MONOTONICITY FAILURE UNDER STV

AND RELATED VOTING SYSTEMS

Nicholas R. Miller
Department of Political Science
University of Maryland Baltimore County
Baltimore, Maryland 21250
nmiller@umbc.edu

Preliminary and incomplete draft prepared for presentation at the
2002 Annual Meeting of the Public Choice Society,
San Diego, March 22-24, 2002

MONOTONICITY FAILURE UNDER STV AND RELATED VOTING SYSTEMS

It is now generally known that the single transferable vote (STV) and related voting systems that entail runoffs of one sort or other are subject to the paradoxical feature that enhanced support for a candidate can cost that candidate electoral victory (or, conversely, that reduced support can give the candidate electoral victory) — that is, that they are subject to so-called "monotonicity failure." However, it remains unclear how extensive this problem is.

Is monotonicity failure merely a logical possibility that is highly unlikely to occur in practice, or is it a problem that may occur with considerable frequency? Is the possibility of monotonicity failure indicated by manifest characteristics of election results, or is it a problem that remains largely hidden from view? In so far as there is a real problem, is it more substantial and/or more evident under STV or under the other related voting systems that are (or might be) used in single-winner elections. In particular, what is the extent of the monotonicity problem under so-called Instant Runoff Voting [IRV] that now has a number of enthusiastic advocates in the United States?

This paper attempts to address questions along these lines by means of a (more or less) complete analysis of single-winner elections with three candidates, in which case all voting systems in this family are logically equivalent and, at the same time, are also distinct from simple plurality voting (SPV), commonly called First Past The Post (FPTP) in the U.K.

1. Political and Theoretical Context

For well over a hundred years, the Electoral Reform Society of the United Kingdom has enthusiastically — indeed, passionately — advocated the Single Transferable Vote (STV) as a method of conducting parliamentary elections in Britain and elsewhere (as is already done in Ireland). For many years, Australia (in addition to electing Senators from statewide multi-member districts under STV), has filled seats in its (much more important) House of Representatives under the so-called Alternative Vote (AV), which is simply STV applied to single-member districts. In the United States, some non-partisan local elections and some party primary elections (especially in the South) are conducted not on the basis of Simple Plurality Vote but with a runoff between the two top candidates (SPV+R) if neither receives an absolute majority (or perhaps some lower threshold) of the votes cast in the first round of voting. A similar system has been used over the past thirty years in French Presidential elections. Advocates of abolition of the Electoral College for electing U.S. Presidents and its replacement with a direct popular vote often recommend that such a vote be conducted with a runoff in the same manner as French presidential elections (though some recommend a lower threshold than 50% to trigger a runoff). Quite recently in the U. S., what has come to be called Instant Runoff Voting (IRV) has been enthusiastically advocated for U.S. Presidential and other executive elections (and sometimes also for legislative elections in single-member districts). Finally, the newly created office of the Mayor of London is filled by means of the so-called Supplementary Vote (SV), which is a variant of IRV.

All these voting systems belong to a still more general family dubbed *point-runoff voting systems* by Smith (1973), in which through two or more rounds of vote counting (or actual voting) candidates are sequentially eliminated as possible winners on the basis of their having the lowest support (by one measure or other) on the previous round. Within the STV family, support is measured with respect to first preference [plus any transfer] support. Under ordinary runoff voting, voters are called back for a second trip to the polls for the runoff election. Under the other systems, voters come to the polls just once but are invited to indicate second and usually lower preferences on the ballot. Runoffs are then "simulated" by transferring ballots for eliminated candidates to surviving candidates based on the second and lower preferences indicated on the ballots. In the simplest non-trivial case (which we focus on here) in which one candidate is to be elected from a field of three candidates, all these systems are logically equivalent.

Single Transferable Vote. STV is an exceptionally complicated voting system for electing *m* candidates in (usually small magnitude) multi-member districts whose basic characteristic will be only sketched out here.

(1) Voter use an ordinal ballot to rank all the candidates in their order of preference.[1]

(2) The total number of ballots cast is determined, on which basis of quota for election is established. Normally this quota is the next integer above the quotient resulting from dividing the total number of ballot cast by one more than the number of candidates to be elected.

(3) The ballots are sorted according their first preferences. Any candidate who meets the quota at this stage is elected. If *m* candidates achieve quota, vote counting stops.

(4) Otherwise some votes are "transferred" and the counting continues. First preference votes received by elected candidate in excess of the quota are "surplus" and are transferred to unelected candidates according to their second (or lower) preference votes.[2] As a result of the transfer of surplus votes, additional candidates may meet the quota and be elected. If *m* candidates now meet the quota, counting stops.

(5) Otherwise more votes are transferred and counting continues. The candidate with the *fewest* (first preference plus transferred) votes is *eliminated* and all of the eliminated candidate's votes are transferred according to their highest preferences for any remaining (non-eliminated and non-elected) candidates. As a result of this transfer, additional candidates may meet the quota and be elected. If *m* candidate now meet the quota, counting stops.

---

[1] In practice, voters are likely to truncate their rankings. Different quotas and vote transferring algorithms are used to deal with this and other complexities.

[2] Since it is arbitrary *which* votes are surplus, votes are usually transferred in proportion to all the second preference votes on the elected candidate's ballots. This means that either fractional votes may be transferred or transferred votes must be rounded to integer totals.

(6) The vote transfer and counting process continues until *m* candidates have met the quota and are elected.

Alternative Vote. AV is simply STV applied to single winner elections. The quota is therefore a simple majority of the total number of ballots cast. If one candidate receives more than half of the first preference votes, that candidate elected. Otherwise, the candidate with the fewest first preference ballots is eliminated and these ballots are transferred on the basis of their second preferences, and so on until some candidate accumulates a majority of (first preference and transferred) ballots and is elected.

