

Voter Learning in State Primary Elections¹

Shigeo Hirano

Department of Political Science
Columbia University

Gabriel S. Lenz

Travers Department of Political Science
University of California, Berkeley

Maksim Pinkovskiy

Department of Economics
Massachusetts Institute of Technology

James M. Snyder, Jr.

Department of Government
Harvard University
and NBER

June, 2012

¹We thank participants at the MIT and Columbia Political Economy Breakfast Seminars, Doug Ahler and David Broockman for comments and suggestions. We would also like to thank Abuchi Agu, Jessica Cui, Mason Fitch, Aaron Kaufman, Karen Li, Margaret Mattes, Laura Mills, Cody Nager, Adria Schulman-Eyink, Jasmine Senior, Nida Vidutis and Miranda Yaver for outstanding research assistance. Hirano and Snyder thank the National Science Foundation for their financial support under grants SES-0617556 and SES-0959200. Pinkovskiy would like to thank the Paul and Daisy Soros Foundation for New Americans for intellectual stimulation, the NSF Graduate Research Fellowship Program and the Institute for Humane Studies for funding.

Abstract

When voters learn about candidates' issue positions during election campaigns, does it affect how they vote? This basic question about voters remains unanswered in part because of a methodological obstacle: learning candidates' issue positions may not only influence voters' vote choice but also their issue positions. To surmount this obstacle, we attempt to answer this question by examining statewide primary elections, which are arguably less vulnerable to this reverse-causation problem because they lack partisan cues and are of much lower salience than presidential elections are. Using both existing polling data and our own panel internet surveys, we find that voters learn about the ideologies of candidates during statewide primary campaigns and that this learning affects their voting decisions in senate and gubernatorial primaries. We fail to find similar results for downballot primaries, raising concerns about voters' ability to make informed judgments for these types of elections.

1. Introduction

Do voters learn about candidates' attributes and issue positions during election campaigns? Does this learning affect how they vote?

The answer to the first question is almost certainly “yes.” Work by numerous scholars shows that voters know more about candidates' relative policy stances and interest group endorsements at the end of campaigns than at the beginning.¹ Summarizing the literature on the 1976 Carter-Ford debates, Sears and Chaffee (1979) found clear evidence of short-term increases in voter knowledge. Alvarez (1997, 2001) finds that voter uncertainty about candidates' issue positions typically falls over the course of presidential and U.S. senate campaigns. Lipsitz (2004) finds that voter knowledge of candidates' issue positions in the 2000 presidential election was higher in battleground states than in non-battleground states. Several studies show that voter knowledge is higher in highly contested U.S. senate and house races than in lopsided races.² Finally, numerous papers find that exposure to news media, advertising, debates, and direct contact with candidates during a campaign is positively related to voter information by election day.³

The answer to the second question is less clear, for two reasons. First, even taken at face value the evidence from previous research is mixed. Many older studies, and as some recent ones, find small or insignificant effects. They also find that most voters decide how to vote before campaigns begin in earnest, or that campaigns mainly reinforce preexisting attitudes.⁴ On the other hand, a number of recent studies find relatively large effects.⁵

Second, methodological issues make the question especially challenging. Ruling out omitted-variable biases and reverse causation is difficult. Consider a typical presidential general election. Suppose a voter prefers one of the candidates (or parties) at the beginning

¹See, e.g., research by Patterson and McClure (1976), Zhao and Bleske (1993), Martinez et al. (2001), Rice (2004), Craig et al. (2005), Henderson (2009), and Lenz (2009).

²Clarke and Evans (1983), Krasno (1994) and Gronke (2000).

³See, e.g., Atkin and Heald (1976), Weaver (1996), Vavreck et al. (2002), Benoit and Hansen (2004), and Druckman (2005).

⁴See, e.g., Lazarsfeld et al. (1948), Berelson et al. (1954), Patterson and McClure (1976), Finkel (1993).

