

Article

The Impact of Input Rules and Ballot Options on Voting Error: An Experimental Analysis

J. S. Maloy * and Matthew Ward

Department of Political Science, University of Louisiana, Lafayette, LA 70503, USA; E-Mails: maloy@louisiana.edu (J.S.M.), matthew.ward@louisiana.edu (M.W.)

* Corresponding author

Submitted: 15 December 2020 | Accepted: 22 March 2021 | Published: 15 June 2021

Abstract

When election reforms such as Ranked Choice Voting or the Alternative Vote are proposed to replace plurality voting, they offer lengthier instructions, more opportunities for political expression, and more opportunities for mistakes on the ballot. Observational studies of voting error rely on ecological inference from geographically aggregated data. Here we use an experimental approach instead, to examine the effect of two different ballot conditions at the individual level of analysis: the input rules that the voter must use and the number of ballot options presented for the voter's choice. This experiment randomly assigned three different input rules (single-mark, ranking, and grading) and two different candidate lists (with six and eight candidates) to over 6,000 online respondents in the USA, during the American presidential primary elections in 2020, simulating a single-winner presidential election. With more expressive input rules (ranking and grading), the distinction between minor mistakes and totally invalid votes—a distinction inapplicable to single-mark ballots—assumes new importance. Regression analysis indicates that more complicated input rules and more candidates on the ballot did not raise the probability that a voter would cast a void (uncountable) vote, despite raising the probability of at least one violation of voting instructions.

Keywords

American politics; election administration; election reform; Ranked Choice Voting; voting behavior; voting experiments

Issue

This article is part of the issue “The Politics, Promise and Peril of Ranked Choice Voting” edited by Caroline Tolbert (University of Iowa, USA).

© 2021 by the authors; licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

When voters, activists, and politicians consider the merits and demerits of election reform, it is natural for them to consult previous experience. They want to know about the past record not only of the status quo but also of any proposed changes of electoral rules or procedures. Innovative proposals, however, have little or no previous experience to recommend them. Unless they want to rule out innovation altogether, democratic societies must be prepared to substitute experiments for experience when issues of election reform are debated. Academic research can contribute experimental insights

to reform debates when observations of past experience seem insufficient on their own.

Here we report results from an experiment designed to shed light on the problem of voting error in the American context. A standard preoccupation of anti-reform discourse in the USA is the danger of disoriented or confused voters. As cities and states around the country consider switching from plurality voting rules or two-round systems to Ranked Choice Voting (RCV), for example, a plausible suspicion suggests that significant numbers of voters would in effect get counted out by making more mistakes on more complicated ballots. RCV has an observable track record in the USA since the early

2000s (in addition to various short-lived applications in the early and middle decades of the twentieth century). Various non-ranking forms of voting—e.g., “range” (grading candidates with more than two possible scores) and “approval” (grading candidates with only two possible scores)—have seen only a handful of implementations in the last few years. Voting experiments can therefore shed light on common intuitions or suspicions about the likely effect of relatively novel reforms on voting error.

RCV and other ballot reforms are proposals for fundamentally changing input rules, or the structure that the ballot imposes on how voters insert their judgments into the count. Another issue that may complicate voters’ task, and lead to more error, is the number of options on the ballot for any given contest. More complicated input rules and more options on the ballot could both theoretically exacerbate problems of voting error. Observational studies have difficulty confirming these relationships because real public elections never offer more than one ballot type or more than one list of candidates (or parties) for the same contest and the same voters. Our experimental analysis of this question is based on random assignment of different conditions in the two independent variables (input rules and ballot options) to examine their effects on the dependent variable (voting error). We find that, when over 6,000 subjects in four American states cast votes in a hypothetical election for US President in March 2020, just prior to the primaries conducted on “Super Tuesday” in multiple states, both factors had a minor impact on error. More complicated input rules and more plentiful ballot options both raised the likelihood that voters would make at least one mistake on their ballots. Yet the increase in minor mistakes did not result in more void (uncountable) ballots. Ballots that allow the ranking or grading of candidates offer more opportunities for political expression and, correspondingly, more opportunities for mistakes by voters—but not necessarily an increase in void votes or disfranchised voters.

Several challenges for conceptualization of the main variables—input rules, ballot options, and voting error—are addressed in Section 2. Next, we review observational and experimental literatures on issues related to voting error, in Section 3. Our hypotheses are presented in Section 4, and details of our experimental design and our analytic approach appear in Sections 5 and 6, respectively. Section 7 analyzes our results.

2. Conceptual Framework: Input Rules, Ballot Options, and Voting Error

Recent theoretical work on election reform has identified a dilemma for alternative types of input rules, featuring a potential zero-sum game between the qualities of expression and accessibility (Maloy, 2019, pp. 90–91). Do more expressive, and therefore more complicated, input rules inevitably produce more confused voters? Our primary intention is to examine this proposition through experimental treatments on input rules, with

a secondary focus on the possibility that the number of options on the ballot may also be a significant factor inducing voting error. Before surveying prior observational and experimental evidence on these questions, several conceptual difficulties with our three main variables—input rules, ballot options, and voting error—require clarification.

2.1. Input Rules

Our primary explanatory variable for voting error is the input rule on the ballot. The Super Tuesday 2020 experiments randomly assigned three different types of input rule and recorded how voters used their ballots with each of the three: exclusive (or single-mark), ranking, and grading. Respectively, these three input rules were called Check, Rank, and Grade within the experiment.

Researchers in electoral studies are familiar with the two types of ballot structure studied by Rae (1967): categorical and ordinal. These correspond to the Check and Rank input rules in the experiments reported here. The Check ballot’s input rule is categorical (or exclusive) because it requires the voter to indicate a single favorite candidate or party to the exclusion of all others. It presents an all-or-nothing choice. The Rank ballot is ordinal in the sense that it allows a hierarchy of preference to be indicated across multiple options on the ballot, in order from a first preference to a second preference to a third preference, and so on down the list.

RCV in the USA, similar to the Single Transferable Vote (STV) in Scotland and the Supplemental Vote (SV) in English cities (Lundberg, 2018), is one example of a recent reform that substitutes ranking for exclusive input rules. RCV is usually called the Alternative Vote (AV) outside the USA. Using this latter label, a public referendum in Great Britain in 2011 rejected AV as a replacement for plurality elections for the primary legislative assembly, the House of Commons. RCV (or AV), STV, and SV differ in certain respects, but what they have in common is a ranking input rule that allows voters to rank more than one candidate for the same office.

