Public Choice (2015) 165:193–210
DOI 10.1007/s11127-015-0303-6

# The paradox of grading systems

**Steven J. Brams**[1] **· Richard F. Potthoff**[2]

Received: 12 September 2015 / Accepted: 18 December 2015 / Published online: 7 January 2016


**Abstract** We distinguish between (i) voting systems in which voters can rank candidates and (ii) those in which they can grade candidates, using two or more grades. In approval voting, voters can assign two grades only—approve (1) or not approve (0)—to candidates. While two grades rule out a discrepancy between the average-grade winners, who receive the highest average grade, and the superior-grade winners, who receive more superior grades in pairwise comparisons (akin to Condorcet winners), more than two grades allow it. We call this discrepancy between the two kinds of winners the paradox of grading systems, which we illustrate with several examples and whose probability we estimate for sincere and strategic voters through a Monte Carlo simulation. We discuss the tradeoff between (i) allowing more than two grades, but risking the paradox, and (ii) precluding the paradox, but restricting voters to two grades.

**Keywords** Voting and elections · Grading systems · Ranking systems · Approval voting · Condorcet winner · Monte Carlo simulation

## 1 Introduction

For more than 60 years, the standard framework for analyzing voting and social choice, due to Arrow (1963 [1951]), has been one in which the voters are assumed to rank candidates, possibly with ties, from best to worst. Consistent with such a framework are

✉ Steven J. Brams
steven.brams@nyu.edu

Richard F. Potthoff
potthoff@duke.edu

[1] Department of Politics, New York University, New York, NY 10012, USA

[2] Department of Political Science and Social Science Research Institute, Duke University, Durham, NC 27708, USA

such well-known voting systems as the Borda count, the Hare system of single transferable voting (STV)—also known as instant runoff voting (IRV) and ranked choice voting (RCV)—and many others.

Beginning with the work of Brams and Fishburn (1978) as well as other theorists about the same time (see Brams and Fishburn 2007 [1983]), an alternative framework, based on grading candidates, was proposed. Brams and Fishburn championed approval voting (AV), in which voters can give only two grades—approve (1) or not approve (0)—to candidates, but other theorists later proposed that voters be allowed to give more grades (Felsenthal 1989; Hillinger 2005; Alcantud and Laruelle 2014 favor three grades, but others, including Balinski and Laraki 2011, favor four or more).

Only AV rules out a discrepancy between the *average-grade (AG) winners*, who are recipients of the highest average grade, and the *superior-grade (SG) winners*, who receive more superior grades in pairwise comparisons (akin to Condorcet winners in ranking systems). But if voters can give more than two grades to candidates, not only may an AG winner not be an SG winner but, in the extreme case, every voter except one may grade the SG winner higher than the AG winner. The ability of AV to rule out the AG-versus-SG discrepancy for grading systems has no counterpart for ranking systems.

If grades are mapped into ranks, with higher grades being given higher ranks, an SG winner is a *Condorcet winner*, who is preferred to every other candidate based on rankings. Just as a winner under the Borda count or the Hare system may not be a Condorcet winner, an AG winner may not, unsurprisingly, be an SG winner when more than two grades can be assigned.

When this discrepancy occurs, we say there is a *paradox of grading systems*, somewhat but not fully analogous to the paradox that arises when the Borda and Condorcet winners differ. More precisely, because there may be tied AG winners, or the SG winners may be in a cycle—in which case we treat all candidates in the cycle as winners—we say that a paradox of grading systems occurs if and only if the set of AG and the set of SG winners differ.

We estimate through simulation the probability of this paradox for a type of *impartial culture*, in which all voters are equally likely to give each of the possible grades to every candidate. We start by assuming three grades, with varying numbers of candidates and voters. Then we allow the number of grades to increase to determine its effect on the probability of the paradox. We base these calculations on a Monte Carlo simulation but show, in a simple case, that the precise probability closely approximates the simulated probability, and the simulated probabilities vary only slightly in different simulation runs.

Implicitly, our impartial-culture model assumes, in its first form, that voters are *sincere*: They need not give their top candidate the highest grade nor their bottom candidate the lowest grade. That is, if there are more than two grades, they may give either one or both of these candidates an intermediate grade.

Besides assuming that voters are sincere, we also assume, in a second form of the model, that voters are *strategic* in the sense that they desire to help and hurt, respectively, their top and bottom candidates—to the maximum degree possible—by giving their top candidate(s) the highest grade and their bottom candidate(s) the lowest grade.

To assess the effects of strategic voters on the probability of the paradox, we constrain the second form of the impartial-culture model to ensure that every voter gives at least one candidate the top grade and at least one candidate the bottom grade. In fact, the strategic probabilities do not deviate much from the sincere probabilities.

The paper proceeds as follows. In Sect. 2, we compare ranking systems, like the Borda count and the Hare system, and grading systems, highlighting their similarities and differences. Although SG winners are akin to Condorcet winners when grades are replaced by ranks, ranking systems and grading systems are not equivalent. In Sect. 3, we analyze grading systems, giving several examples of the paradox of grading systems.

In Sect. 4, we provide a general method for finding SG winners (determining AG winners is straightforward). Then we present our simulation results, showing that the paradox is not an infrequent phenomenon in both the sincere and strategic models, although our results almost surely overestimate what its actual, real-world frequency would be. In Sect. 5, we discuss the tradeoff between (i) allowing more than two grades, but risking the paradox, and (ii) precluding the paradox, but restricting voters to two grades.

## 2 Ranking versus grading systems

Whether one ranks or grades candidates, the ranks or grades can be aggregated in different ways. To illustrate for ranking systems, if there are $c$ candidates, the Borda count awards points to candidates, from 0 for a voter's lowest-ranked candidate to $c - 1$ for a voter's highest-ranked candidate. It sums these points across all voters, and the candidate with the most points, or highest score, wins.

By contrast, the Hare system starts by eliminating candidates with the fewest first-choice votes. It then transfers the votes of their supporters to their next-highest-ranked candidates who remain in the race, continuing this elimination-and-transfer process until one candidate receives a majority of votes and thereby becomes the winner.

Formally, a *grading system* is a voting system in which a voter can give any of $g$ grades, $\{w_1, w_2,\ldots, w_g\}$, to each candidate. Grades need not be equally spaced; for example, $(w_g - w_{g-1})$ could be large in relation to other differences between adjacent grades. Here, however, we assume that the grades are given by the natural numbers, starting with a lowest grade of 0 and ending with a highest grade of $g - 1$. While sincere voters need not give either a lowest or a highest grade to any of the candidates, strategic voters give at least one lowest grade and one highest grade to candidates.

Like ranking systems, grading systems can differ in how the grades are aggregated to determine a winner. The average-grade (AG) system we analyze here is one in which the grades of each candidate are summed across all voters and averaged, with the candidate with the highest average grade becoming the AG winner.

