Title
Overvoting and the Equality of Voice under Instant-Runoff Voting in San Francisco

Journal
California Journal of Politics and Policy, 7(4)

Authors
Neely, Francis
McDaniel, Jason

Publication Date
2015

DOI
10.5070/P2cjpp7428929

Peer reviewed

THE CALIFORNIA *Journal of*
Politics & Policy

Overvoting and the Equality of Voice under Instant-Runoff Voting in San Francisco

Francis Neely and Jason McDaniel
San Francisco State University

The controversy surrounding the 2000 U.S. presidential race fueled a variety of efforts to improve the administration of elections. Activists, benefiting from that momentum, pushed for reform and found some purchase at the local level in San Francisco, California. Proposition A passed in a 2002 March primary and replaced a two-round runoff system with instant-runoff voting (IRV).[1] Since then IRV has been used to elect their local officials.

As the largest and longest-running application of IRV in the States, this serves as both a vanguard on the reform front and a test case for interested parties.[2] One concern in the discussion of any electoral reform is how well the public will understand a new system and what that implies for the equality of political voice. This is our focus. It is a question that continues to echo from the sidewalk cafes to the corridors of city hall in San Francisco. Concerns about the fairness of IRV led at least four jurisdictions to repeal similar reforms shortly after enacting them: Burlington, VT (2006–2009), Cary, NC (2007–2009), Pierce County, WA (2006–2009), Aspen, CO (2009). As the debate on long-term status of IRV in San Francisco continues, efforts to amend or repeal it persist, often citing doubts about its facility of use and functional transparency.

An earlier study found some patterns of use that raised concerns about the evenness of voters' experience with the IRV ballot, but that was a mere three years into the experiment (Neely and Cook 2008). In this paper we aim to do three things: the first is to extend the analysis of IRV over time to include more years and thus more variation in the nature of the elections. The second is to increase the scope of analysis with a comparison to voting tendencies on other portions of the ballot, and the third is to improve the quality of the inferences we can draw by using a more precise estimator. Our fundamental question centers on equity—are voters from varied backgrounds able to express themselves equally on the IRV ballot?

This is an especially useful case in which to examine the impact of election system complexity. Known for its left-leaning ideological orientation, the community is highly varied in other

[1] The city and county of San Francisco are consolidated, and Prop. A amended the charter. A note on terminology: we use the term instant-runoff voting, as did the original ballot measure. Before the system was implemented in San Francisco the Department of Elections changed the name to ranked-choice voting in order to avoid public expectations that the results would be instantaneous on Election Night. In academic comparative studies IRV is one of several types of preferential systems and is called the alternative vote.

[2] Cambridge, Massachusetts has used a single transferable vote (STV) system since 1941 that also requires voters to rank candidates. Unlike IRV it awards seats in multimember districts according to the proportion of votes cast for candidates.

ways. Voters come from a broad range of life experience, identities, countries of origin, and political sophistication. In addition, the San Francisco Department of Elections makes available individual-level data on the IRV portion of the ballot including an indicator of vote-by-mail versus polling-place ballots in recent years (from 2008 on). They also report counts of undervotes and overvotes at the precinct level for non-IRV contests. This prominent case can supply valuable insights on voting behavior generally and help address questions about IRV specifically as other jurisdictions consider this reform.[3]

Following a brief description of San Francisco's IRV system and a review of what we know so far about various elections systems and uncounted votes, we turn to the empirical evidence. We find in the aggregate that much of voters' behavior in San Francisco is fairly regular. That is, the trend that was found earlier—of fewer and fewer voters bothering to rank candidates—has reversed. The rate of IRV overvotes, meanwhile, varies and falls within the range of overvotes in other portions of the ballot. Our more rigorous tests reveal that all overvotes are not created equal. Voided ballots are consistently more common in precincts where more African-American citizens reside, and are often observed at higher rates in precincts that contain more Latino, elderly, foreign-born, and less wealthy residents. Although this is a real concern for the democratic process, the evidence suggests it is not IRV *per se* but rather ballot complexity more generally that leads to such discrepancies in whose votes get counted. In our conclusion we discuss the implications of the IRV reform and assess its impact on the equality of political voice.

Instant-Runoff Voting, San Francisco Style

San Francisco's version of IRV allows voters to rank up to three candidates. All first-ranked choices are counted and a winner is declared by a simple majority. If no majority winner emerges, then the candidate with the fewest first-ranked votes is eliminated and on those ballots the second-place candidate replaces the eliminated one as the voter's new first choice. Ballots are recounted, and if no candidate gets a majority of the newly defined first-place votes then the elimination rounds continue until one candidate obtains a majority of the remaining first-place votes.

Two characteristics of San Francisco's system are notable: ranking candidates is optional, and voters can rank no more than three candidates. The first is straightforward and reflects the type of system promoted in Prop. A; the second is a function of practical choice. Prop. A's authors anticipated the challenge of recording IRV votes, especially in elections with many candidates competing. Prop. A, therefore, included wording that allows the city/county to limit the number of candidates ranked in order to accommodate existing voting machinery, as long as long as voters can rank at least three (City/County Charter Section 13.102[b]). In 2011 the system survived a legal challenge in the U.S. Ninth Circuit based on the constitutionality of truncating voters' preferences (*Dudum v. Arntz*).[4]

---

[3] Those who adopted IRV after San Francisco and currently use it to elect government officials include Berkeley, CA; Minneapolis, MN; San Leandro, CA; Oakland, CA; St. Paul, MN; and Tacoma Park, MD.

[4] Another aspect of the system is the distinction between what are called *exhausted* and *continuing* ballots. Because voters are not required to rank every candidate, and are not allowed to in races of more than three, some ballots are exhausted and not included in the final count—the candidate(s) chosen has been eliminated. Only continuing ballots—those with votes for someone still in the race—are counted in the denominator that defines a majority winner. This often leads to winners who have the majority of continuing ballots but a plurality of votes cast for that of-

Figure 1 displays San Francisco's IRV ballot; it is single-sided and provides arrows that voters mark with a pen or pencil. For polling-place voters it is scanned at the precinct with the voter present. It is usually one of several ballot papers that voters mark. Optical-scanners were in place when the IRV reform passed, and these ballots were designed to fit that system. Generally, the optical-scan method has been shown to produce moderate (Bullock and Hood 2002, Knack and Kropf 2003; Sinclair and Alvarez 2004) to relatively low rates of uncounted votes (Ansolabehere and Stewart 2005; Kimball, Owens, and Keeney 2004).

In some ways, San Francisco's version of IRV is forgiving. Voters can make two types of mistakes without their ballots being voided. First, some errors of omission are accepted. Voters who find only one candidate worthy of their vote could mistakenly leave the first column blank and mark the ballot in the second or third column for that one candidate. The vote would still be counted. Ballots are advanced to the next column when either or both of the first two columns are left blank. Second, voters can incorrectly make duplicate marks without spoiling the ballot. If someone misunderstands the directions and votes for the same candidate in all three columns (or in any two) then the vote in the first marked column is counted and the subsequent redundant vote is ignored. What is not allowed, however, is marking more than one candidate in a single column; ballots like this are considered spoiled, and are what we call overvotes.

A final note on errors—Election Day voters benefit from optical scanners at the precinct polling place that reject any ballot with an error. If that happens, a voter is given an opportunity to change the ballot, get a new one and fill it out, or resubmit the ballot as is. In this way, errors of any type are allowed in the first scan of a ballot without a penalty to voters. Of course, this only applies to voters who turn out in person and not those who vote by mail. In recent years roughly half of the ballots are vote-by-mail.

Literature and Hypotheses

Our question is about representation. The three-column IRV ballot is more complex than plurality or two-round systems (TRS) where voters choose a single candidate from one list. The decision task is also more onerous—ranking one's top three candidates requires more information and cognitive effort than picking a single candidate (Lau and Redlawsk 2006). Are less educated and less informed voters at a disadvantage in the more complex system? Are voters whose first language is not English at a disadvantage as they attempt to understand how to cast their vote? Do cultural differences across varied communities lead to differences in how well voters understand the more complicated system, and thus differences in how well they might express themselves on the ballot?

We turn to studies of local election reforms in the U.S., the impact of ballot formatting, and analyses of the residual vote to inform our expectations. In addition, we draw on the results from a 2008 study mentioned above that examined San Francisco's first few IRV elections (Neely and Cook 2008). While another important literature on preferential voting analyzes the comparative effects in parliamentary systems on the proportionality of outcomes or the number of parties (Farrell and McAllister 2003; Lijphart 1997; Reilly 2004; Zimmerman 1994), narrower questions like ours are rare in that literature (but see Jansen 2004, and Reilly and Maley 2000). While the

fice (Burnett and Kogan 2012). The alternative is to require all voters to rank all candidates, as Australia does. Consideration of this aspect of San Francisco's IRV system is beyond the scope of our analysis.

Figure 1. Sample IRV Ballot from the 2004 Election District 5 Board of Supervisors

Note: The ballot image (Dennis 2004) is cropped for legibility; the full District 5 ballot listed 22 candidates in each column.

broader question of how IRV might affect the descriptive and ideological representation in municipal institutions is also important, that lies beyond the scope of our study.

Further, much of the literature on the residual vote lacks the data necessary to distinguish between various types of uncounted ballots. We rely more heavily on the few studies that do identify both undervotes (ballots unmarked for a given item) and overvotes (disqualifying multiple marks on the ballot paper). Our focus is on overvote errors because the consequence of making such an error is a disqualified ballot, and because they offer a clearer indication of voter intent. While an undervote—i.e., the IRV portion of the ballot left blank—may result from a voter accidentally overlooking that ballot paper, it could also reflect a conscious choice to skip the race. An overvote, by contrast, clearly indicates that the voter intended to voice a preference in that contest. However, by marking the ballot in a way that disqualifies it, the voter's preference is not recorded. We consider these mistakes more vital, and the degree to which they are unevenly distributed across types of voters a clearer liability in democratic elections. Such inconsistencies compromise the equality of voice among voters.