Yet the design of voting experiments today should go beyond Rae’s binary classification of ballot types, which was based on observed variation in input rules in established democracies in the 1960s. Election reform now involves a wider range of input rules to choose from. For example, the Cumulative Vote uses an input rule that gives the voter multiple votes to distribute across as many or as few candidates as the voter chooses, provided that the ballot’s budget of votes (the maximum number to be distributed in one contest) is not exceeded. The Approval Vote and the Range Vote (the latter is sometimes called the Evaluative Vote in Europe or the Grade Point Average [GPA] system in the USA) allow voters to grade as many or as few candidates as they choose on a certain numeric scale. Approval, by definition, offers only two possible levels of support (“approve” or “disapprove”), while the GPA family of input rules offers three

or more levels of support. Thus, after the commonly used exclusive type of ballot, there are not one but three additional types: ranking (Rae's "ordinal"), grading, and cumulative. These three types may be generally classified as multi-mark ballots (MMB) to distinguish them from the single-mark ballots (1MB) which employ exclusive input rules in most actual electoral systems, majoritarian and proportional alike (Maloy, 2019, pp. 86–89).

The Super Tuesday 2020 experiments included ranking and grading input rules, but not cumulative. The reason for the exclusion of cumulative ballots is that our experiments used real candidates' names for the American presidential contest, which is a single-winner election. Ranking has been used for both single-winner (RCV) and multi-winner (STV) contests, and grading input rules are usually proposed for single-winner contests. But cumulative input rules are usually proposed for elections in multi-seat districts. In voting experiments on multi-winner elections, cumulative input rules would certainly merit inclusion.

2.2. Ballot Options

Our second main variable of interest is ballot options. We use this term to refer to the number of options that are available on the ballot for the voter's consideration. While every competitive election must, by definition, have at least two options on the ballot, many elections have more than two. How many more may impact voters' behavior in general and voting error specifically, and our experimental design randomly assigned two unequal lists of candidates to different subjects voting for the same seat.

For single-winner elections in the USA, this variable could also be called "candidate supply," and perhaps with greater clarity. In electoral studies, however, it is well known that different countries' electoral systems may be more party-centric or more candidate-centric, depending on whether various institutional features incentivize (or require) voters to think more of parties or of individual candidates for office as the objects of their choice. Yet the number of options on the ballot may be an important factor in either case (cf. Seib, 2016). Since the term "candidate supply" might suggest the mistaken assumption that the variable under consideration here is only relevant to candidate-centric electoral systems (such as the USA's), we use "ballot options" instead.

2.3. Voting Error

The third key variable—our dependent variable—also presents a varied and potentially challenging conceptual terrain. Depending on the input rule in use, there can be more than one way for a voter to violate the instructions on the ballot, i.e., more than one type of error. Not every type of error has the same effect on how the ballot can be counted. Some errors limit the extent to which the voter's preferences can be incorporated into the count, while others require the ballot to be thrown out altogether.

In the analysis below, we observe the crucial distinction between a "mismarked" and a "void" ballot. The reason is that we are studying MMB voting in addition to 1MB voting, the latter kind of input rule being the status quo in the USA (and most other countries, regardless of district magnitude or allocation formula). As it turns out, the logical structure of 1MB input rules means that every mismarked ballot is by definition invalid. But the logical structures of ranking and grading input rules are different.

By allowing (and encouraging) voters to register judgments about more than one candidate, ranking and grading ballots create the possibility that an error made in how one candidate is marked does not necessarily preclude counting the mark made for another candidate. As we use the terms, a mismarked ballot is one that shows one or more violations of the instructions for a given input rule; a void ballot is one that is mismarked in a particular way, such that no quantitative contribution from that ballot can be used in the count. In other words, researchers interested in voting error with ranking and grading (and cumulative as well) ballots have an extra responsibility to distinguish clearly between invalidating errors and non-invalidating errors—whereas researchers who study voting error with exclusive ballots never encounter that necessity.

The terminology that has grown up around observational studies of voting error with exclusive ballots can still be useful to studies of MMB voting. In previous literature, the concept of "residual" votes includes three components: "under-votes" (i.e., blanks), (intentionally) "spoiled" votes, and (unintentionally) "invalid" votes (Herrnson, Hanmer, & Niemi, 2012, p. 722). Spoiled and invalid votes are two types of "over-vote," when a voter violates exclusive input rules by marking more than one candidate for the same contest. "Wrong" votes are valid votes cast for a candidate or party contrary to voters' intentions. Attempting to measure intentions independently of the ballots cast is another challenging territory, methodologically, and we do not address issues of intentionality—or of spoiled or wrong votes—in this study.

This terminology can be transferred to ranking and grading input rules, to some extent, but must also be expanded. Completely blank ballots cannot be counted under any input rule, of course. But the logical structure of both ranking and grading tolerates partial under-votes (also known as truncated votes), when the voter chooses to mark some options while leaving others blank. In Australia, most RCV (AV) and STV contests require voters to rank every option on the ballot, or else see their ballot voided. By contrast, most ranking systems in the rest of the world, including RCV cities and states in the USA, tolerate truncated votes. Our coding follows the norm outside Australia. Truncated votes are neither mismarked nor void; they are irrelevant to the analysis of voting error.

Ranking and grading can also be more tolerant of over-votes than exclusive input rules can be. One type of

over-vote which can apply to both ranking and grading is redundant marking, when one option on the ballot is marked with two distinct levels of support (e.g., ranked both second and third, or graded both B and C). This may be an invalidating error for that option, but other options may still receive valid levels of support on the same ballot. A second type of over-vote which applies only to ranking ballots is duplicate ranking, when the same ranking is applied to more than one option; a duplicate grade for more than one option does not violate the instructions on a grading ballot. Duplicate rankings may or may not cause difficulties for the count, again depending on how other options are ranked.

For ballots that are marked but cannot be counted at all for a given contest, we use the term “void votes.” The emptiness of the vote, quantitatively, is its most important feature. Our theoretic assumption is that void ballots represent a more consequential form of voting error than mismarked ballots. Whereas the latter may result in an incomplete expression of a voter’s political judgment, the former is tantamount to total disfranchisement for a given contest. We believe that this consideration explains why previous academic research on voting error has been principally concerned with void votes. After all, a totally invalidated ballot yields the same result as if the voter had stayed home altogether.

At the same time, mismarked ballots retain some normative interest because an incomplete expression of a voter’s political judgments, compared to what was structurally possible on a ranking or grading ballot, may still be a matter of regret. We therefore use “void” and “mismarked” as two alternative specifications of voting error as our dependent variable, which are described in greater detail in Section 6.