This grading system is sometimes referred to as range or score voting (Center for Election Science 2015; Center for Range Voting 2015) or evaluative or utilitarian voting (Baujard et al. 2014, and references therein). Other methods for aggregating grades to determine a winner, such as choosing a candidate with the highest median grade, have been proposed (Balinski and Laraki 2011).[1] It turns out that a median-grade system also is vulnerable to the paradox of grading systems, as we will illustrate with an example in Sect. 4, but we focus on AG and how the frequency of the paradox of grading systems—when the AG and SG winners differ—depends on the numbers of candidates, grades, and voters.

[1] Because ties for highest median grade are common when relatively few grades can be assigned, Balinski and Laraki (2011) propose a method for breaking ties in a system they call "majority judgment." Cumulative voting, in which voters can allocate a fixed number of votes to one or more candidates, effectively allows voters to grade candidates, though it is not usually thought of as a grading system. It is used to elect multiple winners, unlike the systems analyzed here, and affords parties or factions the opportunity to elect candidates in proportion to their share in the electorate. For analyses of its properties and experience with it, see Brams (2003 [1975]) and Bowler et al. (2003).

To determine an *SG winner*, we compare each candidate's grades with the grades of all other candidates. Candidate $X$ beats candidate $Y$ if the number of voters who grade $X$ higher than $Y$ exceeds the number who grade $Y$ higher than $X$. If one candidate beats each rival in a pairwise comparison with each, then he or she is the unique SG winner. But as with Condorcet winners, SG winners may be in a *top cycle*, whereby they each beat all candidates not in the cycle—because more voters give them more higher grades—but no smaller set of candidates exists, all of whom beat everyone else. Members of the top cycle can either be beaten or be tied by at least one other candidate in the cycle. A paradox of grading systems arises if and only if the sets of AG winners and SG winners differ.

Before we analyze the paradox of grading systems, we show that AV precludes the paradox:

**Proposition 1** *With two grades, the sets of AG and SG winners are identical.*[2]

*Proof* Because an AG winner—when averaging the candidates' grades of 1 and 0—has, by definition, the highest average grade, he or she must have been approved of by more voters than any other candidate. Thus, in pairwise comparisons, more voters give the AG winner a grade of 1, and other candidates a grade of 0, than vice versa—ignoring 0–0 and 1–1 ties—so an AG winner is also an SG winner. The proof extends easily if the winner sets contain more than one candidate. □

We will shortly suggest an analogue of Proposition 1 (Proposition 2 below) for the median-grade, Borda, and Hare systems. But first we specify precisely how we determine winners under each of these systems:

1. *Median grade* When voters can assign only two grades, 0 and 1, all candidates will have a median grade of either 0 or 1 (or occasionally ½, based on the standard definition of median), so there may often be ties for the median-grade winner, as noted in footnote 1. We assume the method of Balinski and Laraki (2011, pp. 224–225) to break ties,[3] which ensures that the candidate receiving the largest number of 1s is the median-grade winner.
2. *Borda* When a voter ranks candidates as tied under the Borda system, each of them receives the mean of the corresponding strict ranks (see footnote 6 below for an example). Thus, with the Borda system restricted to two grades, if there are $c$ candidates and a voter assigns a 0 (do not approve) to $k$ of them ($1 \leq k \leq c - 1$) and

[2] Merrill and Nagel (1987) distinguish a balloting method, such as one that allows two grades, from a method for aggregating the ballots (what they call "decision rules"), such as one that elects the candidate with more superior grades in pairwise comparisons—or, equivalently, in the case of AV the highest average grade. As we will show, using a different aggregation method can produce a winner different from the AG-SG winner, even when only two grades are possible (the Hare system is an example). However, systems that choose as winners the candidates with the highest median grade or the highest Borda score duplicate AV in always selecting AG–SG winners for the case of two grades.

[3] For any $g$, the method breaks ties as follows: Order the grades for each candidate from low to high; for an even number of grades, treat the grade just below the middle, rather than the average of the two middle grades, as the median; delete each candidate's (tied) median grade; obtain the median for each candidate from the new (smaller) sets of grades; and if a tie still exists, continue to delete grades successively, one at a time, in the same way as before, until the tie is broken. As an example for $g = 2$, suppose the ordered grades from six voters are 0 0 $\underline{1}$ 1 1 1 for candidate $X$ and 0 1 $\underline{1}$ 1 1 1 for $Y$, with the remaining candidates receiving more 0s than 1s (the medians, as just defined, are underscored). Deleting both underscored medians yields 0 0 $\underline{1}$ 1 1 for $X$ and 0 1 $\underline{1}$ 1 1 for $Y$. The new medians are again tied, so they are deleted, giving 0 $\underline{0}$ 1 1 for $X$ and 0 $\underline{1}$ 1 1 for $Y$. The tie is now broken in favor of $Y$ as the median-grade winner, who indeed did receive more 1s.

a 1 (approve) to the remaining $(c - k)$, then the scores will be $(k - 1)/2$—the mean of $(0, 1,\ldots, k - 1)$—for each of the former candidates, and $(c + k - 1)/2$—the mean of $(k, k + 1,\ldots, c - 1)$—for each of the latter.

3. *Hare* When a voter ranks candidates as tied under the Hare system, and $m$ ways of resolving the aggregate of ties are possible, we specify that the voter's ballot is to be counted as a fraction of a ballot, $1/m$, for each of those $m$ ways, e.g., with two grades, $m$ is equal to $k! \times (c - k)!$ if a ballot has $k$ 0s and $(c - k)$ 1s. For instance, under Hare with three candidates and two grades, if a ballot gives a 1 to candidates $X$ and $Y$ and a 0 to candidate $Z$, then it is tallied as half a ballot ranked $(X, Y, Z)$ and half a ballot ranked $(Y, X, Z)$.

**Proposition 2** *With two grades, the sets of median–grade and Borda winners are each identical to the set of AG (and therefore also SG) winners. But the set of Hare winners can differ from the other four sets of winners.*

*Proof* *Median-grade winner* The candidate with the most 1s is the median-grade winner. But that candidate is also the AG winner.

*Borda winner* On a given ballot with $k$ 0s and $(c - k)$ 1s, the score difference between an approved and a non-approved candidate is $[(c + k - 1)/2 - (k - 1)/2]$, or $c/2$. The key to the proof for Borda is that this difference, $c/2$, does not depend on $k$, the number of candidates not approved. Ballots with the AG winner approved, and any other candidate $Z$ not approved, are more numerous than those of the opposite type, so the AG winner's score surplus of $c/2$ for each of the former ballots exceeds $Z$'s surplus of $c/2$ for each of the latter, thereby making the Borda and AG winners the same.

The proofs for median-grade and Borda winners extend easily if the winner sets contain more than one candidate.