First, let us consider the findings from the earlier analyses of San Francisco's IRV elections that best match results in the broader literature. A relatively higher rate of overvotes resulting in spoiled ballots was found in precincts with more black residents (Neely and Cook 2008). This follows work that has examined the residual vote (Bullock and Hood 2002; Darcy and Schneider 1989; Kimball et al. 2004; Kimball and Kropf 2005) and overvotes, in particular (Sinclair and

Alvarez 2004).[5] Higher counts of overvotes were also found, at times, among San Francisco communities with more Latino residents (Neely and Cook 2008), something shown in a similar analysis of voters in Los Angeles (Sinclair and Alvarez 2004), and in areas with more foreign-born residents. Meanwhile, although the Los Angeles study (2004) showed consistently higher rates of overvotes among Asian-American communities, no such evidence emerged in the San Francisco analysis (Neely and Cook 2008). We expect, then, that our results will reveal consistently higher rates of overvotes in areas with more African-American residents, at times more overvotes in Latino neighborhoods and those where more foreign-born voters reside. We do not expect more overvotes in communities with more Asian residents.

Women have been shown to be more likely than men to file uncounted ballots (Stiefbold 1965) and to overvote in elections (Sinclair and Alvarez 2004), possibly a function of lower levels of political knowledge or interest (Carpini and Keeter 2000). However, because those differences were rare in the earlier IRV analysis (Neely and Cook 2008) we expect there to be no significant difference in overvotes between areas with more male or female residents. Other factors have been shown to matter in the rate at which voters cast uncounted ballots: elderly voters (Darcy and Schneider 1989; Kimball and Kropf 2005; Stiefbold 1965), less educated voters (Walker 1966; Bullock and Hood 2002), and poorer voters (Darcy and Schneider 1989; Kimball, Owens, and Keeney 2004; Knack and Kropf 2003) have been identified as more likely to cast a ballot that is not counted. However, the prior IRV study found few of these relationships. Recall that these other studies examine uncounted votes, and not overvotes, specifically. In the IRV analysis of overvotes, education and income were consistently unrelated to those counts (Neely and Cook 2008). In half of the tests, however, the rate of elderly voters in an area was related to the number of overvoted ballots. We expect that pockets of elderly residents in the city may have higher rates of overvoted ballots, and that the two factors of education and income will be unrelated to the counts of such ballots.

To summarize, we expect that certain groups in the electorate will be more likely to file overvoted ballots than others. Those include African Americans, Latinos, the foreign born, and the elderly. Meanwhile, we expect no such differences among Asian Americans, or between men and women. Finally, we expect no differences based on one's level of education or income.

In addition to the characteristics of voters, we might expect overvotes to occur for other reasons. Two that we can test include the length of the slate of candidates running, and the place of voting—either at the polling place on Election Day or on a vote-by-mail (VBM) ballot. Analyses of uncounted votes have shown that ballot complexity can cause voter fatigue and confusion (Bullock and Dunn 1996; Walker 1966). Choices may be affected by the sequence of the names listed (Hecock and Bain 1956; Krosnick, Miller, and Tichy 2004) or the general layout of the information on the ballot paper (Bullock and Hood 2002; Darcy and Schneider 1989; Kimball and Kropf 2005; Walker 1966). The previous study of IRV in San Francisco showed that the counts of overvotes were consistently and positively related to the number of candidates on the slate (Neely and Cook 2008). However, it did not examine the impact of vote-by-mail versus polling place ballots.

We expect that longer lists of candidates will increase the likelihood of voters making errors while marking the ballot papers. Our expectations are less clear on whether such errors will be more likely to occur on ballots cast at the polling place on Election Day than on VBM ballots. It seems reasonable to expect that VBM voters may take more time and mark their ballots in a

[5] Others, however, have found little or no difference between Black voters and others once the type of voting equipment is taken into account (Tomz and Van Houweling 2003).

more comfortable and controlled setting than those who use the polling places. The extra time and the familiarity of their surroundings may lead to fewer disqualifying errors. However, polling-place voters who do make mistakes on their ballots are made aware when they stand next to the scanning machine and the paper is rejected. At that point, these voters can check the ballot and make changes if they desire. In this way, we might expect polling place voters to be less likely to overvote on the IRV portion of the ballot. These competing considerations lead us to refrain from hypothesizing on this factor and instead to take an exploratory stance on the question.

We have three goals: The first is to extend the previous analysis over time. That study looked at the first three years of IRV elections (2004, 2005, and 2006). We replicate that below and then add all six elections that used IRV up through 2011. Importantly, this includes two mayoral races, one of which was competitive. Second, we expand the analysis to compare IRV voting behavior to other portions of the ballot for which we have similar data—a judicial election, a U.S. Senate race, and a Board of Education contest, all in the November 2010 election. Third, as detailed below, we improve on the methodological approach used in the previous analysis of IRV elections by using a multi-level mixed-effects logistic regression estimator. This replaces a negative binomial regression used in the earlier study that required all data to be aggregated at the precinct level; our approach takes advantage of the individual-level ballot data.

From this new perspective we address the question of whether disqualifying mistakes are evenly distributed among types of voters. We then ask if those who tend to overvote on the IRV portion of the ballot are similar to those who make similar mistakes elsewhere. After describing our data and methods we report the results and then discuss the findings.

Data, Measures, and Methods

Our data come from the San Francisco Department of Elections and the U.S. Census. The Statement of the Vote provides aggregate data at the precinct level for non-IRV contests. Individual-level data are reported in a file called the *ballot image* for each IRV election that does not produce a majority winner in the first round. The ballot image file contains three categories of information for every ballot cast—a record of what the voter did in the first, second, and third columns on the IRV section. It also includes information on whether the ballot was cast on Election Day or filed as a vote-by-mail ballot plus the voting precinct number. We use these ballot image data to build the overvote dependent variable central to the analysis. Our operationalization identifies only those overvotes that disqualify the ballot. This differs from past work that counts any overvote (Sinclair and Alvarez 2004; Neely and Cook 2008). Because our concern is the voice that speaks but is not heard in an election, we require both the error and the disqualification based on that error. In addition to overvotes, we aggregate the ballot image data to indicate how many voters ranked three candidates, ranked two candidates, or simply chose one in any given election.

For the comparison to other portions of the ballot we rely on precinct-level reports provided in the Statement of the Vote. Overvotes in these dependent variables are counts per precinct in the 2010 elections for Superior Court Judge, U.S. Senate, and Board of Education. These contests are chosen to represent varying levels of complexity and a range of placement across the ballot cards. A simple task placed down the ballot was the election of the judge, in which two candidates ran and voters chose one. A similar race that appeared early on the ballot was the U.S. Senate contest in which six candidates ran (Barbara Boxer was re-elected). Next, we include the more complex task for Board of Education (BOE) that allowed voters to mark up to three candidates from a list of the 11 running; this is the item closest to what voters experience on the IRV ballot. The Superior Court seat and the BOE item appeared together in that sequence on the ballot. From these cases we can compare IRV to non-IRV overvoting behavior based on the complexity of the task and also consider the effect of ballot placement.[6]

We analyze all 10 IRV contests using a multilevel mixed-effects logistic regression in which the dependent variable is an individual-level indicator of an overvote (coded 1) or not (0).[7] The hierarchical structure of the ballot image data (individuals clustered within voting precincts for ten separate elections) calls for the use of a multilevel modeling strategy in which individual overvotes are modeled with predictors measured at both the individual level and precinct level. Previous research has shown that overvotes tend to be clustered at the precinct level (Neely and Cook 2008). As such, any attempt to model overvotes without accounting for the precinct-level clustering will violate the assumption of uncorrelated residuals. Moreover, single-level models that do not adjust for the effects of clustering will produce incorrectly underestimated standard errors. For precinct-level explanatory variables, the underestimation of standard errors is generally quite severe (Gelman and Hill 2007). A multilevel model provides a way to efficiently adjust for precinct-level clustering, providing correct standard error estimates. Additionally, by incorporating randomly varying intercepts at the precinct-level, the multilevel model allows for investigation of between-precinct variance rather than simply controlling for its potential effects by, for instance, utilizing clustered standard errors in a single-level logistic regression (Steele 2009).

We specify independent explanatory variables in a way that most closely follows the previous study, and reflects the theoretical expectations summarized above. The precinct-level explanatory variables contain demographic characteristics constructed from U.S. Census data. Our analysis spans eight years, from 2004 to 2011, and so we draw on a range of U.S. Census and American Community Survey estimates in order to match estimated population characteristics to each election. Population counts were apportioned according to census block population, then spatially joined to San Francisco voting precincts, and converted to population proportions.

To measure population race-ethnicity, the following precinct-level variables are included in each model: Asian (not Hispanic/Latino), black or African American (not Hispanic/Latino), and Hispanic/Latino. To measure the level of civic resources or skills in the population, each model includes the following variables: female, population 65 years of age or older, educational attainment less than high school (the least educated), and foreign-born population (to indicate potential difficulties based on language and/or culture). In addition to the population proportion measures, each model includes median household income (divided by $10,000 to ease reporting). These demographic measures vary greatly across the precincts (see Appendix for descriptive statistics).

In addition to the demographic variables, we use two other explanatory variables. The first is an individual-level dummy indicator for whether a ballot was cast via vote-by-mail (VBM) or at

[6] The 2010 election used five two-sided ballot cards. Card 1 contained 13 offices from Governor down to State Superintendent of Public Instruction; the U.S. Senate was the ninth item. Card 2 contained 15 offices, the first eleven of which were Yes/No choices to keep sitting Justices. The two elections we use—the Superior Court Judge and the Board of Education—were the twelfth and thirteenth items. Card 3 was the IRV ballot, for Assessor/Recorder, Public Defender, and Board of Supervisors. Card 4 contained nine state ballot measures, and Card 5 had fifteen local ballot measures.

