Instant runoff voting with restricted voting

Author

Gary W. Cox, Professor of Political Science, University of California, San Diego. B.S. with honors, History, California Institute of Technology, 1978. Ph.D., Social Science, California Institute of Technology, 1982. Fellow, American Academy of Arts and Sciences. Former Guggenheim Foundation Fellow. Curriculum vitae attached.

I am the author of over thirty articles in scholarly journals analyzing different electoral systems around the world and of *Making Votes Count: Strategic Coordination in the World's Electoral Systems* (winner of the American Political Science Association's premier prize, the Woodrow Wilson Foundation Award). I have also lectured extensively at universities in the U.S. and abroad on electoral system design. Most of my work focuses on how different voting procedures affect voters' and candidates' strategies and behavior. Among the voting procedures which I have examined are the standard plurality voting system and various alternatives to it, such as approval voting, the single non-transferable vote, cumulative voting, majority run-off, and a wide variety of ranked-preference voting methods (including the alternative vote, also known as instant runoff voting, or IRV for short).

Date

May 22, 2003.

Executive Summary

I was asked to review San Francisco's proposed instant runoff voting (IRV) system, which is slated for its initial use in the November, 2003, municipal election. In my review, I focus on several ways in which San Francisco's version of IRV differs from other ranked-choice voting systems and the implications of those differences. My observations and conclusions follow.

The San Francisco system is *unusual* among IRV and IRV-like systems in both the physical layout of the ballot and in that it restricts voters to ranking only three candidates on the ballot, rather than allowing them to rank as many as they wish.[1] It is *unique* among current IRV and IRV-like systems in that it allows the chief election administrator to decide whether and how much to restrict voters' rankings.[2]

This document makes three main points. First, the ranking restriction can change the election outcome from what it would have been, had voters been allowed to rank as many candidates as they wished. Second, the discretion granted under law to San Francisco's Director of Elections, allowing him or her to decide whether and how much to restrict

[1]. The only other current systems of which I am aware that are similar to IRV and restrict voters' rankings are those used to elect the Mayor of London and the President of Sri Lanka.

[2]. To the best of my knowledge, no previous IRV or IRV-like system has allowed the chief election administrator to decide whether and how much to restrict voters' rankings.

voters' rankings, gives that officer the ability to affect the outcome of the election. Third, the ballot format proposed for San Francisco is more confusing for voters than the format adopted virtually everywhere else that IRV has been used.

How restricting the number of rankings can affect the outcome

Restricting the number of rankings voters can express on the ballot can change the election outcome from what it would have been, had voters been allowed to rank as many candidates as they wished. Indeed, restricting voters' options *will* change the election outcome whenever such restrictions prevent what would otherwise be "late transfers" from affecting the outcome. Before explaining what "late transfers" are, I first provide an example of how restricting voters' rankings can change the outcome (Example 1) and then briefly explain how the San Francisco system diverges from true IRV systems.

Example 1

In this example, there are five candidates, denoted with capital letters: A, B, C, D, and E. Each voter or bloc of voters ranks these five candidates in some order of preference. For example, a voter who ranks the candidates DBCAE most prefers candidate D, ranks B second, and so on. If a voter has clear preferences regarding A and B but no preference regarding C, D and E, then this voter's preference ranking would be written AB or BA (depending on which of the ranked candidates was judged best).

NUMBER OF VOTERS IN BLOC	HOW VOTERS IN THIS BLOC RANK THE CANDIDATES
8,000	ABC
9,000	BA
3,500	CDEAB
2,000	DECAB
1,000	EDCAB

In this example, there are 17,000 voters (in the first two rows) who split between the two major candidates, A and B (their votes go to A and B in the first round and there is no need for transfers). In addition, there are 6,500 (= 3,500 + 2,000 + 1,000) voters whose votes will transfer to one of the top two candidates late—after the minor candidates C, D and E have been eliminated.