*Hare winner* The following counterexample, with three candidates and 27 voters, establishes that the Hare winner set may differ from the AG winner set (and thus also from the SG, median–grade, and Borda winner sets):

*2 voters*: Approve of $A$ only
*6 voters*: Approve of both $A$ and $B$
*6 voters*: Approve of both $A$ and $C$
*7 voters*: Approve of $B$ only
*6 voters*: Approve of $C$ only

The AV totals are $2 + 6 + 6 = 14$ for $A$, $6 + 7 = 13$ for $B$, and $6 + 6 = 12$ for $C$, so average grades are 1/27 of those values, and $A$ is the AG winner. But the number of top ranks (i.e., first-place votes) is $2 + (½ \times 6) + (½ \times 6) = 8$ for $A$, $(½ \times 6) + 7 = 10$ for $B$, and $(½ \times 6) + 6 = 9$ for $C$. Thus, Hare eliminates $A$ on the first round, after which $B$ becomes the Hare winner by beating $C$, $6 + 7 = 13$ votes to $6 + 6 = 12$ (or 14 to 13 if the two voters who voted for $A$ only split evenly between $B$ and $C$). Note that, as must be the case, $A$ is (i) the SG winner, beating $B$ by 8 to 7 and $C$ by 8 to 6 (these numbers disregard indeterminate voters); (ii) the median-grade winner, with a median of 1, versus 0 for both $B$ and $C$; and (iii) the Borda winner, with points equal to 28½, versus 27 for $B$ and 25½ for $C$. □

Thus, when only two grades can be assigned, AG, SG, median-grade, and Borda mimic AV in rendering the same winners, whereas Hare does not.

To forge a link between grading systems like AV or median grade, and ranking systems like Borda or Hare, assume that the higher is the grade a voter gives to a candidate, the higher is his or her rank. If, as under AV, only two grades are possible, voters express *dichotomous preferences*: They prefer approved to nonapproved candidates, but they do not make finer distinctions within their approved and nonapproved subsets.

If voters have dichotomous preferences, under AV each voter has a dominant strategy of approving only the candidates in his or her preferred subset of candidates, which yields an outcome at least as good as and sometimes better than any other strategy. If all voters choose their dominant strategies, then a candidate wins under AV if and only if he or she is a Condorcet winner (Brams and Fishburn 1978). If voter preferences are not dichotomous, winners under the Borda count or the Hare system may differ from Condorcet winners; and AG winners may differ from SG winners, as we will show in Sect. 3.

If a Condorcet paradox occurs because of a top cycle, different Condorcet completion methods have been proposed for choosing one candidate from among the candidates in the top cycle. For example, the Black procedure chooses the Borda winner. Another possibility is to choose the AV winner, using a ballot that captures not only ranks but also approval votes, the latter to be used only if needed (Potthoff 2013, 2014).

We do not propose a "solution" for choosing among candidates in a top cycle when more than two grades can be assigned. Instead, as Proposition 1 shows, AV offers a solution: It renders identical the AG and SG winners—and, per Proposition 2, also the median-grade and Borda winners—in the case of two grades. This is decidedly not true for more than two grades, as we show in the next section, wherein we focus on differences between the AG and the SG winners.

## 3 The paradox of grading systems

We next present several examples of the paradox of grading systems. Except for one example with two candidates, we assume three candidates but varying numbers of voters and grades.

**Example 1** 3 grades, {2, 1, 0}, and 9 voters

*2 A voters*: Grades of (2, 1, 0) to (*A*, *B*, *C*)
*3 B voters*: Grades of (0, 2, 1) to (*A*, *B*, *C*)
*4 C voters*: Grades of (1, 0, 2) to (*A*, *B*, *C*)

*A*, *B*, and *C* voters are those who give the highest grade (i.e., 2) to, respectively, candidates *A*, *B*, and *C*. Multiplying the numbers of *A*, *B*, and *C* voters by the grades they give to each candidate, and dividing by the total number of voters, the AG winner is *C*, whose average grade is

$$\frac{(2 \times 0) + (3 \times 1) + (4 \times 2)}{9} = \frac{11}{9}$$

compared with average grades of 8/9 for *A* and *B*.

To determine the SG winner(s), we ask which candidate(s) receive more higher grades in the three pairwise contests:

- *A v. B*: 2 *A* and 4 *C* voters (total: 6) grade A higher; 3 *B* voters grade *B* higher
- *B v. C*: 2 *A* and 3 *B* voters (total: 5) grade *B* higher; 4 *C* voters grade *C* higher
- *C v. A*: 3 *B* and 4 *C* voters (total: 7) grade *C* higher; 2 *A* voters grade *A* higher

These contests generate a cycle of winners,

$$A > B > C > A,$$

in which a majority of voters grades each of $A$, $B$, and $C$ higher than the candidate to his or her right, in the cycle. Thus, whether a candidate wins or loses in a pairwise contest depends on whom he or she is paired against.

If a subset of candidates is in a top cycle (in Example 1, *all* candidates are in the top cycle, so there are no "bottom" candidates), we consider all candidates in the cycle to be SG winners. Because the set of three candidates differs from the single AG winner ($C$), a paradox of grading arises.[4]

There cannot be a paradox under AV (Proposition 1). To see why in the case of Example 1, assume that each of the $A$, $B$, and $C$ voters collapse their three grades into two by giving the candidate who receives the highest grade of 2 a grade of 1, and the candidate who receives the lowest grade of 0 a grade of 0. As for the candidates who receive the middle grade of 1, voters may or may not approve of them.

For concreteness, assume that the $A$ and $B$ voters approve only of their first choices, but the $C$ voters also approve of their second choice, $A$:

*2 A voters*: Grades of (1, 0, 0) to ($A$, $B$, $C$)
*3 B voters*: Grades of (0, 1, 0) to ($A$, $B$, $C$)
*4 C voters*: Grades of (1, 0, 1) to ($A$, $B$, $C$)

This would occur if the $C$ voters view $A$ as "close enough" to $C$ to be acceptable, too.

Observe that $A$ is the AG winner with an average grade of 6/9, trailed by $B$ and $C$ with average grades of 3/9 and 4/9, respectively (recall that before the dichotomization, $C$ was the AG winner). Because $A$ is graded

- higher than $B$ by the 2 $A$ and 4 $C$ voters (total: 6), and lower by the 3 $B$ voters
- higher than $C$ by the 2 $A$ voters, and the same by the 3 $B$ voters (both 0s) and the 4 $C$ voters (both 1s)

$A$ is also the SG winner, as guaranteed by Proposition 1, although $A$ is the candidate with the fewest first-choice supporters.