⁵See, e.g., Fan and Tims (1989), Roberts (1992), Bartels (1993), Holbrook (1996), Shaw (1999), Martinez et al. (2001), Iyengar and Simon (2000), and Highton (2006). See also Gelman and King (1993), who argue that the increasing stability of the polls over a campaign is the result of voters learning key facts during the campaign and deciding how to vote on the basis of these facts.

of the campaign for non-policy reasons, such as charisma (or partisan attachment). Suppose also that during the campaign this voter learns about the policy or ideological positions of this candidate (party). Then the voter might adjust her own positions so they are more consistent with those of the candidate (party), or even adopt the candidate’s (party’s) positions outright on some issues. The data would then show a strong relationship between the citizen’s vote choice and the “policy distance” between the voter and the candidate, but this correlation would not reflect any causal effect of policy information on vote choice. This relationship might grow stronger during the campaign, too, as the voter learns more about the candidate’s issue positions and changes her own positions accordingly. The increasingly strong relationship might look like learning that affects voting decisions, but, again, it would be spurious. These same patterns could also arise if, rather than shifting her policy stances, the voter selectively retains information that supports her prior positive opinion of the candidate and disregards other information. Numerous studies find that individuals do, in fact, tend to adopt policy positions consistent with their partisanship, the positions of their preferred candidate, or other predispositions.⁶

This paper attacks the questions above by focusing on an under-explored type of election: non-presidential primary elections. In particular, we study primary elections for governorships, U.S. senate seats, and state downballot offices. The vast majority of previous research has studied presidential elections, and the small number of non-presidential studies all focus on general elections.⁷

What are the advantages of studying state primaries? First, except for incumbents most voters probably have very little information about the candidates before campaigning begins. Thus, the potential for learning to occur is large.⁸ Second, the partisan version of reverse-causality discussed above is not likely to be a problem, because the elections are *intra*-party

⁶See, e.g., Berelson et al. (1954), Campbell et al. (1960), Abramowitz (1978), Zaller (1992, 1994), Bartels (2002a, 2002b), Carsey and Layman (2006), and Lenz (2009).

⁷To our knowledge, only one previous paper has examined ideological voting in state primary elections. Tedin and Murray (1981) study the 1978 Texas state races for governor and attorney general. However, they only look at sorting on ideology in a single primary. Their results are consistent with ours – they find sorting in the governor’s race but not in the attorney general race.

⁸In general, this might or might not be an advantage. However, given the strength of the “minimal effects” position in the literature, it is reasonable to look for situations where the potential for leaning is large.

contests. Third, the relatively low salience of the offices sought – at least compared to the presidency – means that the candidate-centered version of reverse causality is also unlikely to be a serious problem. How many people in California changed their ideological positions because of their high (or low) opinions of Bill Simon, Pete Wilson, or Leo McCarthy? We guess the number is very small, most likely close to zero, a guess confirmed by panel analyses we discuss below.

Finally, there are important reasons to study non-presidential primaries for their own sake. One reason is that it is not obvious that we should even have them. Do voters learn enough during the campaigns for these offices to vote in a reasonably informed manner? As noted, state primary elections are low-information affairs, so it would not be surprising if most voters know very little about the candidates even by the end of the campaign. On the other hand, turnout in these elections is low, so perhaps the citizens who do vote are especially interested, attentive, and informed.

Our findings are straightforward. First, using polling data we find evidence consistent with the claim that voters learn candidates’ ideological positions over the course of a campaign and that this learning affects their vote choice. During primary elections without an incumbent, conservative (liberal) respondents increase their support for the conservative (liberal) candidate in surveys conducted close to the primary election date compared to in the surveys conducted early in the primary campaign by more than liberal (conservative) respondents do.

Consistent with this result, we also find evidence that the magnitude of the ideological gap between the candidates, voters’ initial beliefs about the candidates, and the amount of attention given to campaigns appear to also affect the degree to which voters learn about candidates’ ideologies. The amount that voters learn is positively related to the size of the ideological gap between the candidates as well as the degree to which voters had incorrect beliefs about the candidates’ ideologies. We find little effect for “placebo” races – i.e. races with little ideological gap between the candidates – or races with an incumbent – i.e. races where voters are likely to have more information about at least one candidate before the primary campaign. However, these patterns we observe are only significant for top-of-the-

ticket races, i.e. gubernatorial and senate. The pattern is not significant when we focus on downballot races, which is perhaps not surprising given the difference in campaign resources and media attention given to the different races.

