavior under AV, RV, and HA. Under AV, for the participants who gave one or more approvals, the average number of approvals per ballot is 1.24 (out of four candidates) with the short set, and 1.73 (out of nine candidates) with the long set.⁷

In a 1984 AV experiment where 300 Pennsylvanian college students were asked whether to support candidates from lists of eight or nine names—thus close to our long set—Koc (1988) obtained an average number of approvals of 1.8 and observed similarly that it was “low given the results of other AV experiments where 2.0 votes per ballot has been the norm”, referring to Nagel (1984). European experiments have had more candidates: the average numbers of approvals observed in France in similar experiments were 3.15 for sixteen candidates in 2002, 2.23 for twelve in 2007, 2.63 for ten in 2012, and 2.42 for eleven in 2017 (Laslier 2019).

There is a general (and quite intuitive) correlation between the number of candidates running and the average number of approvals. But beyond this observation, compared to US data as a whole, it also seems that French voters are more inclined to give their support to several candidates. This might be due to the different voting rules used in these countries. In France, the official rule to elect the president is a two-round system, and elections often include more than two serious candidates in the first round, so voters are used to facing many significant candidates. That contrasts with the US in both the voting method and the number of candidates. In any case, these relatively low observed averages imply that a fraction of the

⁶US ballot access varies drastically state-by-state based on whether the candidate is within a major party, minor party, or independent. Some states require a filing fee while others can require over 100K signatures for that state alone. Major party requirements are always equal to or easier than for other independent or third-party candidates.

⁷With those who abstain from voting for any candidate, these averages fall to 1.15 and 1.61, respectively.

voters approve of only one candidate (“bullet voting”). When facing the short set, almost three participants out of four approved only one candidate and, for the long set, one out of two approved only one and one out of four approved two candidates. Again, this contrasts with the French experiments, where the percentages of participants who approved of only one candidate never exceeded 25% (Table A1 in the on-line appendix A). Still, even with the relatively high frequency of “bullet voting” in the US experiment, those voters who approved of more than one candidate had a large impact on the winner and on the support measured for other candidates, as will be seen.

Turning to range voting, the participants took advantage of the increased opportunities to express their electoral preferences. The average grades given to candidates are very close for the two sets: 1.78 for the long set and 1.77 for the short one. Further, the distributions of grades for both sets are also remarkably similar (Table A2 and Figure A2 in the online appendix A). This would support the theory that voters behave the same way whether they are facing a 4 or 9-candidate list under RV; this is an important observation that backs up the idea that actual votes under RV might be essentially “independent of eliminated alternatives”. We will come back to this idea when studying the results at the candidate level.

One can further compare the distributions of grades under Range Voting (RV) with the ones obtained as an answer to the evaluation question regarding the running candidates (“honest assessment”, HA), with the same range from 0 to 5. Under HA, the average grades per candidate are 1.41 for the long set and 1.58 for the short one, and the grade distributions are almost identical (Table A3). Also remarkable is the observation that the grade distributions under RV and HA are very similar (Figures A2 and A3). This spotlights the fact that participants do not overall behave according to the rational theory. Indeed, fully strategic voting under RV in a large election would involve giving either maximal or minimal grades but no intermediate ones (Núñez and Laslier, 2014). Here, strategic voting under RV would thus require participants either to give the grades 0 or 5; and the fact that the distributions of RV and HA are similar and are not degenerated support the view that respondents voted sincerely with little, or no, strategic consideration.

These statistics help explain how voters behave under the multi-nominal rules tested. The vote frequency distributions for AV and RV and the average number of approvals/grades per ballot show that voters have used both voting methods to more widely express their electoral preferences. Some remarks are in order:

The fact that the number of approvals per ballot is lower than in previous experiments might be due to the way the AV instructions are stated (Koc 1988). In our survey, however, the wording of the AV ballots ought to have encouraged the participants to vote for more than one candidate, since they were asked “to select as many candidates as [they] want”. It might well be the case that framing effects are not that important, and that “bullet voters” do so for deeper reasons, such as the perceived election dynamics or familiarity with candidates. For instance, a 2016 Gallup Poll⁸ indicated that two-thirds of Americans didn’t know who Gary Johnson and Jill Stein were. This voter ignorance is likely due to a number of causes,

⁸<https://news.gallup.com/poll/194162/third-party-candidates-johnson-stein-largely-unknown.aspx>

including the fact that the Commission on Presidential Debates⁹ excluded these two candidates from every presidential debate.¹⁰

In any case, we observe a nontrivial number of voters who chose more than one candidate. And, as we will see, these votes can have important consequences.

3.2 Results per candidate for the short and long sets

Table 1 provides the observed scores of the candidates for plurality and approval voting as percentages of the population and for range voting as average grades. Standard errors estimated by bootstrapping are provided in parenthesis. The candidates are ranked according to their plurality scores.

Table 1: Candidate scores under three voting rules

| | Plurality | | Approval | | Range | |
|-----------|-----------------|-----------------|-----------------|-----------------|---------------|---------------|
| | short set | long set | short set | long set | short set | long set |
| Clinton | 47.73 (1.63) | 31.38 (1.50) | 50.12 (1.62) | 39.78 (1.51) | 2.33 (.07) | 2.32 (.07) |
| Trump | 40.52 (1.57) | 27.78 (1.25) | 42.01 (1.60) | 33.73 (1.44) | 1.89 (.07) | 2.00 (.07) |
| Sanders | | 19.98 (1.08) | | 39.25 (1.47) | | 2.72 (.07) |
| Cruz | | 9.73 (0.89) | | 21.46 (1.36) | | 1.93 (.06) |
| Johnson | 8.25 (0.93) | 4.54 (0.62) | 20.65 (1.26) | 12.16 (1.02) | 1.59 (.05) | 1.50 (.05) |
| Bloomberg | | 4.41 (0.50) | | 11.65 (0.91) | | 1.67 (.05) |
| McMullin | | 1.73 (0.40) | | 7.60 (0.89) | | 1.36 (.05) |
| Stein | 3.51 (0.65) | 0.26 (0.15) | 11.52 (1.18) | 5.06 (0.72) | 1.30 (.05) | 1.21 (.05) |

⁹ The Commission On Presidential Debates is a private entity that controls US Presidential debates. Major party candidates themselves are free to use a different host, though they have little incentive to. To participate in debates, the Commission requires that candidates poll nationwide at 15% by five different polling agencies using plurality voting. These polls, however, often exclude third-party candidate names.

¹⁰ A legitimate question is thus how it is possible to vote for an unknown candidate. Indeed, respondents might either put unknown candidates at the bottom of their ranking, or put them in the middle (the latter move being suggested by the “feeling thermometers” from the ANES). Anyhow, both behaviors lead to noise in the answers to the survey. In spite of this limit and since one aim of this paper is to study the independence of eliminated alternatives, we have decided to keep all the candidates of the long set in our analyses.

| | | | | | | |
|--------|--|----------------|--|----------------|--|---------------|
| Castle | | 0.19 (0.13) | | 2.18 (0.45) | | 0.95 (.05) |
|--------|--|----------------|--|----------------|--|---------------|

a) Plurality.

Under plurality with the short set of candidates, Clinton is the winner with 47.73%. Trump is second with 40.52% (statistical confidence¹¹ for the statement “Clinton wins” is actually 100%). With the long set of candidates, Clinton wins with 31.38%, Trump is second with 27.78% and Sanders is ranked third with 19.98% of votes (statistical confidence for the statement “Clinton wins” is 96%). No other candidate reaches 10%. There is no contradiction between the rankings of the long and the short sets. The results of plurality are thus in accordance with those of the “popular vote”, providing further evidence that the sample is representative,¹² not only with respect to socio-economic variables but also with respect to politics.

b) Approval

The exact phrasing used in the survey for approval voting was the following:

“If the US presidential election were held today, and your ballot asked you to select all the candidates you wanted, which candidate or candidates would you select? In the approval voting method, the candidate selected the most would win.”

The results of the survey are more thought-provoking under AV than under plurality. Table 1 shows that adding Sanders to the ballot has huge consequences. Indeed, while Trump remained second under AV for the short set, as he was under plurality, he is now far surpassed by Sanders, who is almost tied with Clinton: the observed scores are 39.78% for Clinton against 39.25% for Sanders, this difference is not statistically significant (statistical confidence for a Clinton win is only 60%). The results of AV thus highlight the very strong support Sanders had in 2016, which was partially hidden by the mono-nominal feature of the plurality rule. Of course, when discussing the results of this kind of experiment and as already stressed earlier above, one should keep in mind that, under another official voting rule, the strategies of the political parties and candidates would be different, as indeed might the whole political landscape. This might lead the voters to hold different political preferences. Still, the *prima facie* reading of our data is very simple: everything else being equal, under AV, Sanders could have won the 2016 US election.

¹¹ This means that Clinton is always the plurality winner when we randomly re-sample in the first group of respondents (short set).

¹² The numbers for third parties in our poll are higher than the official count, which is typical. There is normally a drop for third party support between polls and the actual election, possibly due to voters fearing they might throw their vote away on a losing candidate under plurality voting.

c) Range voting

The last two columns of Table 1 display the average grades obtained by the candidates under RV. The exact phrasing used in the survey was the following:

“The next method is called score voting. If the US presidential election were held today, and your ballot asked you to score each candidate on a scale from 0-5, how would you score them? Please enter a number, where ‘0’ means ‘Worse’ and ‘5’ means ‘Better’. In the score voting method, the candidate with the highest total score would win.”

Here the differences between the short and the long sets are even more obvious than under plurality—the winner, indeed, is different. For the short set, Clinton scores highest with a substantial advantage over Trump. For the long set, Sanders beats Clinton with an average grade of 2.72, while Clinton is second with 2.32. Note that, when present, Ted Cruz reaches the level of Donald Trump: their scores are respectively $2.00 \pm .07$ and $1.93 \pm .06$, and the confidence in the statement “Trump is ahead of Cruz” is only .85 from our sample. Another interesting trait of this voting rule—especially compared to AV—is that the average grades of the four candidates present in both sets are rather similar from one set to the other, thus showing a kind of “absolute” preference (“independence of eliminated alternatives”). In other words, the way participants assess these four candidates under RV was not modified by the inclusion of five more candidates. Such was not exactly the case under AV even though the ranking of candidates by approval was consistent between the two sets.

d) Instant runoff

For instant runoff voting (IRV), the exact phrasing used in the survey was the following:

“The next method is called ranked choice voting. If the US presidential election were held today, and your ballot asked you to rank your top three candidates, how would you rank them? Use a ‘1’ for your top choice, a ‘2’ for your second choice, and a ‘3’ for your third choice. In the ranked choice voting method, the winner requires a majority of first-choice votes. If no candidate had this majority, then the candidate with the fewest first-choice votes would be eliminated and those votes would transfer to the next-preferred choices. This would repeat until a candidate had a majority.”

Table 2 reproduces the last rounds of elimination.¹³ The detailed percentages of votes obtained by each candidate after each round of elimination are provided as supplementary material in the appendix (Tables A7 and A8, Figures A3 and A4).

Table 2: Instant Runoff results (long set): % after the last rounds of elimination

| | 5 | 6 | 7 | 8 |
|---------|----|----|----|-----|
| Clinton | 32 | 33 | 54 | 100 |
| Trump | 31 | 39 | 46 | |

¹³See the note on the statistical robustness of these findings in the on-line appendix.

| | | | | |
|---------|----|----|--|--|
| Sanders | 25 | 28 | | |
| Cruz | 13 | | | |

The main transfer to Clinton comes from Sanders’s voters, while Trump benefits from transfers from Cruz’s voters. At the first round, as under plurality, Clinton is ranked first, Trump second, and Sanders third. Then, after smaller candidates are eliminated, Sanders is eliminated, letting Clinton win against Trump.