To illustrate situations in which a candidate like $C$ with the most first-choice support might be displaced under AV, consider the 1912 US presidential election, though it does not mirror Example 1 exactly. It is reasonable to suppose that Theodore Roosevelt, the Progressive ("Bull Moose") candidate who received 27 % of the popular vote, was acceptable to many supporters of the Republican candidate, William Howard Taft, who received 24 % of the popular vote. Roosevelt, after all, had been a Republican when he was president earlier and was, presumably, more acceptable to Taft supporters than Woodrow Wilson, the Democratic candidate who received 41 % of the popular vote. If Roosevelt had been approved of by both Progressive and Republican voters,[5] then he

[4] The Borda count gives rise to the same paradox in Example 1 if the grades of 2, 1, and 0 are interpreted as ranks, because the Borda winner, $C$, is different from the three candidates in the Condorcet cycle. In our later examples, however, the grades do not necessarily correspond to Borda scores, including tied scores, and the Borda and AG winners can differ, illustrating that ranking systems like Borda are categorically different from grading systems.

[5] Taft, also, would have benefited from the approval of both his supporters and Roosevelt supporters, but probably not to the extent of Roosevelt for two reasons: Roosevelt (i) edged him out in the plurality vote and (ii) probably would have drawn more support from Wilson voters, and even supporters of Eugene Debs, the Socialist candidate who received 6 % of the vote, because he generally was perceived to be less conservative than Taft.

would have easily beaten with 51 % approval the Democratic candidate, Woodrow Wilson, the candidate with the most first-choice support.

Example 1 illustrates the occurrence of the paradox of grading systems in the presence of a top cycle, in which a single AG winner ($C$) is different from the set of candidates in the cycle. When, as here, AG and SG winners overlap but do not coincide, we call the condition a *weak paradox of grading systems*. The paradox is starker when the AG and SG winners do not overlap, which we call a *strong paradox of grading systems*, as in our next example:

**Example 2** 4 grades, {3, 2, 1, 0}, and 3 voters

*1 voter*: Grades of (3, 0, 0) to ($A$, $B$, $C$)
*1 voter*: Grades of (2, 3, 3) to ($A$, $B$, $C$)
*1 voter*: Grades of (0, 1, 1) to ($A$, $B$, $C$)

The AG winner is $A$, with an average grade of 5/3, whereas $B$ and $C$ each receive an average grade of 4/3. But $B$ and $C$ receive higher grades than $A$ from the second and third voters, so they are the SG winners.

Example 2 also differs from Example 1 in showing that the paradox can occur with only 3 voters, though we assume in Example 2 that there are 4 rather than 3 grades. Note that the second voter does not give a bottom grade (0), and the third voter does not give a top grade (3) to any of the candidates, so they are sincere—or at least not strategic—in the sense defined earlier.

Of course, the first voter may also be sincere if he or she believes that $A$ genuinely is deserving of a top grade and both $B$ and $C$ of a bottom grade. Although we are not able to distinguish in a case like this whether a voter is acting sincerely or strategically if he or she gives no middle grades, we require in our strategic model that *all* voters give at least one candidate a top grade and one candidate a bottom grade. This is not the case for the second and third voters in Example 2.

Example 2 illustrates how a grading system differs from a ranking system like the Borda count, which in general requires that different ranks be given to each candidate (i.e., the ranking is *strict*).[6] A more significant difference is that, unlike ranking systems, wherein the ranks have the same meaning for all voters when they are aggregated, this is not true of grades.

To illustrate, the top grade of the first two voters (3) has triple the effect in helping a favorite candidate as the top grade of the third voter (1). Similarly, the bottom grade of the first and third voters (0) is much more significant in hurting a least favorite candidate than the bottom grade of the second voter (2). These statements apply only in an absolute sense, though. In a relative sense, the first voter has triple the effect of each of the last two, because $(3 - 0)$ is thrice $(3 - 2)$ or $(1 - 0)$.

While the ranks of a ranking system like the Borda count can be interpreted as grades (see footnotes 4 and 6), a grading system allows voters to express levels of satisfaction or dissatisfaction that a ranking system does not. In this sense, a grading system offers a richer menu of choices for voters.

[6] If a voter's ranking is not strict, the usual convention is to give all tied candidates the mean of the ranks as if the ranking were strict. To illustrate in Example 2, with ranks of 2, 1, and 0 for the three candidates, the Borda scores of the first voter would be (2, ½, ½), of the second voter (0, 1½, 1½), and of the third voter (0, 1½, 1½). In this example, $B$ and $C$ would be the tied winners, and no paradox of grading systems would occur, because these candidates are preferred by two of the three voters. In general, when translating ranks into grades or grades into ranks, ranking systems and grading systems may yield different outcomes.

Of course, AV limits this menu to two grades. Translating the grades of the voters in Example 2 into 1 and 0 in the obvious way (observe that each voter has dichotomous preferences, based on the grades), the AG and SG outcomes yield the same pair of candidates, showing that AV produces no paradox.

It is worth pointing out that even when only two candidates are on the ballot, a strong paradox of grading systems may crop up, as our next example demonstrates:

**Example 3** 3 grades, {2, 1, 0}, and 5 voters

*2 voters*: Grades of (2, 0) to ($A$, $B$)
*3 voters*: Grades of (0, 1) to ($A$, $B$)

The AG winner is $A$, with an average grade of 4/5, whereas $B$ receives an average grade of 3/5. But $B$ receives a higher grade from 3 of the 5 voters and so is the SG winner.

But with only two candidates, strategic voters will always have an incentive to give the top grade to their preferred candidate and the bottom grade to their nonpreferred candidate, making implausible the choice of a middling grade (1 in this example).[7] Accordingly, we assume in our simulation later that there are always at least three candidates in both the sincere and strategic models.

Our next example illustrates an extreme example of the strong paradox—that all voters except one may grade the SG winner higher than the AG winner.[8]

**Example 4** 6 grades, {5, 4, 3, 2, 1, 0}, and 5 voters

*1 voter*: Grades of (5, 0, 0) to ($A$, $B$, $C$)
*4 voters*: Grades of (0, 1, 1) to ($A$, $B$, $C$)

The AG winner is $A$, with an average grade of 5/5, whereas $B$ and $C$ each receive average grades of 4/5. But $B$ and $C$ receive higher grades than $A$ from 4 of the 5 voters, so they are the SG winners.

One can blame the election of $A$ on the 4 voters who give $B$ and $C$, their preferred candidates, only the next-lowest grade of 1. But independent of their choices, this strong paradox of grading systems would be nullified if, as under AV, the maximum grade that the first voter can give $A$ is 1.

A paradox can occur when there are three candidates, three voters, and three grades—even when voters are strategic, and each voter gives at least one candidate a maximum grade of 2, and at least one candidate a minimum grade of 0—as the following example demonstrates:

**Example 5** 3 grades, {2, 1, 0}, and 3 voters

*1 voter*: Grades of (2, 1, 0) to ($A$, $B$, $C$)
*1 voter*: Grades of (0, 2, 0) to ($A$, $B$, $C$)
*1 voter*: Grades of (0, 0, 2) to ($A$, $B$, $C$)

The AG winner is $B$, with an average grade of 3/3, whereas $A$ and $C$ each receive an average grade of 2/3. But one finds a cycle of winners,

[7] Such a dichotomization is effectively what AV forces, and not just for two candidates. In the presence of two candidates only, plurality voting suffices to enable a voter to vote for just his or her preferred candidate.

