



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. 2018. halshs-03926997v1

HAL Id: halshs-03926997

<https://shs.hal.science/halshs-03926997v1>

Preprint submitted on 7 Jan 2019 (v1), last revised 6 Jan 2023 (v2)

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.



PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

WORKING PAPER N° 2018 – 55

Comparing Voting Methods: 2016 US Presidential Election

Herrade Igersheim
François Durand
Aaron Hamlin
Jean-François Laslier

JEL Codes: D72, C93

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



PARIS-JOURDAN SCIENCES ÉCONOMIQUES

48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 80 52 16 00=
www.pse.ens.fr

Comparing Voting Methods: 2016 US Presidential Election

Herrade Igersheim ‡

François Durand*

Aaron Hamlin †

Jean-François Laslier§

November, 20, 2018

Abstract. Before the 2016 US presidential elections, more than 2,000 participants participated to a survey in which they were asked their opinions about the candidates, and were also asked to vote according to different alternative voting rules, in addition to plurality: approval voting, range voting, and instant runoff voting. The participants were split into two groups, a first one facing a short set of four candidates (Clinton, Trump, Johnson and Stein), and a second one facing a long set of nine candidates (the previous four plus Sanders, Cruz, McMullin, Bloomberg, and Castle). The paper studies three issues: (1) How do U.S. voters effectively use these alternative rules? (2) What kind of candidates, in terms of individual preferences, is favored by which rule? (3) Which rules empirically satisfy the independence of irrelevant alternatives? Our results evidence that Bernie Sanders stands out as the “best” candidate in terms of individual preferences (using any standard criterion), 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

1 Introduction

Since November 2016, both the US and worldwide media have emphasized that the sitting Republican president Donald Trump did not get the majority of the votes. Indeed, he lost the “popular vote” by almost three million votes against 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 U.S. president via a different route. It’s an indirect winner-take-all system that adds up votes state by state. Those 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 unpopular elections (nearly one in 14) have occurred—with two occurrences in the last five elections. All these aspects have been already widely addressed by political scientists and social choice theoreticians (among others, see Abbott and Levine 1991, Ball and Leuthold 1991, Lepelley et al. 2011, Lahrach and Merlin 2012, Miller 2012, 2014, Barthélémy 2014, Kurrild-Klitgaard 2017).

But if the U.S. official voting method is often questioned for its two-tier aspect, discussions regarding its “*mono-nominal*” feature—that is, a voter can vote for only one candidate under plurality voting—are raised much less often. “*Multi-nominal*” voting rules (rules that allow voters to vote or grade more than one candidate) are commonly used for decision making in practice, settings like in schools, sports, online applications, etc., and 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 wished, thus increasing their overall satisfaction. But only a few voting theoreticians have analyzed the possible effects of these evaluative processes in general (see Hillinger 2004ab, Gaertner and Xu 2012, Pivato 2013, Macé 2018). The exception to this is the approval voting rule, a particular case of multi-nominal voting 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 (see notably Brams and Fishburn 1978, 1983, Laslier and Sanver 2010).

Experimentally, a stream of research has emerged that is testing approval voting and other multi-nominal rules alongside large-scale official elections. Such experiments have been conducted in France since 2002 in parallel to 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 been organized online also, in France, in Canada, and throughout Europe; see Laslier et al. 2015 and the web-site and vote.imag.fr. Lessons drawn from these studies are: (i) The principles of multi-nominal voting are easily understood by the voters, and they show a slight preference for range voting over approval voting. (ii) Such voting methods can indeed modify the ranking of candidates and can lead to different winners.

The U.S. is one of the not-so-many countries with a presidential regime to elect its president under universal¹ (even if indirect) suffrage. Another such country is France where, as stressed above, we have many studies. But, to the best of our knowledge, there is no research aiming to test multi-nominal voting rules in a U.S. national election context.

¹ For those who are 18 years or older and those who haven’t had a felony and aren’t currently in prison (some states restrict felons). This point is relevant especially given the high per capita incarceration rates in the US.

Compared to other countries, in Europe and elsewhere, the U.S. political system is highly specific, so that results obtained elsewhere might not be valid there. This study thus fills this gap by analyzing how U.S. voters react to multi-nominal voting rules and what the impact of these rules could be on election results. To do so, we use the data of 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 any number of candidates they want; the winning candidate is the one with the largest number of approvals.
- Range voting (RV): Voters give a grade to each candidate on a six-level scale (from 0 to 5, inclusive, with 0 being the worst grade and 5 the best); the winning candidate is the one who receives the largest total number of points.
- 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.

Additionally, participants to the survey were asked to assess honestly the candidates on a six-level scale basis. 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 conceived to mimic the actual 2016 election, and that same election with a more competitive context. Our analyses focus on three issues. (1) How do U.S. voters effectively use these alternative rules? (2) How do candidate options influence the election outcome? (3) Do alternative voting methods influence the outcome compared to plurality voting?

According to our data, Bernie Sanders stands out as the “best” candidate according to voters’ true preferences. As will be seen in section 3.3, if one considers the honest assessment of the candidates by the voters (HA), Sanders wins according to several aggregation rules, be them 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” candidate: range voting would clearly elect Sanders, and approval voting could elect either Sanders or Clinton. (2) Such is not the case for direct plurality and instant runoff voting (both would elect Clinton) and 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 tested voting methods. 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, and a fifth section concludes.

2 Survey and data

The Center for Election Science designed and contracted the survey from the international firm Growth from Knowledge (GfK). The survey ran from November 3, 2016 to the day of the U.S. presidential election on November 8, 2016 (noon). GfK used a panel of 4,181 members of which 2,552 responded and 2,367 participants 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 nationally representative participants, but with no specific target beyond that: general population adults (18+), English- and Spanish-language survey-takers.