A final challenge concerns how to conceptualize blank ballots (total under-votes) in terms of voting error. It is generally assumed that voters who leave their ballots blank do so intentionally, and this assumption seems especially secure in a controlled experiment with a small number of contests in which to vote. If a partial under-vote (blanks for some options) presents no error on a ranking or grading ballot, a total under-vote (blanks for all options) should not be construed as error either. To code blank votes as equivalent to void votes runs the risk of conflating deliberate choices with unintentional mistakes. We therefore exclude blank votes from our analysis. As a result, our term “void votes” is not equivalent to “residual votes,” since the latter term conventionally includes blank votes.

In summary, *valid* votes in our study make a countable contribution for at least one candidate per contest; *mismarked* votes are marked in such a way that violates the instructions on the ballot in at least one respect, while still indicating a quantifiable preference for at least one candidate; and *void* votes are marked in such a way that no countable contribution can be registered for any candidate in a given contest.

3. Voting Error: Previous Research

Intuitions about the impact of input rules on voting error could in theory be tested against previous research observing rates of residual votes in real public elections. As we will now see, it has proved difficult to isolate input rules as a causal factor in voting error; hence the value of an experimental approach.

Observational studies of elections using ranking ballots in Great Britain reveal slightly higher levels of residual (also known as “rejected”) votes compared to 1MB voting. Scotland’s local council elections switched in 2007 from single-seat plurality to STV, which combines ranking input rules with multi-seat districts. The percentage of rejected votes associated with this change rose from 0.8 to 1.8. In Northern Ireland, having used STV for several decades, voters show a residual rate that ranges from 1 to 2 percent; in Ireland, with a century’s experience with STV, it is consistently around 1 percent (Clark, 2013; Denver, Clark, & Bennie, 2009).

In San Francisco, the first local elections with RCV from 2004 to 2006 produced residual rates slightly lower than previous elections with 1MB voting (Neely & Cook, 2008, pp. 538–541). Subsequent analysis of over-votes found considerable variation across plurality and RCV elections, with no clear advantage for one or the other, given the range of other factors that may affect voting error (Neely & McDaniel, 2015, pp. 10–12). Because of data limitations, separate rates of over-voting and under-voting are unknown for pre-RCV elections in San Francisco (see also Neely & Cook, 2008, p. 540). The conceptual difficulty with comparing over-votes under separate 1MB and RCV contests (i.e., different voting methods for different offices, a common feature of post-reform local governments in the USA) which occurred in the same year is that the numbers of candidates, not to mention the salience and visibility of the candidates and offices involved, may not be comparable.

In Minneapolis since 2009, residual votes in general and over-votes in particular have remained about the same with the shift from plurality to RCV elections (Kimball & Anthony, 2018, pp. 108–109).

In Maine in 2018, RCV was used for the first time for US House of Representatives elections. In District 1, the residual vote in 2018 was 2.3 percent; in 2016, prior to the switch to RCV, it had been 3.6 percent. In District 2, where a four-stage RCV count was needed to determine a winner in 2018, the residual vote was 2.1 percent; in 2016 in the same district, it had been 3.5 percent (calculations based on public records held by the Maine Bureau of Corporations, Elections, and Commissions). But the higher proportion of residual votes in 2016, a presidential year, was almost certainly the product of roll-off when some voters marked the presidential contest but left everything else blank.

Overall, it is not obvious that traditional voting in the USA affects voting error differently from RCV, in practice—contrary to the common-sense intuition about

the simplicity of 1MB input rules. Scotland has shown a more definite pattern toward greater ranking-based error, but the STV system there has the additional feature of applying to multi-winner elections. In other words, the switch from exclusive to ranking input rules was not the only thing that changed in Scotland in 2007, and therefore input rules cannot be isolated as the cause of increased error. Experimental data, though, could enable more controlled comparisons and more confident conclusions for this type of question.

Observational studies show similar levels of suggestiveness coupled with uncertainty in the evidence about which voters are more error-prone. With 1MB in the USA, studies tend to find higher residual rates in voting precincts containing higher numbers of Black residents and residents without high-school degrees (Kimball & Kropf, 2005, p. 522). But Black voters' higher residual rates are partly the result of deliberate choice in some contexts (Herron & Sekhon, 2005). Racial discrepancies in voting error have persisted with the switch to RCV ballots in Minneapolis (Kimball & Anthony, 2018, p. 109); in San Francisco, precincts with more Latino, elderly, and less educated residents have often shown higher residual rates as well (Neely & Cook, 2008; Neely & McDaniel, 2015). Observing which precincts show higher error rates, however, is not the same as identifying individual voters that make errors. We cannot observe error at the individual level of analysis in actual public elections because of the secret ballot. But we can use experiments to get at the individual level of analysis while preserving subjects' anonymity.

Among experimental as opposed to observational studies, a rich literature on alternative ballot types in Europe and Canada (e.g., Alos-Ferrer & Granic, 2012; Baujard, Igersheim, Lebon, Favrel, & Laslier, 2014; Blais, Heroux-Legault, Stephenson, Cross, & Gidengil, 2012; Farvaque, Jayet, & Ragot, 2011; Van der Straeten, Laslier, Sauger, & Blais, 2010) lays the foundation for offering different treatments on input rules in voting experiments. But these studies have not directly confronted the theoretic dilemma of expression and accessibility, and the related concept of voting error has barely featured there.

4. Hypotheses

The general hypothesis that our experiments on input rules were designed to test corresponds to common-sense intuitions about a zero-sum relation between expression and accessibility: More complicated rules produce more confused voters.

H1: Fewer voters make mistakes (cast void votes) while using exclusive (1MB) input rules than while using the more complicated input rules of ranking and grading.

As we know, however, ballot structure and contest structure are bound together in relations of mutual influ-

ence. In the American context, it makes sense to hold district magnitude at a value of 1, since the vast majority of federal, state, and local elections are single-winner contests (e.g., for senator, representative, governor, or mayor). But another, overlooked aspect of contest structure may have a special relation to how complicated a ballot appears to a voter, or how disoriented a voter may become. This is the number of options on the ballot. It seems intuitive that a ballot with a larger number of options for the voter's choice would be more likely to induce error, particularly with distributive input rules that (as in both ranking and grading) allow voters to make marks for more than one candidate. Hence, after our general hypothesis, a secondary hypothesis:

H2: While using ranking and grading input rules, fewer voters make mistakes (cast void votes) when confronting a smaller number of options on the ballot.

Experiments offer analytic leverage on our two hypotheses. Our general hypothesis requires us to analyze error rates across three different treatments, corresponding to the three main types of input rule for single-winner elections. Our special hypothesis requires us to vary the number of candidates presented to each subject, thereby introducing a second dimension of treatment.