[8] This was also true in Example 2, but “all except one” meant only 2 of the 3 voters rather than, as in Example 4, 4 of the 5 voters.

$$\mathrm{A} \sim B > \mathrm{C} \sim A$$

where "$\sim$" denotes a tie—that is, candidates on each side of this sign receive higher grades than each other from the same number of voters (one each in this example). A (weak) paradox of grading systems occurs because the AG winner ($A$) and the SG winners ($A$, $B$, $C$) are not identical.

One might argue in this example that $B$ is the most deserving SG winner because he or she is the only candidate to have more higher grades (2) than one of the other candidates ($C$). On the other hand, each candidate can, as a result of the cycle, win or tie against another candidate.

We adopt an inclusive definition of SG winners by including *all* candidates in a top cycle linked by either ties or wins. This definition avoids the problem of determining, especially in complex cycles in which the candidates may have different numbers of ties or wins, which candidates in the cycle deserve to be called the SG winners.

We have given several examples of both the weak paradox and the strong paradox of grading systems. Such paradoxes can afflict any grading systems with more than two grades. While two grades preclude these paradoxes, we recognize that AV does limit a voter's choices—compared with grading systems that allow more than two grades—which is a tradeoff that we discuss in the concluding section.

We turn next to assessing the probability of the paradox under different conditions. But first we describe how we make our calculations and then present the results of our simulation.

## 4 The probability of the paradox of grading systems

Although it is straightforward to determine AG winners, the determination of SG winners requires a more complex procedure. We begin with a list of voters' grades, as illustrated by the sub-table in the top half of Table 1, in which 4 voters—shown along the rows {I, II, III, IV}— give 4 candidates—shown across the columns {$A$, $B$, $C$, $D$}—4 grades {0, 1, 2, 3}, which are the entries in the sub-table. Thus, for example, voter I gives candidate $C$ a grade of 1.

We call this the *data table*. Note that voter II does not satisfy the constraint of the strategic model, because he or she does not give a lowest grade of 0 or a highest grade of 3 to any candidate.

Let $c$ be the number of candidates. From the data table, we construct a $c \times c$ matrix whose entries, ($i$, $j$), are 1 if more voters give the row candidate $i$ a higher grade than the column candidate $j$, ½ if the same number of voters give candidates $i$ and $j$ higher grades, and 0 if more voters give candidate $j$ a higher grade than candidate $i$. (Voters who give $i$ and $j$ the same grade are not counted in this comparison.) We call this the *win matrix*. Its diagonal elements are all 0.

To illustrate the derivation of the win matrix from the data table, consider entry (1, 3) of the win matrix, which is based on comparing how many of the four voters grade $A$ higher than $C$ versus grade $C$ higher than $A$. Voters I, III, and IV grade $A$ higher, and none grade $C$ higher (voter II grades the two candidates the same), so the (1, 3) entry is 1, indicating that more voters grade $A$ higher.

The set of SG winners is derived from the win matrix. One method is as follows. Obtain the column totals in the win matrix—½, 2½, 2½, ½ for candidates $A$, $B$, $C$, $D$ (respectively) in the example. Order the candidates, low to high, according to their column

**Table 1** Data table and win matrix for 4 candidates, 4 voters, and 4 grades

| | Candidate *A* | Candidate *B* | Candidate *C* | Candidate *D* |
|---|---|---|---|---|
| Data table | | | | |
| Voter I | 2 | 0 | 1 | 3 |
| Voter II | 1 | 1 | 1 | 2 |
| Voter III | 3 | 0 | 0 | 0 |
| Voter IV | 3 | 1 | 0 | 0 |
| Win matrix | | | | |
| Candidate *A* | 0 | 1 | 1 | ½ |
| Candidate *B* | 0 | 0 | ½ | 0 |
| Candidate *C* | 0 | ½ | 0 | 0 |
| Candidate *D* | ½ | 1 | 1 | 0 |

totals—*A*, *D*, *B*, *C* here (the order does not matter when there are ties). Rearrange the columns of the data table so that they are in this same order—columns 1–4 for candidates *A*, *D*, *B*, *C*, respectively. Now use this new data table to create a new win matrix (actually, it is just the original one rearranged). Finally, examine successively the $1 \times 1$, $2 \times 2$, $3 \times 3$,… sub-matrices in the upper left corner of the new win matrix until reaching the first such sub-matrix that has nothing but 0s directly below it. The candidates associated with the columns (or rows) of that sub-matrix constitute the SG winner set. In the example, the first qualifying sub-matrix in the upper left corner is $2 \times 2$, so the SG winner set is $\{A, D\}$.

We next analyze the probability of the weak paradox of grading systems for different numbers of candidates ($c$), voters ($v$), and grades ($g$). We begin by assuming $g = 3$ and let $c$ and $v$ vary. Then we let $g$ vary as well.

Our results are based on a computer simulation using Python, in which we ran 10,000 trials to estimate the probability of the weak paradox for different values of ($c$, $v$, $g$).[9] The simulation is necessitated by the fact that even for very small parameter values, an exhaustive calculation can be infeasible. Each of the $cv$ cells in the data table can be filled with any of the $g$ grades. For example, assume that $(c, v, g) = (3, 4, 5)$. Then for each of the 3 candidates, each of the 4 voters can assign him or her one of 5 grades, giving

$$g^{cv} = 5^{3\times4} = 5^{12} = 244,140,625$$

different combinations for sincere voting. (Our simulation samples less than 0.0041 % of these possibilities.[10]) If the number of voters is as few as 20, the number of combinations is astronomical.

We present our simulation results for an impartial culture, in which each voter is assumed to be equally likely to give each possible grade to each of the candidates, which is what we earlier called the sincere model. The strategic model is one in which we constrain each voter to giving at least one candidate a bottom grade and at least one candidate a top grade.

We define and implement the strategic model as follows. Our definition specifies that all voter grade assignments that have at least one top grade and one bottom grade are equally

[9] We are grateful to Sean J. Vasquez for writing the computer program to do the simulation.

[10] The sampling is carried out with replacement, so in general it may include duplicate combinations, but that is unlikely for this example.

likely. Other definitions are possible, but ours seems reasonable, need not cause programming problems, and (at least for moderate parameter values) does not lead to computational infeasibility. To implement, we draw as for the sincere model, but, generally, some of the voter grade combinations that we draw will not satisfy the constraint of the strategic model. Then we replace each of the non-conforming voters by voters who do satisfy the constraint, drawing at random as many voters as is necessary until $v$ voters are found. Because of the need to replace voters who do not give both a bottom grade to one candidate and a top grade to another candidate, probabilities based on the strategic model take considerably longer to compute for the 10,000 trials (though a different technique could possibly reduce the time required).

