ECONOMIC RESEARCH REPORTS

Voter Sovereignty and Election Outcomes

by

Steven J. Brams
&
M. Remzi Sanver

RR#: 2003-07 **September 2003**

C.V. Starr Center for Applied Economics

Department of Economics

Faculty of Arts and Science

New York University

269 Mercer Street, 3rd Floor

New York, New York 10003-6687

Voter Sovereignty and Election Outcomes

Steven J. Brams
Department of Politics
New York University
New York, NY 10003
USA
steven.brams@nyu.edu

M. Remzi Sanver
Department of Economics
Istanbul Bilgi University
80310, Kustepe, Istanbul
TURKEY
sanver@bilgi.edu.tr

Prepared for delivery at the 2003 Annual Meeting of the American Political Science Association, August 28 - August 31, 2003.

Abstract

Voters are *sovereign* to the degree that they can express their approval for any set of candidates and, by so doing, help elect or prevent the election of candidates. While voter sovereignty is maximized under approval voting (AV), AV can lead to

- a plethora of outcomes, depending on where voters draw the line between acceptable and unacceptable candidates; and
- Condorcet losers and other lesser candidates, even in equilibrium.

But we argue that voters' judgments about candidate acceptability *should* take precedence over standard social-choice criteria, such as electing a Condorcet or Borda winner. Among other things, we show that

- sincere outcomes under all voting systems considered are AV outcomes, but not vice versa;
- a Condorcet winner's election under AV is always a strong Nash-equilibrium outcome but not under other systems, including those that guarantee the election of Condorcet winners if voters are sincere.

JEL Classification: D71, D72, C70.

Keywords: approval voting; elections; Condorcet winner/loser; Nash equilibrium.

Voter Sovereignty and Election Outcomes[1]

1. Introduction

Our thesis in this paper is that *several* outcomes of single-winner elections may be acceptable. Perhaps the most dramatic recent example illustrating this proposition is the 2000 US presidential election, in which George W. Bush won the electoral vote—disputed though it was in Florida—and Al Gore won the popular vote. Each of these candidates could claim to be the winner according to one of these criteria, but only the electoral vote mattered in the end.

To be sure, the extreme closeness of this election was unusual. But many elections, especially those with three or more candidates, may have more than one acceptable outcome.

For example, even when there is a Condorcet winner, who can defeat every other candidate in pairwise contests, there may be a different Borda-count winner, who on the average is ranked higher than a Condorcet winner. If there is no Condorcet winner because of cyclical majorities, the Condorcet cycle may be broken at its weakest link to select the strongest candidate in the cycle, who need not be the Borda winner.

That different voting systems can give different outcomes is, of course, an old story. The observation that different outcomes may satisfy different social-choice criteria is also old hat (Nurmi, 1999, 2002, and Brams and Fishburn, 2002, give many examples). What is new here is our claim that in an election with three or more candidates, other outcomes—not just the Condorcet winner, the Borda-count winner, or the strongest

[1] We thank Eyal Baharad, Dan S. Felsenthal, Peter C. Fishburn, Shmuel Nitzan, Richard F. Potthoff, and Ismail Saglam for valuable suggestions. Steven J. Brams acknowledges the support of the C.V. Starr Center for Applied Economics at New York University.

candidate in a cycle—may be *more* acceptable to the electorate. In fact, even a Condorcet loser, who would lose in pairwise contests to every other candidate, may turn out to be the most acceptable candidate.

To justify this last statement, we need to define some measure of "acceptability." If voters rank candidates from best to worst, where they draw the line in their rankings between acceptable and unacceptable candidates offers one such measure. It is precisely this information that is elicited under approval voting (AV), whereby voters can approve of as many candidates as they like or consider acceptable. This gives them the opportunity to be sovereign by expressing their approval for any set of candidates, which no other voting system permits.[2] In so doing, AV better enables voters both to elect and to prevent the election of candidates, as we will prove.

Call a candidate a *Pareto candidate* if there is no other candidate that all voters rank higher. We demonstrate that candidates selected under AV always include at least one Pareto candidate. In fact, AV *dominates* so-called scoring systems, including plurality voting (PV) and the Borda count (BC), with respect to the election of Pareto candidates: A Pareto candidate elected by a scoring system is always elected by AV for some sincere and admissible strategies, but not vice versa. This is also true for ranking systems that do not rely on scoring, including the Hare system of single transferable vote (STV) and the majoritarian compromise (MC), as we will show.

[2] Voter sovereignty should be distinguished from Arrow's (1963) condition of "citizen sovereignty," whereby for any two alternatives *a* and *b*, if all voters prefer *a* to *b*, *a* cannot be prohibited as the social choice. If voters are "sincere," AV satisfies citizen sovereignty, because all voters who approve of *b* will also approve of *a*. Note that voter sovereignty describes the behavior of individual voters whereas citizen sovereignty is a property of a voting system.

But if AV does a better job of finding Pareto candidates, doesn't it open the door to a plethora of possibilities? Isn't this a vice rather than a virtue, as some have argued (e.g., Saari and Van Newenhizen, 1988a; Saari, 1994, 2001)?[3]

This argument might have merit if the plethora of possibilities were haphazard choices that could easily be upset when voters are manipulative. But we show that AV often leads to Nash-equilibrium outcomes, from which voters with the same preferences will have no incentive to depart. Moreover, if voters with different preferences are able to coordinate their choices and none has an incentive to depart, AV *guarantees* the election of a unique Condorcet winner (if one exists).

The latter notion of stability is that of a *strong* Nash equilibrium, which yields outcomes that are invulnerable to departures by *any* set of voters. None of the other voting systems we assay guarantees that a unique Condorcet winner, and only a Condorcet winner, will be a strong Nash equilibrium outcome when voters are sincere. While AV offers this guarantee, however, it also allows for other Nash-equilibrium outcomes, including even a Condorcet loser, who may be the most acceptable candidate, even in equilibrium.

In section 2, we define preferences and strategies under AV and give an example that illustrates the choice of sincere, admissible strategies. In section 3 we characterize AV outcomes, describing the "critical strategy profile" that produces them, and compare these outcomes with those given by other voting systems. Among other things, we show

[3] The critique of AV by Saari and Van Newenhizen (1988a) provoked an exchange between Brams, Fishburn, and Merrill (1988a, 1988b) and Saari and Van Newenhizen (1988b) over whether the plethora of AV outcomes more reflected AV's "indeterminacy" (Saari and Van Newenhizen) or its "responsiveness" (Brams, Merrill, and Fishburn); other critiques of AV are referenced in Brams and Fishburn (forthcoming). Here we argue that *which* outcome is chosen should depend on voters' judgments about the acceptability of candidates rather than standard social-choice criteria, which—as we will show— may clash with these judgments.

that no "fixed rule," in which voters vote for a predetermined number of candidates, always elects a unique Condorcet winner, suggesting the need for a more flexible system.

The stability of outcomes under the different voting systems is analyzed in section 4, wherein we show that Nash equilibria and strong Nash equilibria may vary from system to system. Also, Condorcet voting systems, which guarantee the election of Condorcet winners when voters are sincere, may not elect Condorcet winners in equilibrium.

Nonstrong Nash equilibria might be thought of as possessing a kind of local stability, whereas strong Nash equilibria possess a global stability. These different kinds of equilibria may coexist, which is to say that *which* stable outcome is chosen will depend on which candidates voters consider acceptable and whether they coordinate their choices. In large-scale public elections, coordination is typically done when voters draw inferences from polls, not by face-to-face communication, which is commonplace in smaller settings like committees.

That a Condorcet winner is a globally stable choice under AV should not be surprising. What is more surprising is that such a candidate can be upset if (i) coordination is difficult and (ii) many voters consider another candidate more acceptable.

Speaking normatively, we believe that voters *should* be sovereign, able to express their approval of any set of candidates. Likewise, a voting system *should* allow for the possibility of multiple acceptable outcomes, especially in close elections. That AV more than other voting systems is responsive in this way we regard as a virtue.

That it singles out as strong Nash–equilibrium outcomes unique Condorcet winners may or may not be desirable. We discuss these and other questions related to the nature of acceptable outcomes in section 5, where we suggest that "acceptability" replace the usual social-choice criteria for assessing the satisfactoriness of election outcomes chosen by sovereign voters.