Instant Runoff Voting. IRV proceeds in the same way as AV at the outset. But if no candidate receives a majority of the first preference votes, all candidates except the leading candidate and the runner-up are eliminated all at once and their first preference ballots are transferred to one or other of the two surviving candidates according to which is higher ranked on each ballot.[3]

Supplementary Vote. Under SV each voter casts one "regular" vote for one candidate (as under Simple Plurality Voting) but can also cast a second "supplementary" vote. (In effect, voters are invited to express first and second but no lower preferences.) The ballots are tabulated with respect to the regular (first preference) votes, and a candidate who receives a majority of all such votes is elected. If no candidate receives a majority of regular votes, all candidates except the leader and runner-up are eliminated and the ballots whose regular votes were cast for eliminated candidates are transferred to one or other surviving candidate on the basis of the supplementary vote — or not transferred at all in the event the supplementary vote is for another eliminated candidate. Whichever of the two surviving candidates has the most votes after these transfers (which is likely to be fewer than half of the total ballots) is elected.

Advocates have generally claimed these points in favor of STV and related voting systems.

1. STV allows self-defined opinion groupings in the electorate to secure the election of representatives of their choice in rough proportion to their proportion of the electorate in a multi-member district.[4] (Clearly this characteristic of STV does not extend to its single-winner variants.)

---

[3] As defined here, IRV simulates (and is logically equivalent to) ordinary runoff elections but with two practical differences. First, in ordinary runoff elections the composition of the participating electorate may be different in the first and second rounds of voting. (In U.S. runoff elections, turnout typically declines in the second round.) Second, in ordinary runoff elections, voters who voted in the first round for eliminated candidates can (sincerely or strategically) revise their preferences between the two leading candidates (which in any case they had not yet been called upon to express) prior to the runoff (perhaps taking account of new information, including the vote totals from the first round). It should be noted that some people who advocate what they call IRV are actually advocating what is here called AV.

[4] In contrast to workings of list proportional representation, these self-defined opinion groupings need not be organized political parties. (Indeed, advocacy of STV has a distinctly anti-party favor.)

2. The great sin of Simply Plurality Voting or FPTP is that many votes are "wasted," in that they don't contribute to the election of any candidate, either because (i) they were cast for candidates who lost anyway, or (ii) they were cast for the candidate who would have won anyway. Under STV almost all votes will ultimately be counted for, and (given the transfer of surplus votes) be necessary to the election of, one of the elected candidates. (Under the single-winner variants, wasted votes of the first type, but not of the second, are minimized.)

3. Because in SPV elections votes cast for trailing candidates (expected to place third or lower) are almost certainly going to be "wasted," voters who sincerely prefer such candidates are encouraged vote "tactically" for whichever of the two leading candidates they relatively prefer. Thus the trailing candidates do not get even their (small) "fair share" of vote. However, under STV (and its single-winner variants) such voters can rank their most preferred candidate first knowing that, in the event this candidate is not elected, their votes will transfer to other candidates in way that reflects their preferences.

4. In so far as supporters of trailing candidates who actually have a preference between the two leading candidates resist the temptation to vote tactically, the trailing candidates are potential "spoilers" — that is, while they cannot win in any event, their entry into, or withdrawal from, the election can determine which of the leading candidates does win. Eliminating this "spoiler problem" is one of the main claimed advantages of IRV in the U.S.

5. Until fairly recently, the general claim of STV advocates has been that voters can safely rank the candidates on their ballot in order of their true preferences. Indeed, STV ballots often not only instruct voters to rank candidates in order of preference but indicate that, regardless of how many candidates a voter orders, ranking one candidate of above others may help that candidate's chance of being elected and certainly can never reduce it.[5] The general claim that it is never advantageous to depart from voting one's true preferences is the conjunction of two separate claims: given all other ballots, (i) whether candidate *a* is elected is or not is independent of how the voter ranks other candidates (e.g., whether the voter ranks *b* above *c* or *c* above *b*); and (ii) a voter can never bring about candidate *a*'s election by ranking *a* lower or bring about candidate *a*'s defeat by ranking *a* higher.

Voting and social choice theorists have had reason to question, at least in principle, such claims of "strategyproofness" for STV since Gibbard (1973) and Satterthwaite (1975) published theorems to the effect that no non-dictatorial voting procedure can be strategyproof given three or more alternatives. However, voting procedures most commonly fail to be strategyproof as a result failure of condition (i) described above. The kind of "monotonicity failure" entailed by failure of condition (ii) described above is more surprising and might be thought not to characterize any halfway reasonable voting system in actual use.[6] But, in a quite extraordinary paper, Smith (1973)

[5] For example, see the description of STV ballots in Niemi (1970, p. 92). (Niemi examined ordinal STV ballots from university elections to try to track down instances of cyclical majorities.)

[6] However, it has long been recognized that quorum requirements can induce a kind of monotonicity failure.

showed that so-called "point-runoff" systems were subject to monotonicity failure. Several years later, Doron and Kronick (1977) in effect observed that Smith's class of point-runoff systems includes STV and its single-winner variants, which are therefore subject to monotonicity failure. This finding attracted some attention within U.S. academic political science (especially in Riker, 1982, pp. 49-50; also see Brams and Fishburn, 1983, and Fishburn and Brams, 1983) but evidently little elsewhere.