In Table 2 we present the results of our simulation for three grades ($g = 3$) and different numbers of voters ($v$) and candidates ($c$). It may seem incongruous that the number of candidates can exceed the number of voters, but this situation is not as unusual as it sounds. It occurs, for example, when members of a small governing board or search committee (e.g., $v = 3$ in Table 2) must choose from among several candidates (e.g., $c = 10$). More common for elections that political scientists study is that the number of voters (e.g., $v = 50$ in Table 2) substantially exceeds the number of candidates (e.g., $c = 3$).

We have estimated through simulation the probabilities for the sincere and strategic models in all cases except $v = 25$ and 50. To calculate these probabilities for the strategic model without reducing the number of trials from 10,000 would have required an inordinate amount of computer time. Because the probabilities for the strategic model for $v = 3$–9 do not differ much from those of the sincere model (generally, by less than 10 % of the probability), there is not much lost in omitting the strategic probabilities for $v = 25$ and $v = 50$.

When exhaustive calculations needed to compute exact results can be made, one finds that the simulated probabilities are generally accurate to within ±0.01 of the exact results. This accords with the standard error of a sample probability, which is at most 0.005 (equal to $\sqrt{(1/2 \times 1/2)/10,000}$).[11] For example, when ($c$, $v$, $g$) = (3, 3, 3) for the sincere model, there are $g^{cv} = 3^{3\times3} = 3^9 = 19{,}683$ possible combinations. One can examine them all to determine that the exact probability of a weak paradox of grading systems—to four decimal places—is 4,302/19,683 = 0.2186, compared with a simulated probability of 0.2216, both of which round to 0.22. For the strategic model, the exact probability is 342/1,728 = 0.1979, compared with a simulated probability of 0.2037, both of which round to 0.20.

Because of the random element in the choice of 10,000 trials, the simulated probability will vary from sampling to sampling. Consistent with a standard error not above 0.005 for a simulated probability and not above $\sqrt{2}$ times that for the difference between two such probabilities, we have found that different samples generally produce simulated probabilities well within 0.02 of each other—and within ±0.01 of the exact probability, as noted above—demonstrating that the Monte Carlo simulation yields almost exact values. We summarize our findings from Table 2 for the sincere model when $g = 3$:

1. *Holding $v$ constant, the probability of the paradox increases with the number of candidates $c$.*
2. *Holding $c$ constant, the probability of the paradox increases with the number of voters up to about $v = 5$–8 and then declines.*

[11] For $n$ Bernoulli trials each with success probability $p$ (the binomial distribution), the standard error of the sample proportion of successes is $\sqrt{p(1-p)/n}$, which is maximized at $p = 1/2$.

**Table 2** Probability of weak paradox of grading systems for sincere (strategic) models with three grades ($g = 3$), as a function of the numbers of candidates ($c$) and voters ($v$)

| #Candidates → #Voters ↓ | $c = 3$ | $c = 4$ | $c = 5$ | $c = 6$ | $c = 9$ | $c = 10$ |
|---|---|---|---|---|---|---|
| $v = 3$ | 0.2216 (0.2037) | 0.2797 (0.2995) | 0.3339 (0.3501) | 0.3747 (0.4017) | 0.4340 (0.4490) | 0.4523 (0.4508) |
| $v = 4$ | 0.2443 (0.2874) | 0.3168 (0.3447) | 0.3487 (0.3740) | 0.3933 (0.4081) | 0.4687 (0.4698) | 0.4732 (0.4889) |
| $v = 5$ | 0.2526 (0.2315) | 0.3157 (0.3393) | 0.3663 (0.3757) | 0.4179 (0.4194) | 0.4847 (0.4807) | 0.4970 (0.4909) |
| $v = 6$ | 0.2627 (0.2671) | 0.3321 (0.3391) | 0.3839 (0.3758) | 0.4057 (0.4194) | 0.4897 (0.4909) | 0.5084 (0.5045) |
| $v = 7$ | 0.2633 (0.2394) | 0.3225 (0.3330) | 0.3746 (0.3724) | 0.4129 (0.4103) | 0.4863 (0.4918) | 0.4977 (0.4944) |
| $v = 8$ | 0.2658 (0.2555) | 0.3284 (0.3372) | 0.3854 (0.3708) | 0.4097 (0.4055) | 0.4900 (0.4891) | 0.5016 (0.5131) |
| $v = 9$ | 0.2709 (0.2426) | 0.3249 (0.3314) | 0.3780 (0.3681) | 0.4116 (0.4120) | 0.4857 (0.4880) | 0.4869 (0.5069) |
| $v = 25$ | 0.2518 | 0.3071 | 0.3448 | 0.3873 | 0.4441 | 0.4557 |
| $v = 50$ | 0.2310 | 0.2919 | 0.3253 | 0.3455 | 0.4087 | 0.4273 |

3. *The probabilities range from a low of 22 % ($v = 3$, $c = 3$) to a high of 51 % ($v = 6$, $c = 10$).*

These probabilities of the occurrence of the weak paradox, averaging about 40 %, are high, but recall that they do not mean that the AG and SG winners always will be non-overlapping. While the sets of AG and SG winners are different, some members may be common to the two sets. Only when a strong paradox occurs will there be no overlap in the AG and SG winners (illustrated by Examples 2, 3, and 4), ensuring that the AG and SG winners are completely different.[12]

Because every candidate has the same probability of obtaining each grade, on average he or she will have the same average grade, making the election highly competitive. Elections with three or more such candidates are rare, though we describe one in the concluding section.

Suffice it to say here that the probabilities in Table 2 likely overestimate the probability of a paradox of grading systems substantially in actual elections.[13] Nevertheless, it is probable that they accurately indicate what factors (increases in $c$; increases in $v$ up to 5–8 voters) increase the probability of the paradox.

In Table 3, we show the probability of the weak paradox for different numbers of grades ($g = 3, 5, 7, 9$).[14] We chose those values because they enable a voter to assign a median grade (1, 2, 3 or 4, when grades start at 0) to a candidate that he or she believes falls squarely in the middle. We summarize our findings from Table 3:

[12] We did not calculate the probability of a strong paradox. It seemed to us that the possibility, not the certainty, of different AG and SG winners was the first question to address in inquiring whether a discrepancy posed a serious problem.

[13] Regenwetter et al. (2006) make a similar point about the probability of the Condorcet paradox, showing that the paradox turns up much less often than its theoretical probability, based on the impartial-culture assumption.

[14] The first column of Table 3 when $g = 3$ repeats probabilities from Table 2.