IRV has, in our data, the property that very small candidates cannot spoil the election: once these candidates are eliminated, along the counting process the preferences stated by the voters as to the main candidates remain and are counted as if the eliminated candidates were not present. This explains the remarkable feature that Clinton beats Trump by an identical 54-46 margin under the short and long set conditions of IRV.

The subtlety of the run-off system lies in the way candidates are sequentially eliminated. The vote transfers that are sequentially observed here under IRV appear akin to the mechanism that favors “exclusive” candidates under plurality, i.e., candidates who receive strong support from some voters, but are also often rejected by others. The same process is often seen at work under two-round majority voting when a more “inclusive” candidate that would win any second round is eliminated at the first round by two more “exclusive” candidates. An “inclusive” candidate can be defined as follows: they get widespread support from the voters but with no strong feeling of rejection or attachment (on this issue, see the well-documented “squeezing of the center” often observed in French politics; Blais and Indridason 2007; Baujard et al., 2014).

This is a reminder that, even though instant runoff voting allows rich expression to the individual voter, that the rule’s algorithm doesn’t necessarily use all the information. As such, the observation that Clinton wins over Sanders under IRV could be explained simply by the mechanics of IRV. But we can also see that a psychological effect may be at work, because the IRV data and HA data produce different Condorcet winners¹⁴ (on these two effects, see Duverger 1951). Indeed, IRV and plurality voting show similar results for the first round in both the long and short candidate sets.

3.3 The “honest assessment” exercise

Any voting rule may be subject to strategic or “tactical” voting. To elicit sincere preferences, we therefore simply asked the following question:

“Regardless of their chance of being elected, how much do you honestly want the following to be elected? Please use a ‘0’ for ‘Do not want this person elected’ to ‘5’ for ‘Very much want this person to be elected.’”

The answer to this question is the “honest assessment” variable (HA). Table A3 in the appendix shows the average HA grades. As under RV and for the long set, Sanders is rated first, followed by Clinton and Trump. Sanders can thus be seen as the candidate who maximizes voters’ utilities/assessments. As a whole, the average grades given to the different

¹⁴ See Tables A8 and A10: 51% rank Clinton above Sanders according to IRV data, but 54% prefer Sanders to Clinton according to HA.

candidates under HA are a little bit lower than under RV, but the tendencies are remarkably similar, as were the global statistics for these two.

Taking the HA variable at face value, one can compute any social welfare criterion. For instance, the average scores in Table A3 correspond to a simple utilitarian computation, and Bernie Sanders appears as the “utilitarian optimum” among the long set of candidates. The overall domination of Sanders over all other candidates is clear in Figure 1, which shows the distributions of the grades obtained by each candidate with their de-cumulative functions. Denote by $n_1(x), n_2(x), \dots, n_5(x)$ the number of voters who assign grade $g = 1, 2, \dots, 5$ to candidate x . The figure reports the percentage of voters who evaluate a given candidate x at level g or above, that is:

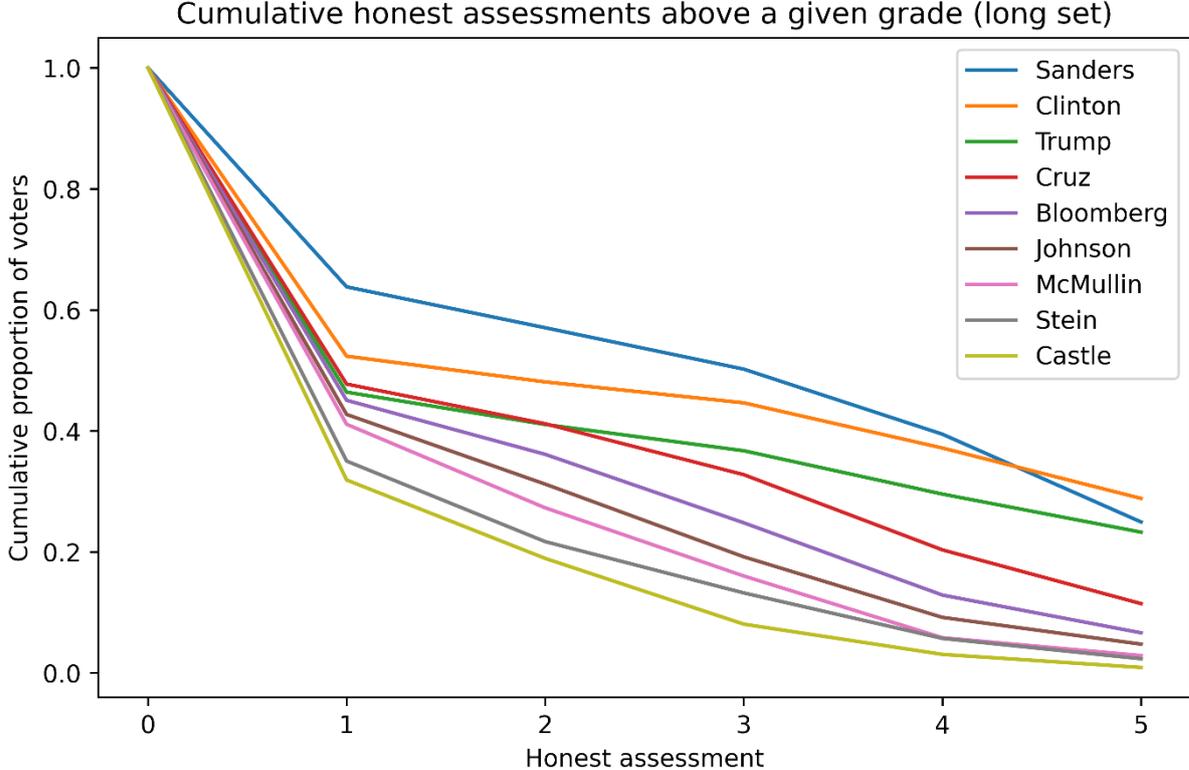
$$m_g(x) = n_g(x) + n_{g+1}(x) + \dots + n_5(x).$$

Recall the standard result: if a distribution for x first-order-dominates the distribution for y , meaning that x 's de-cumulative curve is above y 's, then for any increasing function f :

$$\sum_g n_g(x)f(g) > \sum_g n_g(y)f(g)$$

As seen in Figure 1, the candidates are almost perfectly ordered by first-order stochastic dominance. Therefore this ordering is robust to almost any distortion of the utility scale. The only important exception occurs for Clinton and Sanders, at the upper end of the scale. This means that a distorted utility function that put all its weight on grade 5 would conclude in favor of Clinton instead of Sanders. This observation is in line with the observation that Clinton beats Sanders according to plurality.

Figure 1: Comparing the distributions of honest assessments



Another social criterion, which rests on a different, if not contrary, philosophical basis (Riley 1990), is the notion of a Condorcet winner. From the HA grades one can infer strict preferences: for a given participant, Candidate x is strictly preferred to Candidate y if the HA grade for x is strictly larger than the HA grade for y. Because the grade scale {0,5} is not absolutely precise, it may be that a participant strictly prefers x to y but still grades them alike. Up to that technical difficulty, building the pair-wise comparison matrix among candidates from the HA grades leaves little doubt (Table A8 in the appendix); the honest preferences profile has a Condorcet winner, and it is Sanders.

Computing other social evaluation criteria on the same HA data leads to the same conclusion. Sanders is also ranked first under the Borda and the Bucklin rules. It thus seems fair to qualify him as the “best” candidate according to voters’ true preferences.

For some voting rules, it is also possible to mechanically compute the outcome of the rule using the HA variable as input, thus net of strategic voting. It turns out that “Honest Plurality” elects Clinton (Table A11), and so does “Honest Instant runoff” (Table A12). Notice that it is not possible to do the same exercise for approval voting since the HA data does not split the candidates into two categories (approved/disapproved) in a straightforward manner.

Conversely, the RV and IRV data can provide pairwise comparisons. There, Sanders is the Condorcet winner using RV data and Clinton is Condorcet winner using IRV data.¹⁵

Table 3 sums up our previous remarks and results. This shows the two main dimensions where our extended analyses will proceed. These two dimensions are (1) the chosen set of candidates, which implicitly raises questions about the primaries system, and (2) the chosen voting method with a more specific question regarding the mono-nominal vs. multi-nominal issue.

Table 3: Overall winners

| <i>Voting rules</i> | <i>Short set</i> | <i>Long set</i> |
|-----------------------|------------------|-------------------------------|
| Plurality | Clinton | Clinton |
| IRV | Clinton | Clinton |
| AV | Clinton | Clinton-Sanders ¹⁶ |
| RV | Clinton | Sanders |
| (HA various criteria) | Clinton | Sanders |

3.4 On strategic voting

Taking at face value the “honest assessment” exercise proposed to the participants, we can track down the sincere and strategic features of their votes. Although this issue is not this paper’s main goal, we briefly report on the following tests of strategic voting. We work under

¹⁵ See appendix B for the pairwise comparison matrices under RV and IRV, as well as the results of plurality, and IRV using HA data.

¹⁶From our sample, the approval rates of Clinton and Sanders are not statistically different, thus we cannot predict the winner.

the –reasonable– assumption that the two front runners considered by the participants were Clinton and Trump.

Plurality: The fraction of voters who vote for a candidate whom they do not rank first under HA is $3.7\% \pm 0.6\%$ in the short set and $5.2\% \pm 0.7\%$ in the long set. This direct measure of insincere voting under plurality shows that the insincere voting shown here is a rather marginal phenomenon. Strategic voters should vote for the one they prefer between the two front runners. Among the participants who strictly prefer Clinton to Trump we find that **X%** do not follow the strategic recommendation. Among the participants who strictly prefer Trump to Clinton we find that **Y%** do not follow the strategic recommendation. One can see that for most voters, our data show no contradiction between “sincere” and “strategic” behavior. The most interesting strictly “strategic” effect we can observe is voters who prefer Sanders to Clinton but vote Clinton under Plurality. And the effect is small: it occurs with an estimated frequency of $3.8\% \pm 1.3\%$ relative to the number of voters who honestly rank Sanders above Clinton.

IRV: The fraction of voters who put first a candidate that is not ranked first under HA is $3.0\% \pm 0.6\%$ in the short set and $4.4\% \pm 0.6\%$ in the long set. With respect to the “Clinton-Sanders” dilemma, we note that, among the voters who honestly rank Sanders above Clinton, only a small fraction rank Clinton first under IRV ($3.9\% \pm 1.0\%$). Due to the sequential process of IRV ballot counting, it is not clear how to qualify in theory the behavior of this small group. A more demanding test of sincerity is the following: the fraction of voters who do not vote for their three preferred candidates in the same order is $11.7\% \pm 1.0\%$ in the short set and $24.7\% \pm 1.4\%$ in the long set. Going into more details for these 24.7% can be described as follows:

- 13.3% involve minor candidates that the voter treats differently with IRV and HA, in one way or the other.
- 4.3% can be interpreted as simple strategic moves with respect to the Clinton/Trump contest (putting the candidate you prefer in a higher position or the one dislike in a lower position).
- 2.0% introduce the candidate they dislike among the pair (Clinton, Trump) into their top three. The natural interpretation here is that the voters fear that their rejection of a main candidate is not taken into account if that disliked candidates is not on their ballot. This shows the voter’s misunderstanding of the way IRV ballots are counted.
- 0.5% rank only Clinton, Sanders and Trump, but do it sincerely.
- 4.5% are various other, hardly interpretable, cases.