Let us now describe the six specific questions we will deal with in this paper. Four of them corresponded to the four voting methods mentioned in the introduction: plurality, approval voting, range voting from 0 to 5, and instant-runoff voting. The two remaining questions dealt with the participants' assessment of the experimented voting rules, plus an "honest" assessment of the running candidates from 0 to 5, 5 being the best grade and 0 the worst.

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 sitting US President Donald Trump for the Republican Party. The five remaining candidates of the long set belonged to several political trends, either popular runners 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 nominee of the Constitution Party though he was not on enough state ballots to be able to get elected.² 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 have filled in the survey. After data cleaning—described in the appendix B—we had few inconsistent responses: 10 for the long set and 10 for the short set. What follows is thus based on a set of $1,188 + 1,159 = 2,347$ respondents. This shows that exclusion was rare, suggesting that respondents took the survey seriously.

3 Results and main lessons on voters' behavior

3.1 Global behavior and assessment of the tested voting rules

Global statistics offer 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

² US 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.

candidates) with the long set.³ Former European experiments had between 8 and 16 candidates.

In a 1984 AV experiment where 300 Pennsylvanian college students were asked to support or not 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 similarly observed that it was “low given the results of other AV experiments where 2.0 votes per ballot has been the norm”, thus referring to Nagel (1984). Similarly, the average numbers of approvals observed in France in similar experiments were much higher: 3.15 for sixteen candidates in 2002, 2.23 for twelve in 2007 and 2.63 for ten in 2012 and 2.42 for eleven in 2017.

One can notice first that there is a general (and quite intuitive) correlation between the number of running candidates and the average number of approvals. But beyond this observation, compared to U.S. 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 big candidates at the first round, so voters are used to face more than two important candidates. That contrasts with the U.S in both the voting method and the number of candidates. In any case, these relatively low observed averages imply that a fraction of the voters approve of only one candidate (“bullet voting”). As shown in Table 1, for 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, compared to French experiments, the percentages of participants who approved of one candidate only never exceeded 25% there. 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 measured support for other candidates.

Table 1: Number of approved candidates per ballot

Number of approvals on the ballot	% of ballots, short set	% of ballots, long set
0	7.68	6.85
1	74.13	49.17
2	14.15	26.09
3	3.86	13.07
4	0.19	4.09
5	Not relevant	0.48
6	Not relevant	0.12
7	Not relevant	0.14

Turning to range voting, it is relevant to stress that the participants took benefit of the increased opportunities to express their electoral preferences. The average grades given to

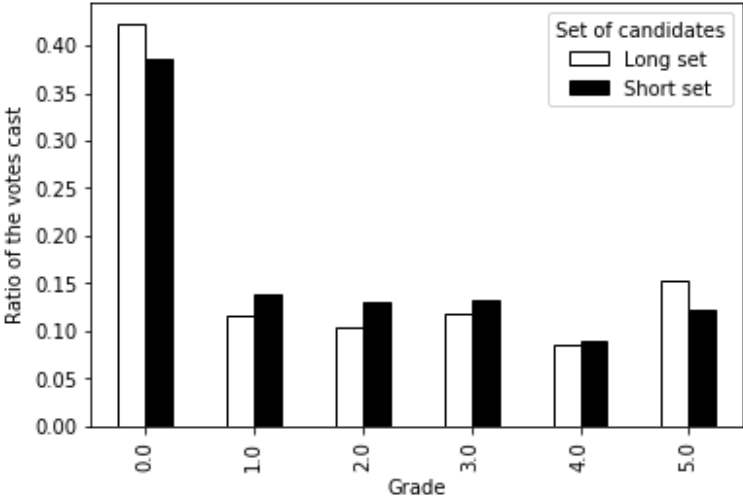
³ With those abstaining from voting for any candidate, these averages fall to 1.15 and 1.61, respectively.

candidates are very close for the two sets: 1.79 for the long set and 1.77 for the short one. Further, as one can see in Table 2, the distributions of grades for both sets are also significantly similar. This means that voters behave the same way whether they are facing a 4 or 9-candidates list under RV; an important observation that backs the idea that actual votes under RV might be essentially “independent of irrelevant alternatives” in the sense of Arrow (1963). We will come back to this idea when studying the results at the candidate level.

Table 2: Distribution of grades under Range Voting

Grade	%, short set	%, long set
0	42.29	38.54
1	11.53	13.84
2	10.44	13.12
3	11.86	13.33
4	8.50	8.89
5	15.39	12.27

Figure 1: Distribution of grades under Range Voting

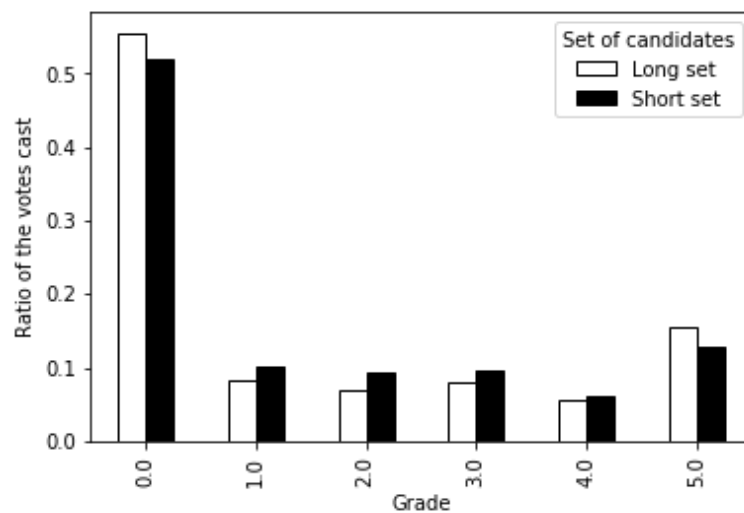


One can further compare the distributions of grades under Range Voting (RV) with the ones obtained as an answer to the evaluation question regarding the running candidates (“honest assessment”, HA), for both range from 0 to 5. Under HA, the average grades per candidate are 1.47 for the long set and 1.46 for the short one. Considering Table 3 which displays the distributions of grades under HA, it can be stressed as well that respondents behave similarly for the long and the short sets. More interestingly, Figures 1 and 2 show that the grade distributions under RV and HA are very similar. This is remarkable because it puts into light that participants do not globally 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 request participants either to give the grades 0 or 5, with no intermediary. Here, the fact that the distributions of RV and HA are comparable means that respondents voted sincerely with little, or no, strategic consideration.