[7] We use the xtmelogit command in Stata 12 to estimate multilevel logistic regression models of individual ballots using precinct-level explanatory variables with intercepts allowed to vary randomly across precincts.

the Election-Day polling place. The information for the individual-level VBM variable is only available for the elections that occurred from 2008 to 2011, and so it is only included in those models. Finally, in the Board of Supervisors elections only, we include a count of the number of candidates appearing on the ballot. We can do this because we pool the data in those election years from across the districts electing supervisors. This creates variation in the number of candidates running (2004 mean = 9.9, SD = 6.3; 2006 mean = 5.0, SD = 2.3; 2008 mean = 5.7, SD = 2.7; 2010 mean = 10.7, SD = 6.7). For the other citywide offices (mayor, district attorney, sheriff, assessor/recorder) we analyze the elections separately and cannot, therefore, specify this variable. All explanatory variables are mean-centered so as to facilitate interpretation of the varying intercept estimates.

We conduct the comparative analysis using data aggregated at the precinct level. We specify negative binomial regression models to estimate the effects in the 2010 Board of Supervisors election, and those mentioned above: the 2010 Superior Court Judge, U.S. Senate, and Board of Education. In all of those models we specify an additional variable to control for the relative size of the precincts by computing the ratio of the number of ballots cast in a precinct to the number of ballots cast in the election, and then multiplied by 10,000 to ease reporting. To be clear, the estimates from the multilevel mixed-effects models are superior for understanding what type of precincts file overvoted IRV ballots; the binomial regression estimates, meanwhile, provide useful comparisons across portions of the 2010 ballot.

The analysis proceeds in three parts: first, we offer some descriptive reports of the elections over time. We look at the tendencies to rank two or three candidates or to choose one, and then we consider overvoting rates in the aggregate. Second, we report the estimated relationships between the probability of overvoting and the various factors listed above. Third, we compare the findings about overvotes in the 2010 IRV election to those in the other 2010 contests.

Results

Table 1 reports basic information on all of the municipal elections using the IRV system from 2004 to 2011, 39 in all. In 21 of those (54%) a majority winner emerged in the first round and IRV was not employed. We analyze the other 18 that did use IRV to produce a winner, with one exception: because of the prominence of the 2007 mayoral election we include it in the analysis even though the outcome did not require IRV elimination rounds. Note that the number of candidates running in the various contests varies widely, from four to 22. For citywide elections (mayor, sheriff, district attorney, assessor/recorder, and treasurer) we analyze the single contest as a separate event. For the Board of Supervisors—a body of 11 representatives from single-member districts—we combine the districts holding elections in a given year and take advantage of variation across the BOS Districts.[8]

Voters in the Aggregate

Figure 2 and Figure 3 report the overall rates of ranking candidates and of overvotes (see the Appendix for tabular results). First, note that the time increments on the horizontal axes are uneven; the lines in Figure 2 are meant to ease the comparison between voters who ranked a different number of candidates and not to characterize year-by-year trends. The solid line indicates the

[8] Representatives in about half of the 11 BOS seats are elected to four-year terms in each even-numbered year.

Table 1. San Francisco Elections Using IRV

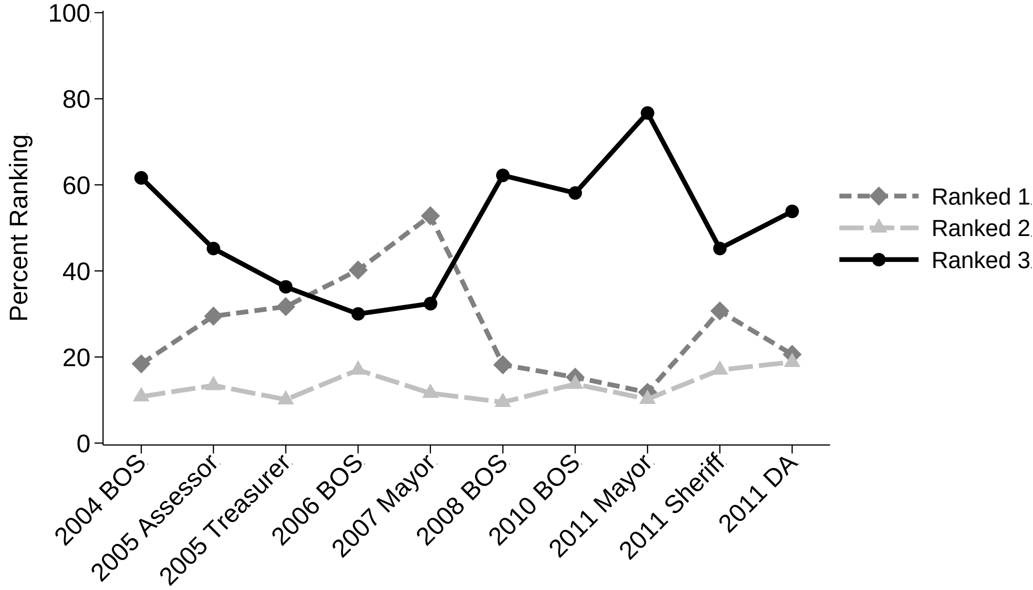
Year	Office	Jurisdiction	Candidates	Leader's % First Tally	Rounds Required	Winner's % Final Round
2004	BOS[a]	District 1	7	41.1	4	54.0
	BOS	District 2	5	61.3	0	61.3
	BOS	District 3	4	62.6	0	62.6
	BOS	District 5	22	28.4	19	50.6
	BOS	District 7	13	33.2	11	56.9
	BOS	District 9	6	50.7	0	50.7
	BOS	District 11	8	32.2	6	58.3
2005	Assessor	Citywide	4	47.4	2	58.1
	Treasurer	Citywide	4	61.4	0	61.4
	City Attorney	Citywide	1	98.1	0	98.1
2006	BOS	District 2	2	80.1	0	80.1
	BOS	District 4	6	26.2	4	52.5
	BOS	District 6	8	48.8	5	50.8
	BOS	District 8	3	66.2	0	66.2
	BOS	District 10	7	56.2	0	56.2
	Assessor	Citywide	1	98.6	0	98.6
	Public Defender	Citywide	1	98.9	0	98.9
2007	Mayor	Citywide	12	73.7	0	73.7
	District Attorney	Citywide	1	98.5	0	98.5
	Sheriff	Citywide	2	73.7	0	73.7
2008	BOS	District 1	9	40.5	2	50.7
	BOS	District 3	9	37.7	7	59.4
	BOS	District 4	3	52.4	0	52.4
	BOS	District 5	3	77.4	0	77.4
	BOS	District 7	3	71.1	0	71.1
	BOS	District 9	7	35.8	3	53.8
	BOS	District 11	8	28.2	4	52.9
2009	Treasurer	Citywide	1	96.8	0	96.8
	City Attorney	Citywide	1	97.1	0	97.1
2010	BOS	District 2	6	41.1	2	50.6[b]
	BOS	District 4	1	98.6	0	98.6
	BOS	District 6	14	31.4	12	54.1
	BOS	District 8	4	42.4	2	55.4
	BOS	District 10	21	11.8	20	52.7[b]
	Assessor	Citywide	2	79.7	0	79.7
	Public Defender	Citywide	1	98.9	0	98.9
2011	Mayor	Citywide	16	30.8	12	59.6
	District Attorney	Citywide	5	41.6	3	62.3
	Sheriff	Citywide	4	38.5	3	53.5

Note: Shaded rows indicate contests in which the first count produced no majority winner and IRV came into play.

[a] BOS = Board of Supervisors, 11 seats in a single-member district system.

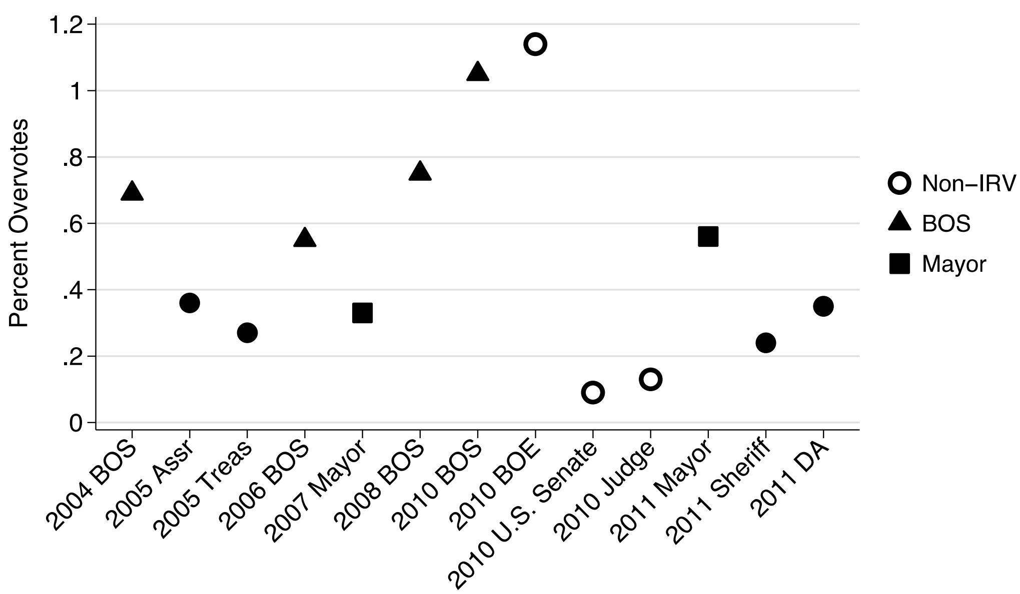
[b] The winning candidate did not lead in the first tally.