Approval: A demanding rational recommendation is here: “among Trump and Clinton, approve the one you prefer and do not approve the other.” The fraction of voters who do not follow this recommendation is $5.6\% \pm 0.7\%$ in the short set and $19.4\% \pm 1.2\%$ in the long set. These 19.4% can be described as follows:

- 12.9% approve neither Clinton nor Trump, and approve exactly one other candidate.
- 5.4% approve neither Clinton nor Trump, and approve several other candidates.
- 0.9% approve both Clinton and Trump plus one or several other candidates.
- 0.1% approve both Clinton and Trump and no others.
- 0.1% approve one of Clinton and Trump in contradiction with their HA.

Here a conclusion can be drawn: in most cases (about 80%) there is no contradiction between sincere and strategic behavior under AV, and when there is contradiction, it is almost always

in the same direction: voters use their approval ballot to express that they approve neither of the two front runners.

Range voting: We analyze the deviations from sincerity by computing differences in the average score of a given candidate between Range Voting and Honest Assessment. The HA grades are generally low (no candidate reaches the middle grade 2.5/5) and the RV grades are slightly larger on average (see Table A14 in the Appendix). One can see here traces of the theory that dictates to overstate the evaluation of the candidates of your camp and to downgrade the opponents (Núñez and Laslier 2014). Starting from low HA grades, the upgrading mechanically tends to be more important than the downgrading. In particular, Sanders and Cruz gain .37 and .40 points.

To go deeper, we wonder whether two candidates of the same camp are treated alike to this respect. Consider for instance the voters who do not give the maximum grade (5) as their honest assessment of Clinton. On average, in this population, the RV grade of Clinton is .27 points higher than her HA grade, and the same computation made on the subset of population who give to Sanders the perfect HA grade (5) yields .33 instead of .27. Table A15 in the Appendix provides the same figures for similar cases. One can see that the figures here are not small: up to .65 points (out of 5) as the average of a difference is a sizable effect. By offering fine-tuned expressive possibilities, Range Voting offers more occasions of observing deviations from sincere evaluations (Baujard et al. 2020). These deviations occur in the direction of the “overstatement” strategic theory.

All this is about individual behavior. The general conclusion at this level is in line with what we know from previous studies (see Laslier 2019). Strategic and non-sincere behavior can be found; the phenomenon is rather marginal for Approval Voting, in particular with respect to the main candidates. But strategic voting is more easily detected under Range Voting, by the nature of the rule, which is based on fine evaluations. Still, even under Range Voting the striking fact is the great similarity between the “Honest Assessment” grades and the votes. Notice that an aggregate measure of the consequences of insincere voting is directly provided by the tables that count HA ballots using different rules (Table A3 for RV, A8 for Pairwise majority, A11 for plurality, A12 for IRV), one can see that the non-sincere voting is of little consequence at the aggregate level, in this data set.

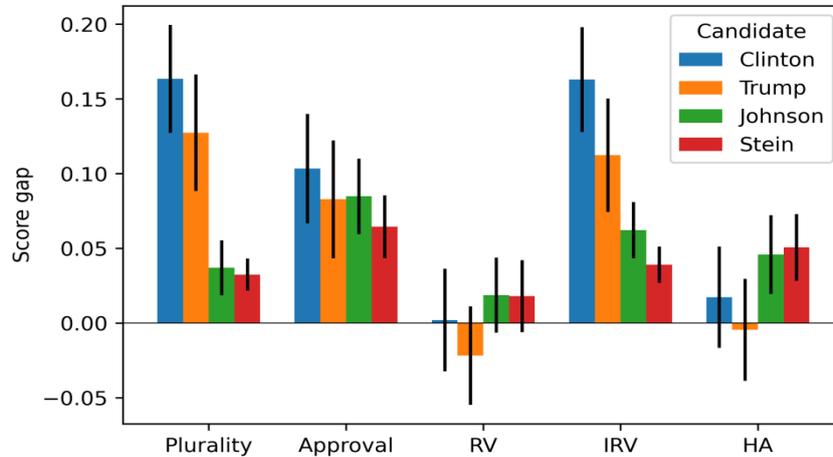
4 Two interconnected issues: which set of candidates for which voting rule?

4.1 From the short set to the long set: the question of “eliminated alternatives”

Our results make clear that the voters’ expressed preferences for a given candidate might differ depending on the number and nature of other candidates running. Obviously, this could have direct consequences on the outcome. Here, while the winner under AV, RV, and HA is Clinton with the short set, Sanders is ranked first with the long set under RV and HA, and he is in a statistical tie with Clinton under AV. Recall that these rules, unlike plurality and IRV, give all voters the possibility to express their support for any candidate independently of the other candidates. To further study this point, consider Figure 2. Figure 2 is built from the

figures in Table 1 and the results of IRV and HA. Figure 2 also shows how normalized scores¹⁷ vary when going from the long to the short set of candidates.

Figure 2: Scores gaps between the short and the long sets



Under plurality and IRV, all the candidates have lower scores with the long set than with the short one. Under AV the effect still exists, but to a much lesser extent. Under RV and HA, the scores remain more or less identical. While one can easily explain why under plurality a longer set of candidates might decrease the number of votes for each candidate, since a voter can give her support to only one of them, the same logic does not apply to AV, where participants can approve of any number of candidates. Our data show that all the four candidates, and Clinton in particular, have lost around 10% of their support with the long set compared to their scores with the short set. This is explained by the fact that even with the long set we still observe that about half of the participants approve of one candidate only under AV. Therefore, both under plurality and AV, the four candidates present in both sets have suffered from a dispersion of the participants’ supports for the long set.

From a theoretical perspective this issue is very important and is widely discussed in the literature. It refers to a condition termed “independence of eliminated alternatives”. Concretely, this condition can be stated as follows: “a candidate that wins an election with m candidates must not lose the election if another candidate is no longer available” (Dougherty and Edward, 2011, p. 79).¹⁸ Clearly vote splitting makes it impossible for plurality to satisfy this condition, and this is what our data confirms. According to Brams and Fishburn (1983), AV should be less affected by this effect, for it allows voters to vote for as many candidates as they want. Accordingly, AV outcomes would be more stable and sincere than under plurality.

¹⁷RV average scores are divided by the maximum grade, 5, and multiplied by 100. IRV “scores” are the percentages of first preferences. See Table A5.

¹⁸ Note that this condition should be distinguished from Arrow’s famous “independence of irrelevant alternatives” condition (Arrow 1963). Although similar in spirit, the conditions are not equivalent. According to Dougherty and Edward (2011, p. 79), Arrow’s condition “requires that the social ranking between any two alternatives should be independent of individual rankings between one or more alternatives that are not part of the pair,” thus “it does not refer to the addition or elimination of alternatives.”

But our results (see Table 1) show that some voters change their decision to approve or not a candidate depending on who the other candidates on the ballot are. How important and consequential is this effect? In the present case, AV does not directly violate the independence of eliminated alternatives condition because the presence of other candidates—while changing absolute approval percentages—does not change the candidates’ winning order compared to the original short list results. Our results would suggest that AV is largely if not totally independent of eliminated alternatives. In the 1984 experiment conducted in Pennsylvania, Koc (1988) made this point and noted a difference when one adds or deletes an important candidate (like Sanders in our case). The differences in the support greatly depended on the “nature of the race” and the “nature of that [added] candidate” (Koc 1988, 704). Contrary to Koc, it appears here that under AV the dispersion of the support is homogeneous for the four candidates of the short set, who all lose comparable percentages of support, as is clear in Figure 2.

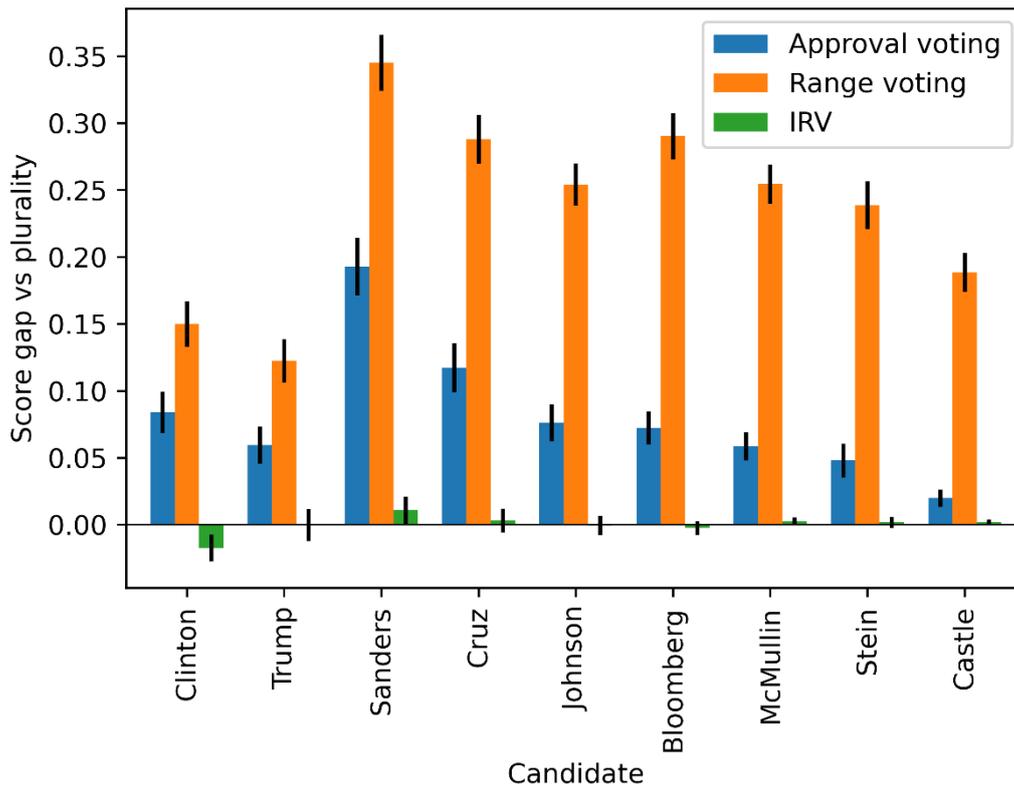
The short and long sets are totally different under plurality. For instance, the negative effects are clear on both Clinton’s votes (due to Sanders) and Trump’s votes (due to Cruz). Indeed, under plurality, the dispersion of votes between the two sets is highly heterogeneous. The same is also true with IRV’s first-round preference results.

4.2 From plurality to multi-nominal voting rules: a change in the winner

In this subsection we focus our analysis on the long set to examine why RV, IRV, and AV produce different winners compared to plurality. More generally, we examine the kind of candidates favored by certain voting rules. Figure 3 represents the gaps in the scores between plurality, AV, and RV for the nine candidates of the long set with normalized scores as in Figure 2. For instance, Clinton got 31.38% of votes under plurality and was approved by 39.78% of participants under AV. Her score gap between these two voting rules is thus equal to 8.40.¹⁹

¹⁹ Note that Table A13 shows the scores of the nine candidates of the long set under plurality, AV, RV, and IRV.

