ONLINE APPENDIX

Robustness Checks and Supplementary Analyses for

The Effects of Ranked Choice Voting on Substantive Representation

Contents

A Evaluating the Translation of Voter Preferences with Simulations

In this section, I examine the hypothesis that RCV improves the translation from voters' preferences to winners. To test the claim, I run simulations of various electorates and candidate sets with randomly generated ideal points. The logic of spatial models suggests that voters' and candidates' positions can be represented in Euclidean space. In these models, voters prefer candidates whose ideal points are closer to their own. The goal of my analysis is to determine which system produces the outcome favored by the median voter.

To do so, I employ a one-dimensional spatial voting model under four electoral systems. Under the first, simple FPTP, each voter votes for the candidate closest to her ideal point, and the candidate with the most votes wins. Thus, if the voter's ideal point is 0.1 and the candidates in the race have ideal points of $A = 0.4$, $B = -0.8$, and $C = 0.3$, she will vote for C. While this helps to examine the plurality winner problem, it does not deal with the question of sincere versus strategic voting.

To address this, I analyze a second system ("restricted FPTP") in which voters only consider casting a ballot for Candidates A and B (the "major party" candidates) and disregard the other candidates. I use this system to mimic the national status quo, in which there are two parties which are considered viable and voters' decide between these two choices, even if other candidates locate closer to the median. Since the candidates' ideal points are randomly drawn from the same distribution, the choice to use the first and second candidates generated is made without loss of generality.[1] The logic here is that the electorate is entering an already-existent equilibrium in which candidates A and B are viewed as viable and the others are not. This electoral rule makes the simplifying assumption that all voters strategically select one of the major party candidates, but my findings remain if a small proportion of voters cast sincere ballots for other "third party" candidates. Thus, this system is meant to simulate partisan elections. Under this rule, our hypothetical voter would vote for A since C is non-viable and she prefers A to B.

The third system offers a refinement to the second system, in which the major party candidates were arbitrarily chosen. We might instead imagine that voters have some knowledge of which candidates are generally preferable to the electorate. Imagine they saw a poll of all candidates, and decided to strategically cast a vote only among those in the top two of polling. A similar way to view this system is in the context of a two-round election; in the first round, each voter votes for her sincere preference, and then the top two vote-getters advance to a runoff in which each voter votes sincerely among the two runoff candidates. Stemming from this latter interpretation, I term the third system the "Runoff" rule. In our example, we might imagine that among the general electorate, B is not especially popular and is eliminated in the first round of balloting; in that case, our voter selects C over A. While the end result is the same whether we interpret this rule as a poll or a runoff, the

[1]This assumption will lead me to underestimate the representativeness of FPTP because the major parties do exhibit some nonzero degree of responsiveness in their platforms to public opinion.

single-round interpretation places a much higher burden on the voters. It requires them to have knowledge of every other voter's *true* preferences—making this the most strategic and cognitively taxing of the three rules discussed thus far. On the other hand, the "runoff" interpretation merely asks voters to vote sincerely in two consecutive elections—thus, it is as sophisticated as the first rule.

Finally, in the fourth system, voters rank a full ballot and the winner is chosen via RCV. In our example, the voter ranks $C \succ A \succ B$.

For these four election rules, I run models with anywhere between 3 and 8 candidates and two different 1,001-voter electorates. The first electorate is a unimodal distribution and the other is a "polarized" bimodal distribution. For the former, I generate 1,001 draws from an $N(0.5, 0.2^2)$ distribution. For the latter, I generate two means, $\mu_l \sim \text{Unif}(0.1, 0.3)$ and $\mu_r \sim \text{Unif}(0.7, 0.9)$. Then, I combine 500 draws from a $N(\mu_l, 0.1^2)$ distribution with 501 from a $N(\mu_r, 0.1^2)$ distribution. By first simulating the means, I let the bimodal distributions be asymmetric. The candidates are randomly drawn from a $\text{Unif}(0, 1)$ distribution; the results are similar if I employ a citizen-candidate model by drawing randomly from the voter pool.

As an example, Figure A.1 shows the results of two separate eight candidate simulations, one of each electorate type. In each panel, the density plot shows the distribution of voters (the top panel is the unimodal electorate and the bottom panel is the polarized electorate). The purple vertical lines represent the candidates' positions in an eight-candidate field. Below each density plot, the red dashed line indicates the location of the median voter, and the labels "FPTP", "FPTP Restr.", "Runoff", and "RCV" indicate the winners under each of those four systems.

For each number of candidates and electorate type, I run 1,000 simulations and record the distance between the median voter and the winner stemming from each of the four election rules. Figure A.2 shows the distance between the median voter and the candidate chosen on average across the 1,000 simulations. Starting with the unimodal electorate in the left panel, the restricted FPTP system performs the worst—the winner it selects is on average 0.17 units away.[2] The other three systems perform comparably. On average across the six unimodal candidate simulations, the runoff winner is 0.003 units closer to the median voter than the RCV winner is, and the RCV winner is 0.009 units closer to the median voter than the FPTP winner is. Nonetheless, the differences in representation are small between FPTP, runoff, and RCV.

Turning to the polarized electorate, we see greater differentiation between the four models. Restricted FPTP again maintains constant degrees of congruence regardless of the candidate count, but in this case, it tends to perform best. Once again, RCV and runoff elections perform comparably. FPTP consistently produces the least congruent winners, which is unsurprising given its likelihood of producing an extreme winner with a mere plurality.

What should we take away from these simulations? Regardless of the model, the dif-

[2]The average distance between the median voter and restricted FPTP is independent of the number of candidates, as the figure shows. To see why this is the case, consider a real-life election in which additional long shot third party candidates enter the field; in that case, those candidates' entry would have no impact on the distance between the median voter and the actual election winner (a major party candidate).

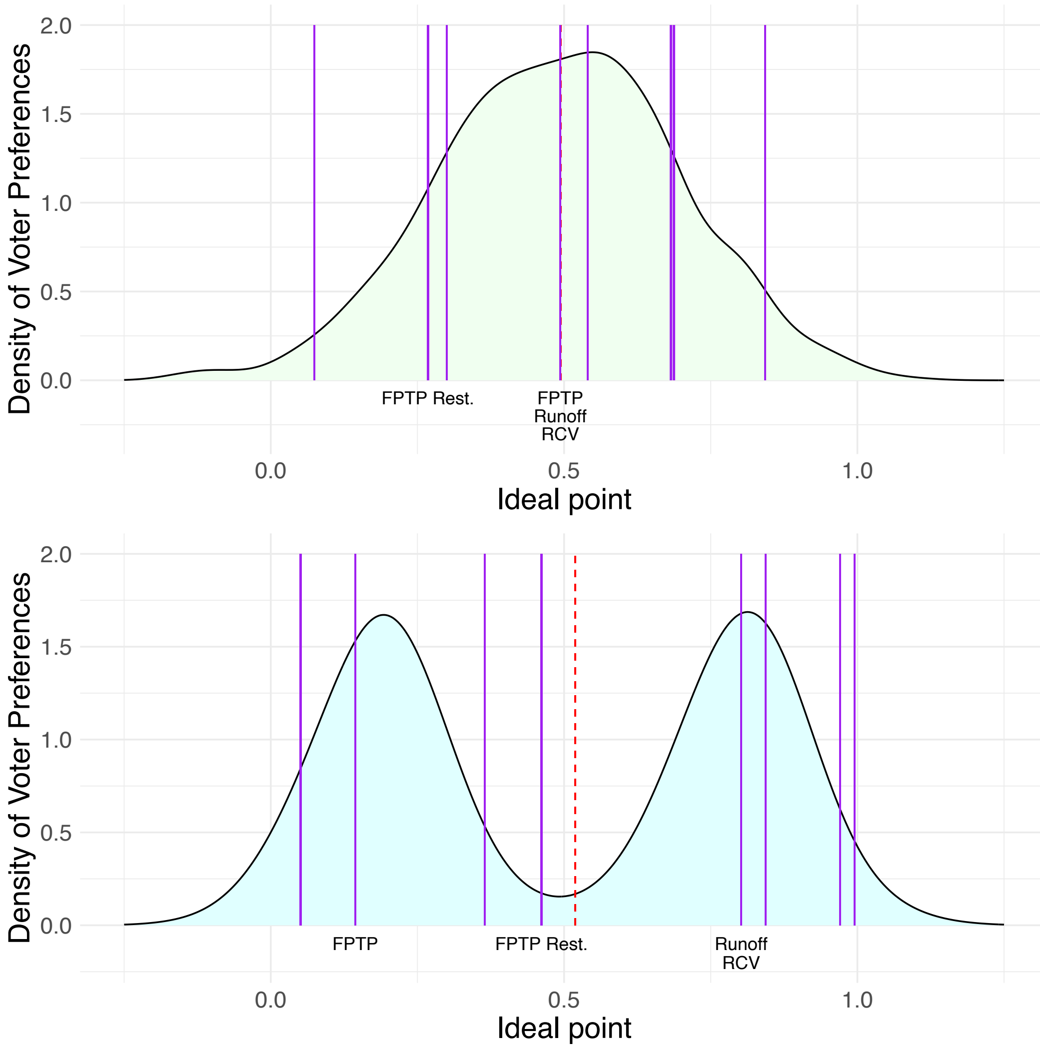

Figure A.1: Two simulations under a unimodal electorate (top panel) and a polarized electorate (bottom panel). The red dashed line marks the location of the median voter.