Figure 2. Ranking Tendencies across IRV Elections



Note: Items on the horizontal axis are not plotted in even time increments.
BOS = Board of Supervisors. DA = District Attorney

Figure 3. Rates of Overvotes across San Francisco Elections



Note: Items on the horizontal axis are not plotted in even time increments.
BOS = Board of Supervisors; BOE = Board of Education; DA = District Attorney.

proportion of voters ranking candidates in all three columns of the IRV ballot. Ranking three candidates is the most common behavior in elections after the 2007 mayoral election. The earlier analysis of elections between 2004 and 2006 voiced concern over the marked decline in fully ranked ballots (Neely and Cook 2008). We see now that no such decline persists.

While a complete analysis of the ranking behavior is beyond our present purpose, it is worth considering this important shift before turning to the question of overvoting. First, consider the highest profile IRV elections—the two mayoral contests. Both occurred in odd-numbered years with the mayoral race the key focus, but they were drastically different races. In 2007 a popular incumbent, Gavin Newsom, ran against 11 challengers and won about 74% of the first-round votes (Table 1). With little competition only 33% of the voters ranked three candidates. Many probably guessed that Newsom would be re-elected, agreed with that sentiment, and saved the bother of marking a second or third column: over 55% chose a single candidate. By contrast, in 2011 Mayor Ed Lee was elected to an open seat in a close contest only after 12 IRV elimination rounds. A full 77% of the voters ranked three candidates in that election. Of course, we can draw few inferences from two data points. But it appears that competition in this case may account for some variation in ranking behavior.

A second factor is simply the number of candidates on the ballot. As voters search for three people worthy of their vote, it is reasonable to expect that the more candidates running the higher the chance that a voter will find three to support. After the mayoral elections, the only repeated contests were for the Board of Supervisors. As mentioned above, in 2004 average of 9.9 candidates ran and in 2010 the mean number was 10.7. More candidates ran in those years than in others (an average of 5.0 ran in 2006 and 5.7 in 2008). The pattern is not perfect, but in 2004 and 2010 a relatively high number of voters ranked three candidates (61.6% and 58.1% did, respectively). These contrast to the lower rates of fully ranked ballots observed in the 2005 assessor (45.2%) and treasurer (36.3%) elections. In both of those contests voters chose from only four candidates.

Again, we hesitate to draw too much from these few cases. But it is possible that the decline noted in the earlier analysis could have been a function of fewer candidates running—for assessor and treasurer in 2005—and a lopsided contest in the case of the 2007 mayor's race.

Let us now turn to our key concern—overvotes. Figure 3 reports those rates with the points marked to group several types of office—Board of Supervisors, mayor, and non-IRV contests. First, note that the Board of Supervisors elections see not only relatively higher rates of ranking three candidates, they also contain higher rates of overvotes (indicated with triangles in the graph). For the 2004, 2006, 2008, and 2010 BOS elections, the overvoting rates were .69%, .55%, .75%, and 1.05% respectively. The squares identify the two mayoral contests. The runaway 2007 race had, unsurprisingly, the second lowest rate of overvotes in any IRV contest we tracked, at .33%. But note that, in the 2011 mayoral election that involved 16 candidates and saw the highest rate of ranking, .56% of the ballots had disqualifying overvotes. That figure is about half the rate of the 2010 BOS elections. We suspect that the prominence of the contest and its placement at the top of the odd-year election ballot matter. Mayoral races in San Francisco are the highest IRV contests and fill a relatively quiet political news space. Meanwhile, the BOS elections are considerably down the ballot in congressional- and presidential- election years.

Finally, the open-circle values are the non-IRV elections that are included to provide a context for comparison. Recall that in the Board of Education portion of the 2010 ballot we have the task that is most comparable to the IRV task—voters were to choose up to three candidates from a list of 11. By contrast, the election of the judge and senator represent simpler marking tasks on the ballot—a single candidate chosen from list. Overvoting in the BOE contest occurred at a rate of 1.14%, similar to the highest rates on the IRV Board of Supervisors ballots and much higher than most IRV contests. Meanwhile, the overvoting rates for judge and senator were .13% and .09%, respectively, lower than any IRV elections we examined. In the aggregate, then, the non-IRV portions of the ballot we examined provide what we had hoped: a comparable and contrasting decision environment. Indeed, the rates of overvotes in the more complex and the simper non-IRV contests bracket the range of overvoting in IRV races.

To appreciate what the rates of overvoting mean in actual elections, we report the counts in the Appendix. In the citywide IRV contests, the number of overvotes ranged from 471 (sheriff 2011) to 1,097 (mayor 2011). Mayor Lee won in 2011 by a margin of 27,297 votes. Clearly, in elections with a couple hundred thousand ballots filed, the numbers of overvotes are relatively small. However, we should also consider the races in the 11 BOS districts, much smaller jurisdictions. A close contest occurred in 2010 in District 10 when 21 candidates ran for an open seat. The initial count of first-ranked candidates was evenly divided with the five leaders each having only 11% to 12% of the vote. In successive rounds of elimination the lead changed and in the end Malia Cohen won by a margin of 442 votes (4,321 to 3,879). In that district, 20,550 IRV ballots were completed and 2.94% of those (605 of them) contained disqualifying overvotes.

While these overvoting rates appear unlikely to tip the scales in most elections, they could make a difference in close elections. And as we look within precincts we see a lot of variation and some surprisingly high rates of overvotes. Table A2 in the Appendix shows that in eight of the 10 IRV elections, the high end of the percentage of overvotes in a single precinct is 3.8% or more; in two of those nine elections at least one precinct had rates over 7.5%. This suggests that the tendency to overvote may not be evenly distributed across the city. In order to better understand this we now turn to the more rigorous analysis.

A Closer Inspection of Who Overvotes

Because we use a negative indicator—overvotes—to examine how well people navigate the IRV task and ballot, we want to be clear about the implications. As we test whether some groups of voters more than others make these disqualifying errors, the normatively desirable results would be null findings throughout.

Table 2 reports the estimated relationships between overvotes and the factors noted earlier. The dependent variable in each model is an individual-level indicator coded 1 if the ballot was overvoted and 0 if not. Because many of our independent variables are measured at the precinct level our interpretations are constrained accordingly. It is important to note that although the individual-level data on overvoting produce better estimates, as explained above, our substantive conclusions are limited to what occurs in the precincts. A final thing to note about this table: the first four columns of results replicate the 2008 study (Neely and Cook 2008), using the new estimator.[9]

Let us first consider how the likelihood of overvoting might have varied based on the race or ethnicity of residents in a precinct. In the second row, the positively signed and statistically significant coefficients indicate a consistently higher probability of overvoting as the proportion of African-American residents in a precinct increases. This occurs in every election examined. The

[9] Also, recall that the five columns of results on the left use data sets that do not distinguish between polling place and vote-by-mail (VBM) ballots and, therefore, those models do not specify that variable.

Table 2. Explaining Overvotes in IRV Contests, 2004–2011

	2004 BOS	2005 Assr	2005 Treas	2006 BOS	2007 Mayor	2008 BOS	2010 BOS	2011 Mayor	2011 Sheriff	2011 DA
Percent Asian	1.75*	.27	-.25	1.26*	.71	.67*	.16	-.66†	-.48	-.40
	(.44)	(.46)	(.52)	(.61)	(.56)	(.31)	(.40)	(.35)	(.51)	(.44)
Percent Black	2.32*	1.25*	1.89*	1.05*	1.83*	2.38*	1.80*	1.95*	1.47*	2.04*
	(.41)	(.37)	(.39)	(.51)	(.41)	(.44)	(.35)	(.34)	(.50)	(.43)
Percent Latino	2.08*	.10	.41	-.34	2.33*	2.00*	.92	1.09*	.04	1.20†
	(.45)	(.49)	(.53)	(.78)	(.56)	(.32)	(.76)	(.54)	(.83)	(.68)
Percent Female	3.31*	.24	.09	2.69*	-1.01	1.74	.12	.47	.87	-.63
	(1.09)	(.76)	(.87)	(1.06)	(.91)	(1.07)	(.70)	(.74)	(1.06)	(.91)
Percent Age 65 plus	1.14*	-.12	1.63*	-.19	1.59*	1.08†	2.83*	3.24*	1.22	1.84*
	(.56)	(.56)	(.52)	(.58)	(.60)	(.64)	(.60)	(.50)	(.79)	(.66)
Education Less HS	-.81†	.30	1.04*	.84	.53	1.63*	.75	.72	.95	-.48
	(.49)	(.47)	(.50)	(.61)	(.55)	(.78)	(.79)	(.76)	(1.15)	(.97)
Percent Foreign Born	1.11†	.84	1.17†	1.06	-.71	.28	1.07†	1.49*	.89	1.82*
	(.62)	(.62)	(.69)	(.84)	(.73)	(.54)	(.65)	(.56)	(.83)	(.72)
Med. Income ($10k)	-.01	-.95*	-.54	-.40	-2.20*	-.01	-.31	-.59*	-.75*	.02
	(.34)	(.44)	(.47)	(.62)	(.55)	(.25)	(.23)	(.23)	(.33)	(.28)
Number of Candidates	1.46*			1.68*		1.49*	1.54*			
	(.13)			(.28)		(.10)	(.18)			
VBM Ballot						-.07	-.26*	.02	-.20*	.01
						(.05)	(.06)	(.06)	(.09)	(.08)
Constant (Individual)	-5.25*	-5.74*	-6.09*	-5.78*	-5.92*	-5.31*	-5.03*	-5.37*	-6.08*	-5.82*
	(.04)	(.05)	(.06)	(.07)	(.06)	(.05)	(.06)	(.05)	(.08)	(.07)
Variance (Precinct)	.17*	.15*	.14*	.17*	.07	.15*	.08*	.12*	.18*	.16*
	(.03)	(.05)	(.06)	(.05)	(.06)	(.03)	(.03)	(.04)	(.07)	(.06)
ICC	.050	.043	.041	.048	.020	.043	.023	.034	.052	.048
Individuals	221,629	224,641	224,641	118,692	148,950	224,816	109,048	196,983	196,983	196,983
Precincts	348	573	573	266	574	336	228	424	424	424

Note: Cell entries are multilevel mixed-effects logistic regression coefficients with standard errors in parentheses. The dependent variable is coded 1 if the ballot contains overvote error and 0 if not. All independent variables range from 0 to 1 (VBM is binary, 0/1).