Figure 3: Scores gaps compared to plurality



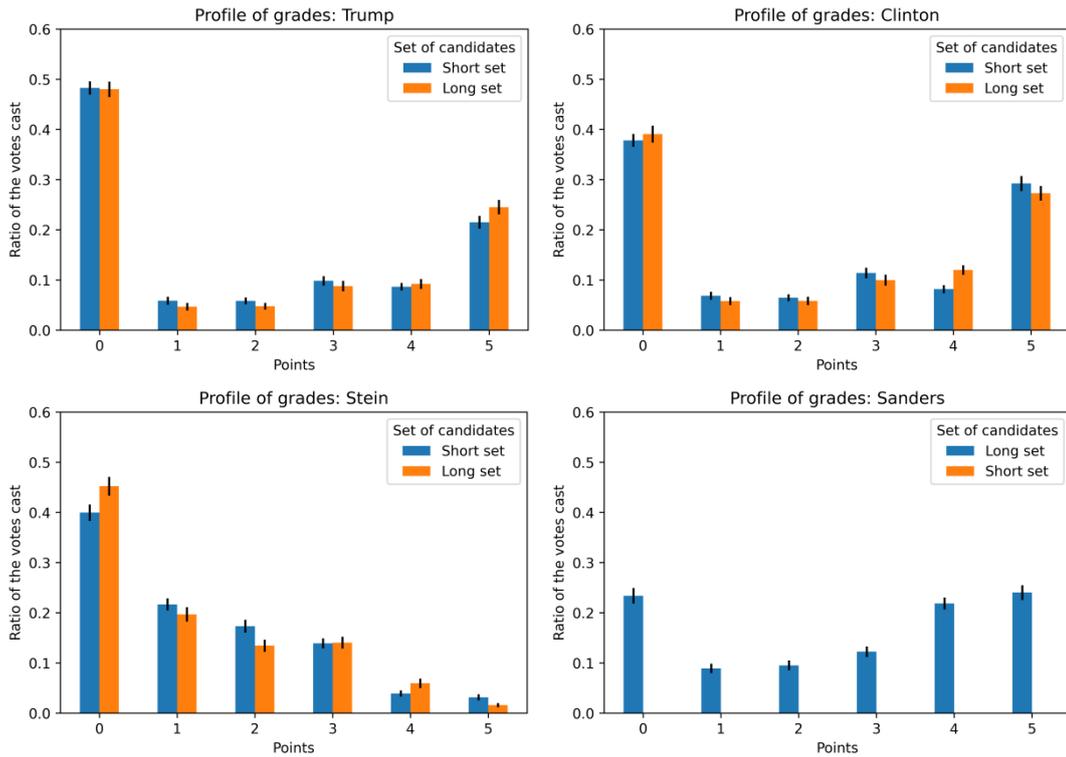
These elements clearly demonstrate that Sanders benefits the most by switching from plurality to evaluative multi-nominal voting rules such as AV and RV. The previously unpopular candidates also make significant improvements under these rules. Conversely, the candidate who is least advantaged by evaluative rules is Trump, thus showing that the latter did not attract much support beyond the supporters who voted for him under plurality. The same applies for Clinton. As previously stressed, the scores of Plurality and IRV first choices are very similar.²⁰

The next stage is to shed light on the characteristics of the candidates who benefited the most and least. We do this to define “types” of candidates beyond their party affiliation. Here, we follow Baujard et al. (2014) and consider the distribution of grades under RV for Trump, Clinton, Stein, and Sanders, four candidates that provide interesting examples of grade profiles (Figure 4).

²⁰ Another way to compare the scores of plurality, IRV, AV and RV is exposed in the Appendix (Figure A6). Figure A6 shows an alternative normalization of the scores so that the total score is 100 for each rule (“normalized relative scores”). For Plurality this is just the vote percentage and for IRV the percentage of voters who place the candidate first, but for Approval it is the percentage, among all approvals, of approvals received by each candidate and for RV, similarly, the percentage of all assigned points. On Figure A6, as in Figure 3, one can see that for all candidates, IRV first choices and plurality scores are hardly distinguishable (within statistical confidence). Above all, Figure A6 evidences the fact that RV, and to a lesser extent, AV produce a flatter image of the political landscape.

First, the profiles of the two main candidates of the 2016 election, Clinton and Trump, share the same characteristics: a U-shape profile, with very high number of grades 0, few intermediate grades, and a rather high number of maximal grades 5. These are the characteristics of a kind of candidate one might term “divisive” or “exclusive” as defined above: they receive strong support, but are also often rejected. In principle, exclusive candidates might be exclusive for at least two different reasons: (1) they might be candidates who generate intense feelings among sincere voters who thus grade them in an “extremist” way; or (2) they might be candidates who have a chance of winning and therefore are voted for in an “extremist” way by strategic voters, even though these strategic voters might otherwise have moderate feelings about them. Note that this second explanation does not fit well in our study, for HA shows the exact same tendencies as RV.

Figure 4: Grade profiles



Sanders’s profile exhibits the reverse features, showing a flatter grade profile. Sanders represents another kind of candidate, whom one may call “inclusive”: as said above, he does benefit from widespread support by the voters, but with lower levels of strong rejection or attachment. This kind of candidate is the one who loses more with plurality voting since this method is unable to take into account moderate support, contrary to the additive multi-nominal voting methods.

Finally, one can observe a third kind of candidate, illustrated by the Stein profile. Candidates such as Stein receive a large number of minimal grades, without benefiting from strong support otherwise. Thus, their grade profile is decreasing from bad to good grades. This third category of candidates typifies the “minor” candidates, who have no chance of winning and do not elicit much expressive support. Contrary to the exclusive and inclusive kinds of candidates, these minor candidates are often unknown to many voters and do not receive significant media exposure. This is especially common in the US because the characteristics of the plurality voting system and the Electoral College lead to the media neglecting minor candidates. In this third category, in addition to Stein, it would be fair to include all the candidates of the long set except the exclusive Clinton and Trump and the “inclusive” Sanders.

To go further, one can describe the proximity between candidates at the voter level. The “agreement” matrix based on AV data is shown in Table 5.²¹ This provides the candidates’ ability to attract voters who also support other candidates. For an election with the nine

²¹ For the sake of readability, we do not report standard errors in this matrix.

candidates of the long set, the agreement matrix for AV has 81 values. Each row in the matrix displays the percentage of the supporters of a specific candidate who approve of another (column) candidate. Consequently, the diagonal is always equal to 100%, but the matrix is not symmetrical. For instance, the proportion of Clinton’s supporters who also support Bloomberg is 13%, while the proportion of Bloomberg’s supporters who also support Clinton is 46%.

The columns of the table show the *cross-over voter support* of a candidate, i.e., a candidate’s ability to attract voters who also support other candidates. The rows, on the other hand, show the *dispersion* of a candidate, i.e., a candidate’s propensity to share supporters with other candidates. From this, one can compute synthetic measures of average cross-over voter support and dispersion for each candidate as the simple average of the columns and rows outside the diagonal. The candidate with the highest cross-over voter support is Sanders with a measured cross-over support of 35%, consistent with his “inclusive” type. Regarding the dispersion, the lowest averages are obtained by Trump and Clinton (9 and 10%, respectively). Again, this is consistent with their “exclusive” type—their supporters are heavily focused on them and tend not to share support with many other candidates.

Table 5: Approval agreement matrix (%)

| (%) | Clinton | Trump | Sanders | Cruz | Johnson | Bloomberg | McMullin | Stein | Castle | Dispersion |
|-------------------|---------|-------|---------|------|---------|-----------|----------|-------|--------|------------|
| Clinton | 100 | 4 | 52 | 5 | 5 | 13 | 4 | 5 | 1 | 10 |
| Trump | 5 | 100 | 13 | 32 | 10 | 7 | 9 | 3 | 2 | 9 |
| Sanders | 52 | 11 | 100 | 12 | 14 | 17 | 3 | 10 | 2 | 13 |
| Cruz | 9 | 50 | 22 | 100 | 18 | 12 | 20 | 5 | 5 | 16 |
| Johnson | 18 | 28 | 44 | 32 | 100 | 12 | 13 | 17 | 9 | 19 |
| Bloomberg | 46 | 21 | 56 | 22 | 13 | 100 | 15 | 5 | 3 | 20 |
| McMullin | 20 | 41 | 18 | 55 | 20 | 24 | 100 | 7 | 13 | 22 |
| Stein | 36 | 17 | 77 | 22 | 42 | 12 | 10 | 100 | 6 | 25 |
| Castle | 18 | 33 | 33 | 48 | 51 | 15 | 44 | 13 | 100 | 28 |
| <i>Cross-over</i> | 23 | 23 | 35 | 25 | 19 | 12 | 13 | 7 | 4 | |

In all, corroborating Baujard et al.’s results (2014), mono-nominal rules such as plurality favor “exclusive” candidates, compared with multi-nominal, evaluative, methods such as AV and RV that benefit “inclusive” candidates.²² In the 2016 US election case, it appears that the latter rules with a large set of candidates would indeed allow the election of the candidate Sanders—whom this study finds to be favored according to several different criteria. Neither direct plurality nor the official result, however, would have led to Sanders winning. One explanation for this situation is that, while Sanders did not win the actual Democratic Party primary, our data indicates that he had large approval voting crossover support with Clinton, Johnson, and Stein. Further, Sanders’s crossover support from Trump voters was nontrivial at 13% (more than double the Trump crossover support that Clinton got).

²²In this comparison, IRV, despite the fact that it allows the voter to rank the candidates, appears closer to plurality. This is a reminder that the IRV counting process is a “single transferable vote” process whose effect might be close to the effects of two-round majority voting.

5. Concluding remarks

Based on a representative sample of more than 2,000 respondents, we found that, according to their stated true electoral preferences, before the 2016 US election Bernie Sanders appears both as the utilitarian optimum and the Condorcet winner. Our results suggest that multi-nominal evaluative voting rules such as RV and, potentially, AV would have led to the election of Sanders among a large set of candidates. This is in contrast to direct plurality and to the official voting rule.

In this study we have focused our inquiry on three main questions: (1) How do US voters use these alternative rules? (2) What is the influence of the chosen set of candidates (short vs. long sets) on the outcome of the election? (3) Do different voting methods differ in their election outcomes?

First, our results are consistent with the interpretation that the respondents really did use the tested multi-nominal rules to express their preferences more widely and sincerely. This is particularly true for range voting whose distribution of grades presents significant similarities with voters' honest assessments (HA). Second, we showed that the number of candidates running might have a large effect on the outcome depending on the voting method used. The multi-nominal rules, and RV in particular, are less sensitive to other "eliminated" candidates. This makes those methods closer to voters' true preferences and thus closer to sincere outcomes. Plurality voting, on the other hand is highly impacted by the number of candidates because of its mono-nominal feature. Third, the multi-nominal rules could indeed elect the "best" candidate according to voters' true preferences. While AV and RV favor the election of inclusive candidates who are appreciated by a large proportion of voters, plurality tends to prefer exclusive candidates who are either strongly supported or strongly rejected. In the 2016 US case, Sanders can be seen as an inclusive candidate while both Clinton and Trump were exclusive. Further, our data show RV and AV capturing much more support for minor party candidates compared to plurality and IRV (Table A5). AV was still able to do capture that minor party support despite most voters choosing only one candidate (Figure A1).

The present study is based on a single poll, made in specific circumstances so that, as explained above, does not allow an explicit inference on what American politics would be under other circumstances. That said, all the observations made are consistent with the empirical literature. It is thus legitimate to take stock of the received knowledge and to state in which direction our result point with respect to the question of electoral reform. According to our results, it appears that changing the official voting rule in the US in favor of multi-nominal methods would be a good reform for American voters because the outcome would be more faithful to their true electoral preferences.