Table 3: Distribution of grades under HA

Grade	%, short set	%, long set
0	55.49	51.91
1	8.34	10.17
2	6.87	9.36
3	8.00	9.56
4	5.72	6.17
5	15.59	12.83

Figure 2: Distribution of grades under HA



These global statistics help explain how voters behave under the tested multi-nominal rules. 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 here in order:

The number of approvals per ballot being lower than in previous experiments might be due to the way instructions of AV are stated (Koc 1988). In our survey, however, the wording of the AV ballots should 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 the familiarity with candidates. For instance, a 2016 Gallup Poll indicated 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, including the fact that the Commission on Presidential Debates⁴ excluded these two candidates from every presidential debate.

⁴ 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

In any case, we observe a nontrivial and material amount 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

a) Plurality.

Under plurality with the short set of candidates, as shown in Table 4, Clinton is the winner with 47.73%. Trump is second with 40.52%. With the long candidate set, Clinton wins with 31.38%, Trump is second with 27.78% and Sanders is ranked third with 19.98% of votes. 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 these of the “popular vote”, providing further evidence that the sample is representative⁵, not only with respect to socio-economic variables but also with respect to politics.

Table 4: Plurality results

	short set %	short set rank	long set %	long set rank
Clinton	47.73	1	31.38	1
Trump	40.52	2	27.78	2
Sanders	absent		19.98	3
Cruz	absent		9.73	4
Johnson	8.25	3	4.54	5
Bloomberg	absent		4.41	6
McMullin	absent		1.73	7
Stein	3.51	4	0.26	8
Castle	absent		0.19	9

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.”

Under AV, the results of the survey are more thought-provoking than under plurality. Table 5 shows that the addition of Sanders in the ballot has huge consequences on the outcome. Indeed, while Trump remained second under AV for the short set as under plurality, he is

nationwide at 15% by five different polling agencies using plurality voting. These polls, however, often exclude third-party candidate names.

⁵ The numbers for third parties in our poll is 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 will throw their vote away on a losing candidate under plurality voting.

largely surpassed by Sanders, who is almost in a tie with Clinton (39.78% for Clinton against 39.25% for Sanders, the difference not being statistically significant) for the long set. The results of AV thus put into light the very strong support Sanders had in 2016, which was partially hidden by the mono-nominal feature of plurality rule. Of course, when discussing the results of this kind of experiments, one should keep in mind that, under another official voting rule, the strategies of the political parties and candidates would be different, as might be the whole political landscape. This might lead the voters to hold different political preferences. Still, the face reading of our data is very simple: everything else being equal, under AV, Sanders could have won the 2016 U.S. election.⁷ Also interesting is that the percentages of approvals for the four candidates present in both sets are systematically lower for the long set. This indicates that voters' approval decisions for a specific candidate can depend on the presence or absence of other candidates

Table 5: Approval results

	short set %	short set outcome order	long set %	long set outcome order
Clinton	50.12	1	39.78	1
Sanders	absent		39.25	2
Trump	42.01	2	33.73	3
Cruz	absent		21.46	4
Johnson	20.65	3	12.16	5
Bloomberg	absent		11.65	6
McMullin	absent		7.60	7
Stein	11.52	4	5.06	8
Castle	absent		2.18	9

c) Range voting.

Third, Table 6 displays 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, compared to plurality are even more obvious—the winner is different. For the short set, Clinton is ranked first, with a significant advance over Trump. For the long set, Sanders beats Clinton with an average grade of 2.72, while Clinton is second with 2.32, the difference being significant. Another interesting trait of

⁷ When discussing the results of this kind of experiments, one should keep in mind that with another official voting rule, the strategies of the political parties and candidates would be different as the whole political landscape, which might obviously lead the voters to have different political preferences.

this voting rule—especially compared to AV—is that the average grades of the four candidates present in both sets are rather similar, thus showing a kind of “absolute” preference (“independence of irrelevant alternatives”). In other words, the way participants assess these four candidates under RV was not modified by the inclusion of five more candidates. This was not exactly the case under AV, even if the ranking of candidates by approval was consistent between the two sets.

Table 6: Range voting results

	short set grade	short set outcome order	long set grade	long set outcome order
Clinton	2.33	1	2.32	2
Trump	1.89	2	2.00	3
Sanders	absent		2.72	1
Cruz	absent		1.93	4
Johnson	1.59	3	1.50	6
Bloomberg	absent		1.67	5
McMullin	absent		1.36	7
Stein	1.30	4	1.21	8
Castle	absent		0.95	9

d) Instant runoff.

Fourth, Tables 7 and 8, and Figures 3 and 4 present the IRV results. 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.”

Tables 7 and 8 show the percentage of votes obtained by each candidate after each round of. The interesting point here is depicted in Figure 4. It shows that the main transfer to Clinton comes from Sanders’ 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.

The vote transfers that are sequentially observed under IRV, here appear akin to the mechanism that favors “exclusive” candidates under plurality, i.e. candidates who receive a large support from some voters, but are also often rejected from some 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” kind of candidates¹⁰. An “inclusive” candidate can be defined as follow: she gets a large support of the voters, but with no strong feeling of rejection or attachment.