5. Experimental Design

The analysis reported in this article is based on voting experiments conducted in four American states in March 2020. In partisan presidential primaries in the USA, states may choose their own date on which to hold such elections, and 14 out of 50 states chose 3 March 2020, also known as "Super Tuesday." Colorado, Tennessee, Texas, and Virginia were among the states voting in the "Super Tuesday" round of presidential nominating primaries, and we leveraged the public salience of those contests by inviting experimental subjects to vote on candidate lists that included real-world candidates for US President.

The Super Tuesday experiments were conducted online in the ten days prior to the actual voting. To be eligible to participate, subjects had to be of voting age (18 years or older) and had to be resident in one of the four selected states. Subjects were recruited by an outside contractor and paid a small consideration for their time, with most subjects taking five to ten minutes to complete the survey (see the Supplementary File for more technical details).

The Super Tuesday studies asked each subject to vote twice, once in a simulated Democratic Party presidential primary and once in a hypothetical "common ballot" contest featuring presidential candidates from multiple parties. A common ballot is an all-party ballot that may or may not function as a primary election. There are by definition no partisan primaries preceding it; as a result, there may be more than one candidate

bearing the same partisan affiliation on the common ballot. Among American states, the common ballot has been employed instead of partisan nominating primaries in Louisiana and Nebraska for over 50 years, and in California and Washington for over 10 years. In Nebraska, no party labels are attached to any candidates on the common ballot. In the other three states, where the common ballot serves as a primary election (also known as “jungle primary” or “top-two primary”) to narrow down the second-round ballot to two leading candidates, it is customary for more than one Democrat and more than one Republican to appear on the common ballot for a particular office, in addition to independent and minor-party candidates. To clarify, then, the Super Tuesday experiments created a hypothetical common ballot for US President with at least two Democrats and at least two Republicans.

To maximize the participation of various types of partisan across the voting-age populations of these American states, subjects were allowed to opt out of voting in the Democratic Party primary (the first voting task) by answering a question about their interest in that intra-party contest. Those who opted out by denying any interest in the Democratic Party primary were immediately presented with the common ballot. Those who opted in also voted on the common ballot, but it was their second voting task after the Democratic Party primary. For this analysis, we are examining results only from the hypothetical common ballot—the one that all subjects participated in, regardless of their level of interest in the Democratic nomination.

Subjects were randomly assigned one of three input rules: Check (single-mark), Rank, or Grade. They were given a brief, two-sentence description of how the ballot works before being shown the instructions on the ballot itself.

The instructions for the Check ballot read as follows: ‘Please indicate your favorite candidate by clicking the box containing their name, leaving all other options blank. Only one candidate can receive your vote.’

The instructions for the Rank ballot read as follows: ‘Please select rankings for one or more of the candidates in your order of preference (first choice, second choice, third choice, etc.). You may choose to rank any number of candidates, including all or only one, but only one ranking can apply per candidate.’

The instructions for the Grade ballot read as follows: ‘Please select a grade or score for each of the candidates with the level of support you wish to give: 4 for a grade of A, 3 for a B, 2 for a C, 1 for a D, or zero (0) for an F. You may choose to grade any number of candidates, but only one grade can apply per candidate.’

After randomly assigning one of these sets of instructions, the experiment did not constrain subjects’ freedom to mark their ballots in any way. The purpose of this design choice was to simulate the freedom of real-world paper ballots. Computerized touch screens are used in some jurisdictions in the USA, but paper bal-

lots remain the norm. There was one minor limitation imposed, for technical reasons, by the online survey platform. The number of available rankings on a Rank ballot was always limited to six, even when the list of eight candidates had been assigned to a particular subject. This kind of limitation also appears in real-world elections in some American jurisdictions that administer RCV elections, when the number of rankings must be capped because of the voting equipment in use. (Other jurisdictions have upgraded equipment to accept a maximum number of rankings which rarely, if ever, is exceeded by the number of candidates on the ballot.)

In addition, there was a second dimension of random assignment. Some subjects voted from a list of six presidential candidates, while others voted from a list of eight candidates. The list of six contained two actually declared Republicans (Donald Trump and William Weld) and four actually declared Democrats (Joe Biden, Mike Bloomberg, Bernie Sanders, and Elizabeth Warren). The list of eight added to the original six candidates one Green Party candidate (Howie Hawkins) and one Libertarian Party candidate (Lincoln Chafee). Figure 1 shows the eight-person common ballot with ranking input rules; Figure 2 shows the same with grading input rules.

Since Rank and Grade are the two MMB alternatives (treatments) to the 1MB status quo (control) in our voting experiments, it may be helpful to describe examples of how a ballot could be mismarked but still valid (not void) under each of these two input rules.

With ranking input rules, a hierarchic ordering of preferences is required: only one candidate per ranking and only one ranking per candidate. Accordingly, in Figure 2, if the voter ranks Bloomberg first and ranks both Sanders and Warren second, the ballot is mismarked but not invalidated. The vote for Bloomberg counts toward the first-round tally, but the second-choice votes for Sanders and Warren are ignored. On the other hand, if the voter ranks both Bloomberg and Warren first and ranks only Sanders second, this kind of mismarking voids the vote altogether. Nothing can be contributed to the first-round count from that voter’s ballot because no single favorite (for a single-winner election) can be ascertained. This coding protocol reflects the standard approach to voiding ballots in American jurisdictions that administer RCV elections (see the Supplementary File for more details).

A similar possibility of mismarked but valid votes exists with the Grade ballot in our experiments. In Figure 1, a voter who tries to give a grade of both B and C to Sanders can contribute nothing to Sanders’ total, but the same voter may still give any one grade to Warren. The instructions would then be violated in one instance, but the ballot would not be invalidated altogether. If the voter tries to give more than one level of support to every candidate on the ballot, only then is it effectively a null ballot under grading input rules.

Apart from the instructions, no other education or information about voting rules was supplied. In the four states covered in our sample, only Colorado has one or

JOE BIDEN, Democrat	1 st choice	2 nd choice	3 rd choice	4 th choice	5 th choice	6 th choice
MIKE BLOOMBERG, Democrat	1 st choice	2 nd choice	3 rd choice	4 th choice	5 th choice	6 th choice
LINCOLN CHAFEE, Libertarian	1 st choice	2 nd choice	3 rd choice	4 th choice	5 th choice	6 th choice
HOWIE HAWKINGS, Green	1 st choice	2 nd choice	3 rd choice	4 th choice	5 th choice	6 th choice
BERNIE SANDERS, Democrat	1 st choice	2 nd choice	3 rd choice	4 th choice	5 th choice	6 th choice
DONALD TRUMP, Republican	1 st choice	2 nd choice	3 rd choice	4 th choice	5 th choice	6 th choice
ELIZABETH WARREN, Democrat	1 st choice	2 nd choice	3 rd choice	4 th choice	5 th choice	6 th choice
BILL WELD, Republican	1 st choice	2 nd choice	3 rd choice	4 th choice	5 th choice	6 th choice

Figure 1. A Rank ballot with eight candidates.