If the election is held under IRV with no restriction on voters' rankings, and if each voter votes sincerely, then the actual rankings put on the ballots will correspond to the preference rankings in the table above.[3] The outcome will be a victory for candidate A. (In the first round, candidate E is eliminated and her votes redistributed, all to D. In the

[3]. A voter votes sincerely if s/he votes first for her/his favorite candidate, second for her/his second-most favorite candidate, and so on, in order of true preference.

second round, candidate D is eliminated and his votes redistributed, all to C. In the third round, candidate C is eliminated and her votes redistributed, all to A.[4])

If the election is held under IRV with each voter restricted to ranking only three candidates, and if each voter votes sincerely, then the actual rankings put on the ballots will be as in the following table:

NUMBER OF VOTERS IN BLOC	HOW VOTERS IN THIS BLOC RANK THE CANDIDATES
8,000	ABC
9,000	BA
3,500	CDE
2,000	DEC
1,000	EDC

The outcome in this case is that candidate B wins. The reason is that, after candidate C is eliminated in the third round, there is no information on the ballot regarding how the late-transfer voters rank the two main candidates. Thus, the 6,500 votes simply extinguish after the third round, and candidate B wins with 9,000 votes to A's 8,000.

By restricting the expression of preference on the ballot, a burden is imposed on voters whose votes would transfer late. They must anticipate the extinguishment of their votes and vote strategically, moving candidate A up to at least third place in their expressed rankings. If the late-transfer voters do not possess the information necessary to anticipate their predicament, their votes will be wasted in the sense that the election outcome ends up being what it would have been had they all abstained. This burden—of anticipating a relatively long chain of calculations accurately; and voting strategically in order to ensure one's vote does influence the ultimate outcome—falls only on the late-transfer voters, not on the early-transfer voters.

Note that, in this example, only 6,500 of the 23,500 voters (28%) would wish to rank more than three candidates on the ballot. Obviously, this number (and percentage) could vary, with different assumptions about the sizes of the voting blocs. Note also that the winner under restricted IRV (i.e., B) wins despite the fact that 14,500 of 23,500 voters (62%) prefer A to B. Put another way, the winner under restricted IRV, B, has the support of only 38% of the voters. This percentage could vary, with different assumptions about the sizes of the voting blocs.

Traditional run-off, IRV and restricted IRV

As Reilly and Maley note, among advocates of IRV, "it has regularly been stressed that voters should be able to choose how many preferences they will indicate."[5] To see why

[4]. The simple pattern of transfers, in which all E's votes transfer to D, and so on, is for ease of exposition only. The essential features of the example are that: (1) there are voters who rank C, D, and E, in some order, in the first three spots on their ballots; and (2) these voters are more likely to rank A fourth than B.

electoral reformers generally do not endorse restricted IRV, such as San Francisco's proposed system, let's compare it to a traditional run-off.

In the latter system, the two candidates who garner the most votes in the first round of voting meet again some weeks later in a run-off election. All voters know who the two finalists are and can vote for whichever they prefer.

In contrast, under an IRV system in which voters can rank at most three candidates, voters must correctly anticipate who the top two candidates will be in the final round of vote-counting of the IRV process, and then vote for one of them. If they fail to correctly anticipate who the top two candidates will be, they will end up casting what they may think of as a "run-off" vote for a candidate who does not in fact make it to the final round. Thus, under a restricted IRV system, some voters who would have had the opportunity to participate in the run-off under a traditional system are effectively debarred from participating in the "run-off" under IRV.[6]

As an example, consider the first-round results from the 1999 mayoral election in San Francisco: Willie Brown 39%; Tom Ammiano 25%; Frank Jordan 17%; Clint Reilly 13%; scattered 6%. Suppose this election had been held under the proposed version of IRV. Consider a voter who preferred at least three other candidates to both Brown and Ammiano. Such a voter might mistakenly have believed that the top two finishers would be Brown and Jordan. Acting on this understandable belief (Jordan had formerly served as Mayor of San Francisco), the voter might have ranked Jordan at least third, in an attempt to avoid wasting their vote. However, their effort would have gone for naught, as in fact Jordan was not one of the top two finalists. The voter's ballot would therefore have extinguished before affecting the real race (between Brown and Ammiano).