Finally, we confirm our findings using panel surveys in which we recruited participants online and interviewed them before and after 2010 primary races. Although the resulting samples are not representative, the overall patterns they exhibit are similar to those in the polling data; and they allow us to conduct important supplementary analyses. First, since we asked participants about their perceptions of candidates' ideological stances, we are able to measure who learns the positions and who does not. Consistent with our polling result, we find that individuals who learn these positions are indeed those who shift their support to the candidate who, they have just learned, agrees with them ideologically. Second, the panel data allow us to further rule out the reverse causation alternative. Indeed, we show that ideology influences vote change even when we measure ideology before voters learn. The panel data also allow us to directly test whether for reverse causation, that is, whether prior vote influences later change in participants' ideologies. Unlike findings in presidential elections, we find no evidence that participants' views about primary candidates shape their ideologies.⁹

Taken together, our findings are reassuring, implying that campaigns are fulfilling their role in senate and gubernatorial primaries. In these races, primary voters are learning policy relevant information and incorporating that information into their voting decisions. On the other hand, our downballot findings are potentially worrisome for democracy, as we fail to find a similar pattern. Voters may therefore be learning too little in these races to make reasonably informed voting decisions.

⁹The panel findings are also more consistent with a learning explanation than a priming explanation (Iyengar and Kinder 1987; Johnston, Blais, Brady, and Crête 1992), as much of the increase in candidate-voter ideological congruence occurs among voters who learned these ideologies. Of course, priming and learning could be occurring simultaneously.

2. Data and Specifications

2.1. Data Sources

Our data comes from three sources. First, we use data from the California/Field Poll for Democratic and Republican primaries in California between 1966 and 2006. The Field Poll is conducted approximately 8 times during every election year in California, with 3-4 of the polls asking questions about primary campaigns. Besides asking its respondents about their voting intentions in the primaries, the Field Poll also collects a large set of richly coded covariates for each respondent, including ideology (coded on a 5-, 7- or even 9-point scale), age, race, gender, income, education and area of residence within California. The Field Poll releases all of its unit record data to the public, so we will be able to make maximal use of these covariates.

Our second data source comes from the press releases of the survey research firm Survey USA (hereafter, SUSA). SUSA has conducted election polls since 1992, but we have found publicly available releases of polls for primary elections in a usable format since 2005. We look at Republican and Democratic primaries in various states since 2005. The publicly available data consists of cross tabulations of ideology (coded on a 3-point scale) and other covariates (usually age, gender, race, income and region of state are included), so we cannot exploit the data as much as we can for the Field Polls. We study all of the available primaries in either source of data for the offices of governor and senator that do not include incumbents, and where we have at least two polls before the primary date (see Table A.1 for a list of these races). We also study some primaries with incumbents and primaries for downballot offices such as lieutenant governor and attorney general.

Our third data source is panel surveys we conducted during the 2010 primaries. We surveyed likely primary voters in Michigan, Pennsylvania, and New Hampshire before and after their gubernatorial and senate primaries, with 1,201 complete panel respondents. Before Election Day, we recruited participants with online ads and through a consumer internet panel (for details, see Appendix D). We then recontacted participants after the election and reinterviewed 60% to 75% of the original respondents. The surveys asked about perceptions

of the candidates' ideologies and about each participant's ideology, vote, party identification, age, race, gender, education, political interests, and political knowledge.

2.2. Specifications

We assume the following simple model. Consider a primary election with only two candidates: one who is more conservative and one who is more liberal. Also suppose that there are only two types of voters in the primary electorate: relatively conservative and relatively liberal. Finally, assume that there have been exactly two polls taken during the primary: an early poll and a late poll. Let $t=0$ denote the early poll and let $t=1$ denote the late poll, and let L_{it} be the indicator for the late poll ($t=1$). Let V_{it} be the preference of respondent i at time t , with $V_{it}=1$ if respondent i prefers the more conservative candidate and $V_{it}=0$ if respondent i prefers the more liberal candidate. Let C_{it} be the respondent's ideology, with $C_{it}=1$ if the respondent is relatively conservative and $C_{it}=0$ if the respondent is relatively liberal. To look at the relative change in support for the conservative candidate among conservatives, we estimate the following regression model:

$$V_{it} = \alpha + \beta_C C_{it} + \beta_L L_{it} + \gamma C_{it} L_{it} + \epsilon_{it} \quad (1)$$

Each coefficient has an intuitive interpretation. The intercept α is the fraction of liberal voters intending to vote for the conservative candidate in the early poll. The coefficient on ideology, β_C , is the amount by which the conservative voters' support for the conservative candidate in the early poll is greater than the liberal voters' support for this candidate. The coefficient on the late indicator, β_L , is the amount by which liberal voters increase their support for the conservative candidate from the early to the late poll. Finally, the interaction coefficient γ is the amount by which conservative voters increase their support for the conservative candidate over the course of the campaign *over and above* the amount by which liberal voters increase their support for the conservative candidate.