JOE BIDEN, Democrat	Grade A	Grade B	Grade C	Grade D	Grade F
MIKE BLOOMBERG, Democrat	Grade A	Grade B	Grade C	Grade D	Grade F
LINCOLN CHAFEE, Libertarian	Grade A	Grade B	Grade C	Grade D	Grade F
HOWIE HAWKINGS, Green	Grade A	Grade B	Grade C	Grade D	Grade F
BERNIE SANDERS, Democrat	Grade A	Grade B	Grade C	Grade D	Grade F
DONALD TRUMP, Republican	Grade A	Grade B	Grade C	Grade D	Grade F
ELIZABETH WARREN, Democrat	Grade A	Grade B	Grade C	Grade D	Grade F
BILL WELD, Republican	Grade A	Grade B	Grade C	Grade D	Grade F

Figure 2. A Grade ballot with eight candidates.

two small towns with any prior implementation of RCV elections. Our assumption, therefore, is that there was a low level of familiarity with RCV across all subjects in our sample.

Demographically, the random assignment of input rules produced roughly comparable treatment groups that were somewhat younger, more female, and Whiter than the American adult population. Table 1 shows summary demographic statistics for the three treatment

groups, compared to the most recent estimates from the US Census Bureau. Table 2 shows the balance of covariates across treatment groups.

In summary, the experimental election that we are analyzing here offered 6,000 subjects in four American states a common ballot or “jungle primary” for US President. Each voter’s ballot combined one of three randomly assigned input rules with one of two randomly assigned rosters of candidates. By collecting information

Table 1. Experimental sample compared with US Census estimates for 2018 (figures represent percentages).

		Check	Rank	Grade	Census
Age	18–24	15.2	14.2	13.6	12.1
	25–34	25.3	23.7	24.0	17.8
	35–44	21.5	21.2	22.6	16.4
	45–54	15.7	16.1	15.1	16.4
	55–64	13.8	13.1	14.1	16.7
	65+	8.5	11.7	10.7	20.6
Race	White	69.1	70.6	68.5	60.2
	Black	15.1	14.8	16.3	12.3
	Latino	10.4	8.9	9.5	18.2
	Asian	2.5	3.7	3.0	5.6
	Am. Ind.	0.9	0.8	1.0	0.7
	Other	2.0	1.3	1.7	3.0
Gender	Male	39.2	40.6	41.3	49.2
	Female	60.8	59.5	58.7	50.8

Table 2. Balance of covariates.

	(1) Check Ballot Assignment	(2) Rank Ballot Assignment	(3) Grade Ballot Assignment
Options	-0.0189 (0.0309)	-0.0109 (0.0310)	0.0297 (0.0309)
Primary	0.0275 (0.0562)	-0.0110 (0.0562)	-0.0163 (0.0561)
Age	-0.00473*** (0.00181)	0.00272 (0.00179)	0.00198 (0.00179)
Female	0.0589 (0.0553)	-0.00888 (0.0551)	-0.0495 (0.0549)
Education	0.00533 (0.0293)	0.00869 (0.0293)	-0.0139 (0.0292)
Income	-0.0335 (0.0267)	0.00406 (0.0267)	0.0292 (0.0266)
White	-0.161 (0.167)	0.311* (0.182)	-0.126 (0.169)
Black	-0.230 (0.178)	0.245 (0.192)	0.00689 (0.179)
Hispanic	-0.0744 (0.186)	0.211 (0.200)	-0.117 (0.188)
Asian	-0.480** (0.231)	0.613*** (0.232)	-0.138 (0.226)
CO	0.0403 (0.0840)	-0.0212 (0.0839)	-0.0194 (0.0839)
TX	0.0455 (0.0835)	-0.00772 (0.0833)	-0.0378 (0.0833)
VA	0.0574 (0.0850)	-0.0188 (0.0849)	-0.0386 (0.0849)
Constant	-0.221 (0.273)	-1.047*** (0.283)	-0.835*** (0.274)
Observations	6,290	6,290	6,290

Notes: * $p < .10$; ** $p < .05$; *** $p < .01$; Tennessee is the reference category for state dummy variables.

about errors that these subjects made on their ballots, therefore, we acquired experimental data that can be used to assess voters' proneness to make mistakes under different ballot conditions.

6. Models and Variables

Our statistical model for voting error employs two alternative specifications of the dependent variable, two independent variables (corresponding to input rules and ballot options), and several control variables suggested by previous studies of voting error in the USA. The unit of analysis in this study is the ballot image for the all-party US presidential contest. A ballot image shows the pattern of marks made by a single voter for a single contest.

Void is the first specification of the dependent variable, representing the most basic and consequential way that voters err: a marked but totally invalid ballot for any given contest, from which no quantitative contribution to the count can be taken. This variable takes a value of 1 if the voter marked the ballot in one of the following ways, depending on the input rule in use: for the Check ballot, more than one candidate was checked; for the Rank ballot, the highest ranking marked was given to more than one candidate; for the Grade ballot, either all candidates received the same score or all candidates were double-scored. If a ballot was left entirely blank, it was not coded as void.

Mismarked is the second specification of the dependent variable. This variable takes a value of 1 for a marked ballot that violated the instructions in at least one respect, regardless of whether a valid vote could still be read off the ballot. If a ballot was left entirely blank, it was not coded as mismarked.

Rank and *Grade* are independent variables measuring the type of input rule used for the hypothetical common ballot for US President. Since the voters using the Check ballot are the control group in our experiment, this variable takes a value of 0 for voters who used that input rule and a value of 1 for *Rank* or for *Grade*, otherwise. Our hypothesis about the effect of MMB rules on voting error (H1) leads us to expect that the regression coefficient for this variable should be significant and positively signed.

Options is the independent variable measuring whether the voter saw the six-candidate or the eight-candidate ballot for US President. This variable takes a value of 6 or 8. Our hypothesis about the effect of ballot options on voting error (H2) leads us to expect that the regression coefficient for this variable should be significant and positively signed.