Note that IRV, in which voters are allowed to rank as many candidates as they wish, solves the "guessing" problem just noted. Any voters who wish to ensure that their votes will count in every stage of the process about which they care simply have to rank all the candidates about whom they care. If they truly do not care about the serious candidates most likely to last into the later rounds, then they can *voluntarily* withdraw from the process. If they do have preferences among the serious candidates, and are allowed to rank as many as they wish, then their votes will always count. But if they have a preference but the ballot only allows three rankings, they may be *involuntarily* removed from the set of voters whose votes count in the later and more crucial rounds.

Both true IRV (in which voters are allowed to rank as many as they wish) and traditional run-off systems seek to ensure that the candidate ultimately elected has the support of a majority of all voters. They do this by conducting one or more run-off elections, either in

---

[5]. Reilly, Benjamin, and Michael Maley. 2000. "The Single Transferable Vote and the Alternative Vote Compared." In *Elections in Australia, Ireland and Malta Under the Single Transferable Vote*, ed. Shaun Bowler and Bernard Grofman. Ann Arbor: University of Michigan Press, p. 43.

[6]. The Electoral Reform Society's web site also notes this feature in connection with the method used to elect the Mayor of London (www.electoral-reform.org.uk/publications/leaflets/london.htm, accessed May 8, 2003).

the traditional sense of holding a separate election some weeks later or in the sense of using voters' rankings to hold the run-off "instantly." In restricted IRV, however, some voters may unwittingly remove themselves from participating in the run-off, simply by virtue of guessing wrong. Thus, a core value of the whole IRV system—that of ensuring the election of a candidate with majority support among all voters who wish to participate—is violated (as Example 1 showed).

Defining early- and late-transfer votes

Having provided an example of how restricted IRV can go wrong and contrasted it to two more common systems—traditional run-off and unrestricted IRV—I turn now to a more technical discussion of the situations under which restricting voters' rankings will affect the outcome. This returns the discussion to the issue of early and late transfers.[7]

To explain what early and late transfers are, consider a race in which there is no majority winner until all but two candidates are eliminated. In an unrestricted IRV election, the final vote for the ultimate winner can be divided into two mutually exclusive categories: (1) the *early-transfer vote*, defined as the number of votes that come from ballots ranking the ultimate winner first, second or third; and (2) the *late-transfer vote*, defined as the number of votes that come from ballots ranking the ultimate winner fourth or lower. Restricting voters to three rankings prevents what would otherwise be the late-transfer vote from being counted, if voters mark their ballots the same way in both the unrestricted and the restricted election. If some voters do figure out that they should change the way they mark their ballots in the restricted election, it is still the case that not all of the late-transfer votes will be counted, unless *every* late-transfer voter *correctly* anticipates who the two leading candidates after the third round will be and ranks one of them in the top three on the ballot. Failure to count the late-transfer vote fully, however, will alter the election outcome whenever the late-transfer vote differs "enough" from the early-transfer vote.

To clarify what is "enough," consider an election in which voters are not restricted and suppose that the first-place candidate gets E1 early-transfer votes ("E" for "early" and "1" for the "first-place" candidate) and L1 late-transfer votes in the final round. Meanwhile, the second-place candidate gets E2 early-transfer votes ("2" for the "second-place" candidate) and L2 late-transfer votes in the final round. Since the first-place candidate by definition wins, we know that $E1 + L1 > E2 + L2$ (that is, the ultimate winner gets more votes than the ultimate runner-up, in the final count). Restricting voters to three rankings can change the election outcome whenever $E1 < E2$. In other words, the election outcome can be changed whenever the second-place candidate has a lead among the early-transfer voters (those ranking at least one of the top two candidates in the first