ferences between RCV and runoff elections are fairly small, so the null findings should not come as much surprise there. There are two sets of beliefs which might lead us to believe RCV will improve over FPTP: 1) the electorate is fundamentally moderate and most voters restrict themselves to choosing between the major-party candidates, or 2) the electorate is fundamentally polarized and most voters vote sincerely for their preferred candidate rather than restrict themselves to strategic choices. If 1) does not hold, however—that is, if the electorate is moderate but voters cast sincere votes—then RCV should make little difference versus FPTP. If 2) does not hold—that is, the electorate is polarized but voters cast strategic votes—then RCV should *worsen* representation. If we move towards the single-round "poll" understanding of the Runoff model in which voters are aware of others' preferences, we can see that the interpretation is further complicated. In the polarized case, for example,

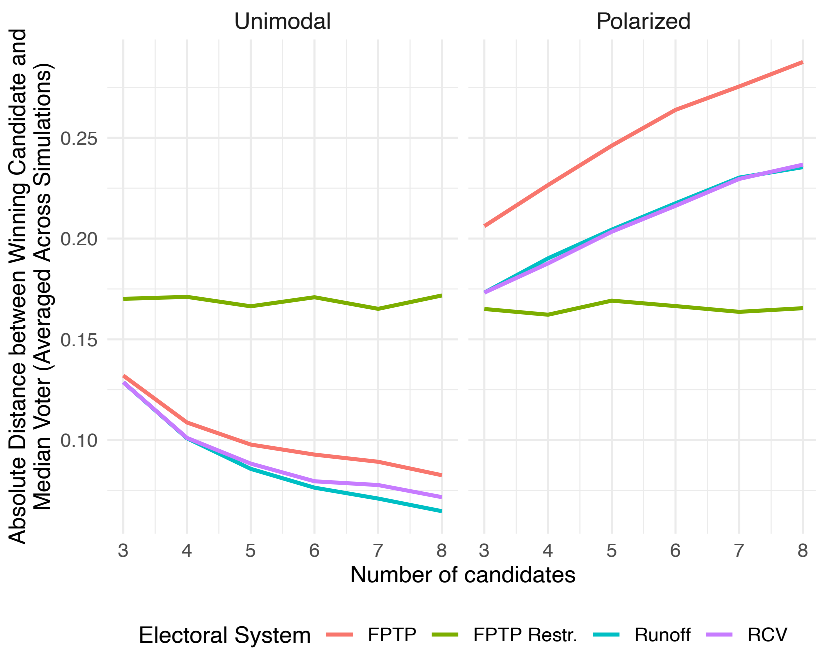

Figure A.2: Success rates by electorate type and number of candidates

it would mean that RCV outperforms FPTP if the electorate is wholly sincere; it loses to FPTP if the electorate is somewhat strategic; and it ties with FPTP if the electorate is very strategic (i.e., able to know and act on everyone else's sincere preferences).

As should be clear by this point, the results of the simulations are sensitive to the specifications—changing the distributions of voters will produce different findings, for example. Perhaps most importantly, it assumes a constant candidate pool. If RCV moderates the candidate pool, then the translation of voter preferences may be improved by that mechanism. While this exercise analyzes the translation of preferences holding these preferences constant, the decision is sensible given the empirical findings in Table G.1 that the candidate pool does not change ideologically post-RCV. In sum, we have no theoretical reason to believe that RCV is necessarily more or less likely to translate voter preferences into median-preferred outcomes.

B Balance Tests

B.1 Covariate Imbalance

Covariate	Ever-treated Mean	Never-treated Mean	Diff.	SE of Diff.	p value
Total Population	169,835	9,572	160,263	10,770	<0.01
Percent Male	48.8%	48.8%	-0.1%	0.1%	0.45
Percent White	60.1%	81.9%	-21.8%	1.1%	<0.01
Percent Black	10.5%	7.4%	3.1%	0.5%	<0.01
Percent Asian	9.7%	1.2%	8.5%	0.5%	<0.01
Percent Hispanic	16.2%	6.9%	9.3%	0.8%	<0.01
Percent College Plus	45.1%	18.4%	26.7%	0.6%	<0.01
Median Income	$54,249	$44,463	$9,785	$896	<0.01
Median Home Value	$338,189	$123,884	$214,304	$11,681	<0.01
Proportion Homeowners	52.7%	72.7%	-20.1%	0.5%	<0.01
Democratic Vote Share (County)	67.3%	42.6%	24.7%	0.6%	<0.01

Table B.1: Raw covariate balance between cities that ever adopted RCV and those that never did. The second and third columns show means for cities that ever adopted RCV and those that never did. The fourth column shows the difference between the second and third columns. The fifth column shows the standard error of the difference. The sixth column shows the p value on the test of equality of group means. Observations are at the city-year level.

Table B.1 shows the raw differences between the cities that ever adopted RCV and those that never did. On every variable other than gender, there are statistically significant differences between the treated cities and untreated cities.

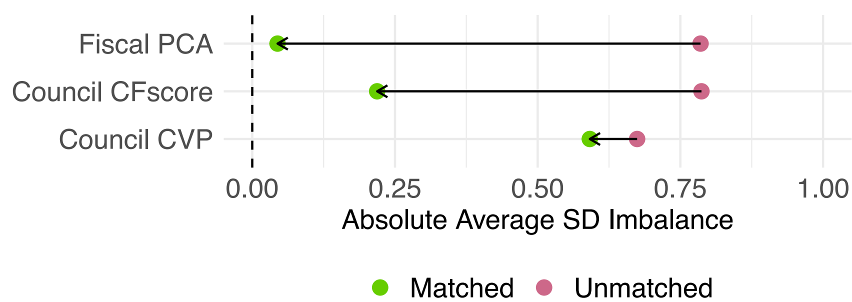

Figure B.1: Covariate balance before and after covariate balancing propensity score weights, spending and ideology variables

Figure B.1 shows the improvement in balance before and after implementing covariate balancing propensity score weights for the spending and ideology variables, averaged across all the covariates. The weights reduce the covariate imbalance by 17.7 times for the fiscal variables and 3.6 times for the city council CFscores. The weights reduce the imbalance by only 1.1 times for the CVP scores, but the lack of reduction here is unsurprising since the CVP scores are not intrinsically scaled across cities; furthermore, there are far fewer control cities to match with in the roll call data. As discussed in the manuscript, the across-city results using roll call data should therefore be considered tentative.

B.2 Imbalance on Other Variables

Variable	Ever-treated Mean	Never-treated Mean	Diff.	SE of Diff.	p value
Standardized 1st PC of fiscal variables	1.151	-0.002365	1.153	0.04692	<0.01
City council conservatism	-0.5004	0.06316	-0.5636	0.02466	<0.01
Representation: standardized 1st PC	-0.5243	-0.5067	-0.01758		0.34
Representation: city council conservatism	-0.335	-0.3501	0.01516		0.35
Holds on-cycle elections	33.3%	21.6%	11.7%	2.7%	<0.01
Has partisan elections	25.0%	16.5%	8.5%	2.5%	<0.01
Has elected mayor	50.0%	37.0%	13.0%	2.8%	<0.01
Has direct democracy	75.0%	69.3%	5.7%	2.5%	0.02
Allowed to have local sales tax	91.7%	62.5%	29.1%	1.6%	<0.01
Has term limits	16.7%	19.2%	-2.5%	2.1%	0.24

Table B.2: Raw balance between cities that ever adopted RCV and those that never did. The second and third columns show means for cities that ever adopted RCV and those that never did. The fourth column shows the difference between the second and third columns. The fifth column shows the standard error of the difference. The sixth column shows the p value on the test of equality of group means. Observations are at the city-year level. For representation variables, p value comes from bootstrapped distribution.

We can also see the imbalance on the key dependent variables and other city-level variables in Table B.2. The adopting cities have higher fiscal expenditure levels and have much more liberal city councils than the other cities in the sample. That said, when looking at the representational differences, they are quite small. The adopting cities are slightly less responsive than the never-treated cities with respect to the fiscal variables, but the difference is not statistically significant. The adopting cities are also a little more responsive in terms of city council ideology, but this difference is also not statistically significant.

I also include other measures of city-level institutional variables. The adopting cities are more likely to hold on-cycle elections, hold partisan elections, elect an executive (mayor), use direct democracy, and be allowed to have local sales taxes. They are less likely than other cities to have term limits, but this difference is not statistically significant.

C Effects of RCV on Individual Fiscal Variables

In this section, I analyze each of the fiscal measures separately rather than as a single composite measure. I exclude the education spending models since not enough treated cities have data for education spending per capita in the sample. I log the taxation and spending per capita measures (De Benedictis-Kessner and Warshaw 2016). Figure C.1 examines the average treatment effects for the full set of municipal revenue and expenditure variables. Since the dependent variables are estimated in logged dollars per capita, the coefficients should be exponentiated to interpret the effect sizes. For example, a coefficient of 0.15 would suggest that RCV increased spending by $e^{0.15} \approx 1.16$, or 16%.