2. Preferences and Strategies under AV

Consider a set of voters choosing among a set of candidates. We denote individual candidates by small letters *a*, *b*, *c*, …. A voter's *strict preference relation* over candidates will be denoted by *P*, so *aPb* means that a voter strictly prefers *a* to *b*, which we will denote by the following left-to-right ranking (separated by a space): *a b*. We assume in the subsequent analysis that all voters have strict preferences, so they are not indifferent among two or more candidates.[4]

We assume that every voter has a *connected* preference: For any *a* and *b*, either *a b* or *b a* holds. Moreover, *P* is *transitive*, so *a c* whenever *a b* and *b c*. The list of preferences of *all* voters is called a *preference profile* **P**.

An *AV strategy S* is a subset of candidates. Choosing a strategy under AV means voting for all candidates in the subset and no candidates outside it. The list of strategies of all voters is called a *strategy profile* **S**.

The number of votes that candidate *i* receives at **S** is the number of voters who include *i* in the strategy *S* that they select. For any **S**, there will be a set of candidates ("winners") who receive the greatest number of votes.

An AV strategy *S* of a focal voter is *admissible* if it is not dominated by any other strategy—that is, if there is no other strategy that gives outcomes at least as good as, and sometimes better than, *S* for all strategy profiles **S** of voters other than the focal voter.

Brams and Fishburn (1978, 1983) show that admissible strategies under AV involve always voting for a most-preferred candidate and never voting for a least-preferred candidate.

An AV strategy is *sincere* if, given the lowest-ranked candidate that a voter approves of, he or she also approves of all candidates ranked higher. Thus, if S is sincere, there are no "holes" in a voter's approval set: Everybody ranked above the lowest-ranked candidate that a voter approves of is also approved; and everybody ranked below is not approved.[5] A strategy profile **S** is said to be admissible and sincere if and only if the strategy S that every voter chooses is admissible and sincere, based on each voter's preference P.

As an illustration of these concepts, assume that there are 7 voters who can be grouped into three different types, each having the same preference P over the set of four candidates $\{a, b, c, d\}$:

Example 1

1. 3 voters: $a\ b\ c\ d$
2. 2 voters: $b\ c\ a\ d$
3. 2 voters: $d\ b\ c\ a$

The three types define the preference profile **P** of all 7 voters. We assume that all voters of each type choose the same strategy S.

---

[4] This restriction simplifies the analysis; its relaxation to allow for voter indifference among candidates has no significant effect on our findings.

[5] Admissible strategies may be insincere if there are four or more candidates. For example, if there are exactly four candidates, it may be rational for a voter to approve of his or her first and third choices without also approving of a second choice (see Brams and Fishburn, 1983, pp. 25-26, for an example). However, the circumstances under which this happens are sufficiently rare and nonintuitive that we henceforth

Voters of type (1) have three sincere, admissible strategies: $\{a\}$, $\{a, b\}$, and $\{a, b, c\}$, which for convenience we write as *a*, *ab*, and *abc*. A typical sincere, admissible strategy profile of the 7 voters is **S** = (*a*, *a*, *a*, *bc*, *bc*, *dbc*, *dbc*), whereby the 3 voters of type (1) approve of only their top candidate, the 2 voters of type (2) approve of their top two candidates, and the 2 voters of type (3) approve of all candidates except their lowest-ranked. The number of votes of each candidate at **S** is 4 votes for *b* and *c*, 3 votes for *a*, and 2 votes for *d*. Hence, AV selects candidates $\{b, c\}$ as the (tied) winners at **S**.

3. Election Outcomes under AV and Other Voting Systems

Given a preference profile **P**, we consider the set of all candidates that *can* be chosen by AV when voters use sincere, admissible strategies. We call this set *AV outcomes*. Clearly, a candidate ranked last by all voters cannot be in this set, because it is inadmissible for any voter to vote for this candidate.

Define an *AV critical strategy profile* for candidate *i* at preference profile **P** as follows: Every voter who ranks *i* as his or her worst candidate votes only for the candidate that he or she ranks top. The remaining voters vote for *i and* all candidates they prefer to *i*.

Let $C_i(\mathbf{P})$ be the AV critical strategy profile of candidate *i*. In Example 1, the critical strategy profile for candidate *a* is $C_a(\mathbf{P})$ = (*a*, *a*, *a*, *bca*, *bca*, *d*, *d*), giving *a* 5 votes compared to 2 votes each for *b*, *c*, and *d*. It can easily be seen that $C_i(\mathbf{P})$ is admissible and sincere.

We next give four lemmata that provide a theoretical foundation for several of our subsequent propositions. They (i) show that under AV candidate *i* cannot do better than

---

suppose that voters choose only sincere, admissible strategies under AV. Relaxing this assumption at $C_i(\mathbf{P})$; (ii) characterize AV outcomes; (iii) characterize outcomes that can never be chosen under AV; and (iv) characterize outcomes that must be chosen under AV.

Lemma 1. *Assume all voters choose sincere, admissible strategies. The AV critical strategy profile for candidate i, $C_i(\mathbf{P})$, maximizes the difference between the number of votes that i receives and the number of votes that every other candidate j receives.*

Proof. Clearly, no other sincere, admissible strategy profile yields candidate *i* more votes than its AV critical strategy profile $C_i(\mathbf{P})$. Now consider the number of votes received by any other candidate *j* at $C_i(\mathbf{P})$. Candidate *j* will receive no fewer and sometimes more votes if there are the following departures from $C_i(\mathbf{P})$:

(i) a voter who ranked candidate *i* last, and therefore did not vote for him or her, votes for one or more candidates ranked below his or her top-ranked choice (possibly including candidate *j*); or

(ii) a voter who did not rank candidate *i* last or next-to-last votes for one or more candidates ranked below *i* (possibly including candidate *j*).

In either case, candidate *j* never gets fewer, and may get more, votes when there are these departures from candidate *i*'s critical strategy profile $C_i(\mathbf{P})$. Because (i) and (ii) exhaust the possible departures from $C_i(\mathbf{P})$ that involve voting for some other candidate *j*, candidate *i* cannot do better vis-à-vis candidate *j* than at $C_i(\mathbf{P})$. Q.E.D.

Using Lemma 1, we give a simple way to determine whether any candidate *i* is an AV outcome:

---

complicates the analysis but does not significantly alter our main findings.

Lemma 2. *Candidate i is an AV outcome if and only if i is chosen at his or her critical strategy profile C_i(**P**).*

Proof. The "if" part is a direct consequence of the fact that C_i(**P**) is sincere and admissible. To show the "only if" part, suppose candidate *i* is not chosen by AV at C_i(**P**). By Proposition 1, C_i(**P**) maximizes the difference between the number of votes that *i* receives and the number of votes that any other candidate *j* receives, so there is no other sincere, admissible strategy profile at which *i* can be chosen by AV. Q.E.D.

Using Lemma 2, we give a characterization of candidates that *cannot* be AV outcomes.

Lemma 3. *Given any preference profile **P** and any candidate i, i cannot be an AV outcome if and only if there exists some other candidate j such that the number of voters who consider j as their best choice and i as their worst choice exceeds the number of voters who prefer i to j.*

Proof. Given any preference profile **P** and any two candidates *i* and *j*, voters can be partitioned into three (disjoint) classes:

(i) those who prefer *i* to *j*;

(ii) those who consider *j* as the best choice and *i* as the worst choice; and

(iii) those who prefer *j* to *i* but do not fall into class (ii).

At critical strategy profile C_i(**P**), the voters in class (i) will vote for *i* but not *j*; those in class (ii) will vote for *j* but not *i*; and those in class (iii) will vote for both *i* and *j*. Setting aside class (iii), which gives each candidate the same number of votes, candidate *i* cannot be selected at $C_i(\mathbf{P})$ if and only if the number of voters in class (ii) exceeds the number of voters in class (i). Hence, by Lemma 2 candidate *i* cannot be an AV outcome. Q.E.D.

In effect, Lemma 3 extends Lemma 2 by saying precisely when candidate *i* will be defeated by candidate *j* and cannot, therefore, be an AV outcome.

Call a candidate *AV-dominant* if and only if, whatever sincere, admissible strategies voters choose, this candidate is the unique winner under AV.