---

[7]. In this section, I consider the conditions under which restricting voters to three rankings can change the outcome. Restricting voters to n rankings can also change the outcome, whenever n is less than one less than the number of candidates.

three ranks on the ballot) but trails when the late-transfer voters (all the rest) are also counted.[8]

The late- and early-transfer votes are likely to differ whenever one side of the left-right political spectrum is more divided, in the sense of fielding more candidates and splitting its vote more widely, than the other. Suppose, for example, that there are 10 leftist candidates for mayor but only 3 centrist and right-of-center candidates. The "real" race is between leftist candidate A and centrist candidate B. The centrist and right-wing votes will transfer to B early, as there are only three candidates from the center and right on the ballot. Thus, the vast bulk of the late-transfer vote will come from leftist voters. If the early-transfer vote gives a close victory to the centrist, but counting all votes would elect the leftist, then preventing the late-transfer vote from being counted will change the outcome.

Example 2

An example illustrating how restricting voters to three rankings can change the election outcome when one side of the political spectrum is more divided than the other is based on the 1999 race for District Attorney in San Francisco. This race featured five candidates in the first round of what was then a traditional run-off process: Gonzalez, Castleman, Hallinan, Fazio, and Schaeffer. The vote totals in the first round were 20,153 for Gonzalez, 17,677 for Castleman, 68,424 for Hallinan, 67,145 for Fazio, and 5,614 for Schaeffer (for a total of 179,013 ballots). In the run-off election, Hallinan beat Fazio by 1,820 votes.[9]

What if San Francisco had used IRV with voters restricted to three rankings in 1999? Could the outcome have changed? The answer is "yes."

To see how, suppose that 5,000 voters ranked Gonzalez, Castleman and Schaeffer, in some order, in the top three places on the ballot. This would represent only 2.79% of the total 179,013 ballots cast. Suppose further that, if given the opportunity, 70% of these voters would have ranked Hallinan fourth, while 30% would have ranked Fazio fourth. (This figure seems plausible, given that both Gonzalez and Castleman were to the left politically of Hallinan, with Fazio to the right. Thus, voters who ranked *both* these candidates in their top three would be considerably more likely to rank Hallinan above Fazio than the reverse.) Suppose finally that, had the election been held under an unrestricted IRV procedure, Hallinan would have beat Fazio by the same margin that he did in the actual run-off (*viz.*, 1,820 votes). Given these three assumptions, restricted IRV would have elected Fazio, not Hallinan. The reason is that, under restricted IRV, the 5,000 ballots ranking neither Hallinan nor Fazio in the top three would extinguish, depriving Hallinan of 5,000×.7 = 3,500 votes and depriving Fazio of 5,000×.3 = 1,500

---

[8]. We say "can" change the outcome, rather than "will," only because it is possible that late-transfer voters will anticipate the extinguishment of their ballots under restricted IRV and take effective action to avoid that outcome. As explained above, however, correct anticipation may be difficult.

[9]. For the data on the S.F. District Attorney's race: http://www.sfgov.org/site/election_page.asp?id=5877, accessed May 9, 2003.

votes. Relative to Hallinan, then, Fazio gains 2,000 votes, which is sufficient to overcome the vote margin by which he lost (1,820).

Obviously, other assumptions can also generate this outcome. For example, if 10,000 ballots ranked Gonzalez, Castleman and Schaeffer, in some order, in the top three places on the ballot; and 60% of them would have ranked Hallinan fourth while 40% would have ranked Fazio fourth; then again restricted IRV would have elected Fazio, rather than Hallinan.

How likely is it that restricting the number of rankings might affect the outcome?

How likely is it in practice that the election outcome in an IRV election will be changed by restricting voters to three rankings? One way to shed light on this question is by looking at the Australian experience with IRV. The Australians have used IRV in elections to their federal House of Representatives since 1919 and they sometimes report election results in a way that allows one to see the vote counts at each stage of the process—the initial count, the second count (after some candidate has been eliminated and his/her votes distributed), and so on to the final count.[10] In some cases, the information provided conclusively shows that the winner of the election would have won, even had voters been restricted to three rankings. In other cases, the possibility remains that the election outcome would be changed by restricting voters to three rankings.