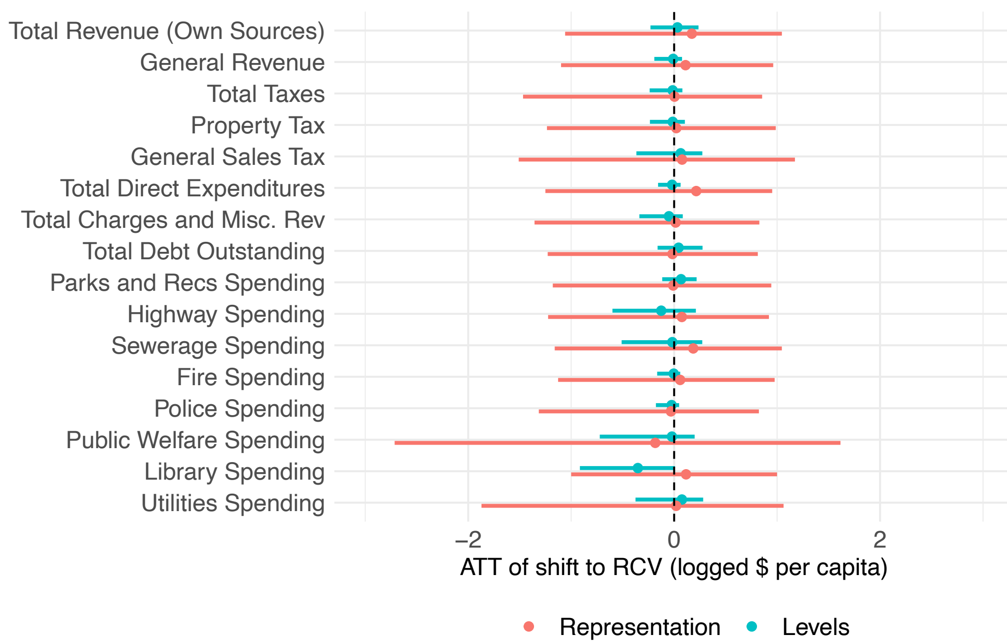

Figure C.1: Coefficient plot of effects of RCV on levels and representation of taxation and spending. The bars indicate bootstrapped 95% confidence intervals.

RCV does not have a statistically significant effect on any of the fiscal variables except for library spending; it appears that RCV reduces library spending ($p = 0.02$). However, after a Holm correction for multiple tests, it appears that the library spending coefficient is not significant—in other words, across sixteen tests, it is likely just a false positive. We also see nulls for representation across all sixteen tests. Although the coefficients are all close to zero, the standard errors are considerably larger here in the representational models, which makes it hard to determine whether the nulls stem from true null effects or an insufficient amount of data. The PCA-based variable used in Table 2 helps to show that fiscal representation in general is unchanged due to RCV.

D Testing Selection Effects at the Legislator Level

In Table 5, I present models with legislator fixed effects in order to test within-legislator change. The legislator-level data used in that analysis, however, also enables us to test whether RCV has effects through replacement at the legislator level.

	Ideology			**Extremism**		
Term	Estimate	CI with clustered errors	Placebo p-value	Estimate	CI with clustered errors	Placebo p-value
Berkeley	-0.02	[-0.11, 0.06]	0.36	-0.06	[-0.12, -0.00]	0.55
Las Cruces	-0.06	[-0.23, 0.11]	0.09	0.01	[-0.12, 0.15]	0.89
Minneapolis	0.12	[0.01, 0.22]	0.00	-0.08	[-0.15, -0.00]	0.47
Oakland	-0.04	[-0.24, 0.17]	0.19	-0.00	[-0.14, 0.13]	0.96
San Francisco	-0.02	[-0.11, 0.07]	0.38	-0.13	[-0.17, -0.08]	0.21
San Leandro	-0.06	[-0.21, 0.10]	0.09	0.08	[-0.03, 0.20]	0.45
Santa Fe	0.04	[-0.11, 0.19]	0.19	0.21	[0.11, 0.31]	0.11
St. Paul	0.04	[-0.24, 0.32]	0.19	-0.03	[-0.20, 0.15]	0.77

Table D.1: Across-legislator estimates of RCV's effects on legislator ideal points with clustered standard errors and p-values from randomization inference

Table D.1 replicates the specification from Table 5 but removes the fixed effects by legislator. Here, we see similar patterns emerge. In some cities, we see more conservative legislators and in others more liberal legislators, but no systematic patterns emerge. The confidence intervals from the models suggest that the Minneapolis city council became more conservative, and the placebo p-value indicates that this is a substantively large change. However, no similar patterns emerge in other cities. It appears that Santa Fe became more extreme while San Francisco, Berkeley, and Minneapolis became less extreme, but none of these effect sizes are distinct from what we observe in control cities as evidenced by the placebo p-values.

	Ideology			**Extremism**		
Term	Estimate	CI with clustered errors	Placebo p-value	Estimate	CI with clustered errors	Placebo p-value
Berkeley	-0.04	[-0.20, 0.12]	0.51	-0.01	[-0.11, 0.10]	0.94
Las Cruces	-0.26	[-0.63, 0.11]	0.03	0.18	[-0.27, 0.62]	0.20
Minneapolis	0.03	[-0.08, 0.13]	0.57	-0.04	[-0.09, 0.02]	0.71
Oakland	-0.14	[-0.49, 0.22]	0.17	0.08	[-0.18, 0.33]	0.57
San Francisco	-0.02	[-0.20, 0.15]	0.57	-0.20	[-0.31, -0.09]	0.17
St. Paul	0.05	[-0.38, 0.48]	0.40	-0.02	[-0.18, 0.14]	0.86

Table D.2: Within-district estimates of RCV's effects on legislator ideal points with clustered standard errors and p-values from randomization inference

To more precisely test the concept of replacement, I run these models but with district level fixed effects (and thus, standard errors clustered at the district level). However, the data identifying members' districts is drawn from the elections dataset, and for legislators who were not merged in, we do not have their districts. As a result, these models exclude San Leandro and Santa Fe (among treated cities) and Madison, Milwaukee, Long Beach, and Chicago (among control cities).

Table D.2 shows that the ideological coefficients are not significant in any of the six cities. However, the coefficient on Las Cruces is substantively meaningful based on the placebo p-value, meaning that legislators' 0.26 point shift leftward was meaningful compared to control cities (although this does not survive a Holm correction). With respect to the extremism results, it appears that San Francisco legislators became less extreme, but the decline is unremarkable compared to other control cities.

E Robustness Checks

E.1 Parallel Trends

Since the main effects are null, there is less of a concern that the parallel trends assumption might be violated. For the assumption to be violated, the cities would already have to be trending in some direction prior to adoption (e.g., representation was improving). One approach to deal with this concern is to include city-level linear time trends, as the second and third out of the three fixed effects specifications in the main article do. Another approach to test the assumption is to employ placebo models with leads of the treatment in order to account for anticipatory effects. I rerun the main across-city models with treatment variables that anticipate the actual treatment date by either one or two years. The results are shown in Figures E.1 and E.2 respectively. The effects are all null, suggesting that the parallel trends assumption is valid.

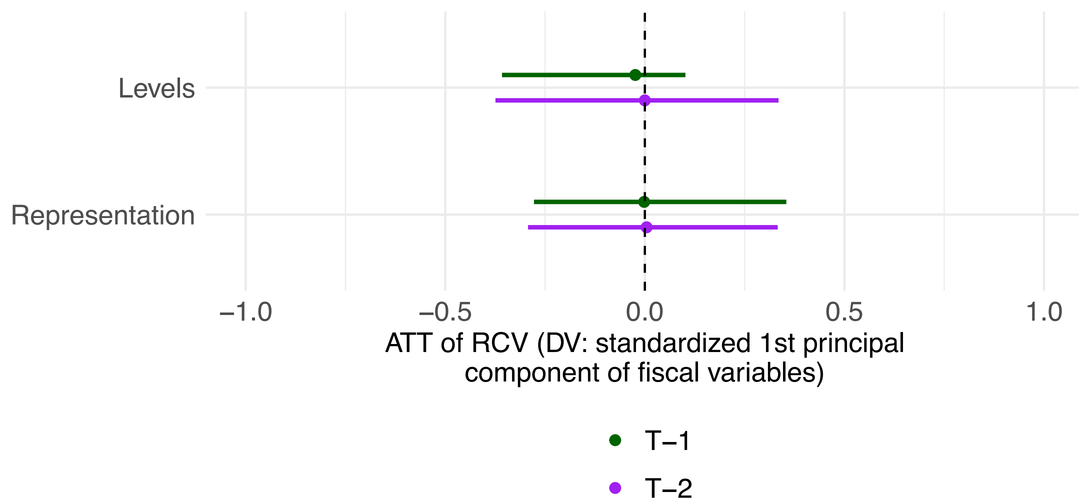

Figure E.1: Coefficient plot of effects of RCV on fiscal levels and representation with one or two leads of the treatment. The bars indicate bootstrapped 95% confidence intervals.

I also run models with the CVP data at the individual legislator data that anticipate treatment by one or two periods. These calculations are not feasible in San Francisco and Minneapolis for two leads because there are only two periods of pre-treatment data at the roll call level. For the most part, we do not see significant effects in Figures E.3 and E.4. However, there are two notable cases where this is not true. In the extremism models in Figure E.4, the coefficients on San Francisco and Santa Fe are significant with one lead. This suggests that the apparent results we find in Table 5 stem from pre-trends.