Our quantity of interest is precisely this coefficient γ . Under the assumption that the only differential trends in support for the conservative candidate among conservatives and liberals arise from voter learning, the coefficient γ also is a measure of how much voters'

voting intentions change because of information that they learn during the campaign. If our hypothesis that campaigns inform voters about candidates' positions in ways that change how voters vote is correct, we should observe $\gamma > 0$; that is, conservatives should increase their support for the conservative candidate over the course of the campaign more than liberals do.

2.3. Classifying Ideological Positions and Timing of Polls

In order to implement this specification empirically, we need to match the rich data on primaries that we have, with multiple polls and many ideological shades, to the stylized description of a primary that we present above. In particular, we need to identify candidates as either liberal or conservative. We also need to classify respondents as either liberal or conservative. Finally, we divide the polls between those that are “early” versus those that are “late.”

For all gubernatorial and senate primaries, we classified candidates as either liberal or conservative using information from newspaper articles and interest group endorsements. In most of the races, this distinction is relatively clear. When there were more than two candidates in a race, we focus on the two with the largest ideological difference who were also among the top three vote getters. In Appendix Table A.1 we list the 37 top-of-the-ticket primaries analyzed, and the candidates identified as being relatively liberal or conservative.

Even though we were able to easily identify candidates' relative ideological positions from various sources, we confirmed our classification using independent coders of candidates' positions based on newspaper coverage of the primary campaigns. Using an automated process, we collected sentences likely to identify candidates' issue positions from newspapers available on Factiva during the three months prior to the primary election. The sentences collected had to mention at least one of the candidates together with certain key words or phrases related to an issue or ideological position, such as abortion, tax cut, etc. (see Appendix A for the full list).¹⁰ For each sentence, the coders attempted to identify the

¹⁰These keywords capture issues of broad national interest and may miss state specific issues. To the extent they do, however, the resulting measurement error will likely lead us to understate learning effects, not overstate them.

ideological position of candidates mentioned in the sentence on a five-point scale based only on the information conveyed within each sentence. Because of limitations in Factiva coverage, this independent coding only covered two-thirds (27 out of 37) of the races we originally coded. Among these two-thirds, however, the independent coders confirmed our liberal / conservative classifications of the candidates in all but one case.¹¹ We describe the procedure for coding the sentences and computing ideology gaps in Appendix A.

The newspaper sentence coding also provides us a quantification of relative candidate ideological position. We use this measure to divide primaries into a “baseline” sample of races with acute ideological distinctions and a “placebo” sample of races where we would expect little ideological learning because of minimal ideological distinctions between candidates. We considering primaries with an ideology gap of 0.5 or more (on the five-point scale) to be “baseline,” and primaries with gaps less than 0.5 to be “placebo” (the results are similar with other cutoffs). For the one third of races that lack independent ratings, we rely on the our classifications.

In our main analyses, we classify voters as either liberal or conservative using their self-reported ideological position in the surveys. For most of our analyses we discard voters in the ideological center of each party in order to focus on voters with relatively distinct self-identified ideological positions. In Democratic primaries, all self-identified conservatives are classed as conservative, and strong liberals (or all liberals on a 3-point scale) are classed as liberal. Weak liberals and moderates are dropped from analyses of the Democratic primaries. In Republican primaries, strong conservatives (or all conservatives on a 3-point scale) are classed as conservatives and all self-identified moderates and liberals are classed as liberal. In these primaries, we are unable to study liberals separately from self-identified moderates, since so few Republican respondents identify themselves as liberal.¹²

We examine the robustness of our findings to whether or not all respondents are included in the sample – i.e. considering all self-declared conservatives to be conservative in Republi-

¹¹The only race where we disagreed with the coders is the 2006 Michigan Republican Senator race, between Bouchard and Butler. We code Bouchard as conservative, while the coders code Butler as conservative. The ideology gap in this race is 0.08, which is one of the smallest ideology gaps for all coded races (it is at the 5th percentile of the ideology gaps).

¹²For the polls using a five point ideology scale weak conservatives are dropped from the analysis of Republicans. Both weak liberals and moderates are dropped from the analysis of Democrats.