95 %

When the ranking restriction will *not* affect the outcome

Let's consider first the situations in which the outcome would *not* change. The simplest such case is one in which a particular candidate wins an outright majority of first-preference votes and is therefore elected at the first count. Restricting voters' rankings in such situations will not affect the outcome, as the first-place rankings alone suffice to determine the outcome.

Other cases in which there would be no change in outcome involve very strong candidates who win by a lot but not necessarily at the first round. An example is the Barton, NSW district in the 10 November 2001 elections. The results from this election are displayed in Appendix 1. As can be seen, the ultimate winner, Robert McClelland, has 35,871 votes in the first count. This exceeds the total number of votes garnered by the second-place candidate in the final round of voting (viz., 32,873). Thus, even if the election had been re-counted, with only voters' first three preferences allowed into the count, McClelland would have won by at least 35,871-32,873 = 2,998 votes. In fact, it can be deduced that he would have won by considerably more than this sum but all that matters for present purposes is that restricting voters' options in this particular race would not have affected the outcome.

---

[10]. I accessed Australian data for their 2001 election from psephos.adam-carr.net/psehpos/index47.html, on May 1, 2003, and May 20, 2003.

When the ranking restriction *could* affect the outcome

Cases in which the outcome *could* be affected by adopting San Francisco's three-ranking restriction are those in which the ultimate winner's vote at the end of the third count is smaller than the ultimate second-place finisher's final total (and voters vote sincerely). To explain why this condition works, recall that, if the ultimate winner gets fewer early-transfer votes than the ultimate runner-up (i.e., E1 < E2), then the three-ranking restriction *will* change the outcome. This is because the three-ranking restriction effectively prevents the late-transfer votes from being put on the ballot and, hence, prevents them from being counted.

To identify cases in which the outcome could be changed, then, we need to identify election outcomes in which it is possible that E1 < E2. As shown in Appendix 3, E1 must be at least as great as the ultimate winner's vote at the third count and no greater than his/her vote at the final count. Similarly, E2 must be at least as great as the ultimate runner-up's vote at the third count and no greater than his/her vote at the final count. Thus, it is possible that E1 < E2 whenever the ultimate winner's third-count vote total (the smallest E1 could be) is less than the ultimate runner-up's final vote total (the largest E2 could be).

As an example consider the results from Adelaide, South Australia, in the 10 November 2001 elections, displayed in Appendix 2. There were a total of 77,513 valid ballots cast in this election, of which 34,258 ranked the ultimate winner, Worth, first; while another 28,732 ranked the ultimate second-place finisher, Stanley, first. Thus, there were 77,513 – 34,258 – 28,732 = 14,523 ballots that we know did not rank either of the top two candidates first. In the second round, Worth and Stanley garnered an additional 598 and 457 votes, respectively, from voters who had ranked Peacock first. These ballots thus clearly rank at least one of the top two candidates in second place. In the third round, Worth and Stanley garnered an additional 519 and 1,635 votes, respectively, from voters who had ranked Osborn first (or had ranked Peacock first and Osborn second). These ballots thus clearly rank at least one of the top two candidates somewhere in the top three places.

Are there any other ballots from Adelaide that we can be sure ranked at least one of the top two candidates in one of the top three places? As it turns out, no. All the other transfers to these candidates could result from information contained in the fourth- and fifth-place rankings. In particular, the 11,314 votes that transfer from Mann in the final count could all come from ballots that rank Mann, Osborn and Peacock (in some order) in the first three places. There is nothing in the election results that allows us conclusively to discount this possibility.