† $p < .10$, * $p < .05$ (two-tailed)

relationship between individual overvotes and the precincts' Latino population is statistically significant in about half of the contests, including the 2007 and 2011 mayoral elections. With regard to the Asian population in the precincts, the results are a bit more mixed. In three of the four BOS elections (2004, 2006, and 2008) the probability of overvoting significantly increases as the proportion of Asian Americans living in a precinct rises. However, in four elections the coefficients for Asian population take on negative values, one of which would pass a one-tailed significance test.

Overall, these results are fairly consistent with our expectations. Overvoting is consistently more likely in precincts with more African-American residents and at times more likely in precincts with more Latino residents. We did not expect to see a higher likelihood of overvoting in areas with more Asian-American residents, yet that was observed about one-third of the time.

In regard to age and gender, it appears that age matters. In seven of the 10 elections the probability of a disqualifying overvote increases with the share of the population that is 65 years or older (one of those seven passes a one-tailed test). Meanwhile, we see little evidence that precincts with more male or female residents differ systematically in the incidence of IRV overvotes. The coefficients testing those differences are often relatively small, and only twice statistically significant.

We are not surprised to see that in two of the 10 tests voters in areas with more foreign-born residents were clearly more likely to cast an overvote on the IRV ballot, and in three other elections that relationship passes a one-tailed test of significance. But we did not expect education and income to matter, and at times they did. Education is related to overvoting in only two elections, with a higher chance of overvoting in precincts containing more residents with less than a high school education. Income levels, meanwhile, were related to the likelihood of an overvote in four of the 10 elections, more than we anticipated.[10]

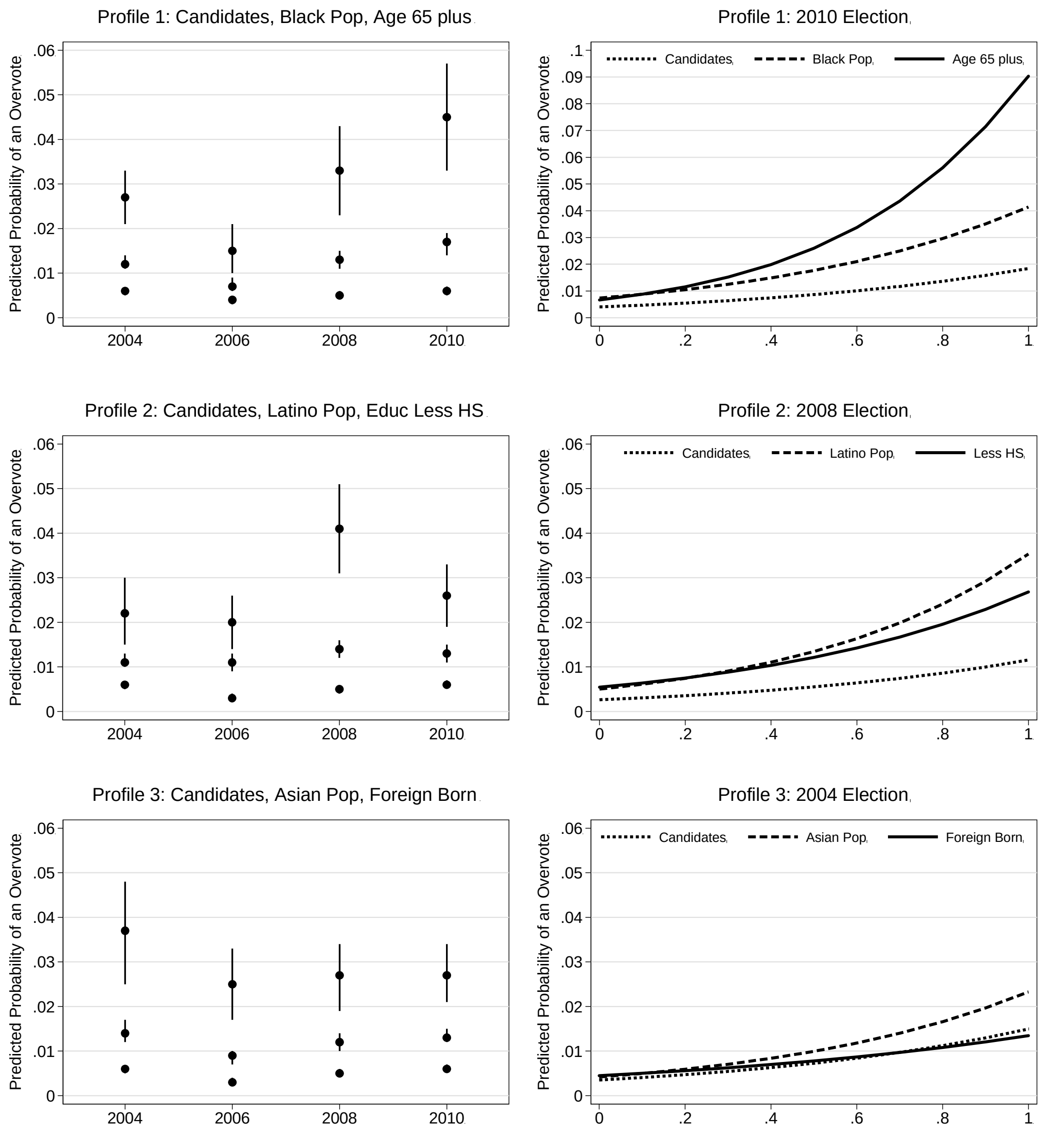
Next, consider factors related to the electoral context rather than the characteristics of the electorate. Perhaps the clearest expectation from studies of uncounted votes is that mistakes like overvotes will be more common on ballots with longer lists of candidates. This is an expectation that is not limited to instant-runoff voting; we would expect a similar effect in plurality elections or the first stage of a two-round runoff system. In each of the four Board of Supervisors election cycles (2004, 2006, 2008, and 2010) we see that, as expected, longer slates are associated with an increased likelihood of overvoting. The coefficients are all positively signed and statistically significant.

The other contextual factor we consider is whether the ballot was cast via vote-by-mail (VBM) or in person at the polling place. Our expectations in this case are less clear. Recall that we see advantages to filling out a VBM ballot at one's leisure, but also appreciate the potential to catch errors when scanning the ballot at the polling places. The results show a difference for the 2010 BOS contest and the 2011 sheriff election, with those casting a ballot by mail less likely to overvote. In the three other elections, no such differences were found. It is possible, then, that voting by mail, at times, reduces the chance of overvoting.

Now, to demonstrate the substantive meaning of the estimates in Table 2 we select several variables of interest and create profiles, reporting effects for the Board of Supervisors elections in Figure 4 and the two mayoral races in Figure 5. (For a full report of the expected changes in

[10] This may seem counterintuitive. Recall that our expectations were informed by the earlier analysis of IRV in San Francisco that consistently found no such relationships (Neely and Cook 2008). It is possible that the findings here differ in nearly half of the elections, especially on income, because of the more accurate estimation technique.

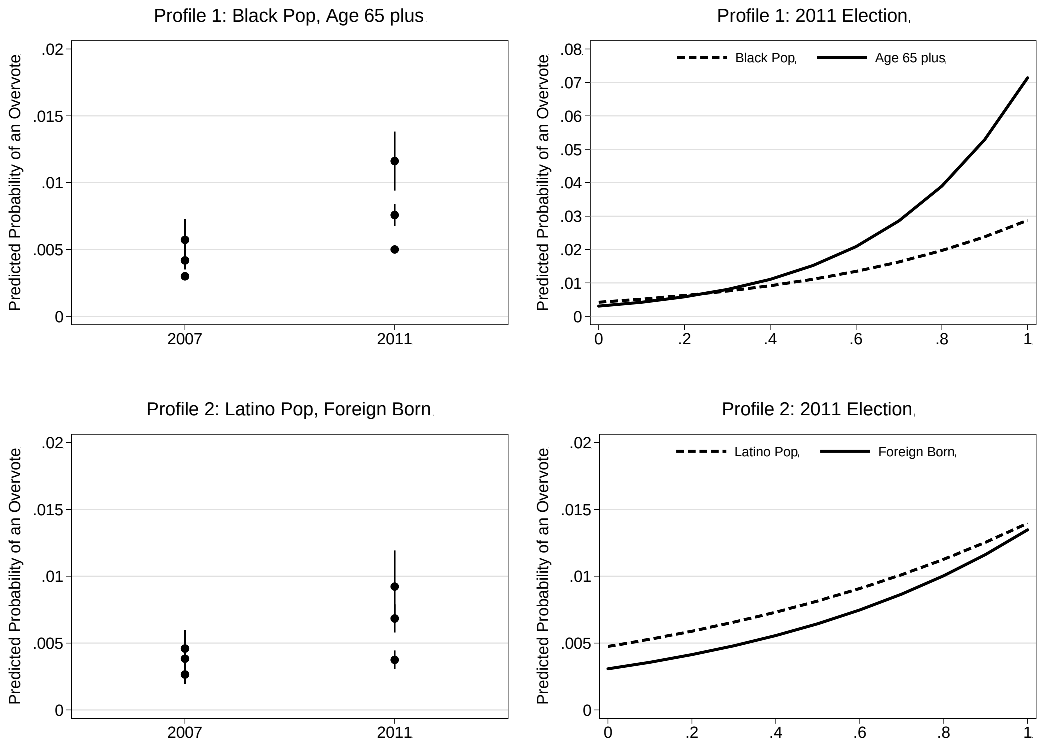
Figure 4. Predicted Probability of Overvoting on Board of Supervisor Election Ballots: Three Profiles