4. *For fixed values of (c, v) for g up to 9, the probability of the paradox increases with the number of grades, at least up to g = 7.*
5. *The probabilities of the paradox range from a low of 22 % to a high of 56 %.*

As in Table 2, the probabilities of the paradox in Table 3 almost surely are overestimates of the probabilities that would be observed in actual elections. But the fact that AG and SG winners may differ for all grading systems except AV highlights the issue of how best to resolve such a discrepancy, which we will turn to in the concluding section.

It is well known that different ranking systems can produce different winners, so it should come as no surprise that the same is true of different grading systems. While we have concentrated on differences between AG and SG winners, these differences extend to median-grade winners, whose election is advocated by Balinski and Laraki (2011). Indeed, median-grade winners may differ from both AG and SG winners, as our final example shows:

**Example 6** 3 grades, {2, 1, 0}, and 9 voters

*2 A-B voters*: Grades of (2, 2, 0) to ($A$, $B$, $C$)
*3 B-C voters*: Grades of (1, 2, 2) to ($A$, $B$, $C$)
*4 A-C voters*: Grades of (1, 0, 1) to ($A$, $B$, $C$)

In fact, there is no overlap at all of the three kinds of winners:

- The AG winner is $A$, with an average grade of 11/9, whereas $B$ and $C$ have average grades of 10/9.
- The SG winner is $C$, because (i) 3 $B$–$C$ voters grade $C$ higher than $A$, whereas 2 $A$–$B$ voters grade $A$ higher than $C$ (the 4 $A$–$C$ voters are indifferent); and (ii) 4 $A$–$C$ voters grade $C$ higher than $B$, whereas 2 $A$–$B$ voters grade $B$ higher than $C$ (3 $B$–$C$ voters are indifferent). Hence, $C$ beats $A$ and $C$ beats $B$.
- The median-grade winner is $B$, with a median grade of 2, whereas $A$ and $C$ each have median grades of 1.

In this example, we can ask, and answer, who would have won under AV. Because all 9 voters, based on their grades (two candidates graded equally, and those higher than the third), have dichotomous preferences, we can translate their grades readily into 1s and 0s. Doing so yields AV vote totals for ($A$, $B$, $C$) of

$$[2 \times (1,1,0)] + [3 \times (0,1,1)] + [4 \times (1,0,1)] = (6,5,7),$$

so $C$ would be the winner, consistent with voters choosing their dominant strategies under AV and thereby electing the Condorcet winner, who is also the AG–SG winner under AV.

We can also ask what the probability of the (weak) paradox of grading is (i.e., AG and SG winner sets not identical) when, as in Example 6, $(c, v, g) = (3, 9, 3)$. From Tables 2 and 3, it is estimated to be 0.2709. We also have estimated through simulation the probability that the sets of AG, SG, and median-grade winners are all different (using the standard definition of the median), which is 0.1084 in this example.

But even the latter probability overestimates the chance of a strong paradox (involving all three of the AG, SG, and median-grade winner sets), because it allows for the overlap of the winning sets, which does not occur in Example 6. In fact, exhaustive examination of all 19,683 possible combinations, using SAS (and again using the standard definition of the median), shows that Example 6, except for the labeling of the candidates, is the *only* example in which the AG, SG, and median-grade winners all differ, without overlap, for subsets of 2, 3, and 4 voters whose members give the same grades (with $g = 3$) to the three

**Table 3** Probability of weak paradox of grading systems for sincere model, as a function of the numbers of candidates ($c$), voters ($v$), and grades ($g$)

| #Grades → #Candidates/#Voters ↓ | $g = 3$ | $g = 5$ | $g = 7$ | $g = 9$ |
|---|---|---|---|---|
| $c = 3$: | | | | |
| $v = 3$ | 0.2216 | 0.3276 | 0.3372 | 0.3344 |
| $v = 4$ | 0.2443 | 0.3763 | 0.4183 | 0.4534 |
| $v = 9$ | 0.2709 | 0.3369 | 0.3614 | 0.3572 |
| $v = 25$ | 0.2518 | 0.3059 | 0.3310 | 0.3378 |
| $c = 4$: | | | | |
| $v = 4$ | 0.3168 | 0.4661 | 0.5103 | 0.5476 |
| $v = 5$ | 0.3157 | 0.4307 | 0.4385 | 0.4385 |
| $v = 9$ | 0.3249 | 0.4196 | 0.4394 | 0.4475 |
| $v = 25$ | 0.3071 | 0.3733 | 0.3962 | 0.4452 |
| $c = 5$: | | | | |
| $v = 5$ | 0.3663 | 0.4947 | 0.5008 | 0.5016 |
| $v = 6$ | 0.3839 | 0.5000 | 0.5275 | 0.5562 |
| $v = 9$ | 0.3780 | 0.4657 | 0.4997 | 0.5021 |
| $v = 25$ | 0.3448 | 0.4399 | 0.4526 | 0.4631 |

candidates. These subsets may be thought of as political parties, whose members grade the candidates in the manner shown in Example 6.

Example 6 throws into sharp relief the normative question of who *should* win when different grading systems yield different winners, especially when there is a strong paradox (even including a different median-grade winner—see Example 6) and, therefore, no overlap among the different winners. We offer some thoughts on this question in the final section.

## 5 Conclusions

Just as different ranking systems, such as the Borda count and the Hare system of single transferable vote, may produce different winners, so may different grading systems if more than two grades can be assigned. We gave several examples in which the average-grade (AG) and superior-grade (SG) winners are different, and one example in which the AG, SG, and median-grade winners not only differ but also have no overlap. When these differences occur, a paradox of grading systems arises, becoming a strong paradox when no overlap exists in the sets of winners (so each system elects a different winner).

The probability of the paradox for the sincere and strategic models increases with the number of candidates; and with the number of voters up to about 5–8, depending on the number of candidates. The probability of the paradox also increases with the number of possible grades, at least up to 7.

The strategic model, in which all voters give at least one candidate a bottom grade and at least one candidate a top grade, seems a more realistic model than the sincere model without such a constraint. Giving bottom and top grades to one's least and most preferred candidates increases the chances of not electing a least preferred, and electing a most preferred, candidate.

But this leaves open how one should grade candidates who fall in the middle, which is analyzed in Brams and Fishburn (2007 [1983], ch. 5). Unless voters dichotomize their grades by grading up some of these candidates to the maximum grade, and grading down the others to the minimum grade—which is in effect to reduce the grades to two (as under approval voting, AV)—then a paradox can occur.

We know of no uses of a grading system in public elections, so there is no basis for judging the relative frequency of a paradox in such elections. But examples of actual elections exist, such as the 1977 Democratic mayoral primary in New York City, in which the paradox would likely have occurred if a grading system had been used. In that election, no candidate received as much as 20 % of the vote, and five candidates received more than 10 % (Edward Koch won with 19.8 %, just beating out Mario Cuomo, who received 18.7 %).