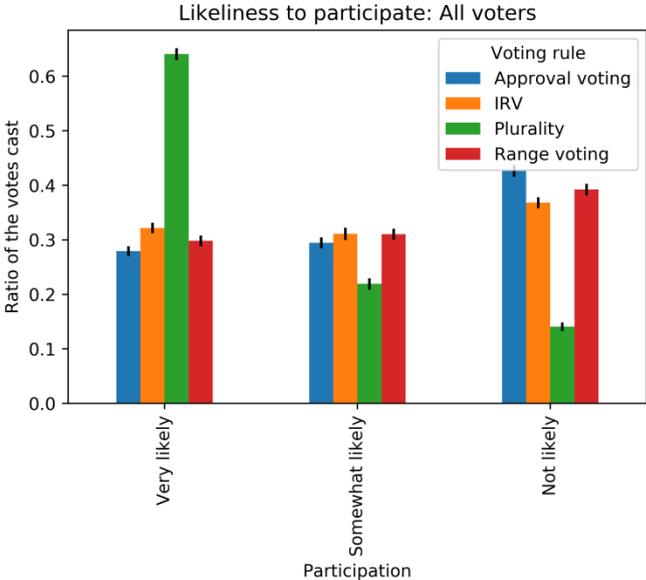
Of course, many important issues remain unaddressed at this stage. One could legitimately wonder whether the American people are prepared to opt for another voting rule. One question in our survey deals with the likeliness to participate in a vote under one or the other voting method, and might be of some help to clarify this point. Figure 5 shows the percentages of voters who would agree to take part in a vote under plurality, AV, RV, and IRV. As one clearly sees, plurality incontestably remains the favorite voting method of the respondents. Complementary studies might put the emphasis on trying to understand the nuance of this phenomenon, which is in tension with the fact that some US cities have indeed adopted alternative voting rules following education campaigns. Given that, it appears that voters need to be educated about voting rules beyond just learning their mechanics.

Defining what it means to “understand a voting rule” is by no means trivial. In a survey on voting experiments Laslier (2011) distinguished between voters’ difficulties in completing the task (choosing, ranking, grading...), understanding of the voting rule per se (how ballots are counted...) and understanding the political consequences of using one rule or the other. He argues that there is generally no problem with the first point, but that the two other points must be treated with care.

The present paper offers some hints there. We indeed spotted no practical difficulties. We found cases that seem to imply that the voter misunderstands the rule itself; these cases are 2% of IRV ballots which are both non sincere and irrational, due to the subtle way IRV iterative counting works. One cannot exclude that these cases are just due to inattentive respondents, so that strong conclusions should be avoided on that point.

AV ballots in general seem both sincere and rational, but we observed a large number of single-name approval ballots (about half of the ballots on the long set of candidates). This is more than what has been described elsewhere (as for the experiments on AV conducted during the French presidential elections from 2002 to 2017, the proportions of single-name approval ballots ranged from 11.1% to 26.2%, see notably Baujard et al. 2016), but is not surprising in the American context with its traditionally small supply of candidates. Lastly, the question of understanding political consequences is beyond the scope of this work. Clearly, further studies in these directions are needed.

Figure 5: Voters’ likeliness to participate



References

Abbott, D. W., Levine J. P., 1991. Wrong Winners: The Coming Debacle in the Electoral College. Praeger, New York.

Alaska Division of Elections, 2020. Ballot measure 19AKBE. An Act Replacing the Political Party Primary with an Open Primary System and Ranked-Choice General Election, and Requiring Additional Campaign Disclosures.

- https://www.elections.alaska.gov/petitions/19AKBE/19AKBE_Ballot_Summary_FINAL.pdf (accessed 15 December 2020).
- Alós-Ferrer, C., Granic, D.-G., 2010. Approval Voting in Germany: Description of a Field Experiment, in Laslier, J.-F., Sanver, M. R. (Eds.), *Handbook on Approval Voting*. Springer, Berlin, pp. 397–411.
- AP-New, 2020a. Maine to use ranked voting for president after repeal fails. <https://apnews.com/article/02eaaa35b5285f208baef64722d4bfb9> (accessed 15 December 2020).
- AP-New, 2020b. Saint Louis voters approve nonpartisan elections. <https://apnews.com/article/election-2020-primary-elections-st-louis-general-elections-elections-cba7eb3251d5479b9375d55db428d429> (accessed 15 December 2020).
- Arrow, Kenneth J. 1963. *Social Choice and Individual Values*. New Haven: Yale University Press.
- Barthélémy, Fabrice, Matthieu Martin, and Ashley Piggins. 2014. “The Architecture of the Electoral College, House Size Effects, and the Referendum Paradox.” *Electoral Studies* 34: 111–118.
- Baujard, Antoinette, and Herrade Igersheim. 2010. “Framed field experiments on approval voting: Lessons from the 2002 and 2017 French presidential elections.” In *Handbook on Approval Voting*, edited by Jean-François Laslier and M. Remzi Sanver, 357–395. Berlin: Springer.
- Baujard, Antoinette, Frédéric Gavrel, Herrade Igersheim, Jean-François Laslier, and Isabelle Lebon. 2014. “Who’s Favored by Evaluative Voting? An Experiment Conducted During the 2012 French Presidential Election.” *Electoral Studies* 34: 131–145.
- Baujard, Antoinette, Herrade Igersheim, and Isabelle Lebon (2020). “Some regrettable grading scale effects under different versions of evaluative voting” *Social Choice and Welfare* (in print) <https://doi.org/10.1007/s00355-020-01300-z>.
- Blais, André, and Indridi Indridason. 2007. “Making candidates count: the logic of electoral alliances in two-round legislative elections” *The Journal of Politics* 69(1): 193–205.
- Bouton, Laurent, Micael Castanheira, and Aniol Llorente-Saguer (2016) “Divided majority and information aggregation: Theory and experiment” *Journal of Public Economics* 134: 114-128.
- Brams, Stephen J., and Peter C. Fishburn. 1978. “Approval voting.” *American Political Science Review* 72, 831–847.
- Brams, Stephen J., and Peter C. Fishburn. 1983. *Approval Voting*. Boston: Birkhäuser.
- Crowder-Meyer, Melody, Shana Kushner Gadarian, and Jessica Trounstein. 2020. “Voting can be hard, information helps.” *Urban Affairs Review* 56: 1 <https://doi.org/10.1177/1078087419831074>
- De Sinopoli, Francesco, Giovanna Iannantuoni, and Carlos Pimienta (2014) “Counterexamples on the superiority of Approval versus Plurality” *Journal of Public Economic Theory* 16: 824-834.
- Dougherty, Keith and Edward, Julian, 2011, *The Calculus of Consent and Constitutional Design*, Springer.
- Duverger, Maurice. 1951. *Les partis politiques*. Paris: Armand Colin. English translation: *Political Parties: Their Organization and Activities in the Modern State*. 1959. New York: John Wiley and Sons.
- Fiorina, Morris P. (1976) “The voting decision: instrumental and expressive aspects” *Journal of Politics* 38: 390-413.

- Forsythe, Robert, Roger B. Myerson, Thomas A. Rietz, and Robert J. Weber (1993) “An experiment on coordination in multi-candidate elections: The importance of polls and election histories” *Social Choice and Welfare* 10: 223-247.
- Gaertner, Wulf, and Youngsheng Xu (2012) “A general scoring rule.” *Mathematical Social Sciences* 63: 193–196.
- Grofman, Bernard, Bernard Dolez, and Annie Laurent, eds. 2011. *In Situ and Laboratory experiments on Electoral Law Reform: French Presidential Elections*. Berlin: Springer.
- Hillinger, Claude. 2004a. “On the possibility of democracy and rational collective choice.” Discussion paper 2004-21, Department of Economics, University of Munich.
- Hillinger, Claude. 2004b. “Voting and the cardinal aggregation of judgments.” Discussion paper 2004-09, Department of Economics, University of Munich.
- Koc, Edwin W. 1988. “An Experimental Examination of Approval Voting under Alternative Ballot Conditions.” *Polity* 20: 688–704.
- Kurrild-Klitgaard, Peter. 2018. “Trump, Condorcet and Borda: Voting paradoxes in the 2016 Republican presidential primaries.” *European Journal of Political Economy*, 55: 29-35.
- Laslier, Jean-François. 2019. *Voter Autrement: Le recours à l'évaluation*. Paris : Presses de l'École Normale Supérieure,
- Laslier, Jean-François, André Blais, Damien Bol, Sona N. Golder, Philipp Harfst, Laura B. Stephenson, and Karine Van der Straeten. 2015. “The Euro Vote Plus Experiment.” *European Union Politics* 16: 601–615.
- Laslier, Jean-François, and M. Remzi Sanver (eds). *Handbook on Approval Voting*. Berlin: Springer.
- Macé, Antonin. 2018. “Voting with Evaluations: Characterizations of Evaluative Voting and Range Voting”. *Journal of Mathematical Economics*, 79: 10–17. <https://doi.org/10.1016/j.jmateco.2018.09.002>.
- Miller, Nicolas R. 2014. “The House Size Effect and the Referendum paradox in U.S. Presidential Elections.” *Electoral Studies* 35: 265–271.
- Nagel, Jack H. 1984. “A Debut for Approval Voting.” *PS: Political Science & Politics* 17: 61–65.
- Núñez, Matias, and Jean-François Laslier. 2014. “Preference Intensity Representation: Strategic Overstating in Large Elections.” *Social Choice and Welfare* 42: 313–340.
- Pivato, Marcus. 2013. “Formal utilitarianism and range voting.” *Mathematical Social Sciences* 67: 50–56.
- Riley, Jonathan. 1990. “Utilitarian Ethics and Democratic Government.” *Ethics* 100: 335–348.
- Spenkuch, Jörg L. (2018) “Expressive vs. strategic voters: and empirical assessment” *Journal of Public Economics* 165: 73-81.
- Stephenson, Laura B., John H. Aldrich ,and André Blais (eds.) (2018) *The Many Faces of Strategic Voting: Tactical Behavior in Electoral Systems Around the World*. University of Michigan Press.
- Vox, 2018. This city just approved a new election system never tried before in America. <https://www.vox.com/future-perfect/2018/11/15/18092206/midterm-elections-vote-fargo-approval-voting-ranked-choice> (accessed 15 November 2020).

Appendix A: Complementary statistics

In the tables, standard errors are in parenthesis and the pictures show the standard error bars.