Lemma 4. *Given any preference profile* ***P*** *and any candidate i, i is AV-dominant if and only if, given any other candidate j, the number of voters who consider i as their best choice and j as their worst choice exceeds the number of voters who prefer j to i.*

Proof. We begin with the "if" part. All voters who consider *i* as their best choice and *j* as their worst choice will vote for *i* and not for *j* under AV. Because this number exceeds the number of voters who prefer *j* to *i*— and would, in the worst situation for *i*, vote for *j* and not for *i*—*i* always receives more votes than *j*. For the "only if" part, assume there exists some *j* such that the number of voters who prefer *j* to *i* equals or exceeds the number of voters who rank *i* as their best choice and *j* as their worst choice. If the voters who prefer *j* to *i* vote for *j* and not for *i*, and the voters who prefer *i* to *j* (without ranking *i* as their best and *j* as their worst choice) vote for both *i* and *j*, the number of votes that *i* receives will not be greater than the number of votes that *j* receives. Consequently, *i* will not be the unique winner under AV. Q.E.D.

AV can generate a plethora of outcomes. Consider again Example 1, in which we showed earlier that AV selects candidate *a* at $C_a(\mathbf{P})$. Similarly, AV selects candidates *b* and $\{b, c\}$, all with 7 votes, at critical strategy profiles $C_b(\mathbf{P}) = \{ab, ab, ab, b, b, db, db\}$

and $C_c(\mathbf{P}) = \{abc, abc, abc, bc, bc, dbc, dbc\}$. However, $C_d(\mathbf{P}) = \{a, a, a, b, b, d, d\}$, so candidate *a* (3 votes) rather than candidate *d* (2 votes) is chosen at candidate *d*'s critical strategy profile.[6] In sum, the set of AV outcomes that are possible in Example 1 is $\{a, b, \{b, c\}\}$. Although none of the three candidates is AV-dominant, candidate *a* would be AV-dominant if there were, for example, 2 *a b c* voters, 2 *a c b* voters, and 1 *b c a* voter. Candidate *a* would always get 4 votes, whereas candidates *b* and *c* would at best get 3 votes each.

As noted earlier, a candidate is a *Pareto candidate* if there is no other candidate that all voters rank higher. Example 1 illustrates three things about the tie-in of Pareto candidates and AV outcomes:

- *a* and *b* are Pareto candidates and AV outcomes;
- *c* is not a Pareto candidate but is a component of an AV outcome (it ties with *b* at $C_c(\mathbf{P})$); and
- *d* is a Pareto candidate but not an AV outcome.

These observations are generalized by the following proposition:

Proposition 1. *The following are true about the relationship of Pareto candidates and AV outcomes:*

(i) At every preference profile $\mathbf{P}$*, there exists a Pareto candidate that is an AV outcome or a component of an AV outcome;*

(ii) Not every Pareto candidate is necessarily an AV outcome; and

[6] That *d* cannot be chosen also follows from Lemma 3: More voters (3) consider *a* as their best choice and *d* as their worst choice than prefer *d* to *a* (2).

(iii) A non-Pareto candidate may be a component of an AV outcome but never a unique AV outcome.

Proof. To show (i), take any preference profile **P**. Assume that every voter votes only for his or her top choice. Then the one or more candidates chosen by AV, because they are top-ranked by some voters, must be Pareto candidates. To show (ii), it suffices to check the critical strategy profile $C_d(\mathbf{P})$ of Example 1, wherein candidate *d* is not an AV outcome but is a Pareto candidate because *d* is top-ranked by the 2 type (3) voters.

In Example 1, we showed that *c* is not a Pareto candidate but is a component of an AV outcome. To show that a non-Pareto candidate can never be a unique AV outcome and prove (iii), consider any **P** at which there exists a non-Pareto candidate *i* that is a component of an AV outcome. Take any sincere, admissible strategy profile **S** where this outcome is selected. Because *i* is not a Pareto candidate, there exists some other candidate *j* that every voter prefers to *i*. Hence, every voter who voted for *i* at **S** must have voted for *j* as well, which implies that *i* and *j* tie for the most votes. Indeed, *all* candidates *j* that Pareto dominate *i* will be components of an AV outcome at **S**. Because at least one of the candidates *j* that Pareto-dominate *i* must be ranked higher by one or more voters than all other candidates *j*, AV picks a Pareto candidate that ties candidate *i*. Q.E.D.

In Example 1, candidate *b* is the *Condorcet winner*, who can defeat all other candidates in pairwise contests, and candidate *d* is the *Condorcet loser*, who is defeated by all other candidates in pairwise contests. Not surprisingly, *b* is an AV outcome but *d* is not. However, consider the following 7-voter, 3-candidate example:

Example 2

1. 3 voters: *a b c*
2. 2 voters: *b c a*
3. 2 voters: *c b a*

Notice that the 2 type (2) and the 2 type (3) voters prefer candidates *b* and *c* to candidate *a*, so *a* is the Condorcet loser. But because the critical strategy profile of candidate *a* is $C_a(\mathbf{P}) = (a, a, a, b, b, c, c)$, *a* is an AV outcome—as are also candidates *b* and *c*, rendering all three candidates in this example AV outcomes.

We summarize the Condorcet properties of AV outcomes with our next proposition:

Proposition 2. *Condorcet winners are always AV outcomes, whereas Condorcet losers may or may not be AV outcomes.*

Proof. If candidate *i* is a Condorcet winner, a majority of voters prefer *i* to every other candidate *j*. This implies that fewer voters rank *j* as their best choice and *i* as their worst choice, which by Lemma 3 implies that candidate *i* is an AV outcome. That a Condorcet loser may not be an AV outcome is shown by candidate *d* in Example 1, whereas candidate *a* in Example 2 shows that a Condorcet loser may be an AV outcome. Q.E.D.

Define a *fixed rule* as a voting system in which voters vote for a predetermined number of candidates. "Limited voting" uses a fixed rule; this system is frequently used in multiwinner elections, such as for a city council, in which voters can vote for, and only for, the number of candidates to be elected.

Proposition 3. *There exists no fixed rule that always elects a unique Condorcet winner.*

Proof. Consider the following 5-voter, 4-candidate example:

Example 3

1. 2 voters: *a d b c*
2. 2 voters: *b d a c*
3. 1 voter: *c a b d*

Vote-for-1 elects {*a*, *b*}, vote-for-2 elects *d*, and vote-for-3 elects {*a*, *b*}. Thus, none of the fixed rules elects the unique Condorcet winner, candidate *a*. Q.E.D.

By contrast, several sincere, admissible strategies, including $C_a(\mathbf{P}) = (a, a, bda, bda, ca)$—in which different voter types vote for different numbers of candidates—elect *a*. Clearly, the flexibility of AV may be needed to elect a unique Condorcet winner.

We next turn to scoring rules and analyze the relationship between the winner they select and AV outcomes. The best-known scoring rule is the Borda count (BC): Given that there are *n* candidates, BC awards $n - 1$ points to each voter's first choice, $n - 2$ points to each voter's second choice, …, and 0 points to his or her worst choice.

In Example 1, the BC winner is candidate *b*, who receives from the three types of voters a Borda score of 3(2) + 2(3) + 2(2) = 16 points. In Example 2, the BC winner is also candidate *b*, who receives from the three types of voters a Borda score of 3(1) + 2(2) + 2(1) = 9 points. In these examples, the BC winners coincide with the Condorcet winners, making them AV outcomes (Proposition 2), but this need not be the case, as we will illustrate shortly.

There are other *scoring rules* besides BC, so we begin with a definition. Given m candidates, fix a non-increasing vector $(s_1,..., s_m)$ of real numbers ("scores") such that $s_i \geq s_{i+1}$ for all $i \in \{1,..., m - 1\}$ and $s_1 > s_m$. Each voter's k^{th} best candidate receives score s_k. A candidate's score is the sum of the scores that he or she receives from all voters.

For a preference profile **P**, a scoring rule selects the candidate or candidates that receive the highest score. A scoring rule is said to be *strict* if it is defined by a decreasing vector of scores, $s_i > s_{i+1}$, for all $i \in \{1,..., m - 1\}$.