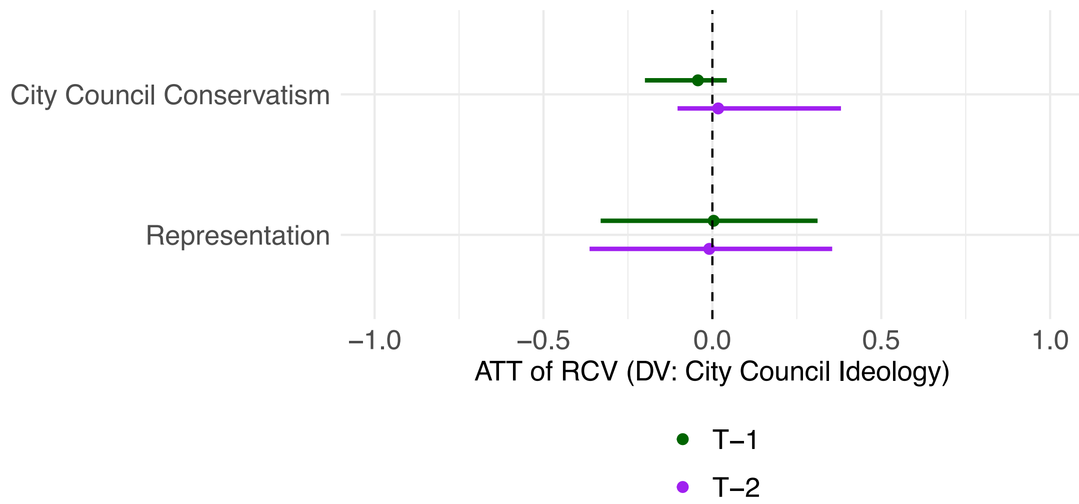

Figure E.2: Coefficient plot of effects of RCV on city council conservatism and representation with one or two leads of the treatment. The bars indicate bootstrapped 95% confidence intervals.

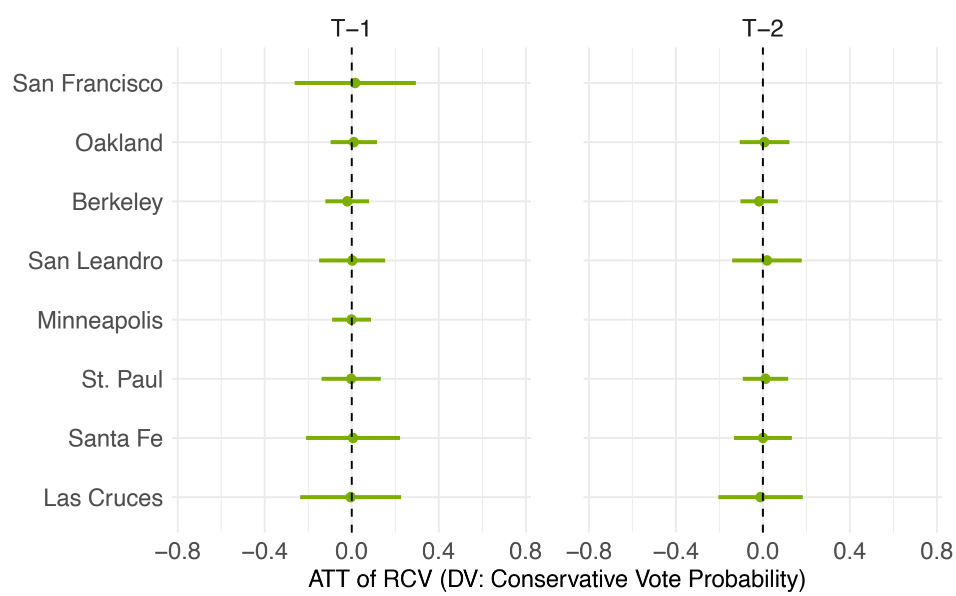

Figure E.3: Coefficient plot of effects of RCV on within-individual shifts of CVP with one or two leads of the treatment. The bars indicate 95% confidence intervals.

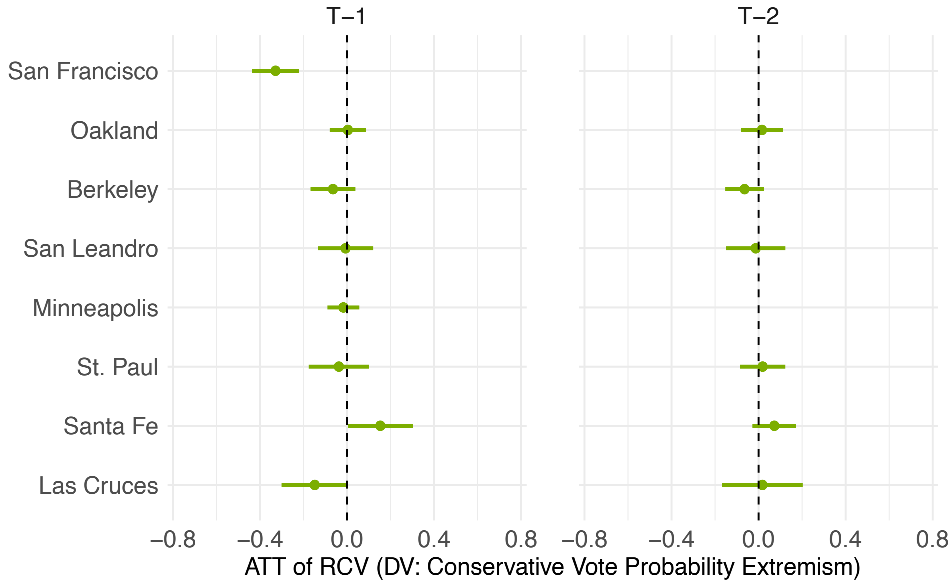

Figure E.4: Coefficient plot of effects of RCV on within-individual shifts of CVP extremism with one or two leads of the treatment. The bars indicate 95% confidence intervals.

E.2 Effects of RCV on Aggregate City Council Ideology using Roll Call Data

In this section, I rerun the city council level models from Table 3 but I use roll call-based ideal points instead of CFscores. Although the models are comparing results within cities, the ideal points are not bridged across cities and so we cannot interpret the results in a common space. Thus, these models should be viewed with caution. Table E.1 shows the results using CVP in Panel A and IRT estimates in Panel B (see Appendix Section E.3 for more detail on the IRT estimation). The four specifications shown are identical to those in Table 3. The effects on CVP range from a 3 point shift leftward to a 2 point shift rightward, and the effect size is significant in one of the models (column 4), suggesting the possibility of a leftward shift. In the IRT models, the first three specifications are not significant but the fourth is. This suggests that RCV may shift ideal points leftward if the fourth specification is correct. In each case, though, the specification just barely achieves statistical significance.

I also use this data to test for polarization. I take the outcome measure as described in Equation 2 and average it at city council-year level. In other words, this model tests whether the average distance to the pre-adoption median increases post-adoption. The results are in Table E.2. Here, we once again do not find any evidence for shifts. The coefficients are largely positive (indicating increased extremity) but none rise to the level of statistical significance.

Table E.1: Effects of RCV on Ideological Outcomes (Using Aggregated Roll-Call Based Ideal Points)

	PanelMatch	FE	FE	FE
DV: City Council CVP				
RCV Election	0.02	-0.02	0.01	-0.03^{*}
	$[-0.02; 0.15]$	$[-0.07; 0.03]$	$[-0.03; 0.05]$	$[-0.06; -0.00]$
DV: City Council IRT				
RCV Election	0.15	-0.03	0.03	-0.27^{*}
	$[-0.25; 0.99]$	$[-0.27; 0.22]$	$[-0.30; 0.36]$	$[-0.54; -0.01]$
City Fixed Effects		✓	✓	
Year Fixed Effects		✓	✓	✓
City-Specific Time Trends			✓	
City-Decade Fixed Effects				✓
City-Decade-Specific Time Trends				✓
N	289	289	289	289

Note: Standard errors clustered at city level in second and third specifications. Standard errors clustered at the city-decade level in the fourth specification. Controls not shown. * indicates $p < 0.05$ (two-tailed tests).

Table E.2: Effects of RCV on Ideological Polarization (Using Aggregated Roll-Call Based Ideal Points)

	PanelMatch	FE	FE	FE
DV: City Council CVP				
RCV Election	0.02	0.00	-0.04	-0.02
	$[-0.01; 0.17]$	$[-0.06; 0.06]$	$[-0.09; 0.02]$	$[-0.07; 0.03]$
DV: City Council IRT				
RCV Election	0.23	0.17	0.14	0.01
	$[-0.26; 0.77]$	$[-0.20; 0.55]$	$[-0.21; 0.50]$	$[-0.23; 0.26]$
City Fixed Effects		✓	✓	
Year Fixed Effects		✓	✓	✓
City-Specific Time Trends			✓	
City-Decade Fixed Effects				✓
City-Decade-Specific Time Trends				✓
N	289	289	289	289

Note: Standard errors clustered at city level in second and third specifications. Standard errors clustered at the city-decade level in the fourth specification. Controls not shown. * indicates $p < 0.05$ (two-tailed tests).

E.3 Within-Legislator Adaptation Using MCMC Dynamic Ideal Points

Although the CVP scores are more easily interpretable, they are not intrinsically bridged over time. As another measure, I estimate dynamic ideal points using an MCMC model with 5,000 simulations (preceded by a burn-in of 4,000 simulations) using the `MCMCdynamicIRT1d` function in the `MCMCpack` package. I thin out the simulations, keeping every tenth one. This gives 500 simulated values of each ideal point. I propagate error as described earlier in the manuscript in analyses using these 500 simulated values. The IRT scores (when averaged across simulations) have an average intra-city correlation with the CVP scores of 0.53, with the CFscores at a rate of 0.43, and with Republican partisanship at a rate of 0.46.