This is a reminder that, even if instant runoff voting allows a rich expression to the individual voter, that information is not necessarily being used. As such, the observation that Clinton wins over Sanders under IRV could be explained by its mechanical effect. But we can also tell that a psychological effect is at work, because IRV data and HA data produce different Condorcet winners¹¹ (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.

Table 7: 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 8: 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								

¹⁰ See the “squeezing of the center” often observed in French politics (Blais and Indriadson 2007). See as well Baujard et al. (2014) on this issue.

¹¹ See Tables 10 and B2 : 51% rank Clinton above Sanders according to IRV data, but 54% prefer Sanders to Clinton according to HA.

Figure 3: Instant Runoff results – short set

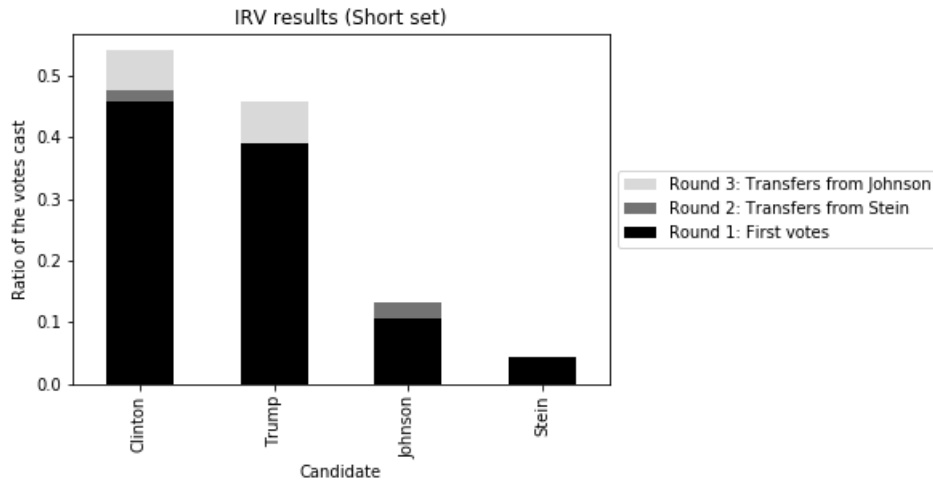
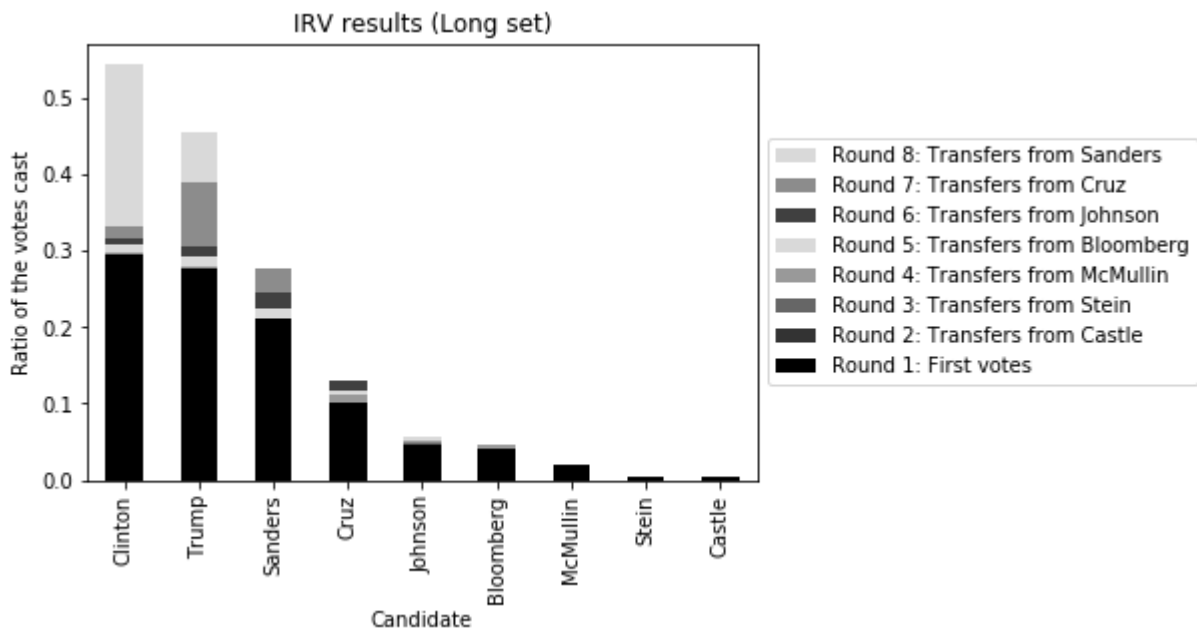


Figure 4: Instant Runoff results – long set



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, in short HA. Table 9 shows the average HA grades obtained by the candidates. 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 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.

Table 9: Average HA grades

	short set grade	short set rank	long set grade	long set rank
Clinton	2.20	1	2.11	2
Trump	1.75	2	1.77	3
Sanders	absent		2.35	1
Cruz	absent		1.53	4
Johnson	1.30	3	1.07	6
Bloomberg	absent		1.26	5
McMullin	absent		0.93	7
Stein	1.03	4	0.78	8
Castle	absent		0.63	9

Taking the HA variable at face value, one can compute any social welfare criterion. For instance the average scores in the previous Table 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 well seen on Figure 5, which reports 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 (see for instance Davison 2008): 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 on Figure 5, 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 would put all its weight on grade 5 would conclude in favor of Clinton instead of Sanders. This observation is well in line with the observation that Clinton wins to Sanders according to Plurality.

Figure 5: Comparing the distributions of honest assessments



Another social criterion, which rests on different, if not opposed, philosophical views (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 maybe 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 (see Table 10): The honest preferences profile has a Condorcet winner: Sanders.