We next show that all scoring-rule winners, whether they are Condorcet winners or not, are AV outcomes, but candidates that are selected by no scoring rule may also be AV outcomes:

Proposition 4. *At all preference profiles **P**, a candidate chosen by any scoring rule is an AV outcome. There exist preference profiles **P** at which a candidate is not chosen by any scoring rule but is, nevertheless, an AV outcome.*

Proof. We begin by proving the first statement. Take any preference profile **P** and any candidate i chosen by a scoring rule at **P**. Let $(s_1,..., s_m)$ be the scoring-rule vector that results in the election of candidate i at **P**. By a normalization of the scores, we can without loss of generality assume that $s_1 = 1$ and $s_m = 0$.

Note that AV can be seen as a variant of a nonstrict scoring rule, whereby every voter gives a score of 1 to the candidates in his or her strategy set S (approved candidates) and a score of 0 to those not in this set. AV chooses the candidate or candidates with the highest score.[7]

---

[7] Of course, AV is not a scoring rule in the classical sense whereby voters give scores to candidates according to the same predetermined vector. The restrictions on the vector that sincere, admissible strategies impose is that (i) the first component (score of the top candidate) be 1, (ii) the m^{th} component

Let $r_k(x)$ denote the number of voters who consider candidate x to be the k^{th} best candidate at **P**. Because candidate i is picked by the scoring rule (s_1,..., s_m), it must be true that

$$s_1[r_1(i)] + s_2[r_2(i)] + ... + s_m[r_m(i)] \geq s_1[r_1(j)] + s_2[r_2(j)] + ... + s_m[r_m(j)] \qquad (1)$$

for every other candidate j.

To show that the scoring-rule winner, candidate i, is an AV outcome, consider i's critical strategy profile C_i(**P**). There are two cases:

Case (i): Voters rank candidate i last. Under a scoring rule, these voters give a score of 0 to candidate i, a score of 1 to their top choices, and scores between 0 and 1 to the remaining candidates. Under AV, these voters give a score of 0 to candidate i, a score of 1 to their top choices, and scores of 0 to the remaining candidates at C_i(**P**).

Thus, candidate i does the same under the scoring rule as under AV (left side of inequality (1)), whereas all other candidates j do at least as well under the scoring rule as under AV (right side of inequality (1)). This makes the sum on the right side for the scoring rule at least as large as, and generally larger than, the sum of votes under AV, whereas the left side remains the same as under AV. Consequently, if inequality (1) is satisfied under the scoring rule, it is satisfied under AV at C_i(**P**).

Case (ii): Voters do not rank candidate i last. Under a scoring rule, these voters give candidate i a score of s_k if they rank him or her k^{th} best. Under AV, these voters give a score of 1 to candidate i at C_i(**P**). Thus, every s_k on the left side of equation (1) is 1 for

---

(score of the bottom candidate) be 0, (iii) all components representing candidates at or above the lowest candidate a voter approves of are 1, and (iv) all components below the component representing this candidate i under AV, which makes the sum on the left side at least as large as, and generally larger than, the sum under a scoring rule. By comparison, the sum on the right side for all other candidates j under AV is less than or equal to the sum on the left side, with equality if and only if candidate j is preferred to candidate i by all voters. Consequently, if inequality (1) is satisfied under the scoring rule, it is satisfied under AV at $C_i(\mathbf{P})$.

Thus, in both cases (i) and (ii), the satisfaction of inequality (1) under a scoring rule implies its satisfaction under AV at candidate i's critical strategy profile, $C_i(\mathbf{P})$. Hence, a candidate chosen under any scoring rule is also an AV outcome.

To prove the second statement, consider the following 7-voter, 3-candidate example (Fishburn and Brams, 1983, p. 211):

Example 4

1. 3 voters: *a b c*
2. 2 voters: *b c a*
3. 1 voter: *b a c*
4. 1 voter: *c a b*

Because candidate b receives at least as many first choices as a and c, and more first and second choices than either, every scoring rule will select b as the winner. But a is the Condorcet winner and, hence, an AV outcome by Proposition 2.[8] Q.E.D.

---

candidate are 0.

[8] Example 4 provides an illustration in which BC, in particular, fails to elect the Condorcet winner.

Note that candidate *b* in Example 4 is not AV-dominant: The number of voters who consider *b* as their best choice and *a* as their worst choice (2 voters), or *b* as their best choice and *c* as their worst choice (1 voter), does *not* exceed the number of voters who prefer *a* to *b* (4 voters) or *c* to *b* (1 voter) (see Lemma 4). Put another way, the critical strategy profile of candidate *a*, $C_a(\mathbf{P}) = (a, a, a, b, b, ba, ca)$, renders *a* the unique AV winner (5 votes), foreclosing the AV-dominance of candidate *b* (3 votes). Likewise, Condorcet winner *a* is also not AV-dominant, because the critical strategy profile of candidate *b*, $C_b(\mathbf{P}) = (ab, ab, ab, b, b, b, c)$, results in *b*'s election (6 votes), foreclosing the AV-dominance of candidate *a* (3 votes).

Call a candidate *S-dominant* if he or she is the unique winner under every scoring rule, as candidate *b* is in Example 4.

Proposition 5. *At all preference profiles **P**, an AV-dominant candidate is S-dominant, but not every S-dominant candidate is AV-dominant.*

Proof. We have just shown that not every S-dominant candidate is AV-dominant in Example 4. To prove the first part of the proposition, note that a necessary and sufficient condition for candidate *i* to be S-dominant is that he or she receives at least as many first choices as every other candidate *j*, at least as many first and second choices as every other candidate *j*, …, and *more* first, second, …, and next-to-last choices as any other candidate *j*; otherwise, candidate *i* would not be assured of receiving more points than candidate *j*. But this condition, while necessary, is not sufficient for a candidate to be AV-dominant (Lemma 4 gives a necessary and sufficient condition). Q.E.D.

In effect, Proposition 5 says that being AV-dominant is more demanding than being S-dominant. Whereas S-dominance counts choices at each distinct level (first, second,…, next-to-last) and requires that an S-dominant candidate never be behind at any level, and ahead at the next-to-last level, AV-dominance counts approval at different levels simultaneously (e.g., in the case of $C_a(\mathbf{P})$ in Example 4, the first level for some voters and the second level for other voters).

We next show the outcomes of two social choice rules that are not scoring rules, the Hare system of single transferable vote (STV) and the majoritarian compromise (MC), are always AV outcomes (at their critical strategy profiles), whereas the converse is not true—AV outcomes need not be STV or MC outcomes.[9] Before proving this result, we illustrate STV and MC with a 9-voter, 3-candidate example:[10]

Example 5

1. 4 voters: *a c b*
2. 2 voters: *b c a*
3. 3 voters: *c b a*

Under STV, candidates with the fewest first-choice—and successively lower-choice—votes are eliminated; their votes are transferred to second-choice and lower-choice candidates in their preference rankings until one candidate receives a majority of votes. To illustrate in Example 5, because candidate *b* receives the fewest first-choice votes (2)—compared with 3 first-choice votes for candidate *c* and 4 first-choice votes for

---

[9] Ideally, of course, it would be desirable to prove this result for all voting systems, but we know of no general definition of a voting system that encompasses all those that have been used or proposed, in contrast to scoring systems and, as we will show later, Condorcet systems (Brams and Fishburn, 2002).

[10] These two voting systems, among others, are discussed in Brams and Fishburn (2002). MC, which is less well known than STV, was proposed independently as a voting procedure (Hurwicz and Sertel, 1997; Sertel and Yilmaz, 1999; Sertel and Sanver, 1999; Slinko, 2002) and as a bargaining procedure under the rubric of "fallback bargaining" (Brams and Kilgour, 2001). As a voting procedure, the threshold for winning is assumed to be simple majority, whereas as a bargaining procedure the threshold is assumed to be unanimity, but qualified majorities are also possible under either interpretation.

candidate *a*—*b* is eliminated and his or her 2 votes go to the second choice of the 2 type (2) voters, candidate *c*. In the runoff between candidates *a* and *c*, candidate *c*, now with votes from the type (2) voters, defeats candidate *a* by 5 votes to 4, so *c* is the STV winner.

Under MC, first-choice, then second-choice, and then lower-choice votes are counted until at least one candidate receives a majority of votes; if more than one candidate receives a majority, the candidate with the most votes is elected. Because no candidate in Example 5 receives a majority of votes when only first choices are counted, second choices are next counted and added to the first choices. Candidate *c* now receives the support of all 9 voters, whereas *a* and *b* receive 4 and 5 votes, respectively, so *c* is the MC winner.