	Ideology			**Extremism**		
Term	Estimate	CI with clustered errors	Placebo p-value	Estimate	CI with clustered errors	Placebo p-value
Berkeley	-0.07	[-0.71, 0.57]	0.77	0.67	[0.23, 1.11]	0.00
Las Cruces	0.03	[-0.61, 0.66]	0.89	0.14	[-0.41, 0.68]	0.32
Minneapolis	0.22	[-0.13, 0.58]	0.40	-0.08	[-0.41, 0.25]	0.51
Oakland	-0.15	[-0.66, 0.37]	0.51	-0.14	[-0.56, 0.28]	0.32
San Francisco	-0.19	[-0.96, 0.57]	0.42	-0.08	[-0.63, 0.48]	0.51
San Leandro	0.25	[-0.34, 0.83]	0.36	-0.20	[-0.73, 0.33]	0.15
Santa Fe	-0.38	[-1.10, 0.33]	0.19	-0.03	[-0.62, 0.55]	0.70
St. Paul	-0.17	[-0.75, 0.42]	0.49	0.05	[-0.47, 0.57]	0.62

Table E.3: Within-legislator estimates of RCV's effects on ideal points with clustered standard errors and p-values from randomization inference, IRT models

Table E.3 shows the equivalent of Table 5 but using the IRT ideal points instead. There do not appear to be any within-individual shifts following RCV adoption with one exception: in Berkeley, legislators became more extreme, which runs contrary to the hypothesized direction. None of the randomization tests are significant except in Berkeley's case, where the degree to which Berkeley legislators became more extreme was statistically significant.

E.4 Alternate City-Decade Specifications

The third fixed effects specification uses city-decade unit fixed effects and city-decade time trends (along with year fixed effects). Following convention, this article operationalizes a "decade" based on the tens place of the year (e.g., 2000-2009 constitutes the '2000s'). I use this to generate city-decade units (e.g., San Francisco in the 2000s, San Francisco in the 2010s). However, this is sensitive to our definition of a decade. The choice of the tens place of the year is an arbitrary artifact for the purposes of these analyses. To see whether the findings are robust to this definition, I construct different versions of the decade variable. In the "D-1" specification, I shift back the endpoints of each decade by a year. For example, 1999-2008 would be a decade, 2009-2018 would be another, and so on. In the "D+1" definition, I shift the endpoints forward such that 2001-2010 is a decade and 2011-2020 is another. Using this approach, I construct six alternate decade definitions shifting the windows up to three years earlier or later.

Table E.4: Effects of RCV on Fiscal Outcomes, City-Decade Fixed Effects (with Different Decade Windows)

	D-3	D-2	D-1	D+0	D+1	D+2	D+3
DV: Fiscal Levels							
RCV Election	-0.06	-0.07	-0.03	-0.04	-0.07	-0.09	-0.03
	$[-0.15; 0.04]$	$[-0.15; 0.01]$	$[-0.11; 0.05]$	$[-0.16; 0.07]$	$[-0.23; 0.09]$	$[-0.23; 0.06]$	$[-0.13; 0.08]$
DV: Fiscal Representation							
RCV Election	0.01	0.00	-0.01	0.00	0.01	-0.00	-0.00
	$[-0.27; 0.30]$	$[-0.28; 0.29]$	$[-0.28; 0.26]$	$[-0.28; 0.29]$	$[-0.29; 0.31]$	$[-0.29; 0.28]$	$[-0.24; 0.23]$
N (Fiscal Levels)	157300	157300	157300	157300	157300	157300	157300
N (Fiscal Representation)	24716	24716	24716	24716	24716	24716	24716

Note: Models include city-decade fixed effects, year fixed effects, and city-decade time trends. Standard errors clustered at the city-decade level in the fourth specification. Controls not shown. * indicates $p < 0.05$ (two-tailed tests).

Table E.5: Effects of RCV on City Council Outcomes, City-Decade Fixed Effects (with Different Decade Windows)

	D-3	D-2	D-1	D+0	D+1	D+2	D+3
DV: City Council CFscore							
RCV Election	0.08	0.12	0.12	0.14	0.00	-0.01	0.04
	$[-0.02; 0.18]$	$[-0.03; 0.26]$	$[-0.02; 0.26]$	$[-0.01; 0.29]$	$[-0.08; 0.09]$	$[-0.09; 0.08]$	$[-0.02; 0.09]$
DV: CFscore Representation							
RCV Election	0.02	-0.00	-0.00	-0.02	0.00	-0.00	0.02
	$[-0.20; 0.23]$	$[-0.25; 0.25]$	$[-0.25; 0.24]$	$[-0.29; 0.26]$	$[-0.27; 0.27]$	$[-0.26; 0.25]$	$[-0.19; 0.23]$
N	7007	7007	7007	7007	7007	7007	7007

Note: Models include city-decade fixed effects, year fixed effects, and city-decade time trends. Standard errors clustered at the city-decade level in the fourth specification. Controls not shown. * indicates $p < 0.05$ (two-tailed tests).

The results are shown in Tables E.4 and E.5. The D+0 column shows the standard decade definition, and so this column replicates the models from column 4 of Tables 2 and 3. The models for fiscal levels are never statistically significant in any of the specifications, although it approaches significance in the "D-2" specification ($p = 0.08$). The fiscal representation models do not approach significance regardless of the decade definition. The ideological models come close to significant effects under the "'D-1" ($p = 0.08$) and "D+0" ($p = 0.06$) specifications but are not close to significance under any of the others. Finally, the representational results appear insignificant regardless of the decade definition. Thus, it does not appear that the null results stem from an arbitrary decade definition.

F Heterogenous Treatment Effects

F.1 Event Study Plots

Another possibility is that the null treatment effects mask heterogeneity in treatment effects over time. For example, it is possible that RCV has short-term effects on outcomes due to the shock in electoral method but then reverts back to expected levels as electoral candidates adapt to the new system. Alternatively, there could be long-term effects but not short-term ones if the policy takes time to display effects.

To test these possibilities, I show event study plots for the main PanelMatch models from Tables 2 and 3. The plots are shown in Figures F.1, F.2, F.3, and F.4. Across all four, there do not appear to be heterogenous treatment effects as a function of treatment time.

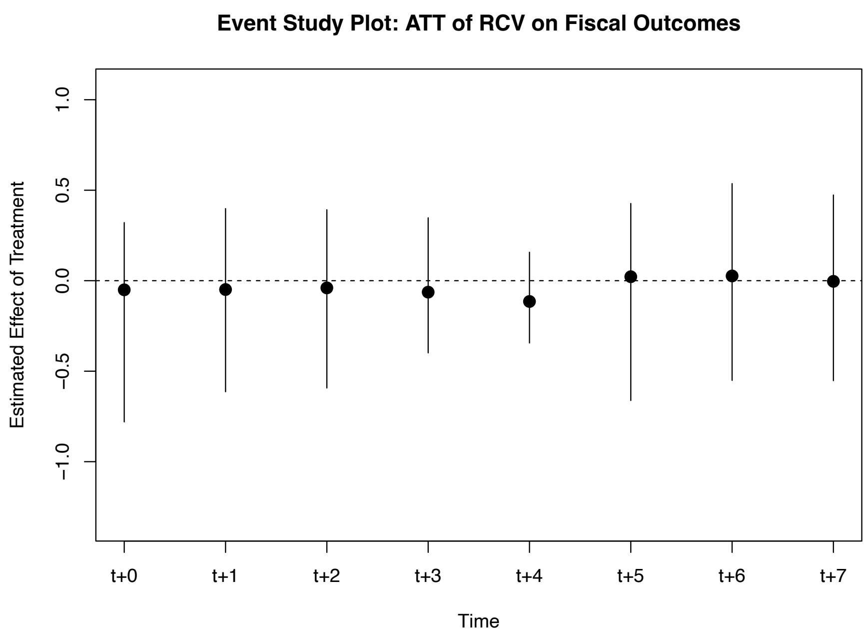

Figure F.1: Event study plot for effects of RCV on fiscal levels. The bars indicate bootstrapped 95% confidence intervals.

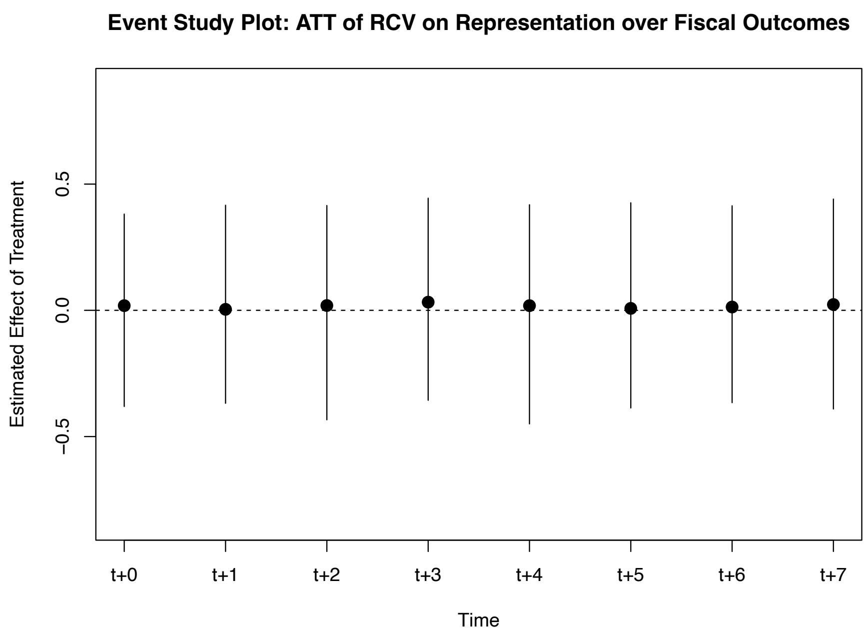

Figure F.2: Event study plot for effects of RCV on fiscal representation. The bars indicate bootstrapped 95% confidence intervals.