Since we do not have access to the actual ballots cast in Adelaide, the most we can say about the number of early-transfer votes that Worth gets is that it is at least his third-round total (of 35,375) and at most his final-round total (of 38,928). Similarly, we know that Stanley's early-transfer vote total, E2, lies somewhere between 30,824 and 38,585. It is thus quite possible, given the published election summary, that E1 < E2, in which case imposing a three-ranking restriction on voters could have changed the outcome.

How many of the Australian outcomes could possibly have been changed, had the ballots restricted voters to three rankings? That is, how many are such that the ultimate winner's third-round total falls short of the ultimate runner-up's final-round total? In the November 2001 elections, there were 32 such cases, out of 150 total districts. Thus, 21.3% of the election outcomes could conceivably have been changed, had the Australians used the proposed S.F. restriction on voters' rankings.

There are two considerations that make this figure loom larger in San Francisco. First, note that the likelihood of the election outcome being changed is greatest in the most closely contested elections, in which the perceived fairness of the electoral process is most crucial. Second, note that all of the Australian constituencies are substantially smaller in population than is the City of San Francisco, with about 70,000 to 85,000 votes cast as opposed to over 200,000 in the 1999 mayoral election. They are also substantially less diverse. Both the size and diversity of the S.F. electorate, compared to the Australian electorates, suggest that the chances of large fields of candidates splitting the vote finely are greater in S.F., which in turn raises the probability that vote restrictions would affect the outcome.

How discretion in deciding whether to restrict the number of rankings can affect the outcome

The previous sections have illustrated the conditions under which restricting the number of rankings voters are allowed to express on the ballot can affect the outcome of the election. This section notes that, because the San Francisco Charter grants discretion to the Director of Elections in determining whether and how much to restrict voters' options, it opens the door to potential manipulation of election outcomes.

Suppose that the next mayoral race in San Francisco features two main candidates, one more liberal (L), one more conservative (C), along with a smattering of minor leftist candidates (A, B, C, D and E). The race, let us say, is relatively close between the top two contenders, L and C. Moreover, suppose also that (1) many of the minor leftist votes would transfer between the minor leftist candidates at least three times before reaching a major candidate, if the minor leftist supporters simply rank candidates on the ballot in order of their true preferences; and (2) a substantial majority of the minor leftist voters prefer L to C. If the incumbent Director of Elections becomes aware of these features of the election—and we shall argue that he is in an excellent position to do so—then he faces a moral temptation, as he could have a clear political incentive either to limit or not to limit.

If the Director favors C, then his political goal (electing C) is best served by imposing the strictest allowable limit on voters' options. This imposes a burden on leftist leaders, who must reach their voters and convince as many as possible that they should rank L at least third, even if their true preferences would dictate otherwise. To the extent that leaders are unsuccessful in reaching all leftist voters with this message (and convincing them), a certain number of leftist voters will cast ballots that exhaust before they transfer to either L or C, and their votes will have no bearing on the ultimate outcome. The decision to

limit voters' options thus helps C, in this case, as it deprives L of a certain number of votes that exhaust (due to the restriction) before they reach him/her.

If the Director favors L, then his political goal (electing L) is best served by imposing no limit on voters' options. This will minimize the number of ballots cast by minor candidates that exhaust before they transfer to one of the major candidates. It will accordingly benefit L, as he is the likely net beneficiary of such transfers.

Note that the example just given is a simply a stylized version of recent mayoral elections in San Francisco. The conditions that need to be met—two main candidates, one more and one less conservative; a smattering of minor candidates, mostly on the left—are already met. Moreover, the Director of Elections will have a good idea about how many candidates there are likely to be, when he must decide whether to restrict voters' rankings; and it is likely that he will be well informed about the political leanings of each candidate. Even with little effort, therefore, he will be in a position to estimate which of the major candidates will be the net beneficiary of late transfers. With some investigation—say a proprietary poll of the San Francisco electorate—the size of the advantage or disadvantage could be calculated more finely.