Proposition 6. *At all preference profiles* ***P****, a candidate chosen by STV or MC is an AV outcome. There exist preference profiles* ***P*** *at which a candidate chosen by AV is neither an STV nor an MC outcome.*

Proof. We start by showing that every STV outcome is an AV outcome. Suppose candidate *i* is not an AV outcome at preference profile **P**. By Lemma 3, there exists a candidate *j* such that the number of voters who rank *j* as their best candidate and *i* as their worst candidate exceeds the number of voters who prefer *i* to *j*. A fortiori, the number of voters who consider *j* as their best candidate exceeds those who consider *i* as their best candidate.

This result holds for any subset of candidates that includes both *i* and *j*. Hence, STV will never eliminate *j* in the presence of *i*, showing that *i* cannot be an STV winner.

Neither can i be an MC winner, because j will receive more first-place votes than i. If this number is not a majority, the descent to second and still lower choices continues until at least one candidate receives a majority. Between i and j, the first candidate to receive a majority will be j, because j receives more votes from voters who rank him or her first than there are voters who prefer i to j. Thus, j will always stay ahead of i as the descent to lower and lower choices continues until j receives a majority.

To show that AV outcomes need not be STV or MC outcomes, consider Example 4, in which the Condorcet winner, candidate c, is chosen under both STV and MC. Besides c, AV may also choose candidate a or candidate b: a is an AV outcome at critical strategy profile $C_a(\mathbf{P}) = (a, a, a, a, b, b, c, c, c)$; and b is an AV outcome at critical strategy profile $C_b(\mathbf{P}) = (a, a, a, a, b, b, cb, cb, cb)$. Q.E.D.

So far we have shown that AV yields at least as many, and generally more, (Pareto) outcomes than any scoring rule and two nonscoring voting systems. To be sure, one might question whether the three possible AV outcomes in Example 4 have an equal claim to being *the* social choice. Isn't candidate c, the Condorcet winner, BC winner, STV winner, and MC winner—and ranked last by no voters—the best overall choice? By comparison, candidate b is only a middling choice; and candidate a, who is the plurality-vote (PV) winner, is the Condorcet loser.[11]

Just as AV allows for a multiplicity of outcomes, it also enables voters to prevent them.

[11] Note that PV is a degenerate scoring rule, under which a voter's top candidate receives 1 point and all other candidates receive 0 points. By Proposition 4, sincere outcomes under PV are always AV outcomes but not vice-versa. As a case in point, candidate a is the sincere PV outcome in Example 5, whereas candidates b and c are also sincere AV outcomes.

Proposition 7. *At every preference profile* ***P*** *at which there is not an AV-dominant candidate, AV can prevent the election of every candidate*, *whereas scoring rules, STV, and MC cannot prevent the election of all of them.*

Proof. In the absence of an AV-dominant candidate, there is no candidate that can be assured of winning, which implies that every candidate can be prevented from winning. To show that scoring rules, STV, and MC cannot prevent the election of all candidates when AV can, consider the following 3-voter, 3-candidate example:

Example 6

1. 1 voter: *a b c*
2. 1 voter: *b a c*
3. 1 voter: *c b a*

It is easy to see that there is no candidate that is AV-dominant in Example 6, based on Lemma 4. But to make perspicuous how AV can prevent the election of every candidate in Example 4—and why the other systems cannot—let "|" indicate each voter's dividing line between the candidate(s) he or she considers acceptable and those he or she considers unacceptable. If the three voters draw their lines as follows,

a | *b c*   *b a* | *c*   *c* | *b a*,

b and *c* will *not* be chosen (*a* will be). If the voters draw their lines as follows,

a | *b c*   *b* | *a c*   *c b* | *a*,

a and *c* will *not* be chosen (*b* will be). Thus the voters can prevent the election of every one of the three candidates under AV because none is AV-dominant.

By contrast, the Condorcet winner, *b*, wins under every scoring system, including BC, and also under MC. Under STV, either *a* or *b* may win, depending on which of the three candidates is eliminated first. Thus, only *c* is prevented from winning under these other systems, showing that AV is unique in being able to prevent the election of each of the three candidates. Q.E.D.

We have seen that AV allows for outcomes that BC, MC, and STV do not (e.g., *c* in Example 6 when there is a three-way tie). At the same time, it may preclude outcomes (e.g., *b* in Example 6) that other systems cannot prohibit. In effect, voters can fine-tune their preferences under AV, making outcomes responsive to information that transcends these preferences.

We next consider not only what outcomes can and cannot occur under AV but also what outcomes are likely to persist because of their stability. While we know that non-Pareto candidates cannot win a clear-cut victory under AV (Proposition 1), might it be possible for Condorcet losers to be AV outcomes *and* stable? To answer this question, we will distinguish two types of stability.

4. Stability of Election Outcomes

As earlier, we assume that voters choose sincere, admissible strategies under AV. Now, however, we suppose that they may not draw the line between acceptable and unacceptable candidates as they would if they were truthful. Instead, they may vote strategically in order to try to obtain a preferred outcome.

To determine what is "preferred," we extend preference to sets. If a voter's preference is *a b*, he or she will prefer *a* to $\{a, b\}$, and $\{a, b\}$ to *b*. If a voter's preference is *a b c*, he or she may prefer any of outcomes *b*, $\{a, c\}$, or $\{a, b, c\}$ to any other. In assessing the stability of outcomes later, we will assume that these are all admissible preferences over sets.

We define two kinds of stability, the first of which is the following: Given a preference profile **P**, an AV outcome is *stable* if there exists a strategy profile **S** such that no voters of a single type have an incentive to switch their strategy to another sincere, admissible strategy in order to induce a preferred outcome.[12] In analyzing the stability of AV outcomes, at least those that do not involve ties,[13] we need confine our attention only to those outcomes stable at $C_i(\mathbf{P})$ because of the following proposition:

Proposition 8. *A nontied AV outcome i is stable if and only if it is stable at its critical strategy profile,* $C_i(\mathbf{P})$.

Proof. The "if" part follows from the existence of a strategy profile, $C_i(\mathbf{P})$, at which outcome *i* is stable. To show the "only if" part, assume candidate *i* is unstable at $C_i(\mathbf{P})$. At any other strategy profile **S'**, candidate *i* receives no more approval votes and generally fewer than at $C_i(\mathbf{P})$ by Lemma 1. Hence, those voters who switch to different sincere, admissible strategies to induce the election of a preferred candidate at **S** can also do so at **S'**. Q.E.D.

---

[12] Treating voters of one type, all of whose members have the same preference, as single (weighted) voters provides the most stringent test of stability. This is because any outcome that can be destabilized by the switch of individual voters (of one type) can be destabilized by the switch of all voters of that type, but the converse is not true.

[13] To illustrate how ties may complicate matters, assume three voters have preferences *a b c, b c a*, and *c a b* and vote only for their first choices, which is not a critical strategy for any of them. Then the resulting tied outcome, {*a*, *b*, *c*}, will be stable if no voter prefers just its second choice to the tie. As this example illustrates, the stability of tied outcomes may depend on comparisons between singleton and nonsingleton subsets; to avoid this comparison, we will assume nontied AV outcomes in several of the subsequent propositions.

The strategies of voters associated with a stable AV outcome at $C_i(\mathbf{P})$ define a Nash equilibrium of a voting game in which the voters have complete information about each others' preferences and make simultaneous choices.[14]

Neither candidate *a* nor candidate *b* is a stable AV outcome in Example 5. At critical strategy profile $C_a(\mathbf{P}) = (a, a, a, a, b, b, c, c, c)$ that renders candidate *a* an AV outcome, if the 2 type (2) voters switch to strategy *bc*, candidate *c*, whom the type (2) voters prefer to candidate *a*, wins. At critical strategy profile $C_b(\mathbf{P}) = (a, a, a, a, b, b, cb, cb, cb)$ that renders candidate *b* an AV outcome, the 4 type (1) voters have an incentive to switch to strategy *ac* to induce the selection of candidate *c*, whom they prefer to candidate *b*.