Table A1: Number of approved candidates per ballot

| Number of approvals on the ballot | % of ballots, short set | % of ballots, long set |
|-----------------------------------|-------------------------|------------------------|
| 0 | 7.68 (0.81) | 6.85 (0.87) |
| 1 | 74.13 (1.30) | 49.17 (1.47) |
| 2 | 14.15 (1.10) | 26.09 (1.26) |
| 3 | 3.86 (0.55) | 13.07 (0.85) |
| 4 | 0.19 (0.11) | 4.09 (0.62) |
| 5 | | 0.48 (0.17) |
| 6 | | 0.12 (0.07) |
| 7 | | 0.14 (0.10) |

Figure A1: Number of candidates approved per ballot

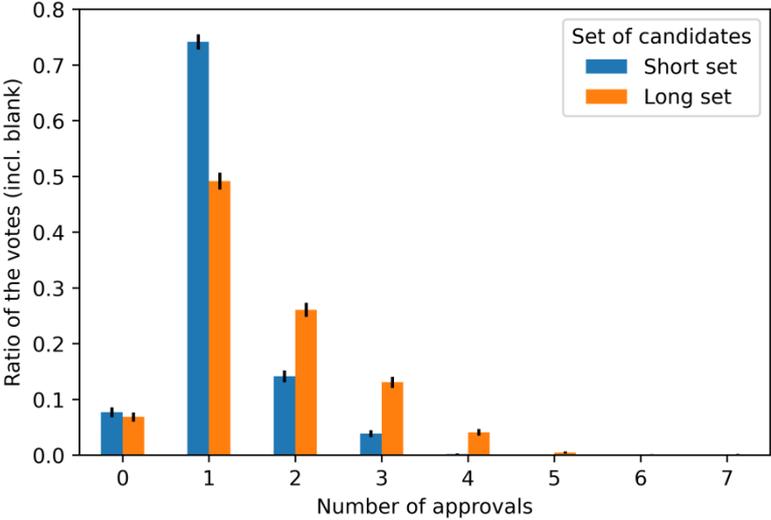


Table A2: Distribution of grades under Range Voting

| Grade | %, short set | %, long set |
|-------|--------------|--------------|
| 0 | 39.88 (0.90) | 39.56 (0.92) |
| 1 | 13.33 (0.54) | 13.06 (0.54) |
| 2 | 12.42 (0.66) | 12.26 (0.46) |
| 3 | 12.56 (0.56) | 13.06 (0.40) |
| 4 | 6.73 (0.39) | 9.84 (0.33) |
| 5 | 15.07 (0.56) | 12.23 (0.36) |

Figure A2: Distribution of grades under Range Voting

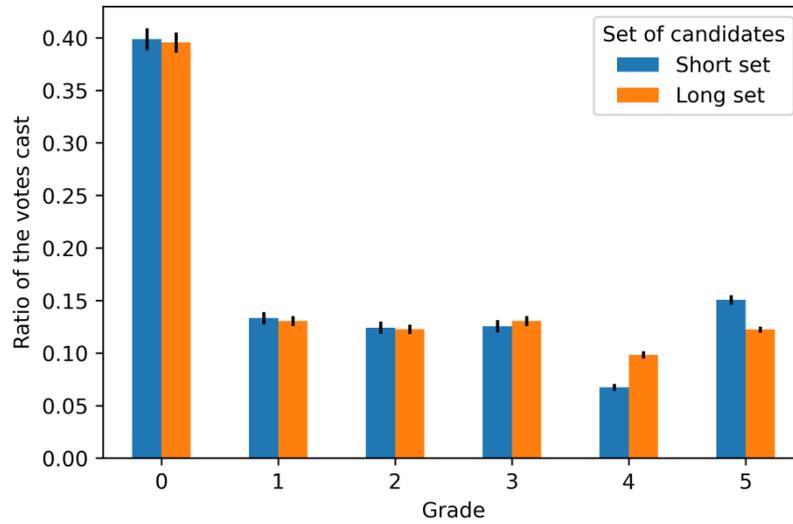


Table A3: Average grades under HA

| | short set grade | long set grade |
|-----------|--------------------|-------------------|
| Clinton | 2.20 (.07) | 2.11 (.07) |
| Trump | 1.75 (.07) | 1.77 (.07) |
| Sanders | | 2.35 (.07) |
| Cruz | | 1.53 (.06) |
| Johnson | 1.30 (.05) | 1.07 (.05) |
| Bloomberg | | 1.26 (.05) |
| McMullin | | 0.93 (.04) |
| Stein | 1.03 (.05) | 0.78 (.05) |
| Castle | | 0.63 (.04) |

Table A4: Distribution of grades under HA

| Grade | %, short set | %, long set |
|-------|--------------|--------------|
| 0 | 49.87 (0.87) | 54.53 (0.94) |
| 1 | 10.68 (0.59) | 9.08 (0.48) |
| 2 | 8.98 (0.46) | 8.40 (0.40) |
| 3 | 8.94 (0.52) | 9.15 (0.36) |
| 4 | 5.12 (0.37) | 6.47 (0.27) |
| 5 | 16.41 (0.45) | 12.36 (0.29) |

Figure A3: Distribution of grades under HA

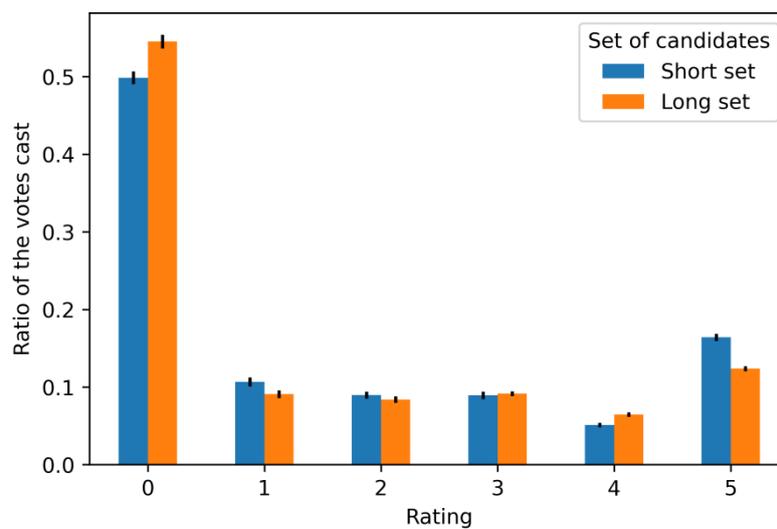


Table A5: Comparing scores between short and long sets (normalized scores)

| | Plurality | | AV | | RV | | IRV | | HA | |
|---------|-----------|------|-------|------|-------|------|-------|------|-------|------|
| | Short | Long | Short | Long | Short | long | Short | Long | Short | Long |
| Clinton | 48 | 31 | 50 | 40 | 47 | 46 | 46 | 30 | 44 | 42 |
| Trump | 41 | 28 | 42 | 34 | 38 | 40 | 39 | 28 | 35 | 35 |
| Johnson | 8 | 5 | 21 | 12 | 32 | 30 | 11 | 4 | 26 | 21 |
| Stein | 4 | 0 | 12 | 5 | 26 | 24 | 4 | 0 | 21 | 16 |

Table A6: Instant Runoff results (short set): % after each round of elimination

| | 0 | 1 | 2 | 3 |
|---------|----|----|----|-----|
| Clinton | 46 | 48 | 54 | 100 |
| Trump | 39 | 39 | 46 | |
| Johnson | 11 | 13 | | |
| Stein | 4 | | | |

Table A7: Instant Runoff results (long set): % after each round of elimination

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----------|----|----|----|----|----|----|----|----|-----|
| Clinton | 30 | 30 | 30 | 30 | 31 | 32 | 33 | 54 | 100 |
| Trump | 28 | 28 | 28 | 28 | 29 | 31 | 39 | 46 | |
| Sanders | 21 | 21 | 21 | 21 | 22 | 25 | 28 | | |
| Cruz | 10 | 10 | 10 | 11 | 12 | 13 | | | |
| Johnson | 4 | 5 | 5 | 5 | 6 | | | | |
| Bloomberg | 4 | 4 | 4 | 5 | | | | | |
| McMullin | 2 | 2 | 2 | | | | | | |
| Stein | 0 | 0 | | | | | | | |
| Castle | 0 | | | | | | | | |

Note on the statistical robustness of IRV results: Working on the long set by bootstrapping our data set, one finds the following. With confidence .962, the first two eliminated candidates are Castle and Stein. With confidence .988, the first three eliminated are Castle, Stein and McMullin: they are followed with full confidence by Bloomberg and Johnson or by Johnson and Bloomberg, and then by Cruz. Therefore, the last three remaining candidates are always Clinton, Trump, and Sanders. Three endings occur: (i) elimination of Sanders then Trump, victory of Clinton (frequency .981); (ii) elimination of Clinton then Trump, victory of Sanders (frequency .151); (iii) elimination of Sanders then Clinton, victory of Trump

(frequency .039). The confidence for the statement “Clinton wins under IRV” is therefore .981.