However, in the early 1990s, the British Labour Party established a commission under the leadership of political scientist Raymond (now Lord) Plant to review alternative voting systems and to make recommendations for possible changes in the British electoral system that might be supported by the party (and that might facilitate cooperation with the Liberal Democrats). The "Plant Report," though it was generally balanced and well informed by academic research, attracted much attention and hostility from British electoral reformers because it rejected STV (while, ironically, speaking favorably of SV) as an option for British parliamentary elections, and it did so largely on theoretical grounds, laying emphasis on the problem of monotonicity failure.[7]

The outraged electoral reformers did not — for they could — dispute the logical possibility of monotonicity failure (which can be definitively established by a single example, such as that provided by Doron and Kronick). They did, however, offer a number of counter-arguments that were potentially sensible and to the point (for example, Editorial, 1993 and Hill, 1994).

1. No voting system is perfect; all have problems of some sort or other. Therefore, one should not summarily reject one voting system on the basis of one type of problem without considering the problems to which other systems are subject. (Invoking the impossibility theorems of voting and social theory in this way is entirely appropriate — but also has a very different flavor from the kinds of claims that previously had been made in behalf of STV.)

2. While the Plant Report rejected STV on the grounds of monotonicity failure, it was "particularly galling" that the Report recommended "the very inferior SV, that contains the very same fault as that for which they rejected" STV (Hill, 1994). (This point seems unanswerable.)

3. The problems to which particular voting systems (including monotonicity failure with respect to STV and related systems) are demonstrated to be subject on the basis of particular examples are merely logical possibilities and, even if the problem is quite serious in the event that it actually occurs (which seems to be true of monotonicity failure), the problem may in fact occur so infrequently that it can be dismissed for all practical purposes.

It is this third counter-argument that we focus on here. STV advocates cite two claims in particular to support this third counter-argument.

---

[7] The report cited the work of Doron and Kronick, Riker, and Brams and Fishburn. The British philosopher Michael Dummett (who has published work on voting procedures [1984] and electoral reform [1997]) also made a direct submission to the Plant Commission that criticized STV on monotonicity (as well as other) grounds (Farrell, 2001, p. 150).

1. In a “hands-on assessment of STV” based on his twenty two years of experience as the Chief Electoral Officer for Northern Ireland since the introduction in 1973 of STV (for MEPs, the Northern Ireland Assembly, and local government elections) in that province, Patrick Bradley (1995) reported that “the experience of the use of STV in Northern Ireland over the past 22 years, involving a range of election types and sizes, reveals no evidence to support *in practice* the lack of monotonicity.” It is difficult to give much credence to this claim, since instances of monotonicity failure do immediately reveal themselves (for example, in the manner the “reversal of winners” problem to which districted electoral system are subject, e.g., the 2000 U.S. Presidential election or the 1951 U.K. general election). Monotonicity failure is not directly apparent from aggregate election results, or even from more detailed tabulations showing the sequence of transferred votes. Rather it is necessary to inspect full “ballot profiles” (as defined in the next section) and then subject them to extensive and tedious analysis and computation.

2. Based on a combination of mathematical analysis, empirical data, and statistical assumptions, Crispin Allard (1995, 1996) estimated that the probability of STV monotonicity failure is so low that “if the U.K. is divided into 138 multi-member constituencies, and assuming quadrennial general elections, *we would expect in the whole country less than one instance every century of monotonicity failure under STV*” (1995, p. 49). (We will review Allard’s arguments in the conclusion of this paper.)

In some degree at least, this conclusion and this evidence supporting it is entering the general literature on voting systems. Thus, in his very useful new book on voting systems, Douglas Amy (2000, p. 55 [with respect to IRV] and p. 105 [with respect to STV]) says: “It is undisputed that nonmonotonicity can theoretically occur in a ... [IRV or STV] election, but most experts believe that the conditions needed for this paradox to occur are so special that it would be an extremely rare occurrence. One statistical study [i.e., Allard] found that if ... [IRV-like or STV] elections were to be held throughout the United Kingdom, a nonmonotonic result would occur less than once a century.” (It should be noted that [non-]monotonicity does not apply to an “election result” but to how elections results may change as ballot profiles change.) And in a recent text on electoral systems, David Farrell (2001, p. 150) cites both Bradley and Allard in claiming that “there is no evidence that it [monotonicity failure] is a common occurrence.”

2. Analysis

From here on out we focus on the case of a single-winner election with exactly three candidates, in which case the voting systems discussed here are logically equivalent (and, at the same time, logically distinct from FPTP). In the concluding section, we shall speculate a bit about how our conclusion might be extended to more general case.

First of all, it will be useful just to present a simple example of how monotonicity failure may occur under voting systems of this type.

Suppose that there are three candidates *a*, *b*, and, *c* and that a highly accurate poll on the eve of the election shows that (i) no candidate commands the first preference support of a majority of the electorate and that (ii) candidate *c* has the least first preference support. Thus the ballots that rank *c* highest will transfer in some mix to *a* and *b* in a (simulated) runoff between them. The poll also elicits information about these second preferences and reveals that (iii) the mix of transferred votes will be sufficiently favorable to *a* that *a* will win the runoff and therefore will be elected.

This is indeed good news for candidate *a* and his supporters, but they want to spare no effort in securing electoral victory and therefore undertake further campaign activities. These last-minute efforts are *unambiguously successful*, in that on election day some voters who previously ranked *a* lower now give first-preference support to *a* (though *a* still falls short of absolute majority first-preference support) and, at the same time, no voters have either moved candidate *a* down in their rankings or changed their ranking of *b* and *c*.