Although AV outcomes *a* and *b* in Example 5 are not stable at their critical strategy profiles, AV outcome *c* most definitely is stable at its critical strategy profile, $C_c(\mathbf{P}) = (ac, ac, ac, ac, bc, bc, c, c, c)$: No switch on the part of the 4 type (1) voters to *a*, of the 2 type (2) voters to *b*, or of the 3 type (3) voters to *cb* can lead to a preferred outcome for any of these types—or, indeed, change the outcome at all (because candidate *c* is the unanimous choice of all voters at *c*'s critical strategy profile).

Not only can no single switch by any of the three types induce a preferred outcome for the switchers at $C_c(\mathbf{P})$, but no *coordinated* switches by two or more types can induce a preferred outcome. Thus, for example, if the *ac*-voters switched from *ac* to *a*, and the *bc* voters switched from *bc* to *b*, they together could induce AV outcome *a*, which the 4 type (1) voters would clearly prefer to outcome *c*. But *a* is the worst choice of the 2 type (2)

[14] For an analysis of Nash equilibria in voting games under different rules and information conditions from those given here, see Myerson (2002) and references cited therein.

voters, so they would have no incentive to coordinate with the type (1) voters to induce this outcome.

That AV outcome *c* is, at the critical strategy profile of candidate *c*, invulnerable to coordinated switches leads to our second type of stability: Given a preference profile **P**, an outcome is *strongly stable* if there exists a strategy profile **S** such that no types of voters, coordinating their actions, can form a coalition *K*, all of whose members would have an incentive to switch their AV strategies to other sincere, admissible strategies in order to induce a preferred outcome.

We assume that the coordinating players in *K* are allowed to communicate to try to find a set of strategies to induce a preferred outcome *for all of them*. These strategies define a *strong* Nash equilibrium of a voting game in which voters have complete information about each others' preferences and make simultaneous choices.

Proposition 9. *A nontied AV outcome i is strongly stable if and only if it is strongly stable at its critical strategy profile,* $C_i(\mathbf{P})$.

Proof. Analogous to that of Proposition 8. Q.E.D.

What we have yet to show is that an AV stable outcome need not be strongly stable. To illustrate this weaker form of stability, consider AV outcome *a* in Example 1 and its critical strategy profile, $C_a(\mathbf{P}) = (a, a, a, bca, bca, d, d)$. The 2 type (2) voters cannot upset this outcome by switching from *bca* to *bc* or *b*, nor can the 2 type (3) voters upset it by switching from *d* to *db* or *dbc*. However, if these two types of voters cooperate and form a coalition *K*, with the 2 type (2) voters choosing strategy *b* and the 2 type (3) voters choosing strategy *db*, they can induce the selection of Condorcet winner *b*, whom both types prefer to candidate *a*. At critical strategy profile $C_a(\mathbf{P})$, therefore, AV outcome *a* is stable but not strongly stable, whereas AV outcome *b* is strongly stable at its critical strategy profile, $C_b(\mathbf{P}) = (ab, ab, ab, b, b, db, db)$.

If an AV outcome is neither strongly stable nor stable, it is *unstable*. Clearly, strongly stable outcomes are always stable, but not vice versa.

Proposition 10. *AV outcomes are strongly stable, stable, or unstable. All three kinds of AV outcomes may coexist.*

Proof. We have just shown that AV outcome *b* is strongly stable, and AV outcome *a* is stable, in Example 1. We now show that candidate *c* in this example is an unstable AV outcome. At the critical strategy profile of candidate *c*, $C_c(\mathbf{P}) = (abc, abc, abc, bc, bc, dbc, dbc)$, candidates *b* and *c* tie with 7 votes each. Candidate *c* might therefore be selected under some tie-breaking rule. But $\{b, c\}$ is not a stable AV outcome: If either the 3 *a*-voters switch to *ab*, the 2 *bc*-voters switch to *b*, or the 2 *dbc*-voters switch to *db*, candidate *b* will be selected, whom all three types of voters prefer to $\{b, c\}$.[15] Q.E.D.

While Proposition 10 shows that strongly stable, stable, and unstable AV outcomes may coexist, it is important to know the conditions under which each kind of outcome can occur.

Proposition 11. *A nontied AV outcome is strongly stable if and only if it is a nontied Condorcet winner.*

Proof. To prove the "if" part, suppose candidate *i* is a unique Condorcet winner at **P**. We will show that *i* is a nontied AV outcome that is strongly stable at its critical strategy profile, $C_i(\mathbf{P})$. Clearly, i is a nontied AV outcome at $C_i(\mathbf{P})$ by Proposition 2. To show its strong stability, suppose there exists a coalition of voters K, comprising one or more types, that prefers some other candidate j to candidate i and coordinates to induce the selection of j. Because candidate i is a unique Condorcet winner, however, the cardinality of K is strictly less than the cardinality of coalition L, whose members prefer i to j. The members of L vote for i but not for j at $C_i(\mathbf{P})$. Hence, whatever sincere, admissible strategy switch the members of K consider at candidate i's critical strategy profile to induce the election of candidate j, j will receive fewer votes than i, proving that i is a strongly stable AV outcome.

To prove the "only if" part, suppose that candidate i is not a Condorcet winner. Consequently, there exists a coalition of voters K, comprising one or more types, that prefers some other candidate j to i and coordinates to induce the election of j. Because i is not a Condorcet winner, the cardinality of K is a strict majority. We will now show that i is not a strongly stable AV outcome at its critical strategy profile, $C_i(\mathbf{P})$, which by Proposition 9 shows that i is not a strongly stable AV outcome. Suppose AV does not elect i at $C_i(\mathbf{P})$. Then i is not an AV outcome and hence not a strongly stable one. Now suppose that AV elects i at $C_i(\mathbf{P})$. Because the members of K can change their strategies to elect j, whom they prefer to i, i is not a strongly stable AV outcome. Q.E.D.

Call a Condorcet winner *tied* if it comprises two or more candidates that tie against each other but defeat all other candidates in pairwise contests. As a simple example, assume that there are two voters, 1 and 2, and two candidates, i and j; voter 1 ranks the candidates $i\ j$, and voter 2 ranks them $j\ i$. Then (i, j) is the critical strategy profile for both

---

[15] As we will show in Proposition 12, unstable AV outcomes do not necessarily include, as here, non-

candidates. Obviously, neither one nor both voters can induce an outcome each prefers to the tied outcome $\{i, j\}$, so this outcome is strongly stable without there being a unique Condorcet winner.[16]

Our next proposition establishes that every AV outcome may be unstable.

Proposition 12. *There may be no strongly stable, or even stable, nontied AV outcomes—that is, every nontied AV outcome may be unstable.*

Proof. That there may be no stable AV outcomes is shown by the following 3-voter, 3-candidate example:[17]

Example 7

1. 1 voter: *a b c*
2. 1 voter: *b c a*
3. 1 voter: *c a b*

Consider the critical strategy profile that selects candidate a, $C_a(\mathbf{P}) = (a, b, ca)$. If voter (2) switches to *bc*, he or she can induce preferred outcome $\{a, c\}$. In a similar manner, it is possible to show that neither candidate *b* nor candidate *c* is a stable AV outcome. Q.E.D.

We next show that Condorcet losers as well as winners may be stable AV outcomes.

Pareto candidates as components. Unique Pareto candidates can also be unstable AV outcomes.

[16] Under a somewhat weaker definition of a strong Nash equilibrium, the equivalence of strong Nash-equilibrium outcomes and Condorcet winners is shown for a large class of voting rules in Sertel and Sanver (forthcoming).

[17] Example 7 is the standard example of the Condorcet paradox, or cyclical majorities, in which there is no Condorcet winner. It is the same example used in ftn. 13 to show that a tied outcome might, under certain circumstances, be stable.

Proposition 13. *A unique Condorcet loser may be a stable AV outcome, even when there is a different outcome that is a unique Condorcet winner (and therefore strongly stable).*

Proof. Consider the following 7-voter, 5-candidate example:

Example 8

1. 3 voters: *a b c d e*
2. 1 voter: *b c d e a*
3. 1 voter: *c d e b a*
4. 1 voter: *d e b c a*
5. 1 voter: *e b c d a*

Candidate *a* is the Condorcet loser, ranked last by 4 of the 7 voters. But at its critical strategy profile, $C_a(\mathbf{P}) = (a, a, a, b, c, d, e)$, candidate *a* is a stable AV outcome, because none of the four individual voters, by changing his or her strategy, can upset *a*, who will continue to receive 3 votes.