Note: The left side shows the estimated combined effect of the three profile variables on the probability of an overvote, setting all three to their means, the means + 1 st. dev., and the means + 2 st. dev. (bars are 95% confidence intervals). The right side separates the effects of those three variables on the probability of an overvote, within one of the four BOS elections. These are calculated as population averaged probabilities with remaining covariates as observed.

the probability of an overvote associated with the estimates in Table 2, see Appendix Table A4; for tabular results used in Figures 4 and 5 see Tables A5 and A6.) The first thing to note is the small absolute size of these estimated effects. This is not surprising, given a dependent variable—an overvote—that is such a rare event in elections; in our case is not unusual for it to occur

Figure 5. Predicted Probability of Overvoting on Mayoral Election Ballots: Two Profiles



Note: The left side shows the estimated combined effect of the two profile variables on the probability of an overvote, setting all three to their means, the means + 1 st. dev., and the means + 2 st. dev. (bars are 95% confidence intervals). The right side separates the effects of those two variables on the probability of an overvote, within the 2011 mayoral election. These are calculated as population averaged probabilities with remaining covariates as observed.

around .5% of the time (i.e., with a probability of .005, see Figure 3). It is useful, then, to also consider the relative change. With such a small starting point, an increased probability of, say, .005 would double that chance of an overvote.

Profle 1 in Figure 4 reports the expected changes in the probability of an overvoted ballot based on three factors: the number of candidates running, and the percentage of African-American and elderly residents in a precinct. The top graph on the left-hand side of Figure 4 shows the combined effects of those three factors (with 95% confidence intervals) for the BOS elections. For each election we plot the expected probability of an overvote with all three factors set to their mean, and then with them set at one and two standard deviations above.[11] In 2010, for example, in the average precinct in regard to these factors (10.7 candidates, 7.6% black residents,

[11] Reporting effects at two standard deviations above the mean would often display unusual circumstances. However, the demographic parameters in San Francisco vary widely. For example, in 2010 the percentage of black residents in San Francisco precincts ranged from 0% to 76.7% with a mean of 7.6% and a standard deviation of 12.4%.

and 14.2% elderly residents) the probability of an overvote was .006 (or a .6% chance). In a precinct with more candidates running, and more black and elderly residents (plus 1 S.D.: 17.4 candidates, 20% black, and 19.6% elderly), the expected probability is .017, nearly three times the probability at the means. Increasing all three factors to two standard deviations above their means leads to a probability of overvoting of .045 (or a 4.5% chance).

In the top graph on the right-hand side of Figure 4 we separate the three factors in Profile 1 and plot their expected effect on the probability of an overvote across their full theoretical range (i.e., 0 to 1).[12] Because the observed range of the demographic variables is smaller than that, these plots are best considered as comparisons of relative effect sizes and not as specific predictors of probabilities, at least not near the ends of the horizontal axes. For Profile 1 we isolate those effects found in the 2010 Board of Supervisors election. Note that the smallest influence among the three factors is the number of candidates on the ballot, while the age of the population is the strongest.

Profile 2 reports the estimated effects for the number of candidates, the percentage of Latino residents, and the percentage in a precinct with less than a high-school education. Again, the point estimates on the left-hand side are the combined effects for those three factors. On the right-hand side we isolate those three influences, this time in the 2008 election. The relative influence of the number of candidates matches the 2010 results—it is less than the other factors, both of which appear to have comparable effects on overvoting. Profile 3 is presented similarly, this time including the number of candidates running, and the percent of Asian-American and foreign-born residents in a precinct. Here, we separate those effects in the 2004 election and see that the percent of Asian residents was more strongly associated with the probability of an overvote than nativity or the length of the slate.

The three profiles and the isolated years we have chosen display some of the strongest effects. But if we consider more modest cases, the results remain meaningful. Among all three profiles and across the four elections, increasing one standard deviation on the three factors leads to probabilities of overvoting between .007 and .017. These range between 2.1 and 2.8 times what was expected at the means. In precincts with two standard deviations above the means, the combined effects lead to probabilities between .025 and .037. In such precincts the estimated chance of an overvote is 4.4 to 7.7 times that of the average precinct.

In Figure 5 we provide two profiles in the two mayoral elections. Profile 1 combines the effects of the percent of African-American and elderly residents in a precinct. On the left we see the point estimates of the probabilities of an overvote. These influences were stronger in the competitive 2011 race for an open seat than in the re-election of the incumbent in 2007. And on the right-hand side of the figure we see that in 2011 overvotes were more strongly associated with elderly than with black populations. Profile two includes the percent of Latino and foreign-born residents. Those combined effects were weaker than the ones specified in Profile 1, as seen in the plots on the left and in the slopes of the lines on the right (note how the scale of the vertical axis differs from the Profile 1 line graph).

Although the effects of these factors were stronger in the 2011 than in the 2007 mayoral election, they are all weaker than what we saw in the Board of Supervisors contests. This is noted, first, by comparing the scale of the vertical axes on the left-hand side of Figure 5, ranging from 0 to .02, to those in Figure 4 that range from 0 to .06. Second, consider Profile 1 in the 2011 mayoral race. The combined effects, comparing precincts with more black and elderly residents (i.e., plus one standard deviation) to an average precinct, shifts the probability of an overvote

[12] All variables are rescaled in our analysis to range from 0 to 1 to ease comparisons.

from .005 to .007. An increase of two standard deviations produces an expected probability of .009. Overvoting is 1.5 and 2.4 times more likely, respectively, in those precincts than in an average one. Profile 2 produces similar results in 2011, if weaker: the probability of an overvote is 1.4 and 1.8 times more likely when the factors are increased to plus one and plus two standard deviations, respectively. Clearly, the relative change in the expected probability of overvoting based on the influences we have considered is much smaller in the mayoral than in the Board of Supervisors elections.

Comparing Overvotes in IRV Elections and Non-IRV Elections

Our final section provides further context by comparing overvote errors in IRV contests with errors in three non-IRV contests. Recall that we contrast the simpler marking task contained in the 2010 contests for Superior Court Judge and U.S. Senate with the case most similar to IRV, the 2010 Board of Education race. This allows a comparison based on complexity (IRV and BOE versus judge) and a comparison of ballot placement (judge versus senator). Because the data for these elections are reported only at the precinct level, that is our unit of analysis. To make the comparison as close as possible we re-estimate the model for the 2010 IRV BOS election with precinct-level counts. The dependent variable is now the number of overvotes in a precinct while the independent variables remain the same.[13] The models were estimated with a negative binomial regression, an appropriate choice when the data are overdispersed. Because we are now estimating counts, we include a variable for the precinct size to control for variation in the number of votes cast.[14]

First, we expect that the if the demographic factors above relate to overvotes in both the IRV and BOE elections then the IRV ballot and task may not present a unique challenge to voters. But if the types of voters who err in IRV elections do not resemble those overvoting in the BOE race then the IRV ballot and task may be intrinsically problematic. Second, if in both cases the nature of the overvotes matches what occurs in the elections for judge and senator, then the tendencies to make these errors may be broadly experienced and not a function of the more complex ballot format and decision task. However, if the factors that relate to overvotes on the IRV ballot are not associated with overvotes on those items, then the complexity could be one reason for those errors in IRV (and possibly BOE) elections. Finally, we compare the factors related to overvotes at the top of the ticket to those further down in order to examine the potential effects of ballot placement.[15] Is it the case that voters are better able to attend to marking their ballots at the start of the process, and become fatigued and more likely to err as they work through the ballot?

Comparing the results in the first two columns of Table 3 we see a fairly good match. In both the BOS and BOE elections, precincts with more African-American residents tended to have more overvotes. The same was true of precincts with more elderly residents (in a one-tailed test) and more foreign-born residents. Two discrepancies emerge: lower income levels mattered in the BOE election (in a one-tailed test) but not for the BOS races, and, while voting by mail was associated with fewer overvoted BOS ballots, no such difference was seen in the BOE election.

[13] The vote-by-mail variable is substantively the same but instead of an individual-level indicator it is now the percent of such ballots in each precinct.

[14] It equals the number of ballots counted in a precinct / the total ballots cast in the election, multiplied by 10,000.

[15] We thank an anonymous reviewer for raising this point.