Table 10: HA pairwise comparison matrix in % - long set¹³

	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

Computing other social evaluation criteria on the same HA data lead 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 B3), and so does “Honest Instant runoff” (Table B4). 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, one can use RV or IRV data to build pairwise comparisons. Then Sanders is the Condorcet winner using RV data and Clinton is Condorcet winner using IRV data.¹⁴

Table 11 sums up our previous remarks and results, thus putting into light the two main dimensions along which our extended analyses will proceed and which would have direct influences on the U.S. presidential election outcome. These two dimensions include (1) the chosen set of candidates which implicitly question the primaries system and (2) the chosen voting method with a more specific interrogation on the mono-nominal vs. multi-nominal issue. Still, it should be explicitly stated that candidate Trump was not the winner of the direct voting rule in this election. And further, our results tend to evidence that Bernie Sanders was the “best” candidate according to voters’ true preferences whenever they were allowed to express them accurately.

¹³ Note that 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.

¹⁴ See the appendix B for the pairwise comparison matrices under RV, and IRV, as well as the results of plurality, and IRV using HA data.

Table 11: Overall winners

<i>Voting rules</i>	<i>Short set</i>	<i>Long set</i>
Plurality	Clinton	Clinton
IRV	Clinton	Clinton
AV	Clinton	Clinton-Sanders ¹⁵
RV	Clinton	Sanders
HA average	Clinton	Sanders

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 “irrelevant alternatives”

Our results make clear that the voters’ expressed preferences for a given candidate might differ depending on the number and nature of other running candidates. 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 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 to any candidate independently of the other candidates. Consider Table 12 and Figure 6 to further study this point. For the four candidates who belong to both sets, Table 12 shows normalized grades for plurality, AV, RV, IRV, and HA¹⁶. Figure 6 represents the (normalized) scores’ gaps between the results obtained in the short and the long sets.

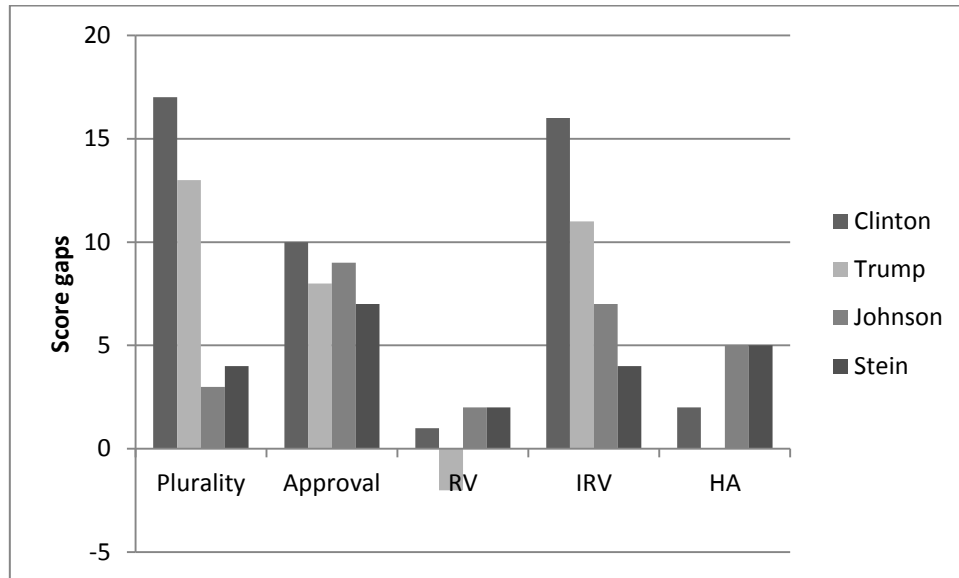
Table 12: Comparing scores between short and long sets (normalized scores)

	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

¹⁵ From our sample, the approval rates of Clinton and Sanders are not statistically different thus we cannot predict the winner.

¹⁶ RV and HA average scores are divided by the maximum grade, 5 and multiplied by 100. IRV “scores” are the percentages of first preferences.

Figure 6: Scores' gaps between the short and the long sets



Under plurality, IRV, and AV, all the candidates systematically have lower scores with the long set than with the short one. Under RV and HA, however, the scores remain more or less identical. If one can easily explain that 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 supports 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.

This issue is theoretically a very important one, much discussed in the literature. It refers to a famous condition defined by Arrow in his landmark impossibility theorem (Arrow 1963). Coined "independence of irrelevant alternatives", this condition states that adding or deleting any candidate should have no influence on the final choice. Arrow's notion, initially defined for ordinal preferences only can be extended to approvals and grades. Clearly vote splitting makes 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.

Arrows' IIA is defined within a formal model of preference aggregation in which the individual preferences to be aggregated do not vary with the voting rule. Our results (see Table 2) show that some voters change they decision to approve or not a candidate depending on who are the other candidate son the ballot. How important and consequential is this effect? In the present case, AV does not directly violate the IIA criterion here 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 irrelevant 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 from the race (like Sanders in our case). The differences of the supports 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 lost comparable percentages of supports, as it is well shown in Figure 6. In this election all four candidates lose sizeable approvals when the longer set set of candidates is put on the ballot.

Conversely, under plurality, adding extra candidates is largely correlated to the present and additional candidates’ platforms. For instance, the negative effects are clear when contrasting the short set from the long set on Clinton’s (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.