To clarify how a candidate with leverage over the Director of Elections might proceed, if unethical, imagine the following. A candidate for mayor commissions a poll of the S.F. electorate shortly before the Director of Elections must decide whether to restrict voters' options or not. The poll asks each respondent to rank as many of the candidates for mayor as s/he wishes, just as s/he would be asked at election time if no restrictions were imposed. Using the poll responses, one can calculate the early- and late-transfer vote for each of the two main candidates. The early-transfer vote for a candidate is simply the sum of all votes that accrue to that candidate by the third round of counting; the late-transfer vote for a candidate is the sum of all votes that accrue to that candidate *after* the third round of counting. Suppose that T percent of late-transfer votes go to the unethical candidate, while 100-T percent go to his or her main rival. Since this is the result of a limited sample of S.F. voters, there will be some uncertainty regarding whether, in the full electorate, the same percentage breakdown would occur. Using standard statistical tests, however, one could classify the poll results as follows: T is significantly below 50%; T is significantly above 50%; or neither. In the first case, allowing the late-transfer votes will likely harm the unethical candidate's chances of winning and s/he should therefore suppress them by instructing the Director of Elections to restrict voters' options. In the second case, allowing the late-transfer votes will likely aid the unethical candidate's chances of winning and s/he should therefore facilitate them by instructing the Director of Elections not to restrict voters' options. In the third case, the late-transfer votes will neither clearly harm nor clearly aid the unethical candidate, and so s/he would be indifferent regarding whether the Director of Elections restricted voting options or not.

It is an easy matter to estimate what proportion of late-transfer votes a given candidate will get, based on proprietary polling information. It is also an easy matter for candidates to see that, if they are likely to get less than 50% of the late-transfer vote, then they should want to suppress it, while if they are likely to get more than 50%, they should

want to facilitate late transfers. The *only* IRV or IRV-like system of which I am aware, past or present, that actually allows candidates potentially to do something about their preferences for allowing or disallowing late-transfer votes is San Francisco's.

Whether S.F. politicians exploit their opportunity or not remains to be seen but it can be noted that politicians in other systems using IRV have long recognized the importance of deciding how many rankings voters can or must express. In Australia, which has used IRV to elect members of its House of Representatives since 1919, it is the right-wing parties that have historically been split (into the National and Liberal parties). The rightist parties have consistently supported Australia's requirement that voters rank *all* candidates on the ballot. The reason for their support is that mandatory ranking of all candidates ensures, in districts where both a National and a Liberal candidate run, that all votes will transfer. Both parties issue how-to-vote cards to their supporters, instructing them to vote National first, then Liberal (or vice versa). In this way, neither party pays a cost from running separate candidates. Recognizing the value of mandatory ranking to its opponents, the Labour party in Australia has often (although not always) opposed the provision or argued for a return to plurality voting. Had the party in government in Australia been given the right to decide whether rankings were mandatory or not, then one would probably have seen a strong pattern (Labour removing the requirement, National/Liberal imposing it). Similarly, in San Francisco, one might find Directors of Elections who favor the more conservative leading candidate imposing a limit on how many rankings voters can list on the ballot, while Directors favoring the more liberal leading candidate impose no such limits.

A morning-after scenario in San Francisco

Consider the following morning-after scenario for the next San Francisco mayoral election, if it is held under the proposed restricted-IRV system. At the final vote count, the ballots will be sorted into five mutually exclusive categories. There will be F1 ballots that, in the final count, provide votes for the winning candidate; F2 ballots that, in the final count, provide votes for the runner-up; X1 ballots that extinguished and ranked only one candidate; X2 ballots that extinguished and ranked exactly two candidates; and X3 ballots that extinguished and ranked exactly three candidates. All valid ballots will fall into one of these five categories.

Suppose that $F1 < F2 + X3$. It will then be possible for the losing candidate and his/her supporters to argue that, had voters not been restricted to three rankings, the outcome of the election would have been different. If F1 and F2 are nearly equal, and X3 substantially greater than the margin of victory, $F1 - F2$, then the argument would be plausible. Unfortunately for public confidence in the electoral process, there would be no way to prove or disprove the claims made on behalf of the losing candidate. The legitimacy of the victor's claim to office would thus be permanently in doubt.