Consider the critical strategy profile of candidate *b*, $C_b(\mathbf{P}) = (ab, ab, ab, b, cdeb, deb, eb)$, who receives 7 votes, compared with 3 votes each for *a* and *e*, 2 votes for *d*, and 1 vote for *c*. Again, no single type of voter can upset this outcome, nor can any coalition, because candidate *b* is the unique Condorcet winner, making him or her strongly stable. Q.E.D.

Whether a Condorcet loser, like candidate *a* in Example 8, "deserves" to be an AV winner—and a stable one at that—depends on whether voters have sufficient incentive to unite in support of a candidate like Condorcet winner *b*, who is the first choice of only one voter. If they do not rally around *b*, and the type (1) voters vote only for *a*, then *a* is arguably the more acceptable choice.

To compare the stability of outcomes under AV and other voting systems, we need stability definitions for these other systems. A *stable outcome* under each of these systems is one in which no single type of voter has an incentive to switch its ranking to another ranking in order to induce a preferred outcome in a voting game in which voters have complete information about each others' preferences and make simultaneous choices. A *strongly stable outcome* is one in which no types of voters, coordinating their actions, can form a coalition *K*, all of whose members have an incentive to switch their rankings in order to induce a preferred outcome.

Proposition 14. *Stable or strongly stable AV outcomes need not be stable or strongly stable scoring-system, STV, or MC outcomes. Conversely, stable or strongly stable scoring-system, MC, or STV outcomes—while always AV outcomes—need not be stable or strongly stable AV outcomes.*

Proof. In the proof of Proposition 4, we indicated that there is no scoring system that selects candidate *a* in Example 4, so obviously *a* cannot be a stable or strongly stable outcome under a scoring system. Because *a* is the unique Condorcet winner in this example, however, it is a strongly stable, and therefore a stable, AV outcome. Likewise, candidate *a* is a stable AV outcome in Example 1 (see proof of Proposition 9), but it is not the STV or MC outcome (candidate *b* is) so cannot be a stable outcome under these systems.

Next we show that both a scoring system and MC may select a stable outcome that is not a stable AV outcome. Consider the following 6-voter, 3-candidate example:

Example 9

1. 2 voters: *a c b*
2. 3 voters: *b c a*
3. 1 voter: *c a b*

Under BC, *c* gets 7 points, *b* gets 6 points, and *a* gets 5 points. Moreover, *c* is stable (and strongly stable): If either the 2 type (1) voters or the 3 type (2) voters interchange *c* and their last choice, the last choice wins, which is worse for them. Likewise under MC, *c* wins by getting 6 votes to 3 votes for *b* and 2 votes for *a* at level 2. Moreover, neither the type (1) nor the type (2) voters can obtain a preferred outcome by giving a different preference ranking.

By contrast, under AV, *c* gets 6 votes at its critical strategy profile, $C_c(\mathbf{P}) = (ac, ac, bc, bc, bc, c)$, whereas *b* gets 3 votes and *a* gets 2 votes. But by voting only for *b*, the 3 type (2) voters can induce $\{b, c\}$, which they prefer to *c* alone, so *c* is not a stable (or strongly stable) AV outcome.

Finally, we show that STV may elect a stable outcome that is not a stable AV outcome. Consider the following 6-voter, 4-candidate example:

Example 10

1. 3 voters: *a b d c*
2. 2 voters: *c b d a*
3. 1 voter: *d c b a*

Under STV, first *d* is eliminated, after which the type (3) voter's vote is transferred to *c*, which ties with *a* (3 votes each) to give $\{a, c\}$ as the outcome. It is easy to see that neither the type (3) voter—whose second or fourth choice will be elected—nor the type (1) or type (2) voters can give (false) preference rankings that lead to a preferred outcome.

Under AV, there are two strategy profiles that elect $\{a, c\}$: (i) (a, a, a, c, c, dc) and (ii) (a, a, a, c, c, dcb). If the 3 type (1) voters prefer $\{a, b, c\}$ to $\{a, c\}$ in the case of (i), or b to $\{a, c\}$ in the case of (ii), they will have an incentive to switch from strategy a to strategy ab. Hence, $\{a, c\}$ is not stable under all admissible preferences over such sets, proving that a stable STV outcome may not be a stable AV outcome. Q.E.D.

Thus, no system, including AV, dominates all others with respect to producing stable outcomes: All the systems we have analyzed so far may produce stable or strongly stable outcomes, but this stability is not necessarily duplicated under other systems. We next ask whether voting systems that guarantee the election of Condorcet winners when voters are sincere do any better.

A *Condorcet voting system* is one that always elects a Condorcet winner, if one exists, when voters are sincere. This candidate, however, may not be a Nash-equilibrium outcome, much less a strong one (as under AV).

Proposition 15. *No Condorcet voting system ensures the election of a unique Condorcet winner as a Nash-equilibrium outcome.*

Proof. Consider the following example, in which there is no Condorcet winner:

Example 11

1. 2 voters: *a d b c*
2. 2 voters: *b d c a*

3. 1 voter: *c a b d*

In the absence of a Condorcet winner, we assume that different candidates may be chosen by a Condorcet voting system (about which we will be more specific shortly).

We first show that by changing the preference ranking of each of the three voter types, one at a time, in Example 11, we can render different candidates Condorcet winners. However, if a Condorcet voting system chooses a candidate or candidates preferred by this type in Example 11, then the Condorcet winner is not a Nash-equilibrium outcome. To prevent this from happening, we must preclude the possibility of choosing the preferred candidate(s). It turns out that the Condorcet winners we show can occur, and whose choice as an equilibrium outcome can be upset by some voter type changing its ranking to that in Example 11, preclude the possibility of a Condorcet voting system's choosing *a*, *b*, *c*, or *d* in this example.[18]

To begin, assume the preference ranking of the 2 type (2) voters in Example 11 is *b d a c*, but the other two types have the same preferences as shown in Example 11. Then candidate *a* is the Condorcet winner. If a Condorcet voting system would choose either candidate *b* or *d* in Example 11, then it would be in the interest of the type (2) voters to switch to *b d c a*, as given in Example 11, to obtain a preferred outcome.

Assume the preference ranking of the type (3) voter is *c d a b*. Then candidate *d* is the Condorcet winner. If a Condorcet voting system would choose candidate *c* in

[18] These choices do not preclude the possibility of the Condorcet voting system's choosing most subsets of candidates in Example 11, such as $\{b, c\}$, which may or may not be preferred to a Condorcet winner. Thus, Proposition 15 is applicable to social choice functions, which are "resolute" social choice rules that choose single candidates, and not to social choice correspondences, which may choose nonsingleton subsets of candidates.

Example 11, then it would be in the interest of the type (3) voter to switch to *c a b d*, as given in Example 11, to obtain a preferred outcome.

Finally, assume the preference ranking of the 2 type (1) voters is *a c d b*. Then candidate *c* is the Condorcet winner. If a Condorcet voting system would choose candidate *c* in Example 11, then it would be in the interest of the type (1) voters to switch to *a d b c*, as given in Example 11, to obtain a preferred outcome.

In summary, we have shown that three of the four candidates in Example 11 can be rendered a Condorcet winner by changing the preference ranking of one voter type. If this is the true ranking of these voter(s), it is always in their interest to misrepresent their preferences to those shown in Example 11, given that a Condorcet voting system chooses the candidates we postulated in each of the above cases. But to prevent this from happening in all the cases, we must preclude the possibility of the Condorcet voting system's choosing all four candidates in Example 11—and thereby undermining the Nash-equilibrium status of the unique Condorcet winners in the modifications of Example 11. Q.E.D.

Proposition 15 casts doubt on the efficacy of Condorcet voting systems, such as those of Black or Copeland (Brams and Fishburn, 2002), to do what they purport to do *in equilibrium*. By contrast, AV always renders Condorcet winners strong Nash-equilibrium outcomes.

The (nonstrong) stability of outcomes under the different systems that we have analyzed so far is very much a local property: It depends on the *inability* of voters of one type, by misrepresenting their preferences, to change the sincere outcome to their advantage.