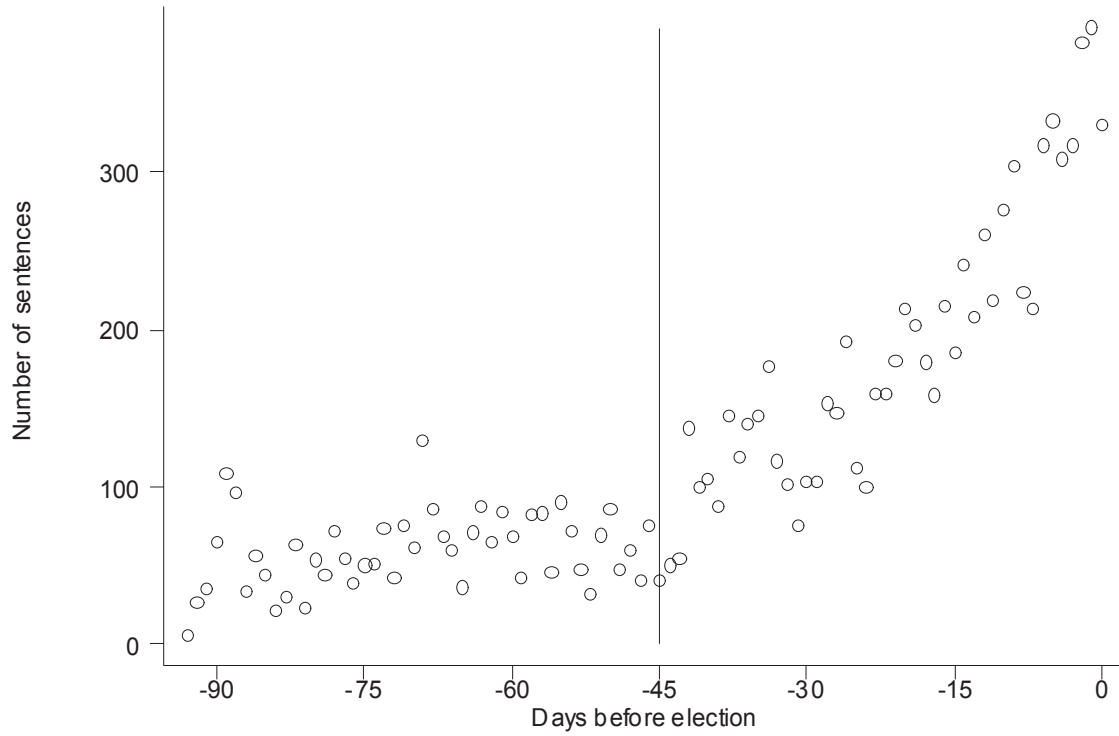
can primaries, all self-declared conservatives and moderates to be conservative in Democratic primaries, and all other voters to be classed as liberal. We also examine whether dropping moderate voters in Democratic primaries rather than classifying them as conservative affects our findings. We do this because Democratic primaries tend to have large blocks of moderate voters – often forming the majority of Democratic voters – that occupy the center of their party. The main substantive findings remain relatively unchanged to these different means of classifying voters as either liberal or conservative.

Finally, in our main analysis we focus on only the first (early) and the last (late) polls. We would expect that the difference between these polls to best reflect the amount of learning that has occurred during the campaign. We also examine whether our results are sensitive to the exclusion of the intermediate polls or the variation in the dates of the first and last polls across primary races. More specifically, we conducted additional analyses where we classify all polls 45 days or more before the primary as early, and all polls less than 45 days before the primary as late. The *45-Day-Cutoff* procedure is motivated by Figure 1, which plots the number of sentences about the primary per day as a function of the number of days remaining before the primary. As the graph reveals, the coverage remains flat until about 45 days before the primary, at which point it rises linearly until the primary date.¹³

¹³We also tried a *30-Day-Cutoff* and the results were substantively unchanged.

Figure 1

Newspaper Articles Covering Primaries Over Time



Note: This figure plots the number of sentences about primaries per day as a function of the number of days remaining before the primary. The source is the same as that described used for the ideological content coding, described in the appendix.

Note that we must drop from the analysis all respondents who do not state a preference for any candidates in a poll. Since there are more of these respondents in early polls than in late polls, and since these respondents tend to be less well informed in the early polls, our estimates likely understate the total amount of learning, as well as the overall impact of learning on voters’ choices. We can address this issue directly using the data from our panel surveys. We discuss this in more detail in Section 5.