Several control variables are suggested by previous studies' findings on the correlates of voting error in the USA. *Age* measures the voter's age, from 18 to 99, and previous research suggests that older voters may be more error-prone (positively signed coefficient). *Female* captures self-reported gender, and previous studies suggest that females' lower intensity of interest in pol-

itics in general may lead to more error (positively signed coefficient). *Education* and *Income* measure voters' self-reported levels of educational attainment and self-reported annual household income, respectively, on an ascending five-point scale; and previous research suggests that less educated and less wealthy voters may be more error-prone (negatively signed coefficients). Race is measured through a series of dummy variables for self-reported *White*, *Black*, *Asian*, and *Latino* subjects (with *Other* as the reference category). Previous research suggests that White voters may be less error-prone (negatively signed coefficient) while non-whites may be more so (positively signed coefficient).

Finally, we have included a control variable for *Second Vote*, reflecting a peculiarity of the structure of these experiments which may have affected voting error. Subjects who opted to vote first in a simulated Democratic Party primary were voting on the common (all-party) ballot for US President as their second election in the experiment, while those who opted out of the Democratic Party primary saw the common ballot as their first election. Those who cast the ballot being analyzed here as their second voting task, for whom the *Second Vote* variable takes a value of 1, may have been less prone to error (negatively signed coefficient) because they had already familiarized themselves with some of the candidates' names that appeared on the ballot being analyzed here. It is important to note, however, that these second-time voters in the experiment saw a different input rule for the common ballot from the one they had previously used for the Democratic primary. Therefore, any error-reducing effect could come from familiarity with the candidates but not from familiarity with the input rule.

7. Results

To test H1 and H2, we modeled the probability of either mismarking or casting a void ballot as a function of ballot type using logistical regression with state-based fixed effects. The results are presented in Table 3.

The model presented in column 1 (void votes) indicates no statistical difference in the probability of casting a void Rank or a void Check ballot. In column 2 (mismarked votes), by contrast, the coefficient of the variable *Rank* is statistically significant and positive, indicating a higher probability of casting a mismarked Rank ballot than a mismarked Check ballot. Interestingly, the coefficient for *Grade* in column 3 (void votes) is statistically significant at the 10-percent level, but with the opposite sign from what was expected. As with the Rank ballot, however, there was a higher probability of casting a Grade ballot with at least one error compared with the Check ballot, as indicated by the positive and statistically significant coefficient on *Grade* in column 4 (mismarked votes).

In three of the four models presented in Table 3, the variable *Options* is statistically significant and positive,

Table 3. Regression analysis of two types of voting error.

	Probability Void	Probability Mismarked	Probability Void	Probability Mismarked
Rank	-0.120 (0.126)	0.862*** (0.106)		
Grade			-0.247* (0.129)	0.330*** (0.114)
Options	0.0180 (0.146)	0.385*** (0.117)	0.258* (0.151)	0.275** (0.132)
Age	-0.0246*** (0.00473)	-0.0184*** (0.00363)	-0.0299*** (0.00495)	-0.0240*** (0.00427)
Female	-0.352*** (0.127)	-0.343*** (0.103)	-0.354*** (0.130)	-0.233** (0.115)
Education	-0.0685 (0.0682)	-0.0446 (0.0552)	-0.183*** (0.0699)	-0.0553 (0.0618)
Income	-0.0530 (0.0634)	0.00185 (0.0505)	0.0207 (0.0640)	0.127** (0.0556)
White	0.0294 (0.434)	0.202 (0.380)	-0.0391 (0.404)	0.335 (0.429)
Black	0.708 (0.445)	0.874** (0.391)	0.361 (0.422)	0.908** (0.439)
Asian	0.854* (0.506)	0.995** (0.436)	0.822* (0.487)	1.024** (0.496)
Latino	0.339 (0.460)	0.475 (0.403)	0.497 (0.430)	0.735 (0.450)
Second Vote	-0.212 (0.131)	-0.112 (0.106)	-0.565*** (0.135)	0.0748 (0.121)
Texas	0.284 (0.189)	0.166 (0.160)	-0.0734 (0.204)	0.195 (0.175)
Virginia	0.0475 (0.200)	0.126 (0.164)	-0.0977 (0.208)	-0.142 (0.186)
Colorado	-0.149 (0.207)	-0.110 (0.169)	-0.0805 (0.206)	-0.332* (0.194)
Constant	-1.253** (0.496)	-2.048*** (0.433)	-0.592 (0.474)	-2.277*** (0.490)
Observations	4, 183	4, 183	4, 199	4, 199

Notes: * $p < .10$; ** $p < .05$; *** $p < .01$.

indicating that a higher number of candidates on the ballot is associated with a higher probability of casting a mismarked Rank and Grade ballot, as well as a void Grade ballot. This lends partial support to H2, which predicted that the number of candidates would increase the probability of mismarking MMB ballots. The number of candidates was not significant in predicting a void Rank ballot.

To investigate further the relation between ballot type and the number of options supplied on the ballot in predicting the dependent variable, Figure 3 depicts the probability of casting a void Check, Rank, and Grade ballot conditional on the number of candidates. We find that the number of candidates did not generally affect the probability of a void ballot. However, the probability for Check ballots was statistically significantly higher

than Rank ballots when there were eight candidates listed ($p = 0.07$).

Figure 4 depicts the probability of mismarking (rather than voiding) Check, Rank, and Grade ballots conditional on ballot options. There was a higher probability of mismarking a Rank ballot compared with a Check ballot regardless of the number of candidates. For ballots with eight candidates, however, the 11.15 percentage-point difference in the predicted probability between mismarked Rank and Check ballots is much larger than for six candidates, and statistically significant ($p < .01$). Given that the Rank ballot type in our experiments limited subjects to six rankings, even when eight candidates were on offer, it is likely that some subjects violated the instructions with duplicate sixth-choice rankings for