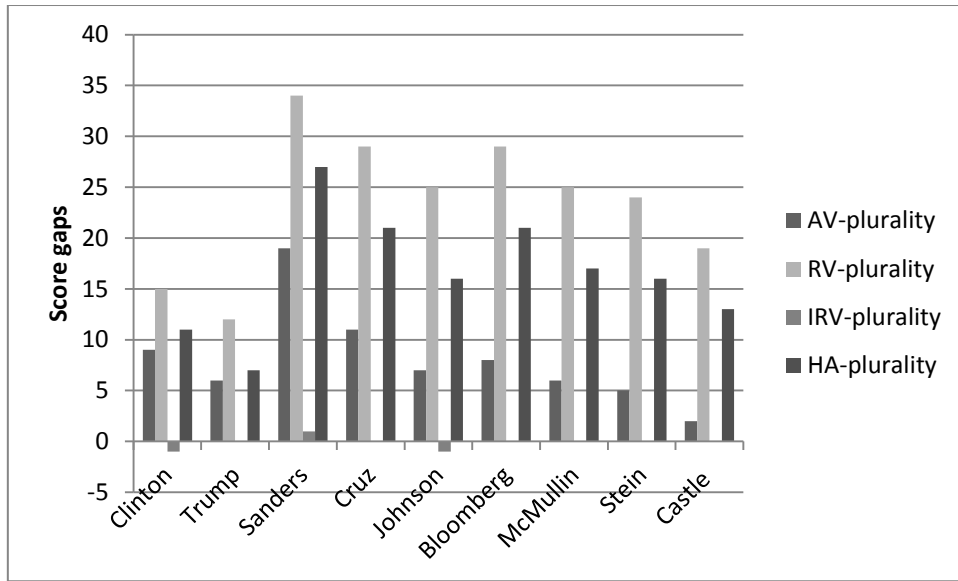
4.2 From plurality to multi-nominal voting rules: a modification of the winner

In this subsection, we focus our analyses on the long set, in order to examine why RV, HA, and AV may produce different winners compared to plurality. More generally, we aim to show the kind of candidates favored by certain voting rules. To emphasize this issue, Table 13 shows the scores of the nine candidates of the long set under plurality, AV, RV, IRV, and HA while Figure 7 represents the gaps of scores between plurality, AV, RV, and HA for the nine candidates of the long set. As in Table 12, the scores of the candidates are normalized (i.e., RV and HA scores are divided by the maximum grade (5)). 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.

Table 13: Comparing rules on the long set (normalized scores)

	Plurality	AV	RV	IRV	HA
Clinton	31	40	46	30	42
Trump	28	34	40	28	35
Sanders	20	39	54	21	47
Cruz	10	21	39	10	31
Johnson	5	12	30	4	21
Bloomberg	4	12	33	4	25
McMullin	2	8	27	2	19
Stein	0	5	24	0	16
Castle	0	2	19	0	13

Figure 7: Absolute Scores gaps compared to plurality



These elements demonstrate that Sanders benefited the most by switching from plurality to evaluative multi-nominal voting rules such as AV and RV. The same remains true when participants have been asked to assess honestly the candidates (HA). Conversely, the candidate who is least advantaged by evaluative rules is Trump. Interestingly, Trump also got the lowest score gap between plurality and honest assessment (HA), thus showing that Trump did not attract much support beyond the supporters who voted for him under plurality. Besides, Clinton’s score gap between plurality and HA is also rather low, which shows that she did not have much more support than Trump.

The next stage is to put into light the characteristics of the candidates who benefited the most, and those who benefited the least in order to define “types” of candidates beyond their party affiliation. To do so, we follow Baujard et al (2014) and consider the distribution of grades under RV for Sanders, Trump, Clinton, and Stein, four candidates that provide interesting examples of grade profiles (Figures 8-11).

Figure 8: Sanders profile

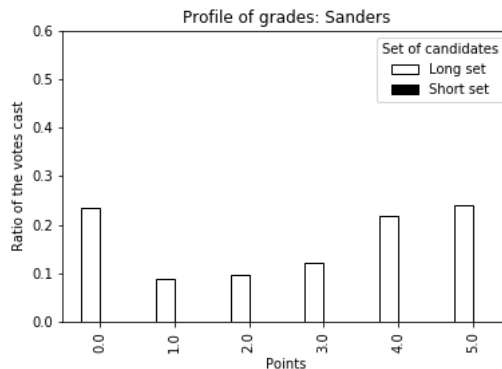


Figure 9: Trump profile

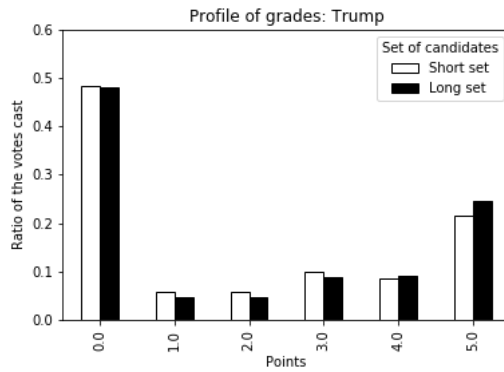


Figure 10: Clinton profile

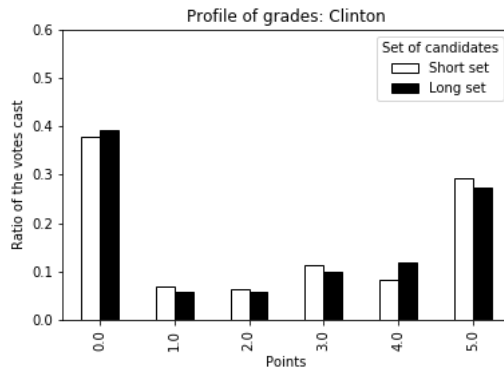
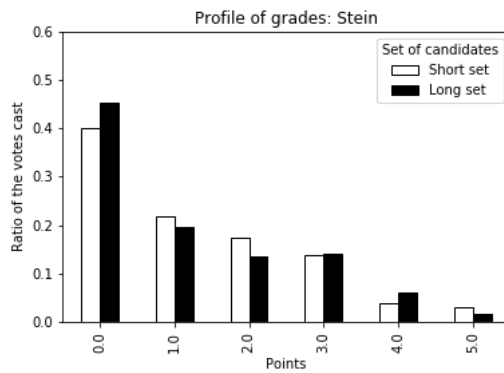