3. Results for Top-of-the-Ticket Races Using Polling Data

In this section, we focus on primaries for governor and senator. These tend to be the statewide primary elections most salient to voters. Moreover, given the amount of newspaper coverage and campaign expenditures, voters are more likely to be exposed to information about candidates for these offices as compared to downballot offices over the course of the primary campaign. Thus, we would expect considerable voter learning about the candidates. Moreover, we use the information in the substantial newspaper coverage of these races to quantify the magnitude of the ideological gap between the candidates. With this measure we examine whether the size of the ideological gaps between candidates corresponds with the amount of information that voters “learn” over the course of the primary campaign.

3.1. Main Results

Do liberal voters shift their support towards liberal candidates and conservative voters shift their support towards conservative candidates over the course of an electoral campaign? To answer these questions, we estimate a saturated version of equation 1 in which we allow all coefficients to vary by primary:

$$V_{ijt} = \alpha_j + \beta_{jC}C_{ijt} + \beta_{jL}L_{ijt} + \gamma_j C_{ijt}L_{ijt} + \epsilon_{ijt} \quad (2)$$

with j indexing primaries.¹⁴ The statistic of interest – how much conservative voters shift towards conservative candidates relative to liberal voters – is the average interaction coeffi-

¹⁴Note that to run this regression, the sufficient statistics are the crosstabs of voter ideology across vote choice for each poll since all variables are binary. Hence, we can run this regression equally well for the SUSA data as for the Field Poll data.

cient:

$$\gamma_{\text{avg}} := \frac{1}{n} \sum_{j=1}^n \gamma_j \tag{3}$$

In Table 1 we present our estimate of γ_{avg} for the 28 “baseline” gubernatorial and senate primary races – i.e. races where campaigns should especially help citizens differentiate between the candidates because they lack incumbent candidates and because the candidates differed ideologically (the independent coders found an ideological difference between the candidates larger than 0.5 on a five-point scale). The table also presents the estimates for races where we would expect electoral campaigns to have less of an effect on the congruence between the ideological position of voters and the ideological positions of the candidates they support – i.e. races where the candidates do not differ in their ideological positions or races with an incumbent.

The first column of Table 1 finds that citizens do indeed appear to learn in these “baseline” races. The average interaction coefficient is positive and statistically significant. Moreover, the coefficient is large in magnitude: 0.10. Conservatives increase their support for conservative candidate over the course of the campaign by an average of 10 percentage points more than liberals do. The congruence thus increases between the ideological position of the voters and the candidates they support. Figure 2 presents a histogram of the interaction coefficients for each of these races (i.e. γ_j). It reveals that the average 0.10 effect reflects the central tendency of the distribution and is not skewed by outliers.