Figure A4: Instant Runoff results – short set

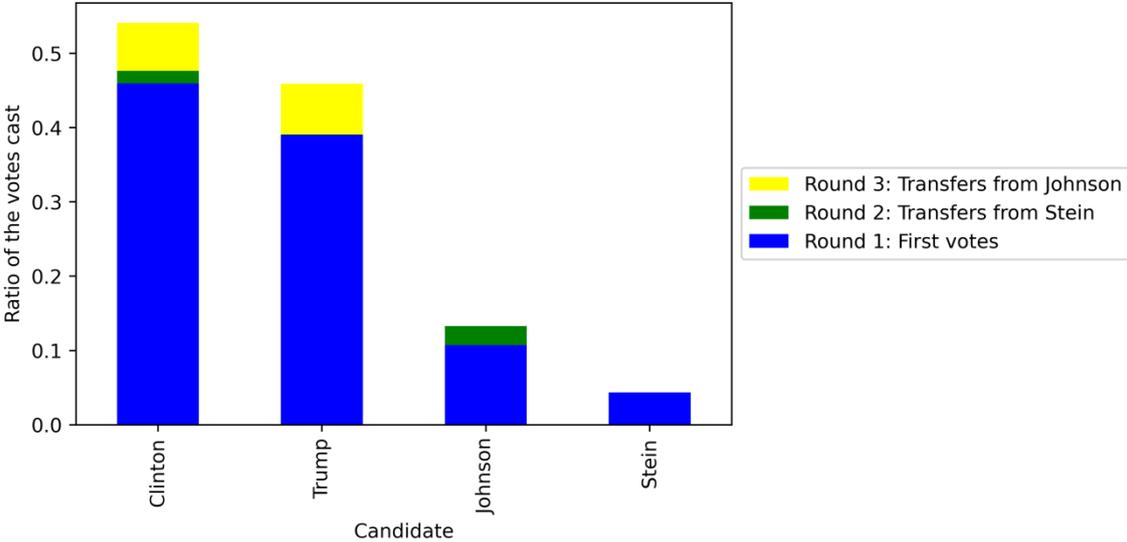


Figure A5: Instant Runoff results – long set

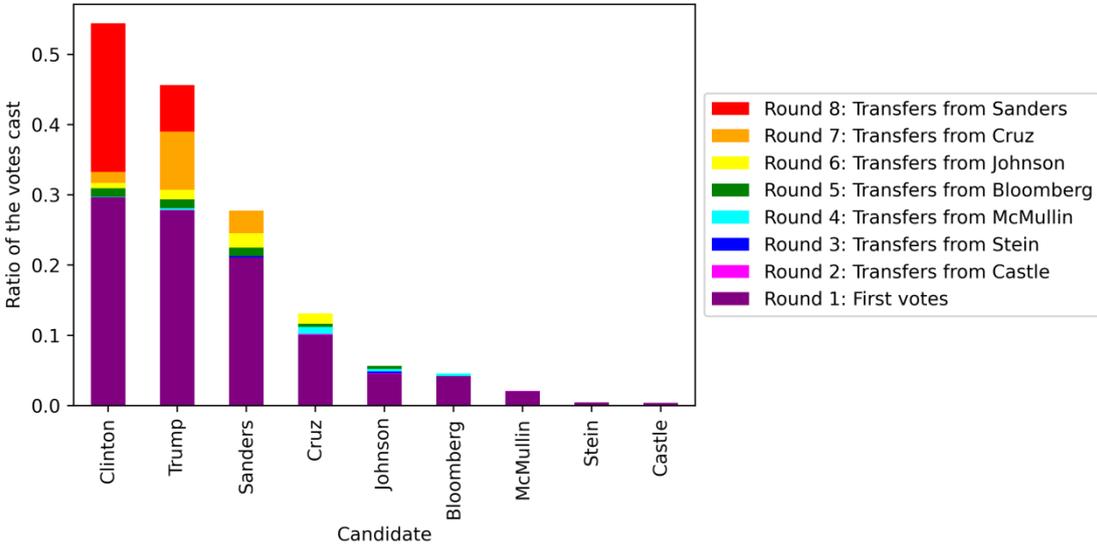


Table A8: HA pairwise comparison ordered matrix in % – long set²³

| | Sanders | Clinton | Trump | Cruz | Bloomberg | Johnson | McMullin | Stein | Castle | <i>Victories</i> |
|-----------|---------|---------|-------|------|-----------|---------|----------|-------|--------|------------------|
| Sanders | | 54 | 57 | 64 | 74 | 77 | 74 | 89 | 83 | 8 |
| Clinton | 46 | | 54 | 59 | 68 | 67 | 68 | 76 | 76 | 7 |
| Trump | 43 | 46 | | 57 | 59 | 62 | 68 | 68 | 74 | 6 |
| Cruz | 36 | 41 | 43 | | 56 | 61 | 71 | 67 | 79 | 5 |
| Bloomberg | 26 | 32 | 41 | 44 | | 55 | 60 | 65 | 72 | 4 |
| Johnson | 23 | 33 | 38 | 39 | 45 | | 56 | 61 | 72 | 3 |
| McMullin | 26 | 32 | 32 | 29 | 40 | 44 | | 55 | 69 | 2 |
| Stein | 11 | 24 | 32 | 33 | 35 | 39 | 45 | | 61 | 1 |
| Castle | 17 | 24 | 26 | 21 | 28 | 28 | 21 | 39 | | 0 |

Table A9: RV pairwise comparison ordered matrix in % – long set

| | Sanders | Clinton | Trump | Cruz | Bloomberg | Johnson | McMullin | Stein | Castle | <i>Victories</i> |
|-----------|---------|---------|-------|------|-----------|---------|----------|-------|--------|------------------|
| Sanders | | 56 | 58 | 63 | 72 | 74 | 73 | 86 | 82 | 8 |
| Clinton | 44 | | 54 | 56 | 63 | 63 | 65 | 73 | 74 | 7 |
| Trump | 42 | 46 | | 53 | 54 | 57 | 64 | 61 | 69 | 6 |
| Cruz | 36 | 44 | 47 | | 55 | 59 | 68 | 64 | 76 | 5 |
| Bloomberg | 28 | 37 | 46 | 45 | | 53 | 60 | 64 | 71 | 4 |
| Johnson | 26 | 37 | 43 | 41 | 47 | | 56 | 61 | 73 | 3 |
| McMullin | 27 | 35 | 36 | 32 | 40 | 44 | | 54 | 67 | 2 |
| Stein | 14 | 27 | 39 | 36 | 36 | 39 | 46 | | 62 | 1 |
| Castle | 18 | 26 | 30 | 24 | 29 | 27 | 33 | 38 | | 0 |

²³ The usual order of the candidates has been changed in this Table in order to stress the victories of Sanders and these of Bloomberg over Johnson.

Table A10: IRV pairwise comparison ordered matrix in % – long set

| | Clinton | Sanders | Trump | Cruz | Bloomberg | Johnson | McMullin | Stein | Castle | <i>Victories</i> |
|-----------|---------|---------|-------|------|-----------|---------|----------|-------|--------|------------------|
| Clinton | | 51 | 54 | 57 | 75 | 72 | 75 | 87 | 89 | 8 |
| Sanders | 49 | | 57 | 61 | 78 | 78 | 77 | 92 | 90 | 7 |
| Trump | 46 | 43 | | 63 | 65 | 70 | 84 | 75 | 90 | 6 |
| Cruz | 43 | 39 | 37 | | 61 | 67 | 83 | 72 | 88 | 5 |
| Bloomberg | 25 | 22 | 35 | 39 | | 53 | 62 | 63 | 78 | 4 |
| Johnson | 28 | 22 | 30 | 33 | 47 | | 57 | 61 | 78 | 3 |
| McMullin | 25 | 23 | 16 | 17 | 38 | 43 | | 53 | 73 | 2 |
| Stein | 13 | 8 | 25 | 28 | 37 | 39 | 47 | | 67 | 1 |
| Castle | 11 | 10 | 10 | 12 | 22 | 22 | 27 | 33 | | 0 |

Table A11: Plurality results with HA data

| | short set % | short set rank | long set % | long set rank |
|-----------|-----------------|----------------|-----------------|---------------|
| Clinton | 44.85 (1.34) | 1 | 27.60 (1.36) | 1 |
| Trump | 36.89 (1.48) | 2 | 25.18 (1.51) | 2 |
| Sanders | | | 21.54 (1.26) | 3 |
| Cruz | | | 10.37 (0.08) | 4 |
| Johnson | 11.80 (1.02) | 3 | 4.59 (0.56) | 6 |
| Bloomberg | t | | 5.59 (0.59) | 5 |
| McMullin | | | 2.89 (0.48) | 7 |
| Stein | 6.47 (0.69) | 4 | 1.37 (0.26) | 8 |
| Castle | | | 0.09 (0.19) | 9 |

Table A12: IRV results with HA data

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----------|----|----|----|----|----|----|----|----|-----|
| Clinton | 27 | 27 | 27 | 28 | 29 | 31 | 35 | 55 | 100 |
| Trump | 23 | 24 | 24 | 25 | 26 | 27 | 34 | 45 | |
| Sanders | 23 | 23 | 23 | 23 | 25 | 27 | 31 | | |
| Cruz | 11 | 12 | 12 | 12 | 13 | 14 | | | |
| Bloomberg | 6 | 6 | 6 | 7 | 7 | | | | |
| Johnson | 5 | 5 | 5 | 5 | | | | | |
| McMullin | 3 | 3 | 3 | | | | | | |
| Stein | 1 | 1 | | | | | | | |
| Castle | 1 | | | | | | | | |

Table A13: Comparing rules on the long set (normalized scores in %)

| | Plurality | AV | RV | IRV |
|-----------|-----------|----|----|-----|
| Clinton | 31 | 40 | 46 | 30 |
| Trump | 28 | 34 | 40 | 28 |
| Sanders | 20 | 39 | 54 | 21 |
| Cruz | 10 | 21 | 39 | 10 |
| Johnson | 5 | 12 | 30 | 4 |
| Bloomberg | 4 | 12 | 33 | 4 |
| McMullin | 2 | 8 | 27 | 2 |
| Stein | 0 | 5 | 24 | 0 |
| Castle | 0 | 2 | 19 | 0 |

Table A14: Strategic grading: global

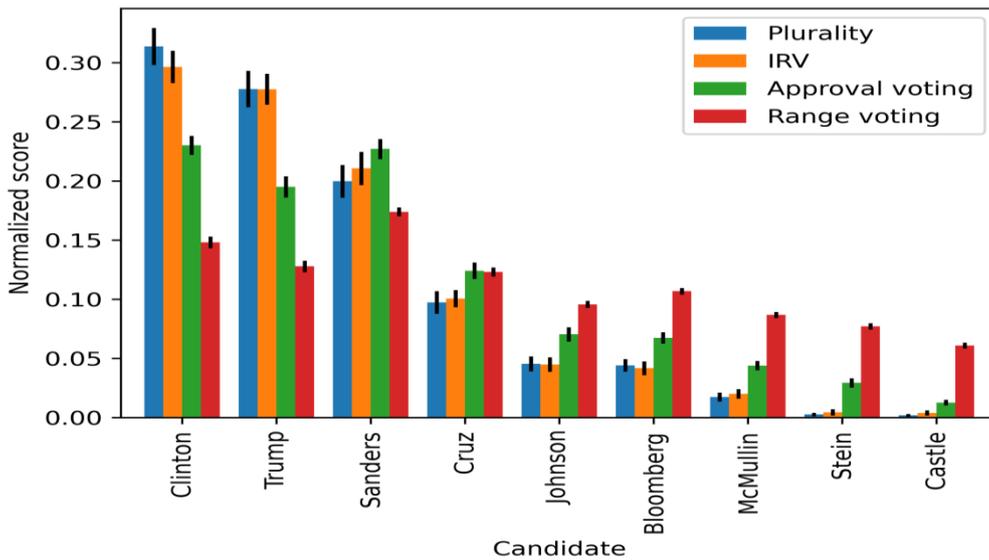
| | HA grade | RV grade | RV - HA |
|---------|----------|----------|---------|
| Clinton | 2.11 | 2.33 | .11 |
| Trump | 1.77 | 2.00 | .15 |
| Sanders | 2.35 | 2.72 | .30 |
| Cruz | 1.53 | 1.93 | .33 |

Note: The difference RV-HA is computed on the voters who provided both grades.

Table A15: Strategic grading: details

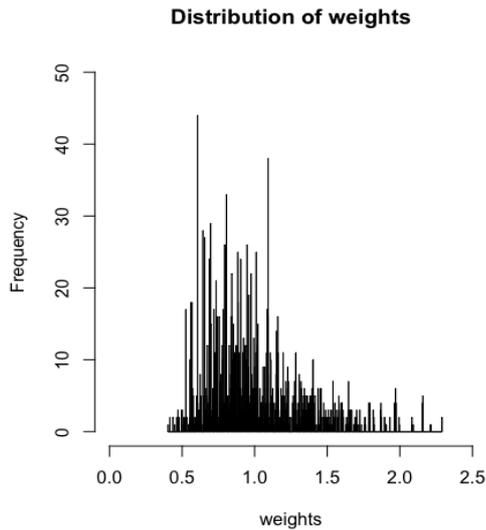
| | |
|--|---|
| <p>RV(Clinton) – HA(Clinton) :</p> <ul style="list-style-type: none"> ▪ .27 when HA(Clinton) < 5. ▪ .33 when HA(Sanders) = 5 and HA(Clinton) < 5 | <p>RV(Trump) – HA(Trump) :</p> <ul style="list-style-type: none"> ▪ .29 when HA(Trump) < 5. ▪ .52 when HA(Cruz) = 5 and HA(Trump) < 5 |
| <p>RV(Sanders) – HA(Sanders) :</p> <ul style="list-style-type: none"> ▪ .50 when HA(Sanders) < 5. ▪ .65 when HA(Clinton) = 5 and HA(Sanders) < 5 | <p>RV(Cruz) – HA(Cruz) :</p> <ul style="list-style-type: none"> ▪ .46 when HA(Cruz) < 5. ▪ .64 when HA(Trump) = 5 and HA(Cruz) < 5 |

Figure A6: Normalized relative scores



Appendix B: Sample quality and data management

The survey was undertaken by GfK, the source of the sample being the GfK Knowledge Panel. Additional information on this contract is available on demand. The survey was conducted in the period November 3 to 8, 2016. The general methodology is that the GfK Knowledge Panel is globally representative of the American population so that a sample taken from this panel requires limited re-weighting. The following picture depicts the weights that are applied to our voting sample. These weights range between .40 and 2.29 and achieve statistical congruence with the target population (that is the general adult population, age 18 and plus, English and Spanish language speakers) with respect to age, sex, education, income, ethnicity, and region in the country.