But *a*'s additional first-preference support must have come at the expense of one or other or both of the other candidates. In fact, it came mostly at the expense of candidate *b*, whose first-preference support now falls below *c*'s. As a result, the runoff is now between *a* and *c*, and its outcome depends on the mix of support for *a* and *c* in the second preferences of the remaining *b* voters. And in fact this mix is (as it had been all along) sufficiently favorable to *c* that *c* wins the runoff is therefore elected. Thus we have monotonicity failure — candidate *a*'s added support has cost *a* electoral victory.

We now aim to pin down more precisely the conditions under which such monotonicity failure can occur.

2.1 Plurality and Ballot Profiles

There are three candidates *a*, *b*, and *c*. Voters mark ballots according to one of six strong orderings, where *P* with a subscript designates the proportion of voters with preferences as shown below.

P_a		P_b		P_c	
P_{ab}	P_{ac}	P_{ba}	P_{bc}	P_{ca}	P_{cb}
a	*a*	*b*	*b*	*c*	*c*
b	*c*	*a*	*c*	*a*	*b*
c	*b*	*c*	*a*	*b*	*a*

Clearly $0 \leq P \leq 1$ for all subscripts and $P_a + P_b + P_c = P_{ab} + P_{ac} + P_{ba} + P_{bc} + P_{ca} + P_{cb} = 1$. For notational and analytic convenience, we assume that all voters mark complete (or *non-truncated*) ballots and that the number of voters is sufficiently large that we can rule out plurality or majority ties (i.e., that always $P_a \neq P_b$, $P_{ab} \neq P_{ac}$, etc.).

A *plurality profile P* is a list (P_a, P_b, P_c) indicating the relative popularity of each candidate with respect to voters' first preferences. We normally label the candidates so that $P_a > P_b > P_c$ (at least until we modify preferences so as to try to produce monotonicity failure). Thus we assume that candidate *a* is the *plurality winner* and also that candidate *a* is the *majority winner* if there is one (i.e., in the event that $P_a > ½$)

A *ballot profile B* is a list (P_{ab}, P_{ac}, P_{ba}, P_{bc}, P_{ca}, P_{cb}) indicating the relative popularity of each preference ordering (or ballot). We use the notation *a B b* to mean that a majority of ballots in *B* rank *a* over *b* (so that, in a runoff between *a* and *b*, *a* wins). If *a B b* and *a B c*, *a* is the *Condorcet winner* at *B*; and if *b B a* and *c B a*, *a* is the *Condorcet loser* at *B*.

We define γ_{ab} as $(P_{ca} - P_{cb}) / P_c$ and likewise for other arrangements of candidates. Thus $\gamma_{ab} \times P_c$ is the *net relative advantage* that *candidate a* realizes when the *c* votes are redistributed between *a* and *b* on the basis of second preferences. For example, if $P_a = .45$, $P_b = .4$, $P_c = .15$ and $\gamma_{ba} =$ a, candidate *b* gains a $\times .15 = .05$ on candidate *a* in a runoff, just enough to close the first preference gap between *b* and *a*. Clearly $-1 \le \gamma_{ab} \le +1$ for all pairs of candidates. If $\gamma_{ab} = -1$, all *c* voters prefer *b* to *a*; if $\gamma_{ab} = 0$, *c* voters split equally between *a* and *b*; if $\gamma_{ab} = +1$, all *c* voters prefer *a* to *b*. In any event, $\gamma_{ab} = -\gamma_{ab}$ and likewise for other pairs of candidates.

Given two ballot profiles *B* and *B′*, we say that *B′* is *more favorable* to *a* than *B* is (or *B* is *less favorable* to *a* than *B′* is) if the two profiles are identical except that some ballots rank *a* higher in *B′* than in *B*. We write this as $B' \succ_a B$.

OBSERVATION 1. If $B' \succ_a B$, then *a B b* implies *a B′ b* and *b B′ a* implies *b B a*.

Given two plurality profiles *P* and *P′*, we say that *P′* is *more favorable* to *a* than *P* is if $P'_a > P_a$. Note that, while each ballot profile *B* entails a unique plurality profile *P*, a more favorable ballot profile need not produce a more favorable plurality profile (because the more favorable ballots may move *a* only from third to second place). Conversely, a more favorable plurality profile need not derive from a more favorable ballot profile (because some ballots may move *a* into first place while others move *a* down). But if the more favorable *B′* also results in a more favorable *P′*, we say that *B′* is *strictly more favorable* to *a* than *B* is. We write this as $B' \succ\succ_a B$.

We say that ballot profile *B* entails *forward monotonicity failure* with respect to candidate *a* in the event that *a* wins at profile *B* but loses at some other profile *B′* that is more favorable to *a* than *B* is. Conversely, we say that ballot profile *B* entails *backward monotonicity failure* with respect to candidate *a* in the event that *a* loses at profile *B* but wins at some other profile *B′* than is less favorable to *a* than *B* is. Given the voting systems under consideration (under which candidates survive or are eliminated on the basis of first-preference support), the fact that the pairing of candidates in the runoff changes when the ballot profile changes from *B* to *B′* implies that the resulting plurality profile changes as well, so monotonicity failure can occur only with respect to ballot profiles that are strictly more (or less) favorable to the focal candidate.

OBSERVATION 2. Given $P_a > P_b > P_c$, forward monotonicity failure can occur with respect to either candidate *a* or candidate *b*. However backward monotonicity failure can occur with only respect to candidate *a*.

Since candidate *c* must lose at *B*, monotonicity failure can occur only with respect to candidates *a* or *b*. That forward failure can occur with respect to either *a* or *b* was illustrated by the informal example that opened this analysis section (in which it was specified only that *c* had the least first-preference support). However, backward failure cannot occur with respect to *b*. Suppose otherwise. Since B' is less favorable to *b* than *B* is, it must be that $P'_a \geq P_a$ and $P_b > P'_b$ but, in conjunction with $P_a > P_b$, this implies that $P'_a > P_b$, so there is no way *b* can get into a runoff with *c* (which is the only way *b* could win).