Figure 11: Stein profile



First, the profiles of the two main candidates of the 2016 election, Clinton and Trump, share the very same characteristics: very high number of grades 0, small number of intermediate grades, and rather high number of maximal grades 5. These are the

characteristics of a kind of candidates one might coin “divisive” or “exclusive” as already defined above: they receive a large support, but are also often rejected. In principle, exclusive candidates might be so for at least two different reasons: 1) they can be candidates who generate intense feelings for sincere voters who thus grade them in an “extremist” way; 2) they can also be candidates who have a chance of winning and therefore are voted in an “extremist” way by strategic voters, even if these strategic voters might otherwise share moderate feelings about them. Note that this second general explanation does not fit well in our study, for HA shows the exact same tendencies as RV.

Conversely, Sanders’ profile has the reverse features, showing a flat grade profile. Sanders represents another kind of candidates, one can call “inclusive”: as said above, he does benefit from a large support of the voters, but with lower levels of strong rejection or attachment. This kind of candidates is the one who loses more with plurality voting since this method is unable to take into account moderate support, contrary to multi-nominal voting methods.

Finally, one can put into light a third kind of candidates, illustrated by Stein profile. Candidates such as Stein receive indeed a large number of minimal grades, without benefiting from important support otherwise. Thus, their grade profile is decreasing from bad to good grades. This third category of candidates gathers typically “minor” candidates, who have no chance of winning and do not bring about much expressive support. Contrary to exclusive and inclusive kinds of candidates, these minor candidates are often unknown to most voters and do not receive important media exposure. This is especially common in the U.S. because the characteristics of the plurality voting system and Electoral College have media often focus on only two 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 proximity between candidates at the voter level. The “agreement” matrix based on AV data is shown in Table 14. It provides the candidates’ ability to attract voters who also support other candidates. For an election with the nine 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 (row) candidate who approve of another (column) candidate. Consequently, the diagonal is always equal to 100%, but the matrix is not symmetric. For instance, the proportion of Clinton’s supporters who also supported Bloomberg is 13%, while the proportion of Bloomberg’s supporters who also supported Clinton is 46%.

Thus, in a column, one can assess the *cross-over voter support* of a candidate, i.e. his or her ability to attract voters of other candidates. While in rows, one can read the *dispersion* of a candidate, i.e. 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 average of the columns and rows minus 1 (1, or 100%, being her own “score” in the matrix). The candidate with the highest cross-over voter support is Sanders with a measured cross-over support of 3.27, that is consistent with his “inclusive” type. Regarding the dispersion, the lowest averages are obtained by Trump and Clinton. 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 14: Agreement matrix (%)

	Clinton	Trump	Sanders	Cruz	Johnson	Bloomberg	McMullin	Stein	Castle	<i>Disp.</i>
Clinton	100	4	52	5	5	13	4	5	1	<i>0.88</i>
Trump	5	100	13	32	10	7	9	3	2	<i>0.80</i>
Sanders	52	11	100	12	14	17	3	10	2	<i>1.21</i>
Cruz	9	50	22	100	18	12	20	5	5	<i>1.40</i>
Johnson	18	28	44	32	100	12	13	17	9	<i>1.72</i>
Bloomberg	46	21	56	22	13	100	15	5	3	<i>1.81</i>
McMullin	20	41	18	55	20	24	100	7	13	<i>1.97</i>
Stein	36	17	77	22	42	12	10	100	6	<i>2.21</i>
Castle	18	33	33	48	51	15	44	13	100	<i>2.54</i>
<i>Cross-over</i>	<i>2.03</i>	<i>2.04</i>	<i>3.14</i>	<i>2.27</i>	<i>1.72</i>	<i>1.12</i>	<i>1.18</i>	<i>0.65</i>	<i>0.40</i>	

In all, corroborating Baujard et al.'s results (2014), mono-nominal rules such as plurality would favor exclusive candidates, compared with multi-nominal, evaluative, methods such as AV and RV that would benefit to inclusive candidates.¹⁷ In the 2016 U.S. election case, it appears that the latter rules with a large set of candidates would indeed allow electing the candidate who seems preferred by voters, i.e. Sanders. Neither direct plurality nor the official result would have led to Sanders winning. One explanation for Sanders being preferred is that, while he did not win the actual Democratic Party primary, our data indicates that he had large approval voting crossover support with Clinton, Johnson, and Stein. Additionally, Sanders' crossover support from Trump voters was nontrivial at 13% (more than double the Trump crossover support that Clinton got).

5. Concluding remarks

Based on a representative sample of more than 2,000 respondents, we found that, according to American voters' true electoral preferences, the "best" candidate of the 2016 U.S. election was Bernie Sanders. He appears both as the utilitarian optimum and the Condorcet winner. Our results show that multi-nominal evaluative voting rules such as RV and, potentially, AV would have led to elect Sanders among a large set of candidates, 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 U.S. voters effectively 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 alternative voting methods influence the election outcome?