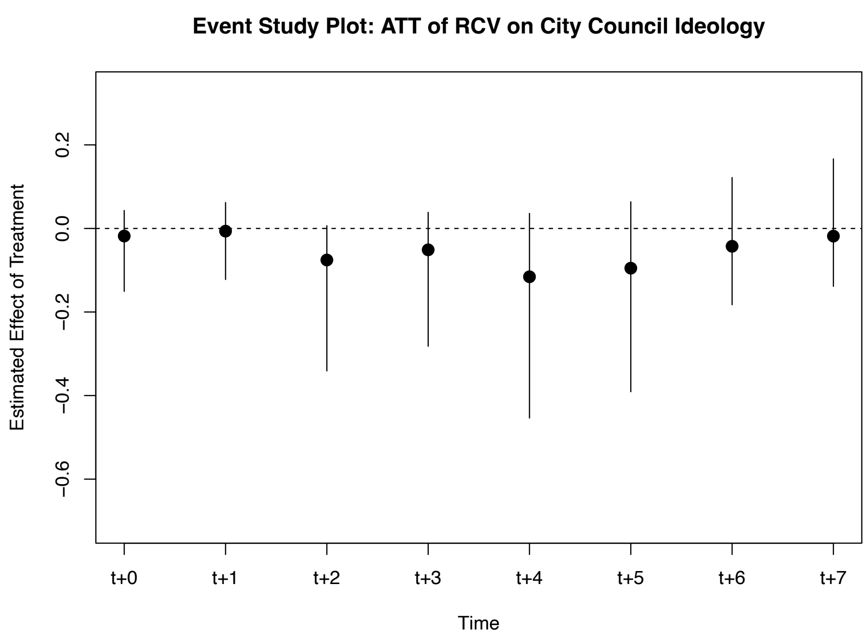

Figure F.3: Event study plot for effects of RCV on ideological levels. The bars indicate bootstrapped 95% confidence intervals.

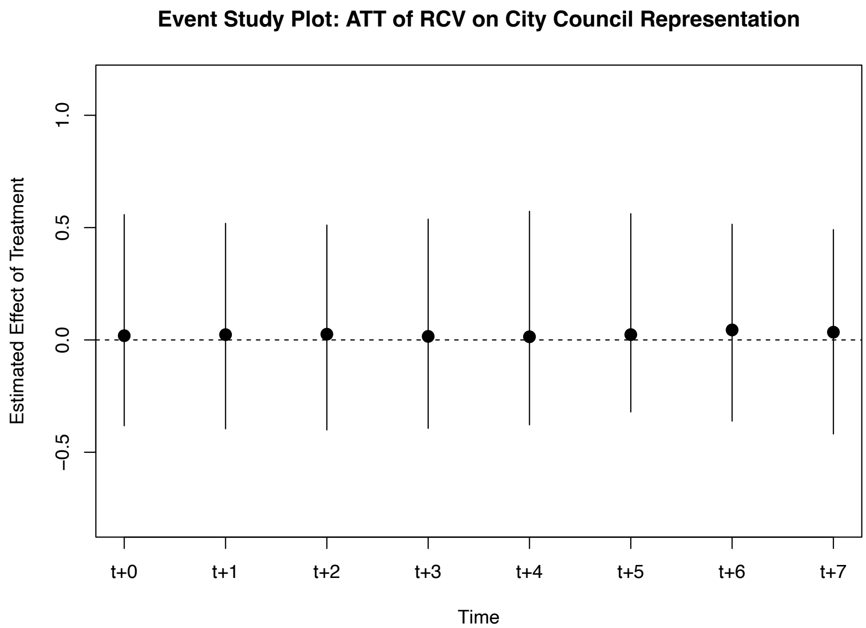

Figure F.4: Event study plot for effects of RCV on ideological representation. The bars indicate bootstrapped 95% confidence intervals.

F.2 Electoral Timing

One possibility is that there are systematic differences in off-cycle versus on-cycle cities, as Dynes, Hartney, and Hayes (2021) suggest. They find that on-cycle cities are more responsive (cross-sectionally) to public opinion than off-cycle cities are. Thus, the null findings in the manuscript might be attributed to heterogeneity between on-cycle and off-cycle cities. If this is the case, we might expect representational gains to be concentrated in off-cycle cities relative to on-cycle cities. There are two reasons this may take place. First, since off-cycle cities start from a lower baseline level of responsiveness, they have more room to improve. Second, it is easier for concentrated interest groups to mobilize in off-cycle elections in favor of their own interests versus the diffuse majority's (Anzia 2011). RCV may force campaigns to appeal more broadly beyond these interests.

I use the coding of cities from Dynes, Hartney, and Hayes (2021) into on-cycle and off-cycle to split the sample into two groups. Of the cities in the sample, 1,182 held off-cycle elections, 329 held on-cycle elections, and there was no data for 18,095 of them. Of the RCV-adopting cities, Berkeley, Oakland, San Francisco, and San Leandro hold elections on-cycle, while Portland, Minneapolis, St. Paul, Cary, and Burlington hold elections off-cycle (there is insufficient fiscal data for San Leandro or city council data for Burlington).

Figures F.5 and F.6 show the results split out by election timing. The effects are null for both on-cycle and off-cycle cities. In addition, the point estimates are quite similar for each type, suggesting that there are no policy differences or representational gains between

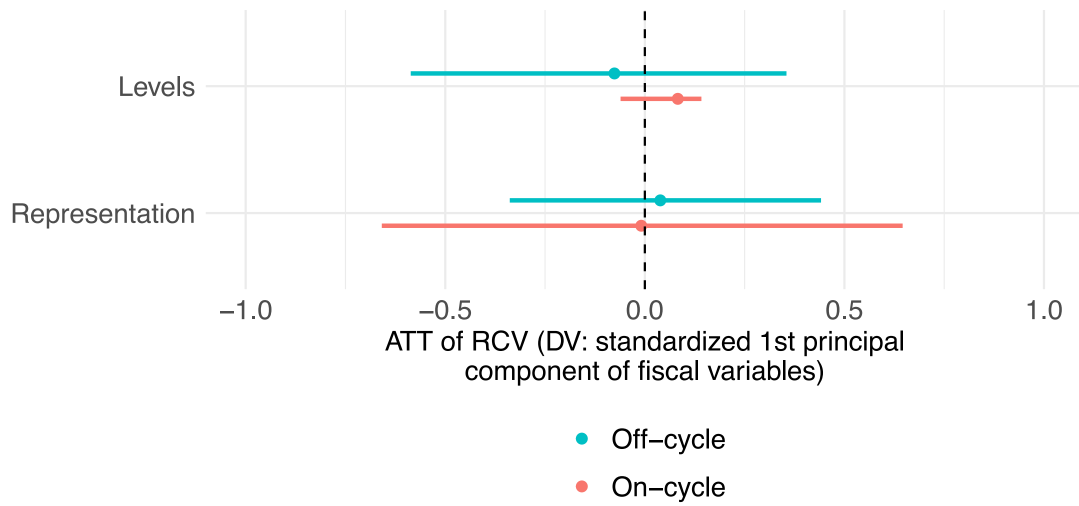

Figure F.5: Coefficient plot of effects of RCV on fiscal levels and representation, separated out by timing of elections. The bars indicate bootstrapped 95% confidence intervals.

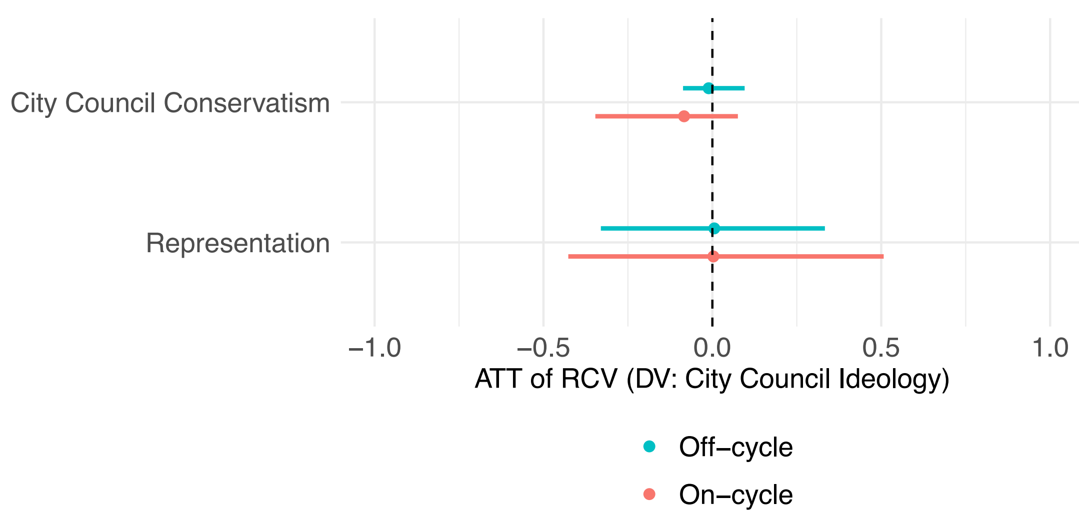

Figure F.6: Coefficient plot of effects of RCV on city council conservatism and representation, separated out by timing of elections. The bars indicate bootstrapped 95% confidence intervals.