OBSERVATION 3. Suppose there is a forward (backward) monotonicity failure such that candidate *a* wins (loses) at some profile *B* but loses (wins) at profile B'' that is more (less) favorable to *a* than *B* is. Then either: (i) there is a profile B' "between" *B* and B'' — that is, $B'' \succ_a B' \succ_a B$ ($B \succ_a B' \succ_a B''$) — at which *a* loses (wins) such that $P'_{ca} = P_{ca}$ and $P'_{cb} = P_{cb}$ or (ii) B'' is self is such a profile, i.e., $P''_{ca} = P_{ca}$ and $P''_{cb} = P_{cb}$.

This follows because monotonicity failure with respect to *a* depends on the *balance* of first preference support between *b* and *c* and not on the absolute level of either *b*'s or *c*'s first preference support. We say that such a shift from *B* to B' "holds *c* harmless."

2.2 The Space of Plurality Profiles

It is well known (Viviani's Theorem) that the sum of the distances from any point in an equilateral triangle such as that shown in Figure 1 (including points on the edges and vertices) is equal to the height of the triangle and is therefore the same for every such point. Thus the sets of points in such a triangle can conveniently represent all possible sets of three positive real numbers that sum to a constant amount and, in particular, can represent the space of all plurality profiles with three candidates.

The three vertices represent the three profiles in which one candidate gets all the first preferences. The point in the center of the triangle represents the profile in which the three candidates split first preferences equally. Lines parallel to an edge represent sets of profiles in which one candidate receives a constant proportion of first preferences (i.e., the candidate is "held harmless"). (The edge itself represents all profiles in which one candidate receives a constant zero proportion of first preferences.) Corresponding to such lines are equations such as $P_a = .4$ (or, equivalently, $P_b + P_c = .6$). Lines not parallel to any edge represent sets of profiles in which the proportion of first preferences received by one candidate is a linear function of the proportion received by another such as $P_a = P_b + .5\,P_c$ (which can be rewritten in various ways to exclude one of the three *P* s). Selected points and lines are labeled in Figure 1.

We aim to specify regions in this space within which a profile may be subject to (forward or backward) monotonicity failure. In so far as we assume that $P_a > P_b > P_c$, we focus on the one sixth of full triangle with that constitutes the right triangle with vertices (1,0,0), (½ ,½ ,0), and (a , a , a).[8]

[8] Note that we are demarcating the region of possible *B* profiles. To produce monotonicity failure, the corresponding B' profiles must lie outside of the $P_a > P_b > P_c$ region, since *c* must get into the runoff (i.e., $P'_c > P'_b$).

We first establish *necessary* conditions for monotonicity failure in terms of *bounds on plurality profiles*.

It is immediate that in the event of (forward or backward) monotonicity failure, we must have ½ $>$ P_a . Suppose to the contrary that P_a $>$ ½, so a wins at B without a runoff. If $B' \succ_a B$, then $P'_a \geq P_a >$ ½, so a also wins at B' without a runoff, and there can be no forward monotonicity failure. And there can be no backward monotonicity failure at B as a already wins at B.

We now consider bounds on P_c and P_b in the event of forward failure. Consider Table 1A, which shows the relationship between plurality profiles associated with B and B' in the event of forward monotonicity failure associated with candidate a (where B' is selected to hold c harmless). Let Δ be the shift in relative pluralities as we move from B to B', i.e., $\Delta = P'_a - P_a = P_b - P'_b$. Since a loses at B' , it must be that ½ $> P'_a$. Thus ½ $- P_a > \Delta$. Since $P_c = P'_c > P'_b$, it must also be that $\Delta > P_b - P_c$. Thus know that ½ $- P_a > \Delta > P_b - P_c$. Substituting $(1 - P_b - P_c)$ for P_a , we get ½ $- (1 - P_b - P_c) > P_b - P_c$, which simplifies to $P_c >$ ¼ or, equivalently, ¾ $> P_a + P_b$. Since $P_a > P_b$, we must have d $> P_b$.

Table 1A: Forward Monotonicity Failure with Respect Candidate a

P	½	>	P_a	>	P_b	>	P_c	>	¼
			$\wedge$		$\vee$		$\Vert$		
P'	½	>	P'_a	>	P'_b	<	P'_c	>	¼

Now consider Table 1B, which shows the relationship between plurality profiles associated with B and B' in the event of forward monotonicity failure associated with candidate b. By the same reasoning as above, ½ $- P_b > \Delta > P_a - P_c$, which likewise implies that $P_c >$ ¼ and d $> P_b$.

Table 1B: Forward Monotonicity Failure with Respect Candidate b

P	½	>	P_a	>	P_b	>	P_c	>	¼
			$\vee$		$\wedge$		$\Vert$		
P'	½	>	P'_a	<	P'_b	>	P'_c	>	¼

.))))))))))) <))))))))))))-

We now consider the bounds on P_b and P_c in the event of backward failure. Consider Table 1C, which shows the relationship between plurality profiles associated with B and B' (where B' is selected to hold b harmless) in the event of backward monotonicity failure associated with candidate a. Now $\Delta = P_a - P'_a = P'_b - P_b$. Since a loses at B, it must be that b B a, and since $B \succ_a B'$, it follows that b B' a. Since a wins at B′, a must be paired with (and beat) c in the runoff. Thus it must be that $P'_a > P'_b = P_b$ and also $P'_c > P'_b = P_b$. Thus we must have $P_a - P_b > \Delta > P_b - P_c$. Substituting $(1 - P_a - P_b)$ for P_c, we get $P_a - P_b > P_b - (1 - P_a - P_b)$, which simplifies to a $> P_b$ or, equivalently, $P_a + P_c >$ b . Since always ½, $> P_a$ we must have $P_c > {}^1/_6$. Since ½ $> P_a$, P_b

$+ P_c > ½$ and since $P_b > P_c$ therefore $P_b > ¼$.