Table 3. Explaining Overvotes on IRV and Non-IRV Ballots in the 2010 General Election

	2010 BOS[a] (4 contests)		2010 BOE[b] (Citywide)		2010 Judge (Citywide)		2010 U.S. Senator (Citywide)	
Percent Asian	.45	--	-.20	--	.26	--	.47	--
	(.41)		(.36)		(.54)		(.72)	
Percent Black	1.70*	32%	1.93*	27%	1.56*	21%	1.71*	24%
	(.38)		(.41)		(.60)		(61)	
Percent Latino	1.07	--	.52	--	.36	--	2.79*	27%
	(.76)		(.47)		(.74)		(.92)	
Percent Female	.001	--	.38	--	-1.41	--	-1.70	--
	(.73)		(.68)		(1.14)		(1.42)	
Percent Age 65 plus	3.03*	25%	.92†	8%	.62	--	4.26*	40%
	(.58)		(.54)		(.89)		(1.21)	
Percent Less HS	.54	--	-.39	---	-.99	--	.04	--
	(.76)		(.81)		(1.21)		(1.47)	
Percent Foreign Born	1.33*	20%	2.15*	36%	1.79*	29%	2.03†	33%
	(.60)		(.57)		(.89)		(1.15)	
Med. Income ($10K)	-.01	--	-.02†	-7%	-.002	--	-.06*	-20%
	(.01)		(.01)		(.02)		(.02)	
Number of Candidates	.09*	81%	--	--	--	--	--	--
	(.01)							
Vote-by-Mail %	-1.12*	-12%	-.43	--	-.39	--	-2.25*	-22%
	(.51)		(.46)		(.71)		(.86)	
Precinct Size	.02*	37%	.07*	36%	.06*	33%	.04*	22%
	(.004)		(.01)		(.02)		(.02)	
Constant	.50		.51		-1.51*		-.95	
	(.33)		(.34)		(.52)		(.59)	
Model χ^2	699.8*		111.6*		35.6*		114.1*	
N	227		572		572		572	

Note: The coefficients are negative binomial regression estimates (robust standard errors). The unit of analysis is a precinct. The dependent variable is the number of ballots in a precinct with an overvote. The percentages report the expected change in the amount of ballots with an overvote, given a one-standard-deviation increase in the independent variable. The BOE election was to seat three members; voters could choose up to three candidates from a single list of the eleven who ran. Voters were to choose one candidate for Superior Court Judge (of two who ran) and U.S. Senate (of six who ran). † $p \leq .10$; * $p \leq .05$; [a] BOS = Board of Supervisors; [b] BOE = Board of Education.

The column of percentages reports the expected change in the amount of overvoted ballots associated with a one-standard-deviation increase in the explanatory variable.[16] In the 2010 Board of Supervisors elections, comparing a precinct with 7.6% black residents (the mean) to one with 20% (adding 1 standard deviation), we would expect 32% more overvotes to occur in the latter.

The similarity of results in the first two columns suggests, then, that the IRV ballot and decision task do not present a unique challenge for voters. Instead, the complexity of it and of the decision environment in the Board of Education election appears to lead voters in some precincts more than others to make disqualifying errors.

What about a simpler candidate selection? Turning to the elections of a judge and U.S. senator, we see more similarities than differences in the factors related to overvoting. In each case, precincts with more African-American and foreign-born residents submitted higher counts of overvoted ballots. In the senate race those errors were also more common in precincts with more Latino, elderly, and low-income residents. This evidence suggests that the discrepancies in the type of precincts with overvotes persist in the easier task of choosing one candidate. These results also show no support for the idea that some voters more than others may become fatigued and more likely to err in down-ballot races. Indeed, the differences across precincts are fewer, not greater, in selecting a judge than in choosing a senator. In the closing section we consider the implications of these results, along with those reported above.

Discussion

The earlier examination of the first years of IRV voting in San Francisco revealed a couple of troubling findings: A precipitous drop in the proportion of IRV ballots on which three candidates were ranked, and overvotes that were too common and disproportionately found in areas where more African-American and foreign-born residents live (Neely and Cook 2008). The results reported here show that part of the story has changed and another aspect remains.

First, the changes: from the additional years of data we see that rates of ranking three candidates are not in decline. Instead, the highest rate of ranking three candidates occurred in 2011. We expect voters are more likely to rank three when race is close—the rate in the competitive 2011 mayoral election was more than double what it was in the 2007 landslide. As outcomes are less certain, voters should care more about expressing second and third choices. A second factor is simply the number of the candidates running—with longer slates increasing the likelihood of voters finding three worthy of support. While these are speculations drawn from a few observations, what we can say with certainty is that the pronounced drop in ranking rates during the first few years of IRV in San Francisco was not indicative of a trend.

As for the overall rate of IRV overvotes, the new comparison to non-IRV elections is instructive. Asking voters to do more than pick a single candidate from a list leads to an increase in disqualifying errors. The IRV portion of the ballot appears to be no more problematic than another similarly complicated contest. Further, these errors are probably not due to the down-ballot placement of the items and voter fatigue: in the plurality election for a Superior Court Judge placed just prior to the BOE contest the overvoting rate was extremely low. This provides an important new clarification that will help as citizens and policymakers weigh the evidence on IRV and assess its usefulness. In addition, while we knew from the earlier work that some voters

[16] Estimates were produced using Long and Freese's (2001) post-estimation routine in Stata.

more than others were likely to submit voided IRV ballots, we learn now that similar discrepancies occur in non-IRV contests. Even among votes cast in a top-of-the-ticket race like U.S. Senate, overvoting occurs in an uneven fashion. In regard to overvotes, then, these two findings reveal IRV as less of an outlier and more in line with other portions of the ballot.

What has not changed is the nature of the discrepancies in who tends to overvote: consistently, precincts where more African-Americans reside are more likely to collect overvoted, voided ballots. And this often occurs where more Latino, elderly, foreign-born, and less wealthy folks live. The additional years of data show no meaningful increase or decline in these tendencies but rather bolster the earlier study's findings.

In all of the elections we examined, some voters were more at risk than others of making disqualifying errors. That inequality is exacerbated in systems that require more of voters, such as San Francisco's Board of Education and IRV contests. If those tendencies are fairly durable—and they appear to be—then the question becomes how high is too high? Are the overvoting rates for these more complicated portions of the ballot problematic or not? Earlier, we examined a couple of IRV elections and looked at the margins of victory compared to the rate of overvoted ballots. In very close contests, the differences could matter; in most elections, however, the rates of disqualifying errors we observed would not alter the outcome. Does that suggest that all is well? We think not. As the rate of errors increases and as those occur disproportionately across precincts, the equality of voice suffers. While the aggregate outcome of most elections may not change, the evenness of voters' say in who wins or loses is at risk.

Clearly, San Francisco's IRV system has its challenges, as do all election systems. We have examined but one slice of the proverbial pie. Questions about IRV's effects on campaign strategies, its outcomes in terms of diversity of the winners, and the question of exhausted ballots and winners' margins are all issues we have left for others to pursue. Our findings on voters' abilities to navigate the system and record their preferences must be weighed against those and other factors. It is possible, for example, that the risk of one group's voice fading due to disqualifying errors on the IRV ballot is outweighed by other positive consequences IRV provides for that group.

Finally, our results should also be considered within a broader set of limits. We have examined one jurisdiction and one version of Instant-Runoff Voting. It is possible that voters' experience elsewhere will differ. However, we suggest that the high level of demographic diversity among the residents in San Francisco and the fact that IRV has been in place since 2004 make this an especially useful case to examine.

What we can say is that voters in San Francisco, as everywhere, who bother to turn out and cast ballots deserve to have their votes count. Because some precincts more than others tend to collect overvoted ballots, we see the situation as fluid. Especially for those populations who can be geographically identified, it would seem wise to increase outreach efforts. In those precincts with relatively concentrated populations of African-Americans, elderly, Latino, foreign-born, or low-income residents, additional information for both the voters and the poll workers could make a difference. But this should not be isolated to education about the IRV portion; instead it should focus on the basics of filing any vote choice. Electing people to office is a human endeavor that will always involve error. It is the aim of those interested in equitable and fair systems to minimize the consequences of those errors whenever possible. Our hope is that the above report might guide such efforts, improving their efficiency and informing their intent.

References

Ansolabehere, Stephen, and Charles Stewart. 2005. "Residual Votes Attributable to Technology." *The Journal of Politics* 67(2): 365–89.

Bullock, Charles S., and M. V. Hood. 2002. "One Person—No Vote; One Vote; Two Votes: Voting Methods, Ballot Types, and Undervote Frequency in the 2000 Presidential Election." *Social Science Quarterly* 83(4): 981–93.

Bullock, Charles S., and Richard E. Dunn. 1996. "Election Roll-Off: A Test of Three Explanations." *Urban Affairs Review* 32(1): 71–86.

Burnett, Craig M., and Vladimir Kogan. 2012. "Ballot (and Voter) Exhaustion under Instant Runoff Voting: An Examination of Four Ranked Choice Elections." Presented at the 70th Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Carpini, M. X. Delli, and Scott Keeter. 2000. "Gender and Political Knowledge." In *Gender and American Politics*, ed. S. Tolleson-Rinehart and J. J. Josephson. Armonk, NY: ME Sharpe.

City/County of San Francisco Charter. Accessed April 21, 2014, http://www.amlegal.com/nxt/gateway.dll?f=templates&fn=default.htm&vid=amlegal:sanfrancisco_ca

Darcy, Robert, and Anne Schneider. 1989. "Confusing Ballots, Roll-Off, and the Black Vote." *The Western Political Quarterly*: 347–64.

Dennis, Greg. 2004. *Ballot-Marking Errors in the First San Francisco Instant Runoff Election*. Unpublished manuscript. Accessed March 27, 2013, http:// www.gregdennis.com/voting/.

Dudum v. Arntz. 10–17198 (9th Cir. 2011) (R. Seeborg).

Farrell, David, and Ian McAllister. 2003. "Electoral Systems." In *The Cambridge Handbook of Social Sciences in Australia*, ed. Ian McAllister, S. Dowrick, and R. Hassan. New York: Cambridge University Press.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Hecock, Donald Sumner, and Henry M Bain. 1956. *The Arrangement of Names on the Ballot and Its Effect on Voter's Choice*. Detroit: Wayne University Press.

Jansen, Harold J. 2004. "The Political Consequences of the Alternative Vote: Lessons from Western Canada." *Canadian Journal of Political Science/Revue canadienne de science politique* 37(3): 647–69.

Kimball, David C., and Martha Kropf. 2005. "Ballot Design and Unrecorded Votes on Paper-Based Ballots." *Public Opinion Quarterly* 69(4): 508–29.