The data set provided by GfK Custom Research was of good quality, but it is important, independent of the quality of the survey, to check to what extent the questions themselves were correctly filled in by the respondents. To do so, the consistency of the voters' answers has been assessed through the four tested voting rules plus HA. Accordingly, voters were split into five groups: (1) OK (auto): the data on the voters were automatically checked (all pairwise covariances between two voting rules were nonnegative); (2) OK (manual): the data on the voters were manually checked and no correction was necessary; (3) Corrected: some simple correction was applied and the result makes sense with no doubt; (4) Dubious: whether some correction was applied or not, the data remain dubious. (5) Excluded: the voter is completely excluded.

All in all, as shown in Table B1, more than 95% of the voters belong to categories (1), (2) and (3). To maximize the quantity of data, we have decided to include Category (4) into our analyses, while a very small percentage of our data has been excluded: 10 voters for the short set and 10 for the long set, which corresponds to about 1% of our observations. Let us offer a few examples of each category, which leads to a better understanding of the data, their meaning and scope.

Table B1: Assessing voter's consistency

| <i>Voter status</i> | % |
|---------------------|-------|
| OK (auto) | 87.46 |
| OK (manual) | 2.77 |
| Corrected | 5.13 |
| Dubious | 3.64 |
| Excluded | 1.00 |

Table B2 illustrates the case of a voter in the long set whose votes are consistent all along the four tested voting rules plus HA with no need for manual checking or extra corrections. This voter supports Clinton under plurality, approves of Clinton and Sanders under AV, ranked Clinton first, Sanders second and Bloomberg third under IRV and gives the maximum grade to Clinton and Sanders under RV and HA. Table B3 also shows a consistent voter of the long set, but with manual checking. There is no need for corrections, for their approval of McMullin under AV and their ranking under IRV might be caused by strategic considerations. Table B4 illustrates a voter of the short set whose electoral choices required some corrections. Indeed, the maximum grades she gave to Clinton and Stein under RV and to Clinton under HA show that they misunderstood the instructions of IRV. Thus, we corrected their votes under IRV by reversing her ranking, which makes all their responses consistent. Table B5 presents a dubious case. Here, the voter's choice of Cruz under plurality (and AV) might seem weird, but remains plausible. Thus, the data from this voter are kept with no further correction. Eventually, Table B6 illustrates the case of a voter of the short set who has been excluded from our analyses on the grounds that her responses are systematically inconsistent.

Now, regarding the importance and the kind of corrections we did, Table B7 clearly demonstrates that some voters have encountered difficulties understanding the instructions of IRV, RV, and HA for they reversed their ranking and grades as in Table B4 for IRV. Under RV and HA, they gave for instance the grade 1 to their favorite candidates instead of 5. Under IRV, a single ballot gave the rank 3 twice, so that we have removed both of them for it seems to be a simple mistake. Besides, some ballots were difficult to understand, but here this does not appear to be a misunderstanding of the instructions, we thus kept the data as they were. All in all, Table B7 shows that the quasi-entirety of our data are reliable, which have led to a very small percentage of exclusions.

Table B2: Assessing voters' consistency – Category 1

| | Original | | | | | Corrected | | | | |
|------------|-----------|----|-----|----|----|-----------|----|-----|----|----|
| | Plurality | AV | IRV | RV | HA | Plurality | AV | IRV | RV | HA |
| Clinton | 1 | 1 | 1 | 5 | 5 | 1 | 1 | 1 | 5 | 5 |
| Trump | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Sanders | 0 | 0 | 2 | 5 | 5 | 0 | 0 | 2 | 5 | 5 |
| Cruz | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Johnson | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Bloomberg | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| McMullin | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Stein | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Castle | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Correction | | | | | | | | | | |

Table B3: Assessing voters' consistency – Category 2

| | Original | | | | | Corrected | | | | |
|------------|-----------|----|-----|----|----|-----------|----|------|----|----|
| | Plurality | AV | IRV | RV | HA | Plurality | AV | IRV | RV | HA |
| Clinton | 0 | 0 | 3 | 2 | 0 | 0 | 0 | 3 | 2 | 0 |
| Trump | 1 | 1 | 2 | 4 | 4 | 1 | 1 | 2 | 4 | 4 |
| Sanders | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Cruz | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Johnson | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Bloomberg | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| McMullin | 0 | 1 | | 0 | 0 | 0 | 1 | | 0 | 0 |
| Stein | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Castle | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Correction | | | | | | | | keep | | |

Table B4: Assessing voters' consistency – Category 3

| | Original | | | | | Corrected | | | | |
|------------|-----------|----|-----|----|----|-----------|----|---------|----|----|
| | Plurality | AV | IRV | RV | HA | Plurality | AV | IRV | RV | HA |
| Clinton | 1 | 1 | 3 | 5 | 5 | 1 | 1 | 1 | 5 | 5 |
| Trump | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Johnson | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| Stein | 0 | 0 | 2 | 5 | 0 | 0 | 0 | 2 | 5 | 0 |
| Correction | | | | | | | | inverse | | |

Table B5: Assessing voters' consistency – Category 4

| | Original | | | | | Corrected | | | | |
|------------|-----------|----|-----|----|----|-----------|----|-----|----|----|
| | Plurality | AV | IRV | RV | HA | Plurality | AV | IRV | RV | HA |
| Clinton | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Trump | 0 | 0 | 1 | 5 | 5 | 0 | 0 | 1 | 5 | 5 |
| Sanders | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Cruz | 1 | 1 | 2 | 0 | 0 | 1 | 1 | 2 | 0 | 0 |
| Johnson | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| Bloomberg | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| McMullin | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Stein | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Castle | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | 0 |
| Correction | | | | | | keep | | | | |

Table B6: Assessing voters' consistency – Category 5

| | Original | | | | | Corrected | | | | |
|------------|-----------|----|-----|----|----|-----------|---------|---------|---------|---------|
| | Plurality | AV | IRV | RV | HA | Plurality | AV | IRV | RV | HA |
| Clinton | 0 | 0 | 2 | 5 | 5 | 0 | 0 | 2 | 5 | 5 |
| Trump | 0 | 1 | 1 | 5 | 0 | 0 | 1 | 1 | 5 | 0 |
| Johnson | 0 | 1 | | 0 | 0 | 0 | 1 | | 0 | 0 |
| Stein | 1 | 0 | 3 | 0 | 0 | 1 | 0 | 3 | 0 | 0 |
| Correction | | | | | | exclude | exclude | exclude | exclude | exclude |

Table B7: Kind of corrections per voting rule

| | Plurality | AV | IRV | RV | HA |
|-------------------|--------------|--------------|--------------|--------------|--------------|
| Exclude | 1.04 | 0.99 | 1.24 | 1.47 | 1.44 |
| Nothing to report | 98.73 | 98.92 | 96.58 | 94.86 | 96.12 |
| Difficult | 0.23 | 0.09 | Not relevant | 0.07 | 0.11 |
| Inverse | Not relevant | Not relevant | 2.15 | 3.60 | 2.33 |
| Remove 3 | Not relevant | Not relevant | 0.03 | Not relevant | Not relevant |

Appendix C: Survey questions

The order of the questions regarding the voting methods was randomized. Depending of the set to which they belong (short or long), respondents have been asked to make a choice between the candidates (four or nine) randomly ordered. At the end of the survey, respondents were requested to answer some supplementary socio-demographic questions.

Main survey questions (here, for the long set beginning with plurality voting)

There are a number of different methods by which people can vote in a democracy. We will next ask you how you would vote using 4 different voting methods, considering only the candidates presented.

1) The first method is called plurality voting. If the US presidential election were held today, and your ballot asked you to select only one candidate, who would you select? In the plurality voting method, the candidate with the highest number of votes would win. Please select ONE only.

- A. Hillary Clinton, Democratic Party
- B. Gary Johnson, Libertarian Party
- C. Jill Stein, Green Party
- D. Donald Trump, Republican Party
- E. Evan McMullin, Former Republican Policy Director for US House
- F. Bernie Sanders, Democratic Party Runner Up
- G. Ted Cruz, Republican Party Runner Up
- H. Michael Bloomberg, Former New York City Mayor
- I. Darrell Castle, Constitution Party

2) The next method is called approval voting. If the US presidential election were held today, and your ballot asked you to select all the candidates you wanted, which candidate or candidates would you select? In the approval voting method, the candidate selected the most would win. Please select ALL that apply.

- A. Hillary Clinton, Democratic Party
- B. Gary Johnson, Libertarian Party
- C. Jill Stein, Green Party
- D. Donald Trump, Republican Party
- E. Evan McMullin, Former Republican Policy Director for US House
- F. Bernie Sanders, Democratic Party Runner Up
- G. Ted Cruz, Republican Party Runner Up
- H. Michael Bloomberg, Former New York City Mayor
- I. Darrell Castle, Constitution Party

3) The next method is called range voting. If the US presidential election were held today, and your ballot asked you to score each candidate on a scale from 0-5, how would you score them? Please enter a number, where '0' means 'Worse' and '5' means 'Better'. In the range voting method, the candidate with the highest total score would win.

- A. Hillary Clinton, Democratic Party
- B. Gary Johnson, Libertarian Party
- C. Jill Stein, Green Party
- D. Donald Trump, Republican Party
- E. Evan McMullin, Former Republican Policy Director for US House
- F. Bernie Sanders, Democratic Party Runner Up
- G. Ted Cruz, Republican Party Runner Up
- H. Michael Bloomberg, Former New York City Mayor
- I. Darrell Castle, Constitution Party

4) The next method is called ranked choice voting. If the US presidential election were held today, and your ballot asked you to rank your top three candidates, how would you rank them? Use a '1' for your top choice, a '2' for your second choice, and a '3' for your third choice. In the ranked choice voting method, the winner requires a majority of first-choice votes. If no candidate had this majority, then the candidate with the fewest first-choice votes would be eliminated, and those votes would transfer to the next-preferred choices. This would repeat until a candidate had a majority.

- A. Hillary Clinton, Democratic Party
- B. Gary Johnson, Libertarian Party
- C. Jill Stein, Green Party
- D. Donald Trump, Republican Party
- E. Evan McMullin, Former Republican Policy Director for US House
- F. Bernie Sanders, Democratic Party Runner Up
- G. Ted Cruz, Republican Party Runner Up
- H. Michael Bloomberg, Former New York City Mayor
- I. Darrell Castle, Constitution Party

5) How likely would you be to vote in this presidential election if one of the following voting methods were used? For each of the voting methods listed below, how likely would you be to vote if it were the voting method that was being used?

- 1. Very likely
- 2. Somewhat likely
- 3. Not likely

- A. Plurality voting, where you chose only one
- B. Approval voting, where you could choose as many as you wanted
- C. Score voting, where you scored candidates on a scale
- D. Ranked choice voting, where you ranked your top 3 candidates

6) Regardless of their chance of being elected, how much do you honestly want the following to be elected? Please use a '0' for 'Do not want this person elected' to '5' for 'Very much want this person to be elected.'

- A. Hillary Clinton, Democratic Party
- B. Gary Johnson, Libertarian Party
- C. Jill Stein, Green Party
- D. Donald Trump, Republican Party

- E. Evan McMullin, Former Republican Policy Director for US House
- F. Bernie Sanders, Democratic Party Runner Up
- G. Ted Cruz, Republican Party Runner Up
- H. Michael Bloomberg, Former New York City Mayor
- I. Darrell Castle, Constitution Party