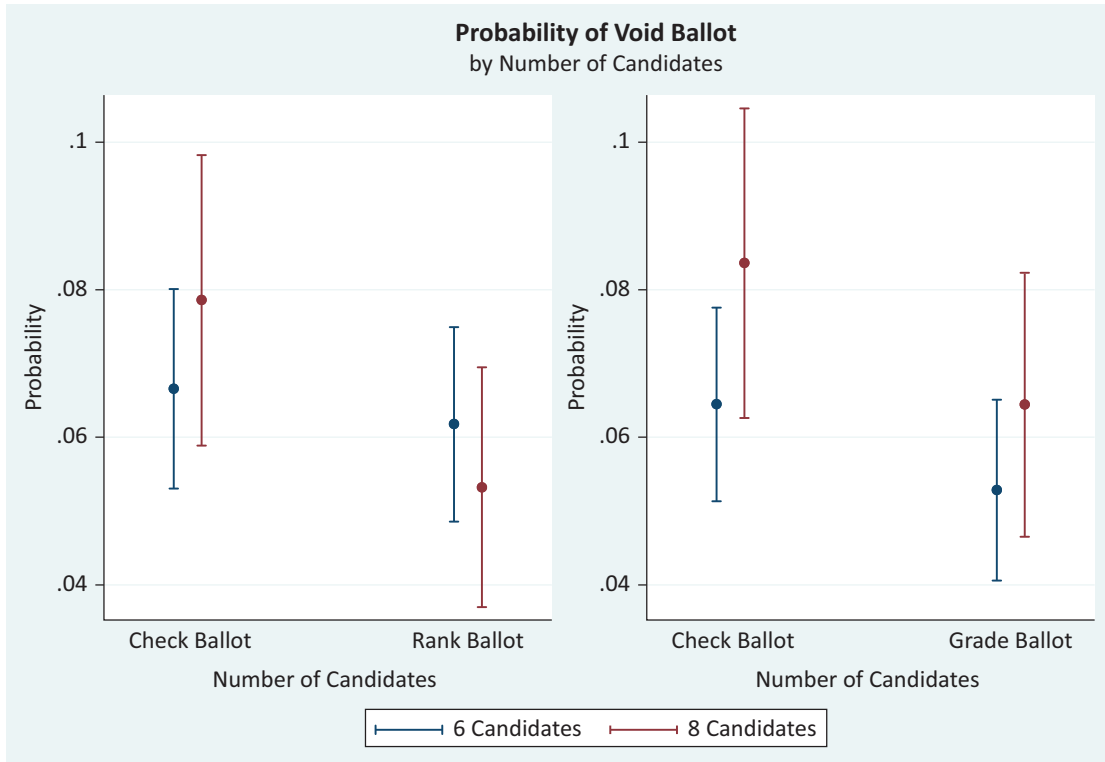


Figure 3. Probability of void ballot.

more than one candidate when they wished to indicate disapproval. This type of ballot image would be a prime example of a “mismarked” but not “void” vote. Differences in the probability of mismarking a Check and a Grade ballot were smaller, but the statistically signifi-

cant coefficient on *Grade* in Table 1 appears to be driven primarily by ballots with eight candidates.

Though a larger number of candidates does generally increase the probability of mismarking a more complex ballot, the interactions plotted in Figure 3 suggest that

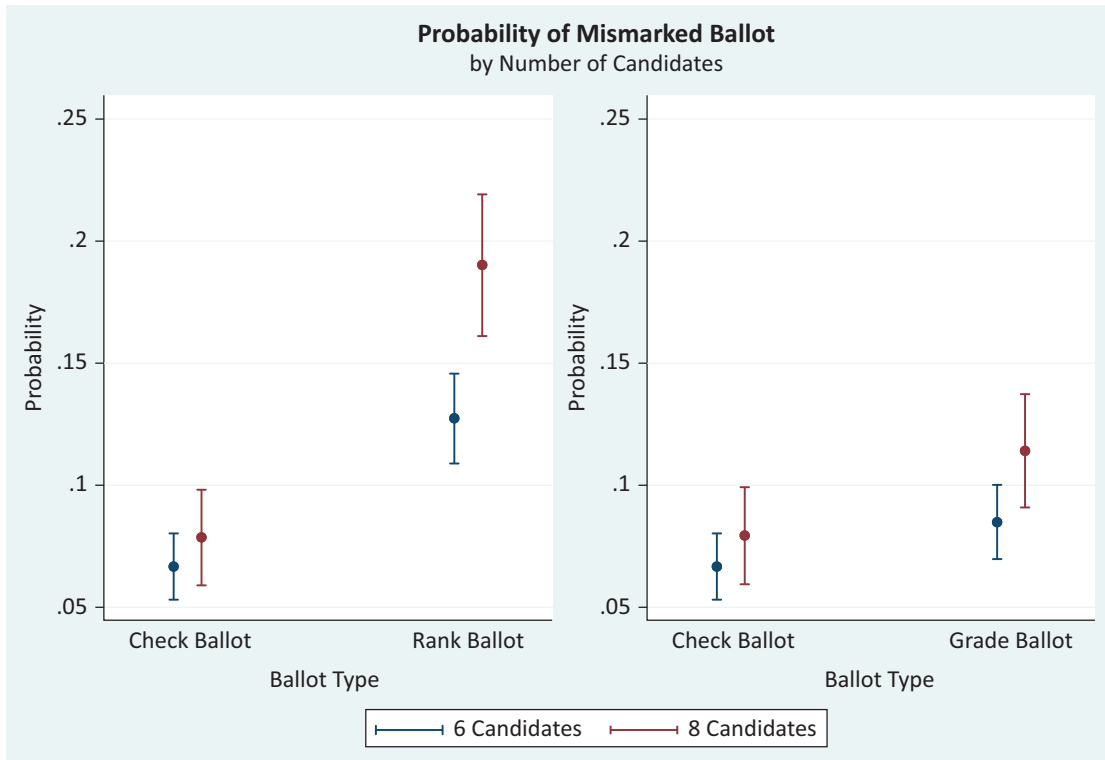


Figure 4. Probability of mismarked ballot.

the number of candidates on each ballot had little statistical or substantive impact on the probability of casting a void ballot.

We also would like to note the statistical significance of several control variables in Table 3. Contrary to conventional expectations about voting error, being female and being older were associated with a lower probability of casting either a mismarked or a void ballot in the Super Tuesday 2020 experiments. The result on age is consistent with an emerging literature that finds lower understanding of and satisfaction with RCV among older Americans (Donovan, Tolbert, & Gracey, 2019; McCarthy & Santucci, in press).

Results on racial variables raise questions for future study. Recent surveys in California cities found no significant racial discrepancies in self-reported understanding of RCV ballot instructions (Donovan et al., 2019). In the Super Tuesday experiments, however, respondents who identified as Asian were somewhat more likely to cast both mismarked and void ballots with the Rank input rule, while Black respondents were more likely to cast mismarked but not void ballots. Importantly, Black voters have been found more likely to make errors under 1MB when there is no Black candidate on the ballot (Herron & Sekhon, 2005). In this connection, we note that both Cory Booker and Kamala Harris had quit the US presidential race prior to Super Tuesday, leaving no Black candidates in the candidate lists of our experiments by the time they were launched.

8. Conclusion

The results of this experiment suggest that more complicated input rules do not have a significant impact on the casting of void (totally invalid) ballots, compared to the familiar all-or-nothing input rules of the 1MB status quo in the USA. But ranking and grading ballots did raise the probability that a voter in our experiments would commit at least one violation of the instructions. More opportunities for expression go hand-in-hand with more opportunities for error, though minor mistakes on the ranking and grading ballots tested here were usually compatible with counting the voter's support for at least one favored candidate.