Ballot format

Another unusual feature of the proposed San Francisco IRV process is the ballot format. Most IRV systems employ a very simple ballot. The Australians, for example, list each candidate's name once with a box next to it, and the voter is instructed to write the

number 1 in the box next to his first choice, the number 2 in the box next to his second choice, and so on (see Exhibit 1). The Irish, too, use this format in their by-elections (see Exhibit 2). Another format (Exhibit 3) lists the names once and invites the voter to indicate whether they wish to vote for this candidate "as their first choice," "as their second choice," and so on. San Francisco, in contrast, proposes to print all candidates' names three times. The voter, if they read and understand the information at the top of each block of names, then votes three times: the first time for their first choice, the second time for their second choice, and the third time for their third choice. The S.F. ballot thus has far more words, and more confusing words, than does the Australian or Irish ballots—where the IRV system has been used the longest.

San Francisco's ballot format will presumably increase the ballot spoilage rate and the mayoral "roll-off" rate for all voters.[11] Previous research, however, shows that the impact of confusing ballot formats is greater on the less educated, minority and non-English-speaking segments of the population. Kousser, for example, explains how the introduction of a complex eight-ballot procedure in South Carolina in 1882 was intended, and had the effect of, disfranchising illiterates (with the primary target being former slaves).[12] More recently, Darcy and Schneider show how more confusing ballots widened the gap in participation between blacks and whites in Oklahoma in the mid-1980s.[13] Given these previous findings—and common sense—one would expect San Francisco's peculiar ballot format for the 2003 IRV election to have the effect of diluting minority groups' influence on the outcome.

The reason that San Francisco chose such an unusual ballot format is presumably that the election administrators feared that a hand-count of ballots in which voters could rank as many candidates as they wished would be technically overwhelming. There are two points to make in this regard.

First, the Papua New Guineans successfully administered a hand-counted system nationwide from 1964-72, while allowing voters to express full preferences. Similarly, the Irish have successfully administered hand counting of by-election results since the 1920s, again allowing voters to express full preferences. In both nations, there were sometimes fields of candidates as large as that expected in the San Francisco mayoral race. For example, in the Dublin North by-election of 23 October 1998, there were 18 candidates and it took 14 counts before the election was decided. Thus, it is technically feasible to hand count IRV elections, even when voters are allowed to express as many rankings as they wish.

---

[11]. When a voter shows up at the polls, votes for some offices or propositions on the ballot, but does not vote in the mayoral race, they are said to "roll off" the ballot for that particular race. Put another way, one calculates the number of all voters who do cast valid votes in the mayoral race, and divides this by the number of all voters who cast valid ballots (with a valid vote for at least one item on the ballot). The inverse of this percentage is the roll off rate for the mayoral election.

[12]. Kousser, J. Morgan. 1974. *The Shaping of Southern Politics: Suffrage Restriction and the Establishment of the One-Party South, 1880-1910.* New Haven: Yale University Press, p. 50.

[13]. Darcy, R., and Anne Schneider. 1989. "Confusing Ballots, Roll-of, and the Black Vote." *Western Political Quarterly* 42(3):347-64.

Second, the fear of a technically difficult task will be realized only in the circumstances in which restricting voters' rankings is most likely to affect the outcome. Suppose that voters are allowed to rank as many candidates as they wish but so few rank more than three that the outcome is the same as it would have been had they been restricted to three rankings. In this case, the administrative difficulty of counting the ballots is not in fact much greater with full ranking than it would be with restricted ranking. If, on the other hand, the administrative difficulty of counting the ballots is greater with full than with restricted ranking, this can only be because many voters would choose to rank more than three, in which case restricting them is likely to affect the outcome. Thus, the counting costs can be avoided only at the likely expense of affecting the outcome.