Table 1C: Backward Monotonicity Failure with Respect Candidate a

P	½	>	P_a	>	P_b	>	P_c	>	$^1/_6$
			$\vee$		$\parallel$		$\wedge$		
P'	½	>	P'_a	>	P'_b	<	P'_c	>	¼

We can summarize these conclusions in the form of a proposition.

PROPOSITION 1. Given that $P_a > P_b > P_c$:

(a) forward monotonicity failure occurs only if $½ > P_a > P_b > P_c > ¼$ (which also implies that d $> P_b$) ; and

(b) backward monotonicity failure occurs only if $½ > P_a >$ a $> P_b > P_c > {}^1/_6$ (which also implies that $P_b > ¼$).

Using Proposition 1, we can demarcate the regions of the plurality profile space in which monotonicity failures may occur.

Given that we are in the $P_a > P_b > P_c$ region of Figure 1, forward monotonicity failure can occur only on the $P_a < ½$ side of the $P_a = ½$ line and only on the $P_c > ¼$ side of the $P_c = ¼$ line, i.e., within the right triangle with vertices (½,¼,¼), (d,d,¼), and (a,a,a). Allowing for all permutations of candidate labels, forward failure can occur only within horizontally marked regions of Figure 2.

Given that we are in the $P_a > P_b > P_c$ region, backward monotonicity failure can occur only on the $P_a < ½$ side of the $P_a = ½$ line and only on the $P_b >$ a side of the $P_b =$ a line, i.e., within the right triangle with vertices (½,¼,¼), (½, a ,$^1/_6$), and (a,a,a). Allowing for all permutations of candidate labels, forward failure can occur only within the vertically marked regions in Figure 2.

A better understanding of Figure 2 can be provided by an example. The horizontal line P_b = 0.3 is shown. Moving across this line from left to right indicates what happens as candidate *c*'s first preference support increases at *a*'s expense, while candidate *b* is held harmless at 30% of the vote). The story is told across the line at the bottom of Figure 2. From the profile (.7, .3, 0) to (.5, .3, .2), *a* wins without a runoff; from (.5, .3, .2) to (.4, .3, .3) *a* goes into a runoff with *b*, the outcome of which depends on the second preferences of the *c* voters. From (.4, .3, .3) to (.3, .3, .4) *a* goes into a runoff with *c*, the outcome of which depends on the second preferences of the *b* voters. From (.3, .3, .4) to (.2, .3, .5) *c* goes into a runoff with *b*, the outcome of which depends on the second preferences of the *a* voters Finally, from (.2, .3, .5) to (0, .3, .7) *c* wins without a runoff.

In the (.5, .3, .2) to (.4, .3, .3) interval, *a* goes into a runoff with *b*, which *b may* win (especially where *a*'s first preference advantage is reduced toward the right end of the interval). But in the (.4, .3, .3) to (.3, .3, .4) interval, *a* goes into a runoff with *c*, which *a may* win (especially where *a* retains some first preference advantage at the left end of the interval). Thus the shift in preferences that moves the profile across the (0.4, 0.3, 0.3) threshold *may* entail backward monotonicity failure as *a*'s decreasing support converts a win by *b* into a win by *a*. Likewise the shift in preferences across the (0.3, 0.3, 0.4) mark may entail forward monotonicity failure with respect to *c*. And of course, if such failures occur as *c*'s support increases at *a*'s expense, the reverse failures occur as *a*'s support increases at *c*'s expense.

Looking at the full profile space in Figure 2, the triangles constituting the regions of possible forward and possible backward monotonicity failure can be paired off. For example, a profile *B* in the right triangle with vertices (½,¼,¼), (d,d,¼), and (a,a,a) may be subject to forward monotonicity failure with respect to candidate *a* resulting from candidate *a*'s gaining first preference support at *b*'s expense (while *c* is held harmless). This will entail a "southwest" shift from *B* along a line parallel to the $P_c = 0$ edge to some profile B' located in the right triangle with vertices (½,¼,¼), (½,$^1/_6$,a), and (a,a,a), i.e., the region susceptible to backward monotonicity failure with respect to *a* given that $P_a > P_c > P_b$. In this event, a shift from B' back to *B* will entail backward failure with respect to *a*. So these two triangular regions are paired off.

But a profile *B* in the triangle with vertices (½,¼,¼), (d,d,¼), and (a,a,a) may also be subject to forward monotonicity failure with respect to (the second ranked) candidate *b*. This will entail a "northeast" shift from *B* again along a line parallel to the $P_c = 0$ edge to some profile B' located in the right triangle with vertices (¼,½,¼), ($^1/_6$,½,a), and (a,a,a), i.e., the region susceptible to backward monotonicity failure with respect to *b* given that $P_b > P_c > P_a$. (Note that the $P_b > P_a > P_c$ region is skipped over.) In this event, a shift from B' back to *B* will entail backward failure with respect to *b*. So these two triangular regions are also paired off.