The number of options presented on the ballot for the voter's choice also affected the likelihood of error, but again not as strongly as expected. Experimental subjects were somewhat more likely to make minor mistakes with all three ballot types when they had eight rather than six candidates to choose from, with the Rank ballot showing the biggest increase in mismarked ballots. The effects were weaker on void votes. Overall, both Rank and Grade ballots were voided less often than Check ballots, regardless of the number of candidates. In fact, the eight-candidate ballot (with a Green and a Libertarian added to the list) actually seemed to make experimental subjects using the Rank ballot less likely to submit void votes than the six-candidate list (with only

Democratic and Republican candidates). The difference in ballot options captured by our experimental design, between six and eight candidates, was far from dramatic, and future studies should be designed to implement a wider range of treatments on ballot options. A range from three candidates to ten would be realistic for many sub-national elections in the USA, for American states that use a common (all-party) ballot in primary elections, and for some nominating contests for national offices. It would also be worthwhile, in an experimental setting, to address variation in ballot options in party-centric rather than candidate-centric contexts. Cross-national and cross-cultural comparisons are as yet poorly understood on the question of voting error.

In summary, we find support for our two hypotheses about voting error in the mismarking of ballots but not in the voiding thereof. This is to some extent a disappointing result for the intuition behind anti-reform arguments in the American context, since void ballots carry greater normative weight than mismarked ones. Effective disenfranchisement (for a particular contest) is more serious than incomplete expression. The mismarking of ballots is still worrisome, but it can theoretically be alleviated by the actions of local election administrators and media prior to the implementation of new election reforms. Different treatments as to voter information were not part of this experimental analysis but should be considered a priority for future experimental research on election reform. The possibility that voting error can be reduced in the case of MMB input rules by familiarization and education may help to explain why, in the real world, observational studies of jurisdictions that switch to ranking ballots in the USA often show little or no increase in voting error.

Acknowledgments

J. S. Maloy acknowledges financial support from the Electoral Reform Research Group (ERRG), sponsored by New America, and from the Kaliste Saloom Endowed Chair at the University of Louisiana, Lafayette, in the collection of the data analyzed here.

Conflict of Interests

The authors declare no conflict of interests.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

- Alos-Ferrer, C., & Granic, D.-G. (2012). Two field experiments on approval voting in Germany. *Social Choice and Welfare*, 39, 171–205.
- Baujard, A., Igersheim, H., Lebon, I., Favrel, F., & Laslier,

- J.-F. (2014). Who's favored by evaluative voting? An experiment conducted during the 2012 French presidential election. *Electoral Studies*, 34, 131–145.
- Blais, A., Heroux-Legault, M., Stephenson, L. B., Cross, W., & Gidengil, E. (2012). Assessing the psychological and mechanical impact of electoral rules: A quasi-experiment. *Electoral Studies*, 30, 829–837.
- Clark, A. (2013). Second time lucky? The continuing adaptation of voters and parties to the single transferable vote in Scotland. *Representation*, 49(1), 55–68.
- Denver, D., Clark, A., & Bennie, L. (2009). Voter reactions to a preferential ballot: The 2007 Scottish local elections. *Journal of Elections, Public Opinion, and Parties*, 19(3), 265–282.
- Donovan, T., Tolbert, C. J., & Gracey, K. (2019). Self-reported understanding of ranked-choice voting. *Social Science Quarterly*, 100(5), 1768–1776.
- Farvaque, E., Jayet, H., & Ragot, L. (2011). French presidential election: A field experiment on the single transferable vote. In B. Dolez, B. Grofman, & A. Laurent (Eds.), *In situ and laboratory experiments on electoral law reform* (pp. 55–68). New York, NY: Springer.
- Herrnson, P. S., Hanmer, M. J., & Niemi, R. G. (2012). The impact of ballot type on voter errors. *American Journal of Political Science*, 56(3), 716–730.
- Herron, M. C., & Sekhon, J. S. (2005). Black candidates and black voters: Assessing the impact of candidate race on uncounted vote rates. *Journal of Politics*, 67(1), 154–177.
- Kimball, D. C., & Anthony, J. (2018). Ranked choice voting: A different way of casting and counting votes. In T. Donovan (Ed.), *Changing how America votes* (pp. 100–112). Lanham, MD: Rowman & Littlefield.
- Kimball, D. C., & Kropf, M. (2005). Ballot design and unrecorded votes on paper-based ballots. *Public Opinion Quarterly*, 69(4), 508–529.
- Lundberg, T. C. (2018). Electoral systems in context: United Kingdom. In E. S. Herron, R. J. Pekkanen, & M. S. Shugart (Eds.), *The Oxford handbook of electoral systems* (pp. 627–650). Oxford: Oxford University Press.
- Maloy, J. S. (2019). *Smarter ballots: Electoral realism and reform*. New York, NY: Palgrave Macmillan.
- McCarthy, D., & Santucci, J. (in press). Ranked-choice voting as a generational issue in modern American politics. *Politics and Policy*.
- Neely, F., & Cook, C. (2008). Whose votes count? Undervotes, overvotes, and ranking in San Francisco's instant-runoff elections. *American Politics Research*, 36(4), 530–554.
- Neely, F., & McDaniel, J. (2015). Overvoting and the equality of voice under instant run-off voting in San Francisco. *California Journal of Politics and Policy*, 7(4). <https://doi.org/10.5070/P2cjpp7428929>
- Rae, D. W. (1967). *The political consequences of electoral laws*. New Haven, CT: Yale University Press.
- Seib, J. D. (2016). Coping with lengthy ballots. *Electoral Studies*, 43, 115–123.
- Van der Straeten, K., Laslier, J.-F., Sauger, N., & Blais, A. (2010). Strategic, sincere, and heuristic voting under four election rules: An experimental study. *Social Choice and Welfare*, 35, 435–472.

About the Authors



J. S. Maloy is Professor of Political Science and holder of the Kaliste Saloom Endowed Chair at the University of Louisiana, Lafayette. His recent book on democratic theory and election reform is *Smarter Ballots* (Palgrave Macmillan, 2019). He previously authored two books on the history of democratic ideas in Europe and North America, most recently *Democratic Statecraft* (Cambridge, 2013), as well as articles on electoral accountability, democracy measurement, and comparative democratic institutions.



Matthew Ward is a Visiting Assistant Professor of Political Science at the University of Louisiana, Lafayette. His research interests include Middle Eastern politics, intolerance, democratization, measurement and survey design, and experimental methods. His research agenda is located at the intersection of political psychology and comparative politics, focusing on democratization, the formation of democratic attitudes, the dynamic effects of intergroup relations on social and political intolerance, and support for democratic values.