Kimball, David C., Chris T. Owens, and Katherine M. Keeney. 2004. "Unrecorded Votes and Political Representation." In *Counting Votes: Lessons From the 2000 Presidential Election*. Gainseville, FL: University Press of Florida.

Knack, Stephen, and Martha Kropf. 2003. "Voided Ballots in the 1996 Presidential Election: A County-Level Analysis." *The Journal of Politics* 65(3): 881–97.

Krosnick, Jon A., Joanne M. Miller, and Michael P Tichy. 2004. "An Unrecognized Need for Ballot Reform: Effects of Candidate Name Order." In *Rethinking the Vote: the Politics and Prospects of American Election Reform*, ed. Ann Crigler, Marion Just, and Edward J McCaffery. New York: Oxford University Press.

Lau, Richard R., and David P. Redlawsk. 2006. *How Voters Decide: Information Processing in Election Campaigns*. New York: Cambridge University Press.

Lijphart, Arend. 1997. "Unequal Participation: Democracy's Unresolved Dilemma." *The Ameri-*

can Political Science Review 91(1): 1–14.

Long, J. Scott, and Jeremy Freese. 2001. *Regression Models for Categorical Dependent Variables Using Stata*. Thousand Oaks, CA: Sage.

Neely, Francis, and Corey Cook. 2008. "Whose Votes Count? Undervotes, Overvotes, and Ranking in San Francisco's Instant-Runoff Elections." *American Politics Research* 36(4): 530–54.

Reilly, Ben, 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: Reflections on an Embedded Institution*, ed. Shaun Bowler and Bernard Grofman. Ann Arbor, MI: The University of Michigan Press, 37–58.

Reilly, Benjamin. 2004. "The Global Spread of Preferential Voting: Australian Institutional Imperialism?" *Australian Journal of Political Science* 39(2): 253–66.

Sinclair, D.E.B., and R. M. Alvarez. 2004. "Who Overvotes, Who Undervotes, Using Punchcards? Evidence from Los Angeles County." *Political Research Quarterly* 57(1): 15–25.

Steele, Fiona. 2009. "Multilevel Models for Binary Responses (Concepts)." *Centre for Multilevel Modelling, University of Bristol*. http://www.cmm.bris.ac. uk/lemma/ (August 20, 2013).

Stiefbold, Rodney P. 1965. "The Significance of Void Ballots in West German Elections." *American Political Science Review* 59: 391–407.

Tomz, Michael, and Robert P. Van Houweling. 2003. "How Does Voting Equipment Affect the Racial Gap in Voided Ballots?" *American Journal of Political Science* 47(1): 46–60.

Walker, Jack L. 1966. "Ballot Forms and Voter Fatigue: an Analysis of the Office Block and Party Column Ballots." *Midwest Journal of Political Science* 10(4): 448–63.

Zimmerman, Joseph F. 1994. "Alternative Voting Systems for Representative Democracy." *PS: Political Science and Politics* 27: 674–77.

Appendix

Table A1. Descriptive Statistics: Demographic Variables

	Mean	SD	Min.	Max.
2004 to 2006				
Black	8.0	12.9	0.1	75.1
Latino	13.1	13.4	1.0	76.6
Asian	29.2	18.6	4.5	92.3
Female	49.0	5.6	29.0	64.3
Age 65 plus	10.0	6.0	0.3	57.8
Less than HS	17.1	12.8	0.2	72.7
Med. Income ($)	62,171	23,140	9,994	174,456
Foreign	33.6	15.7	5.3	79.2
2007 to 2011				
Black	7.6	12.4	0	76.7
Latino	9.9	8.5	0	57.0
Asian	27.0	18.8	1.5	93.5
Female	38.4	9.3	9.6	92.6
Age 65 plus	14.2	5.4	2.3	36.9
Less than HS	12.7	8.9	0	50.3
Med. Income ($)	80,503	26,952	11,646	173,557
Foreign	33.5	10.5	10.5	71.3

Note: The units are precincts. Cell entries for all variables except income are percentages. Data in the top half are from the 2000 U.S. Census, except for income and education (from the American Community Survey 2005, 3-year). Data in the bottom half are from the 2010 U.S. Census, except for income (in 2007 analyses, from the ACS 2005 3-year; in the 2008–2011 analyses, from the ACS 2009 3-year).

Table A2. Descriptive Statistics: Overvotes on IRV and Non-IRV Ballots

	Precinct-level: Aggregated within precincts					Overall: Aggregated across the City/County		
	Mean %	Min. %	Max. %	Median %	N	Overvoted Ballots	Total Ballots	Overvoted Ballots %
2004 BOS	0.7	0	5.3	0.5	363	1,550	223,837	.69
2005 Assr.	0.4	0	2.4	0.3	578	814	225,370	.36
2005 Treas.	0.3	0	4.6	0.3	578	608	225,370	.27
2006 BOS	0.7	0	4.9	0.3	276	659	119,906	.55
2007 Mayor	0.4	0	4.8	0.3	579	499	149,465	.33
2008 BOS	0.8	0	4.2	0.5	349	1,707	227,045	.75
2010 BOS	1.3	0	8.2	0.5	240	1,182	113,069	1.05
2010 BOE	1.2	0	9.0	0.9	577	3,144	275,911	1.14
2010 Judge	0.1	0	1.6	0.0	577	369	275,911	.13
2010 US Sen.	0.1	0	1.5	0.0	577	259	284,625	.09
2011 Mayor	0.6	0	7.7	0.5	428	1,097	197,242	.56
2011 Sheriff	0.3	0	2.3	0.2	428	471	197,242	.24
2011 DA	0.4	0	3.8	0.2	428	682	197,242	.35

Note: The left side reports precinct-level rates of overvoting; the N is the number of precincts voting in a given election. The right side reports overvoting rates computed from the individual-level records for the entire City/County, ignoring precinct lines (or in 2004, 2006, 2008, and 2010, that portion of the City/County that cast IRV ballots).

Table A3. Descriptive Statistics: Candidate Ranking Tendencies

	Ranked One Candidate	Ranked Two Candidates	Ranked Three Candidates	
	Percent	Percent	Percent	Total votes
2004 BOS	18.4	10.8	61.6	223,837
2005 Assessor	29.5	13.4	45.2	225,370
2005 Treasurer	31.7	10.1	36.3	225,370
2006 BOS	40.2	17.0	30.0	119,906
2007 Mayor	52.8	11.6	32.4	149,465
2008 BOS	18.2	9.5	62.2	227,045
2010 BOS	15.3	13.7	58.1	113,069
2011 Mayor	11.8	10.2	76.7	197,242
2011 Sheriff	30.7	17.0	45.2	197,242
2011 DA	20.6	18.8	53.8	197,242

Table A4. Change in the Probability of an Overvote, IRV Contests, 2004–2011

	2004 BOS	2005 Assr	2005 Treas	2006 BOS	2007 Mayor	2008 BOS	2010 BOS	2011 Mayor	2011 Sheriff	2011 DA
Percent Asian	.011*	.0009	-.0006	.006*	.002	.005*	.002	-.003	-.001	-.002†
Percent Black	.015*	.004*	.005*	.005*	.006*	.016*	.018*	.011*	.003*	.007*
Percent Latino	.013*	.0003	.001	-.002	.008*	.013*	.009	.006†	.0001	.004*
Percent Female	.021*	.0008	.0002	.014*	-.003	.012	.001	.003	.002	-.003
Percent Age 65 plus	.007*	-.004	.004*	-.001	.005*	.007†	.028*	.017*	.003†	.006*
Percent Less HS	-.005†	.001	.003*	.004	.002	.011*	.008	.004	.002	-.002
Percent Foreign Born	.007†	.003	.003†	.005	-.002	.002	.011†	.008*	.002	.007*
Med Income ($10k)	-.00002	-.003*	-.001	-.001	-.007*	-.0001	-.003	-.003	-.002*	-.0003*
Number of Candidates	.010*			.010*		.010*	.015*			
VBM Ballot						-.0005	-.003*	-.0001	-.0004*	.0001

Note: Cell entries represent change in probability of an overvote associated with 1-unit change in the explanatory variable. Because all variables are coded 0-to-1 (or 0/1 in the VBM variable) this indicates the full range of each variable.

† $p < .10$, * $p < .05$ (two-tailed)

Table A5. Predicted Probabilities of Overvoting on Board of Supervisors Ballots: Three Profiles (Tabular Results Used in Figure 4)

Profile 1: Number of Candidates, % Black, % Elderly				
	2004	*2006*	*2008*	*2010*
At means	.006	.004	.005	.006
Plus 1 SD	.012	.007	.013	.017
Plus 2 SD	.027	.015	.033	.045
Profile 2: Number of Candidates, % Latino, % Less than HS				
	2004	*2006*	*2008*	*2010*
At means	.006	.003	.005	.006
Plus 1 SD	.011	.011	.014	.013
Plus 2 SD	.022	.020	.041	.026
Profile 3: Number of Candidates, % Asian, % Foreign-born				
	2004	*2006*	*2008*	*2010*
At means	.006	.003	.005	.006
Plus 1 SD	.014	.009	.012	.013
Plus 2 SD	.037	.025	.027	.027

Note: Cell entries are the predicted probabilities of an overvote when considering the combined effect of the three factors in each of the profiles (See the text and Figure 4).

Table A6. Predicted Probabilities of Overvoting on Mayoral Election Ballots: Two Profiles (Tabular Results Used in Figure 5)

Profile 1: % Black, % Elderly		
	2007	*2011*
At means	.003	.005
Plus 1 SD	.004	.008
Plus 2 SD	.006	.012
Profile 2: % Latino, % Foreign-born		
	2007	*2011*
At means	.003	.005
Plus 1 SD	.004	.007
Plus 2 SD	.005	.009

Cell entries are the predicted probabilities of an overvote when considering the combined effect of the two factors in each of the profiles (See the text and Figure 5).