RCV-adopting cities in on-cycle locations versus those in off-cycle locations. Of course, this does not mean that shifting to on-cycle elections does not improve representation; rather, it suggests that the (null) effects of RCV appear consistently in both types of timing regimes.

F.3 Nonpartisan versus Partisan Elections

Another important source of institutional variation in local elections is whether the elections are held with partisan labels or not. As discussed in the text, RCV in a partisan context may produce representational gains through partisan selection. Thus, insofar as the findings of this article extend to the state level context, the partisan election context may be a more apt comparison. On the other hand, in a nonpartisan election, there may be a wider range of candidates, which would mean that RCV has more of an ability to discriminate between candidates with divergent platforms. Moderate candidates might also have more success in these nonpartisan elections since they do not have to worry about adopting policies consistent with a party platform. In sum, there are a variety of reasons why we may care about the differences in treatment effects between partisan and nonpartisan elections.

I use the same dataset from Dynes, Hartney, and Hayes (2021), which has information on cities' electoral systems. Of the cities in the sample, 1,223 held non-partisan elections, 242 held partisan elections, and there was no data for 18,141 of them. Of the RCV-adopting cities in the PanelMatch analyses, Berkeley, Oakland, San Francisco, San Leandro, Portland, and Cary have nonpartisan elections while Minneapolis, St. Paul, and Burlington have partisan elections.

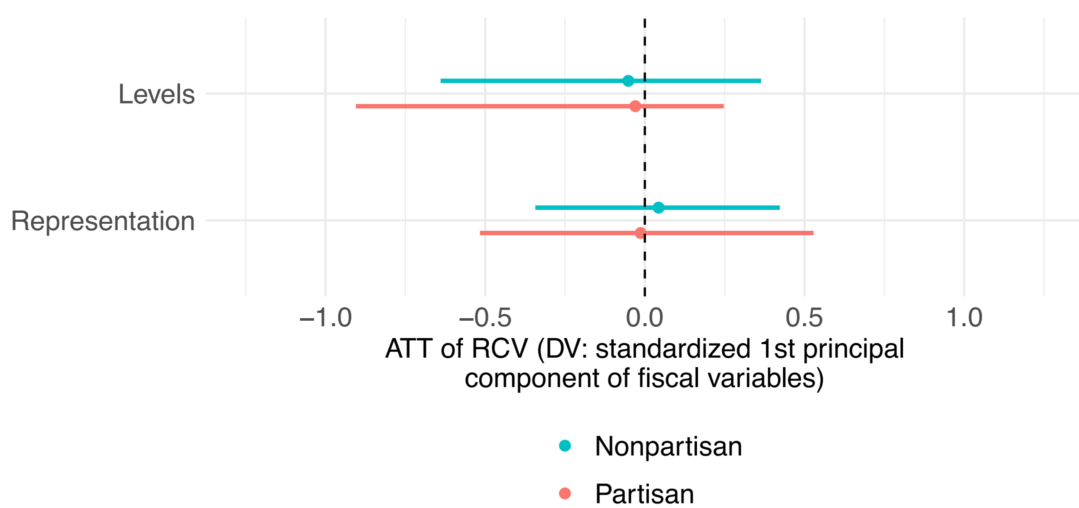

Figure F.7: Coefficient plot of effects of RCV on fiscal levels and representation, separated out by partisan versus nonpartisan elections. The bars indicate bootstrapped 95% confidence intervals.

Figures F.7 and F.8 show the results. With respect to fiscal outcomes, there do not appear to be differences between cities that have partisan elections and those that do not.

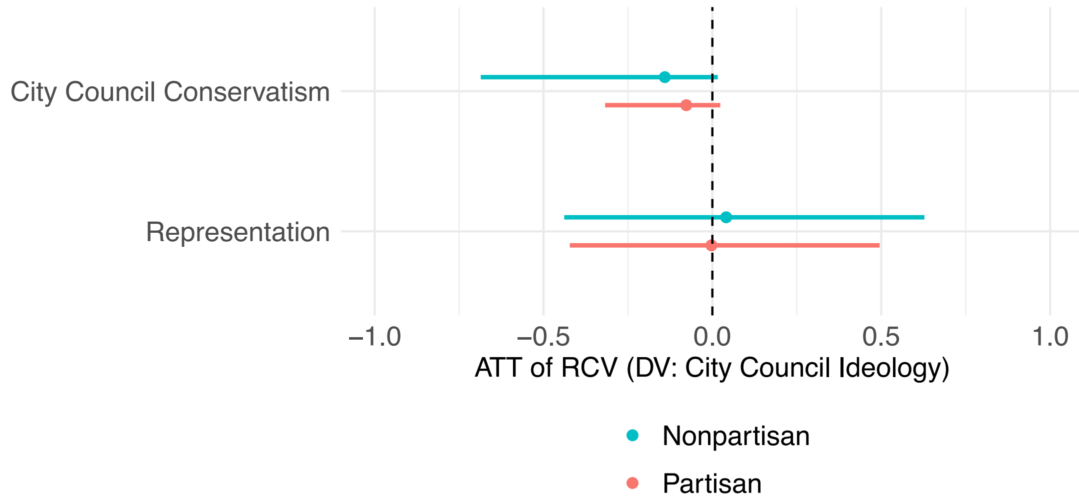

Figure F.8: Coefficient plot of effects of RCV on city council conservatism and representation, separated out by partisan versus nonpartisan elections. The bars indicate bootstrapped 95% confidence intervals.

Turning to the ideological composition of city councils, it does not appear that RCV's effects on city council ideology are different from zero in either electoral regime. The null results also hold in the representational outcomes. Of course, given that there are only three cities with partisan elections, the analysis warrants further investigation on a broader set of partisan elections in cities like New York City.

F.4 Plurality versus Runoff Elections

One final source of institutional variation in local elections is whether elections are held with a plurality versus runoff rules. As discussed in Appendix Section A, we would expect smaller treatment effects on representation in cities switching from runoffs to RCV than cities switching from plurality to RCV.

I code each RCV-adopting city based on their electoral system prior to RCV adoption on the basis of electoral returns and conversations with city government officials. A research assistant and I also classified a large sample of control cities based on their electoral system over the 2000-2020 period from electoral returns and news articles on the basis of whether they use FPTP or runoff rules.[3] In all, I have 302 cities with plurality methods and 188 with runoff methods in the sample. Of the RCV adopting cities, Berkeley, Las Cruces, Santa Fe, Portland, Burlington, Takoma Park, and Telluride previously employed FPTP; on the other hand, San Francisco, Oakland, Minneapolis, St. Paul, San Leandro, Cary, St. Louis Park,

[3]Some cities employ multi-candidate elections. Because similar dynamics are at play, I group together FPTP and plurality block voting cities together as "plurality" rule elections. On the other hand, I classify cities with primary elections in which multiple candidates advance to the general as equivalent to runoffs.

and Hendersonville previously employed runoffs.

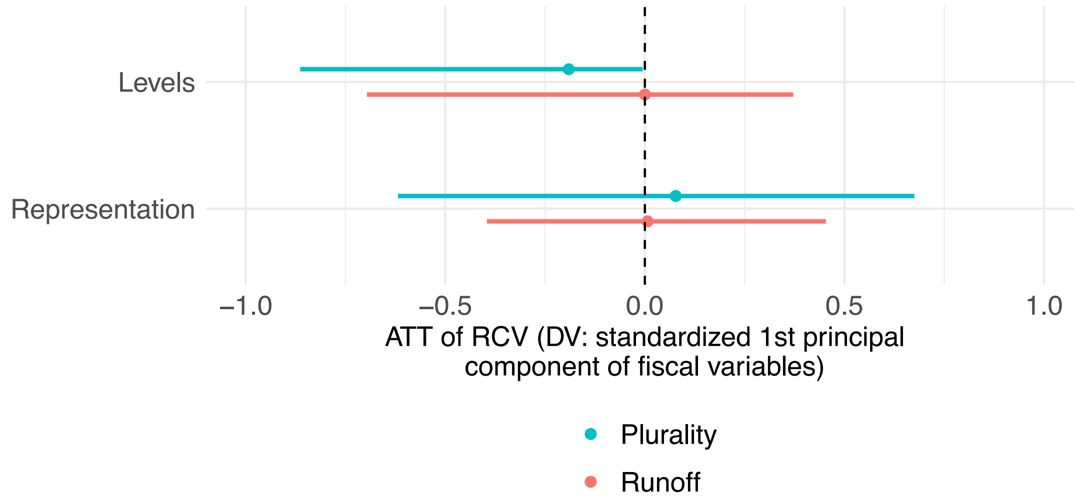

Figure F.9: Coefficient plot of effects of RCV on fiscal levels and representation, separated out by plurality versus runoff elections. The bars indicate bootstrapped 95% confidence intervals.