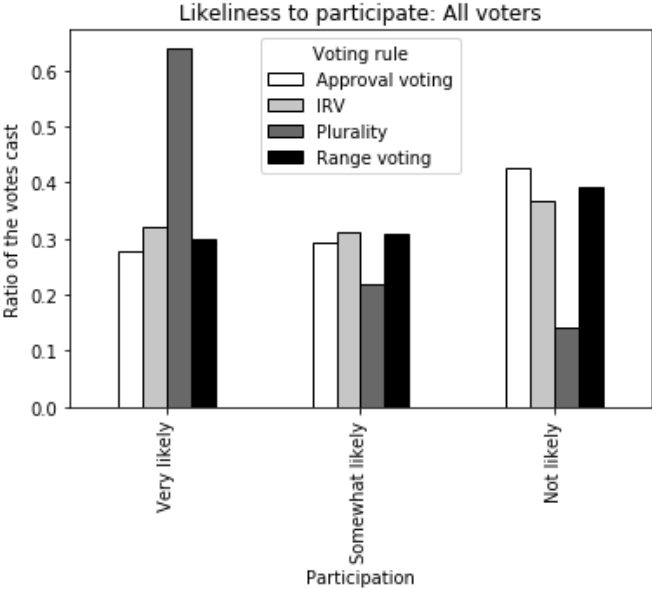
First, our results showed that the respondents really used the tested multi-nominal rules to express their preferences more widely and sincerely. This is particularly true for range voting,

¹⁷ In this comparison, instant runoff, despite the fact that it allows the voter to rank all candidates, appears closer 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.

whose distribution of grades present high significant similarities with voters’ honest assessments (HA). Second, we put into light that the size of the running candidates might have a large effect on the ballot depending on the voting method we are dealing with. While the multi-nominal rules are less sensitive to other “irrelevant” candidates, RV in particular, thus being closer to true preferences and thus to sincere outcomes, plurality is highly impacted by them 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 part of the voters, plurality tends to rank first exclusive candidates, who are either strongly supported or strongly rejected. In the 2016 U.S. case, Sanders can be seen as an inclusive candidate, while both Clinton and Trump were exclusive.

To conclude, according to our results, it appears that changing the official voting rule in the U.S. in favor of multi-nominal methods would be a good reform for American voters because the outcome of the ballot would be more faithful to their true electoral preferences. Of course, many important issues still remain at this stage. One could legitimately wonder whether American people are ready to opt for another voting rule or not. One question in our survey is dealing 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 12 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 remains with no doubt the favorite voting method of the respondents. Complementary studies might put the emphasis on the understanding of this phenomenon, which contradicts the fact that US cities have adopted alternative voting rules following education campaigns. Given that, it appears that voters must be educated about voting rules beyond just learning their mechanics.

Figure 12: Voters’ likeliness to participate



References

- Abbott, David W., and James P. Levine. 1991. *Wrong Winners: The Coming Debacle in the Electoral College*. New York: Praeger.
- Alós-Ferrer, Carlos, and Dura-Georg Granic. 2010. "Approval Voting in Germany: Description of a Field Experiment." In *Handbook on Approval Voting*, edited by Jean-François Laslier and M. Remzi Sanver, 397–411. Berlin: Springer.
- Arrow, Kenneth J. 1963. *Social Choice and Individual Values*. New Haven: Yale University Press.
- Ball, William J., and David A. Leuthold .1991."Estimating the Likelihood of an Unpopular Verdict in the Electoral College." *Public Choice* 70: 215–224.
- 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 ad 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.
- 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.
- 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.
- Davidson, Russel. 2008. "Stochastic dominance" In: *The New Palgrave Dictionary of Economics*, Steven N. Durlauf et Lawrence E. Blume (Eds.), 2008-02, pp. 921-925, Nature Publishing Group,
- 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.
- 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. 2017. "Trump, Condorcet and Borda: Voting paradoxes in the 2016 Republican presidential primaries." *European Journal of Political Economy*, in press.
- Lahrach, Rahhal, and Vincent Merlin. 2012. "Which Voting Rules Minimizes the Probability of the Referendum Paradox: Lessons from French Data." In *Electoral Systems, Studies in Social Choice and Welfare*, edited by Dan Felsenthal and Moshé Machover, 129–150 Berlin: Springer.
- 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.
- Lepelley, Dominique, Vincent Merlin, and Jean-Louis Rouet. 2011. "Three Ways to Compute Accurately the Probability of the Referendum Paradox." *Mathematical Social Sciences* 23: 227–257.
- 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, Nicholas R. 2012. "Why the Electoral College is Good for Political Science (and Public Choice)." *Public Choice* 150: 1–25.
- 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.

Appendix A: Complementary tables

Table A1: RV pairwise comparison matrix in % - long set

	Sanders	Clinton	Trump	Cruz	Bloomberg	Johnson	McMullin	Stein	Castle	Victories
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

Table A2: IRV pairwise comparison matrix in % - long set

	Clinton	Sanders	Trump	Cruz	Bloomberg	Johnson	McMullin	Stein	Castle	Victories
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 A3: Plurality results with HA data

	short set %	short set rank	long set %	long set rank
Clinton	44.85	1	27.60	1
Trump	36.89	2	25.18	2
Sanders	Absent		21.54	3
Cruz	Absent		10.37	4
Johnson	11.80	3	4.59	6
Bloomberg	Absent		5.59	5
McMullin	Absent		2.89	7
Stein	6.47	4	1.37	8
Castle	Absent		0.09	9

Table A4: 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								

Appendix B: Data cleaning

The data set provided by GfK Custom Research was of good quality, but an important point, independent of the quality of the survey is to check to what extend the votes questions themselves were correctly filled 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 of the voter were automatically checked (all pairwise covariances between two voting rules were nonnegative); (2) OK (manual): the data of the voter 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). In order to maximize the number 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 propose 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 of the long set whose votes are consistent all along the four tested voting rules plus HA with no need of a 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 a manual checking. There is no need for corrections though, for her approval of McMullin under AV and her 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 she misunderstood the instructions of IRV. Thus, we corrected her voter under IRV by reversing her ranking, which makes consistent all her responses. 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 of 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 for 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 to understand 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 twice the rank 3 so that we have removed both of them for it seems to be a simple mistake. Besides, some ballots were difficult to understand, but it 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 exclusion.

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