2.3 The Underlying Ballot Profiles

The conditions on plurality profiles discussed in the previous section state only necessary conditions for monotonicity failure. Such failure actually occurs or not *depending on the second preferences of the voters whose most preferred candidate does not make it into the runoff*. Thus we have to look beyond plurality profiles and look at (some aspects of) the full orderings given by ballot profiles. However, monotonicity failure with respect to candidate *a* (for example) depends only the parameters γ_{ab} (or γ_{ba}) and γ_{ac} (or γ_{ca}) — the second preferences of the *a* voters play no role.

Suppose as usual that $P_a > P_b > P_c$ and that there is a ballot profile *B* that entails backward monotonicity failure, necessarily with respect to candidate *a*. This means that *B* puts candidate *a* into a runoff with candidate *b*, which *a* loses. Thus we have *b B a* or:

$$\begin{aligned} P_b + P_{cb} &> P_a + P_{ca} \\ P_b + (P_{cb} - P_{ca}) &> P_a \\ P_b + \gamma_{ba} \times P_c &> P_a . \end{aligned}$$

Candidates *a* and *b* tie in the event that:

$$P_a = P_b + \gamma_{ba} \times P_c \qquad \text{(Equation 1)}$$

Figure 3 shows these lines:

(i) $P_b = ½ = P_a + P_c$, which can be rewritten as
$2 P_a + 2 P_c = 1$ or
$P_a + P_c = 1 - P_a - P_c$
$P_a + P_c = P_b$
$P_a = P_b - P_c$, i.e., Equation 1 in the event $\gamma_{ba} = -1$

(ii) $P_a = P_b$, i.e., Equation 1 in the event that $\gamma_{ba} = 0$; and

(iii) $P_a = ½ = P_b + P_c$ or [proceeding as in (i)]
$P_a = P_b + P_c$, i.e., Equation 1 in the event $\gamma_{ba} = +1$.

In general, the line representing the Equation 1 $P_a = P_b + \gamma_{ba} \times P_c$ passes through the point (½,½,0) and falls between the (ii) $P_a =$ ½ and (iii) $P_b =$ ½ lines depending on the value of γ_{ab} . Given that $P_a > P_b > P_c$, we are concerned only with case (iii) in which γ_{ba} is positive. In Figure 3, the line $P_b = P_a + \gamma_{ba} \times P_c$ is drawn in for $\gamma_{ba} = ½$. (By straightforward calculation, the line intersects the $P_b = P_c$ line at $a = {}^3/_7$.[9]) Thus, given $\gamma_{ba} = ½$, it follows that *b* wins at all plurality profiles "northeast" of this line.

For backward monotonicity failure, we must find some B' such that:

(i) $B \succ\!\succ_a B'$ (the new profile is strictly less favorable to *a*);
(ii) $P'_a > P'_b$ and $P'_c > P'_b$ (the runoff is between *a* and *c*); and
(iii) $a\, B'\, c$ (*a* wins the runoff).

In order for *a* to win the runoff against *c*, we must have:

$$P'_a + P'_{ba} > P'_c + P'_{bc}$$
$$P'_a > P'_c + (P'_{bc} - P'_{ba})$$
$$P'_a > P'_c + \gamma_{ca} \times P'_b .$$

We now rewrite this condition in terms of the original plurality profile, where as before Δ is the shift in first preferences from candidate *a* to candidate *b* that characterizes the shift from *B* to B' :

$$P_a - \Delta + P_{ba} > P_c + \Delta + P_{bc}$$
$$P_a + \gamma_{\mathrm{ac}} \times P_b > P_c + 2\,\Delta .$$

We know that Δ must be large enough to tip the balance of first preferences between candidates *b* and *c*, i.e., that $\Delta > P_b - P_c$. Thus we have:

$$P_a + \gamma_{ac} \times P_b > P_c + 2\,P_b - 2P_c .$$

Replacing P_c with $1 - P_a - P_a$ and combining we get:

---

[9] Since $P_a = P_b + P_c \,/\, 2$ and $P_b = P_c$, $P_a = 3\,P_b \,/\, 2$. Since $P_b = 1 - P_a - P_c$ and $P_b = P_c$, $P_b = (1 - P_a) \,/\, 2$. Thus $P_a = 3\,(1 - P_a) \,/\, 4$, which simplifies to $P_a = {}^3/_7$.

$$\gamma_{ac} \times P_b \; > \; 3\,P_b - 1$$

$$\frac{1}{(3 + \gamma_{ac})} > P_b \,.$$

The equation

$$P_b \; = \; \frac{1}{(3 + \gamma_{ac})} \qquad \text{(Equation 2)}$$

is represented by a horizontal line that lies below or above the P_b = a line depending on whether γ_{ac} is positive or negative. We have *a B′ c* at all profiles that lie above this line.

Putting everything together, we can state the following:

PROPOSITION 2. Given that $P_a > P_b > P_c$, backward monotonicity failure occurs (with respect to *a*) at profile *B* if and only if these conditions hold:

(1) ½ > P_a > a > P_b > P_c > $^1/_6$;
(2) $P_b \; + \; \gamma_{ba} \times P_c > P_a$; and
(3) $\dfrac{1}{(3 + \gamma_{ac})} > P_b$

Suppose that γ_{ba} = ½ (i.e., with respect to their second preferences, 75% of the *c* supporters prefer *b* over *a*, so that the net relative increment in *b*'s support over a resulting from votes transferred from *c* is equal to one-half of those votes) and γ_{ac} = ¼ (i.e., with respect to their second preferences, 62.5% of *b* supporters prefer *a* over *c*). Given these parameters γ_{ba} and γ_{ac} , backward monotonicity failure with respect to *a* at any profiles in the area shown in Figure 3a.