AV probably allows for more of this kind of stability than any other system—in addition to rendering Condorcet winners strongly stable. But we have seen that there are preference profiles in which other systems give (strong) stability when AV does not (Proposition 14), so the stability picture is somewhat mixed.

5. Summary and Conclusions

We began by suggesting that more than one outcome of an election may be acceptable, which we illustrated with the 2000 US presidential election. But our focus was on elections with three or more candidates, wherein acceptability is determined by where voters draw the line between acceptable and unacceptable candidates.

This criterion of acceptability underlies the choices that voters make under AV. We summarize both our positive findings and our normative conclusions about how voter sovereignty, expressed through AV, affects election outcomes:

1. AV enables voters to indicate those points in their preference rankings above which candidates are acceptable, which we believe is information that *should* determine their social choices. This information may either rule in or rule out candidates that would be chosen under other systems.

2. AV may select a multitude of candidates—even all candidates in a race—at their critical strategy profiles. This may include even Condorcet losers, whom most social choice theorists would condemn as egregious choices, especially if there is also a Condorcet winner in the race. Our view is different: The choice of even a Condorcet loser may, on occasion, be justified—and more so if voters are divided over the choice of another candidate.

3. We do not disparage Condorcet winners, which no fixed rule may elect. Furthermore, there may be a candidate or candidates who stand in between a Condorcet winner and a Condorcet loser that are also AV outcomes, rendering several candidates in a race viable.

4. Grounding social choices on the notion of acceptability rather than on traditional social-choice criteria is a radical departure from the research program initiated by Borda and Condorcet in late 18th-century France (McLean and Urken, 1995). While we do not eschew these criteria, they should not be the be-all and end-all for judging whether outcomes are acceptable or not. Rather, the pragmatic judgments of sovereign voters about who is acceptable and who is not should be decisive.

5. Voters' judgments will be affected by the stability of different AV outcomes. Thus, a Condorcet loser may not be a stable AV outcome—much less a strongly stable one—so this candidate's viability is less than were he or she a stable outcome.

6. When there is a Condorcet winner, this candidate is always a strongly stable AV outcome. Thereby AV preserves the majority will, at least if there is some kind of strategic coordination among voters, through sincere voting.

7. In large-scale elections, this coordination is possible to a limited extent from information provided by polls (Brams and Fishburn, 1983, ch. 7). But strategizing by voters may not be perfect, allowing Condorcet winners, even under AV, to be defeated on occasion.

8. This failure may sometimes be salutary, especially when a BC winner differs from a Condorcet winner and "majority tyranny" is a concern (Baharad and Nitzan,

2002). In such a situation, the BC winner may be a more acceptable candidate, even if he or she is not, like a Condorcet winner, a strongly stable AV outcome.

9. While the stability and strong stability of outcomes facilitates their selection, even unstable AV outcomes should be considered acceptable, especially if there are no stable outcomes because of a Condorcet paradox (Miller, 1983).

10. Speaking normatively, AV provides a better way of finding consensus choices than do other voting systems because of the information that it both suppresses (preference rankings) and expands upon (who is acceptable and who is not in the rankings):

• It is simpler to use than the alternative voting systems, except possibly PV, because it does not require that the voters rank candidates (often an arduous task if there are more than about five candidates).

• It may actually make choices easier than PV for a voter who (i) is relatively indifferent among more two or more candidates or (ii) favors a candidate that is not competitive (e.g., Ralph Nader in 2000) and, hence, may also want to vote for a more viable second choice (e.g., Al Gore).

11. While different outcomes may be strongly stable, stable, or unstable under different systems, AV probably endows outcomes with more stability, on average, than do its competitors. Condorcet systems, in particular, do not always elect Condorcet winners in equilibrium, rendering them vulnerable when voters are strategic.

12. The local stability of Nash equilibria and the global stability of strong Nash equilibria indicate that acceptability may be stabilized at different levels. No level is sacrosanct. Thus, we see no reason to insist that the strong stability of a Condorcet

winner under AV should supersede the local stability of a different Borda winner, or even the possible instability of a Condorcet loser.

13. Ultimately, acceptability depends on the judgments of voters. AV provides a compelling means for them to exercise their sovereignty, both for and against candidates.

References

Arrow, Kenneth J. (1963). *Social Choice and Individual Values*, 2nd ed. New Haven, CT: Yale University Press.

Baharad, Eyal, and Shmuel Nitzan (2002). "Ameliorating Majority Decisiveness through Expression of Preference Intensity." *American Political Science Review* 96, no. 4 (December): 745-754.

Brams, Steven J., and Peter C. Fishburn (1978). "Approval Voting." *American Political Science Review* 72, no. 3 (September): 831-847.

Brams, Steven J., and Peter C. Fishburn (1983). *Approval Voting*. Boston: Birkhäuser.

Brams, Steven J., and Peter C. Fishburn (forthcoming). "Going from Theory to Practice: The Mixed Success of Approval Voting." *Social Choice and Welfare*.

Brams, Steven J., and Peter C. Fishburn (2002). "Voting Procedures." In Kenneth J. Arrow, Amartya K. Sen, and Kotaro Suzumura (eds.), *Handbook of Social Choice and Welfare*, vol. 1. Amsterdam: Elsevier Science, pp. 175-236.

Brams, Steven J., Peter C. Fishburn, and Samuel Merrill III (1988a). "The Responsiveness of Approval Voting: Comments on Saari and Van Newenhizen." *Public Choice* 59 (November): 121-131.

Brams, Steven J., Peter C. Fishburn, and Samuel Merrill (1988b). "Rejoinder to Saari and Van Newenhizen." *Public Choice* 59 (September): 149.

Brams, Steven J., and D. Marc Kilgour (2001). "Fallback Bargaining." *Group Decision and Negotiation* 10, no. 4 (July): 287-316.

Fishburn, Peter C., and Steven J. Brams (1983). "Paradoxes of Preferential Voting." *Mathematics Magazine* 56, no. 4 (September): 207-214.

Hurwicz, Leonid, and Murat R. Sertel (1999). "Designing Mechanisms, in Particular for Electoral Systems: The Majoritarian Compromise." In Murat R. Sertel (ed.), *Econmic Design and Behaviour*. London: Macmillan.

McLean, Ian, and Arnold B. Urken (eds.) (1995). *Classics of Social Choice*. Ann Arbor, MI: University of Michigan Press.

Miller, Nicholas R. (1983). "Pluralism and Social Choice." *American Political Science Review* 77, no. 3 (September): 734-747.

Myerson, Roger B. (2002). "Comparison of Scoring Rules in Poisson Voting Games." *Journal of Economic Theory* 103, no. 1 (March): 219-251.

Nurmi, Hannu (1999). *Voting Procedures and How to Deal with Them*. Berlin: Springer-Verlag.

Nurmi, Hannu (2002). *Voting Procedures under Uncertainty*. Berlin: Springer-Verlag.

Saari, Donald G. (1994). *Geometry of Voting.* New York: Springer-Verlag.

Saari, Donald G. (2001). *Chaotic Elections! A Mathematician Looks at Voting.* Providence, RI: American Mathematical Society.

Saari, Donald G., and Jill Van Newenhizen (1988a). "The Problem of Indeterminacy in Approval, Multiple, and Truncated Voting Systems." *Public Choice* 59 (November): 101-120.

Saari, Donald G., and Jill Van Newenhizen (1988b). "Is Approval Voting an 'Unmitigated Evil': A Response to Brams, Fishburn, and Merrill." *Public Choice* 59 (November): 133-147.

Sertel, Murat R., and M. Remzi Sanver (1999). "Designing Public Choice Mechanisms," in Imed Limam (ed.), *Institutional Reform and Development in the MENA*

Region. Cairo, Egypt: Arab Planning Institute, pp, 129-148.

Sertel, Murat R., and M. Remzi Sanver (forthcoming). “Strong Equilibrium Outcomes of Voting Games are the Generalized Condorcet Winners. *Social Choice and Welfare*.

Sertel, Murat R., and Bilge Yilmaz (1999). “The Majoritarian Compromise Is Majoritarian-Optimal and Subgame-Perfect Implementable.” *Social Choice and Welfare* 16, no. 4 (August): 615-627.

Slinko, Arkadii (2002). “The Majoritarian Compromise in Large Societies.” *Review of Economic Design* 7, no. 3 (November): 341-347.