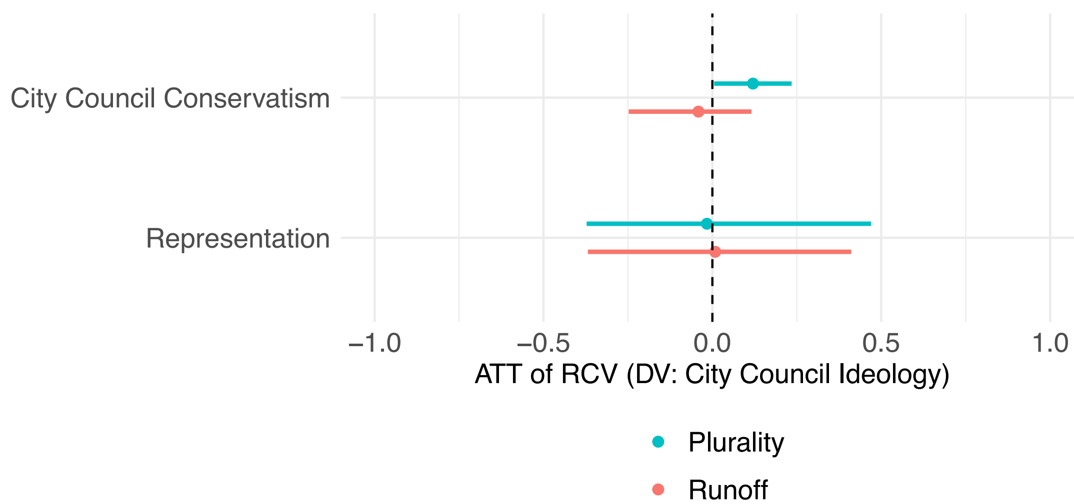

Figure F.10: Coefficient plot of effects of RCV on city council conservatism and representation, separated out by plurality versus runoff elections. The bars indicate bootstrapped 95% confidence intervals.

Figures F.9 and F.10 show the results. We do not see statistically significant effects of RCV in cities which previously employed runoffs in either the fiscal or city council models. However, in the cities that switched from FPTP, we see that fiscal levels declined and city council conservatism increased, although each just barely passes the $\alpha = 0.05$ standard. Despite these rightward shifts, we do not see evidence for gains in representation in moving from FPTP to RCV—suggesting that the method is not bringing FPTP cities any closer in line with public opinion. Furthermore, the apparent effects among FPTP-to-RCV cities are likely a function of the cities which the PanelMatch method drops due to insufficient pre-treatment period data—in the fixed effects specifications (which do not drop these cities), the p-values of treatment on city council conservatism are not significant, suggesting that this effect likely is null as well.

F.5 Variation in Legislator-Level Electoral Competitiveness

The findings in Table 5 suggest that legislators did not change their voting behavior following the adoption of RCV. One possibility is that many legislators did not feel much electoral pressure to adapt because they maintained a strong electoral hold on their seats. If this is the case, then we might expect to see changes in behavior primarily among those who were in close elections.

To test this, I subset the data to all legislators in RCV-adopting cities for whom a) we have roll-call based ideal points in the first period after RCV, b) we have an ideal point prior to RCV, and c) we have her election results prior to RCV. I create a dummy variable that indicates whether the legislator received less than 60% of the vote (out of the share of the top two candidates) in her last election prior to RCV. If so, I mark her as being electorally marginal. I construct two versions of the dependent variable. The first is simply the difference in her CVP between the RCV session and the pre-RCV session. However, it is possible that some marginal legislators move in a more conservative direction and others in a more liberal direction. To account for this, I also construct a dependent variable with the absolute value of the CVP change (akin to the extremism variable described in Equation 2, but the baseline is her own score in the pre-adoption period, not the city average).

Table F.1: Difference-in-Differences Model Between Competitive and Non-Competitive Winners

	Dependent variable:			
	CVP		IRT	
	Change	Abs. Val. of Change	Change	Abs. Val. of Change
Close Election Prior to RCV	−0.008	0.017	0.088	0.054
	(0.050)	(0.026)	(0.211)	(0.126)
City Fixed Effects	Yes	Yes	Yes	Yes
Observations	42	42	42	42
R^2	0.129	0.262	0.169	0.169

Note: Standard errors clustered at the city level. * indicates $p < 0.05$ (two-tailed tests).

Table F.1 shows the results of these regressions. There do not appear to be detectable shifts in the ideal points of legislators following RCV based on whether they previously won a close election or not. Those who previously won in a narrow election have a CVP that is 0.8 points lower than those who won in a less competitive context, but the effect is not statistically significant. It may be that some of the close election winners moved rightward and others moved leftward. However, the second column shows that the average absolute movement in CVP of 1.7 points is also not statistically significant. The models using IRT data also fail to uncover effects.

Thus, it does not appear that RCV is more likely to have an effect on an incumbent after she faces a close challenger. This does not cover other types of competitive races though. RCV might have effects in cases where the incumbent had previously narrowly won against multiple challengers, for example. However, there are only a handful of races that meet this criteria (for example, there are only four races where the first place and third place candidates were separated by less than 15% of the vote), and so we cannot test whether RCV has an impact in such cases. The paucity of races that satisfy this criteria is not surprising. Even at the national and state levels, incumbents rarely face a competitive challenger—let alone multiple. Thus, even though this data does not allow us to test whether RCV has effects in such races, we would only be dealing with a fraction of real-world cases even if RCV did improve representation in such cases. Put differently, in the two types of cases that typify most American elections featuring an incumbent—an uncompetitive race and a race with a single competitive challenger—we do not observe treatment effects.

G Changes in the Candidate Pool

Although RCV does not alter policy outcomes or city council representation, the third theory suggests that it may alter the candidate pool. I conduct four statistical tests to evaluate this claim. I start by looking at whether the mean, variance, or distribution of the candidate pool shifted following adoption using the CFscores for the candidates. Beyond these three tests, I also test whether the number of candidates that run for election (including those that do not have CFscores) changed following RCV adoption. For the analyses in this section, I look at candidates in the twenty-year window around adoption for each city. I only include all cities with data for more than ten candidates both before and after adoption.

City	Diff. in Means	Ratio of Variances	K-S Statistic	No. of candidates per election (post, pre)
Berkeley	0.09	0.41*	0.25	(11.8, 10.2)
Cary	0.07	0.94	0.22	(8.0, 6.9)
Minneapolis	-0.11	3.65*	0.17	(44.3, 25.0)
Oakland	-0.02	1.92	0.22	(16.0, 9.5)
San Francisco	0.08	2.65*	0.25	(40.8, 21.0)
San Leandro				(7.5, 6.8)
Santa Fe				(8.0, 8.8)
St. Paul	0.10	1.2	0.29	(21.7, 15.0)

Table G.1: Tests of changes in the candidate distribution. The second column shows the difference in the post-RCV mean and the pre-RCV mean. The third column shows the variance of the post-RCV candidate pool over the variance of the pre-RCV pool. The fourth column shows the D-statistic from the Kolmogorov-Smirnov test that the pre- and post-RCV distributions are the same. The fifth column shows the number of candidates that ran per election post-RCV and pre-RCV. * indicates $p < 0.05$ (two-tailed tests).

I start by exploring whether the mean of the candidate pool shifted in column 2 of Table G.1. I use two-sample t-tests of the mean of the pool before and after adoption. These tests do not produce any evidence that the mean of the candidate pool changed following adoption. Thus, not only do we fail to conclude that the ideological composition of the winners is unchanged (Table 3), we also cannot detect any ideological shifts in the candidate pool.

While the means do not appear to have shifted, I move now to a more direct test of advocates' claim that the candidate pool moderates following RCV adoption. To do so, I look at the *variance* of the candidate pool. I use a two-sided F-test of the sample variances in candidate ideology within-city before and after adoption of RCV. Under the null hypothesis, this ratio is 1. If the candidate pool has moderated, we should expect to see lower variance after adoption (and thus a ratio less than 1). If the candidate pool becomes more extreme—another possible effect of RCV—we would see a ratio greater than 1.

As column 3 shows, the shifts in candidate pool variance are not significant in three of these six cities. The variance increases in Minneapolis and San Francisco but decreases in Berkeley. While the size of these shifts is notable (from a 365% increase in Minneapolis to a 59% decrease in Berkeley), the results taken as a whole do not point to any systematic shift in ideological variance.

Third, I consider shifts in the distribution of each city's candidates using a Kolmogorov-Smirnov test in column 4. Rather than testing the mean or variance alone, the K-S test explores whether the entire distribution changes. For every city in the sample, we cannot reject the null hypothesis that the distributions do not change.

These tests suggest that RCV does not change the ideological composition of the candidate pool. However, it is possible that RCV encourages *more* candidates to run. Column 5 shows the average number of candidates per election. In general, it appears that more candidates decide to run after adoption, although none of the differences are statistically significant. These changes are substantively meaningful, however; the average number of candidates per election increases by over 50% in Minneapolis, Oakland, and San Francisco following RCV adoption. Even if more candidates decide to run under RCV, however, the other three tests suggest that the new, larger candidate pools are ideologically similar to the pools that existed prior to RCV adoption. It does not appear that voters are achieving more ideological diversity in their candidate choices than they were prior to RCV.