If second preferences are distributed "favorably" enough, these bounds enclose the entire triangle with vertices (½,¼,¼), (½, a ,$^1/_6$), and (a , a , a), in which case monotonicity failure will actually occur at all profiles that are potentially subject to it.

COROLLARY 2.1. Given that $P_a > P_b > P_c$ and also that γ_{ba} = 1 and $\gamma_{ac} \leq 0$, backward monotonicity failure occurs (with respect to *a*) at every profile *B* satisfying the conditions specified in Proposition 1(b).

If γ_{ba} = 1, the $P_a = P_b + \gamma_{ba} \times P_c$ line becomes simply the horizontal $P_a = P_b + P_c$ (or P_a = ½) line and, if γ_{ac} = 0, $P_b = 1 \,/\, (3 + \gamma_{ac})$ = a , so the entire triangle with vertices (½,¼,¼), (½, a ,$^1/_6$), and (a , a , a) lies between these lines. Put more substantively, in the event that every *c* supporter prefers *b* to *a* and no more than half of the *b* supporters prefer *a* to *c*, every profile B such that ½ > P_a > a > P_b > P_c > $^1/_6$ is subject to backward monotonicity failure with respect to candidate *a*.

On the other hand, if second preferences are distributed "unfavorably" enough, these bounds enclose the entire triangle with vertices (½,¼,¼), (½, a ,$^1/_6$), and (a , a , a) in which profiles are potentially subject to monotonicity failure and such failure therefore in fact cannot occur at any profile.

COROLLARY 2.2. Given that $P_a > P_b > P_c$ and also that $\gamma_{ba} = 0$ or $\gamma_{ac} = 1$, backward monotonicity failure cannot occur at any profile *B*.

The region of monotonicity failure shrinks to zero area just as the three lines $P_a = P_b + \gamma_{ba} \times P_c$, $P_b = 1 / (3 + \gamma_{ac})$, and $P_b = P_c$ intersect at a common point, as shown in Figure 3b.

We can readily find a formula for the point at which the $P_a = P_b + \gamma_{ba} \times P_c$ line intersects the $P_b = P_c$ by substituting P_b for P_c and $1 - 2P_b$ for P_a. Thus we have:

$$1 - 2P_b = P_b + \gamma_{ba} \times P_b$$
$$3P_b + \gamma_{ba} \times P_b = 1$$
$$(3 + \gamma_{ba})\, P_b = 1$$
$$P_b = \frac{1}{(3 + \gamma_{ba})}$$

The region of monotonicity failure thus shrinks to a point of zero area on the $P_b = P_c$ line when $\gamma_{ac} = \gamma_{ba}$, and its location on the $P_b = P_c$ line depends on the common γ parameter value.

Combining these considerations with previous conclusions, we have the following:

PROPOSITION 2′. Given that $P_a > P_b > P_c$, backward monotonicity failure occurs (with respect to *a*) at profile B if and only if $\frac{1}{2} > P_a > 1/(3 + \gamma_{ac}) > P_b > 1/(3 + \gamma_{ba}) > P_c > {}^1/_6$.

COROLLARY 2.3. Given that $P_a > P_b > P_c$, backward monotonicity failure occurs (with respect to *a*) only if $\gamma_{ac} > \gamma_{ba}$..

References

Allard, Crispin (1995). “Lack of Monotonicity — Revisited,” *Representation*, 33/2:48-50.

Allard, Crispin (1996). “Estimately the Probability of Monotonicity Failure in a U.K. General Election,” *Voting Matters*, Issue 5: 1-7.

Amy, Douglas (2000). *Behind the Ballot Box: A Citizen’s Guide to Voting Systems*. Westport, CN: Praeger.

Bradley, Patrick (1995). “STV and Monotonicity: A Hands-on Assessment Assessment,” *Representation*, 33/2: 46-47.

Brams, Steven J., and Peter Fishburn (1983). “Some Logical Defects of the Single Transferable Vote,” in Arend Lijphart and Bernard Grofman, eds. *Choosing an Electoral System*. New York: Praeger, pp. 147-151.

Bullock III, Charles, and Loch K. Johnson (1992). *Runoff Elections in the United States*. Chapel Hill: University of North Carolina Press,

Doron, Gideon, and Richard Kronick (1977). “Single Transferable Vote: An Example of of Perverse Social Choice Function,” *American Journal of Political Science*, 21: 303-311.

Dummett, Michael (1984). *Voting Procedures*. Oxford: Clarendon Press.

Dummett, Michael (1997). *Principles of Electoral Reform*. Oxford: Oxford University Press.

Editorial (1993). “Lack of Monotonicity,” *Representation*, 31/116: 70-71.

Farrell, David M. (2001). *Electoral Systems: A Comparative Introduction*. Hampshire: Palgrave.

Fishburn, Peter, and Steven J. Brams (1983). “Paradoxes of Preferential Voting,” *Mathematics Magazine*, 56: 207-214.

Gibbard, Allan (1973). "Manipulation of Voting Schemes: A General Result," *Econometrica*, 41: 587-601.

Hill, I.D. (1994). “Where Plant Got It Wrong,” *Representation*, 32/117: 20-22.

Niemi, Richard G. (1970). “The Occurrence of the Paradox of Voting in University Elections,” *Public Choice*, 8: 91-100.

Riker, William H. (1982). *Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice*. San Francisco: W.W. Freeman.

Satterthwaite, Mark A. (1975). “Strategyproofness and Arrow’s Conditions: Existence and Ciorrespondence Theorems for Voting Procedures and Social Welfare Functions,” *Journal of Economic Theory*, 10: 187-217.

Smith, John H. (1973). “Aggregation of Preferences with Variable Electorate,” *Econometrica*, 41: 1027-1041.