In close elections in which candidates bunch at the top, it may be desirable to let voters make judgments of who, among their middling candidates, merits approval and who does not. In that sense, AV is "responsive" to the views of each voter (Brams et al. 1988), though others—including Saari and Van Newenhizen (1988), who favor the Borda count—disagree. Still other theorists have suggested from three (Felsenthal 1989; Hillinger 2005; Alcantud and Laruelle 2014) to six (Balinski and Laraki 2011) to more grades be used, but they disagree as to whether it is better to choose an AG winner, an SG (or Condorcet) winner, or a median-grade winner.

Among grading systems, we favor AV ($g = 2$), not only because it ensures that the AG and SG winners are the same, so no paradox can arise, but also because it is sensitive to how individual voters dichotomize their middling candidates. Whether voters approve of many or few such candidates, their choices influence the choice of a winner. As a case in point, if the first voter in Example 5 approves of the candidate graded 1 ($B$), $B$ wins under AV, but if not a three-way tie among $A$, $B$, and $C$ occurs, in which case one candidate presumably would be chosen at random. Likewise, depending on how the $A$, $B$, and $C$ voters in Example 1 dichotomize their 1 votes, any of the three candidates could be the unique AV winner.

If voters can assign more than two grades, they will have a strong incentive to give a maximum grade to only one candidate—at least if he or she is viewed as a contender—and a minimum grade to all the other candidates. If a voter's favorite is not viewed as a contender, then he or she will be motivated to give a maximum grade to his or her (noncompetitive) favorite *and* a favorite among the contenders, and a minimum grade to everybody else. In either case, these strategies are tantamount to those that would be used under AV.

To be sure, in some nonpublic elections, such as in judging certain sporting events (e.g., figure skating) and rating wines, grading systems are used and generally work well. In these contests, there typically is not a single winner. Three players usually receive medals in sports contests, though not all have the same prestige, and more than one wine may be rated highly. However, the award(s) go to the AG winners, not the SG winners, though a case can be made that the SG winners are just as deserving when a paradox arises.

To conclude, we have identified a paradox of grading systems, which does not just mirror the well-known differences that crop up in aggregating votes under ranking systems. Unlike those systems, for which there is no accepted way of reconciling which candidate to choose when, for example, the Hare, Borda, and Condorcet winners differ, AV provides a solution when the AG and SG winners differ. We believe this to be an eminently sensible way of reaching a consensus on who, normatively speaking, *should* win.

Although AV may not be the ideal system to use in figure-skating contests or wine-tasting competitions—wherein voters have little or no incentive to dichotomize their grades—we think it has much to commend in political contests in which more than two grades can result in the paradox, even when voters are strategic. AV has been advocated for other reasons, including its simplicity and practicality,[15] but our intent in this paper has been to show how it resolves a troublesome paradox.

**Acknowledgments** We thank two reviewers and the editor for valuable comments.

## References

Alcantud, J. C. R., & Laruelle, A. (2014). Dis&Approval voting: A characterization. *Social Choice and Welfare*, *43*(1), 1–10.

Arrow, K. J. (1963 [1951]). *Social choice and individual values*. New Haven, CT: Yale University Press.

Balinski, M. L., & Laraki, R. (2011). *Majority judgment*. Cambridge, MA: MIT Press.

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

Bowler, S., Donovan, T., & Brockington, D. (2003). *Electoral reform and minority representation: Local experiments with alternative elections*. Columbus, OH: Ohio State University Press.

Brams, S. J. (2003 [1975]). *Game theory and politics*. Mineola, NY: Dover.

Brams, S. J. (2008). *Mathematics and democracy: Designing better voting and fair-division procedures*. Princeton, NJ: Princeton University Press.

Brams, S. J., & Fishburn, P. C. (1978). Approval voting. *American Political Science Review, 72*(3), 831–847.

Brams, S. J., & Fishburn, P. C. (2005). Going from theory to practice: The mixed success of approval voting. *Social Choice and Welfare, 25*(2), 457–474.

Brams, S. J., & Fishburn, P. C. (2007 [1983]). *Approval voting*. New York: Springer.

Brams, S. J., Fishburn, P. C., & Merrill, S. III. (1988). The responsiveness of approval voting: Comments on Saari and Van Newenhizen; Rejoinder to Saari and Van Newenhizen. *Public Choice*, *59*(2), 121–131; 149.

Brams, S. J., & Sanver, M. R. (2009). Voting systems that combine approval and preference. In S. J. Brams, W. V. Gehrlein, & F. S. Roberts (Eds.), *The mathematics of preference, choice, and order: Essays in honor of Peter C. Fishburn* (pp. 215–237). Berlin: Springer.

Camps, R., Mora, X., & Saumell, L. (2014). Choosing by means of approval-preferential voting: The revised approval choice. Preprint, http://mat.uab.cat/departament/Publ/prep/p21_14.pdf.

Center for Election Science (2015). http://www.electology.org.

Center for Range Voting (2015). http://rangevoting.org.

Felsenthal, D. S. (1989). On combining approval with disapproval voting. *Behavioral Science, 34*(1), 53–60.

Felsenthal, D. S., & Machover, M. (Eds.). (2012). *Electoral systems: Paradoxes, assumptions, and procedures*. Berlin: Springer.

Hillinger, C. (2005). The case for utilitarian voting. *Homo Oeconomicus, 22*(3), 295–321.

Laslier, J.-F., & Sanver, M. R. (Eds.). (2010). *Handbook on approval voting*. Berlin: Springer.

Merrill, S. III., & Nagel, J. (1987). The effect of approval balloting on strategic voting under alternative decision rules. *American Political Science Review*, *81*(2), 509–524.

Potthoff, R. F. (2013). Simple manipulation-resistant voting systems designed to elect Condorcet candidates and suitable for large-scale public elections. *Social Choice and Welfare, 40*(1), 101–122.

Potthoff, R. F. (2014). Condorcet completion methods that inhibit manipulation through exploiting knowledge of electorate preferences. *Games, 5*(4), 204–233.

Regenwetter, M., Grofman, B., Marley, A. A. J., & Tsetlin, I. (2006). *Behavioral social choice: Probabilistic models, statistical inference, and applications*. Cambridge, UK: Cambridge University Press.

[15] Some uses of, and experience with, AV in elections are discussed in Brams and Fishburn (2005), Laslier and Sanver (2010), Felsenthal and Machover (2012), Baujard et al. (2014), and references therein. Brams (2008), Brams and Sanver (2009), Sanver (2010), and Camps et al. (2014) suggest ways in which ranking and grading systems can be combined or otherwise reconciled.

Saari, D. G., & Van Newenhizen, J. (1988). Is approval voting an 'unmitigated evil': A response to Brams, Fishburn, and Merrill. *Public Choice*, *59*(2), 133–147.

Sanver, M. R. (2010). Approval as an intrinsic part of preference. In J.-F. Laslier & M. R. Sanver (Eds.), *Handbook on approval voting* (pp. 469–481). Berlin: Springer.