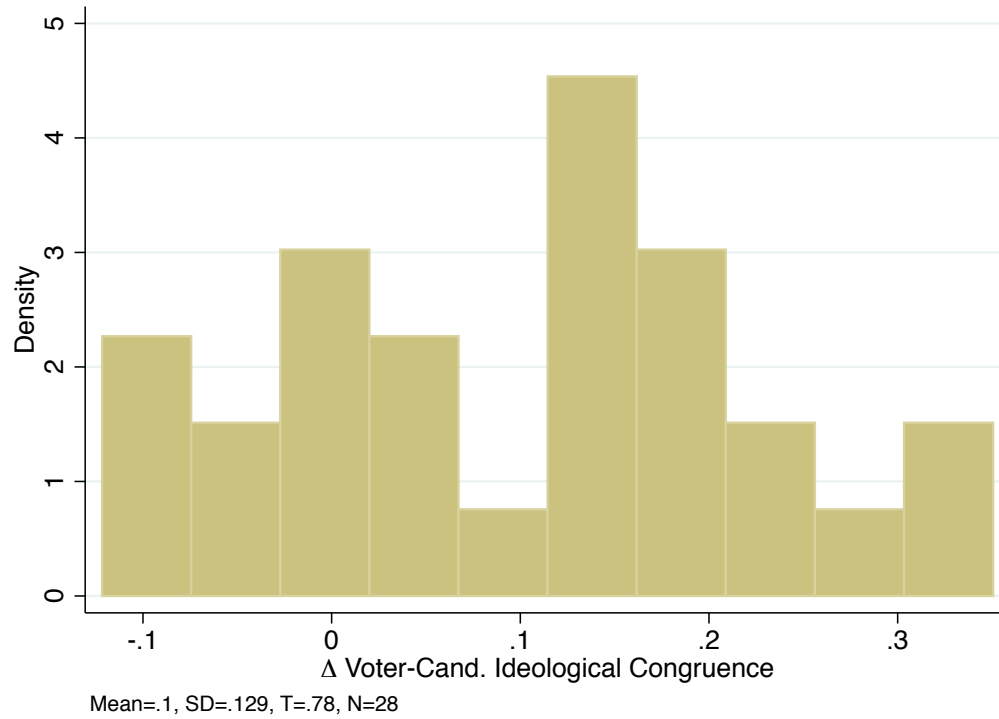
Table 1: Gubernatorial and Senate Races

	Baseline	Placebo	Incumbents
Δ Voter-Cand. Ideological Congruence (γ_{avg})	0.100** (0.021)	-0.036 (0.034)	0.022 (0.044)
Observations	12,099	4723	2894
Number of Primaries	28	9	5
P-value	3.39e-08	0.774	0.199

Robust standard errors in parentheses. P-value is for the F-statistic for testing that all interaction coefficients are zero. Δ Voter-Candidate Ideological Congruence is the average interaction coefficient γ and estimates the amount by which conservative voters increase their support for the conservative candidate over the course of the campaign *over and above* the amount by which liberal voters increase their support for the conservative candidate. ** p<0.01, * p<0.05, + p<0.1

Figure 2

Histogram of Δ Voter-Candidate Ideological Congruence Coefficients



Notes: Δ Voter-Candidate Ideological Congruence is the interaction coefficient γ and estimates the amount by which conservative voters increase their support for the conservative candidate over the course of the campaign *over and above* the amount by which liberal voters increase their support for the conservative candidate. The histogram shows these coefficients for baseline races.

The first estimate uses the Extremist Only classification and the first and last poll, but the estimate is robust to other coding decisions. In results reported in Tables B.1 and B.2, we find a similar effect when we include voters who classify themselves as “moderate” and include all polls but divide them between those taken more or less than 45 days prior to the primary. Not surprisingly, the effect attenuates slightly as we expand the sample to include moderate voters and/or intermediate polls. We also find some variation across surveys and across political parties. The γ_{avg} are weakest for the SUSA polls for the Democrats, which may reflect the weakening divisions within Democratic party in recent years.

According to these findings, voters appear to be learning about candidates’ ideologies and primaries and changing their support accordingly. If these associations do indeed reflect learning, we should observe several other patterns in these data. In particular, we should fail to find the same effect in the “placebo” races – those with minimal ideological differences between candidates. In the second column of Table 1, we present our estimate of γ_{avg} for the nine “placebo” races for governor and senator.¹⁵ For these races, our estimate of the interaction coefficient is close to zero and has the wrong sign, -0.036, indicating little effect of any learning about the ideological positions of the candidates among the voters in these races. The F-test fails to reject the null hypothesis that all the interaction coefficients for all of the races are zero ($p > 0.33$) and the difference between the “baseline” and “placebo” races is significant at $p < 0.01$).¹⁶

Finally, in the third column of Table 1, we present the results for five races where an incumbent governor or senator was challenged in their primary.¹⁷ In these races, we suspect that voters are better informed about the incumbents’ ideological positions before the campaigns. As with the “placebo” races, the coefficient is indeed smaller: 0.022. The F-test again fails to reject the null hypothesis that all the interaction coefficients for all of the races are zero (p-value of 0.20) and the coefficient is significantly smaller than the “baseline” estimate ($p < 0.10$).¹⁸

¹⁵See Table A.1 for a list of placebo races.

¹⁶Figure 3 shows the interaction coefficients for these races (those to the left of 0.5).

¹⁷The races included in this analysis are the 1966 California Democratic gubernatorial primary, 1992 California Republican senate primary, 2006 Alabama Republican gubernatorial primary, 2006 Oregon Democratic gubernatorial primary, and the 2007 Kentucky Republican gubernatorial primary.

¹⁸We should note that races with incumbents may suffer from a similar endogeneity problem as studies of

3.2. More Evidence That It's Learning: Magnitude of the Ideological Gap between Candidates

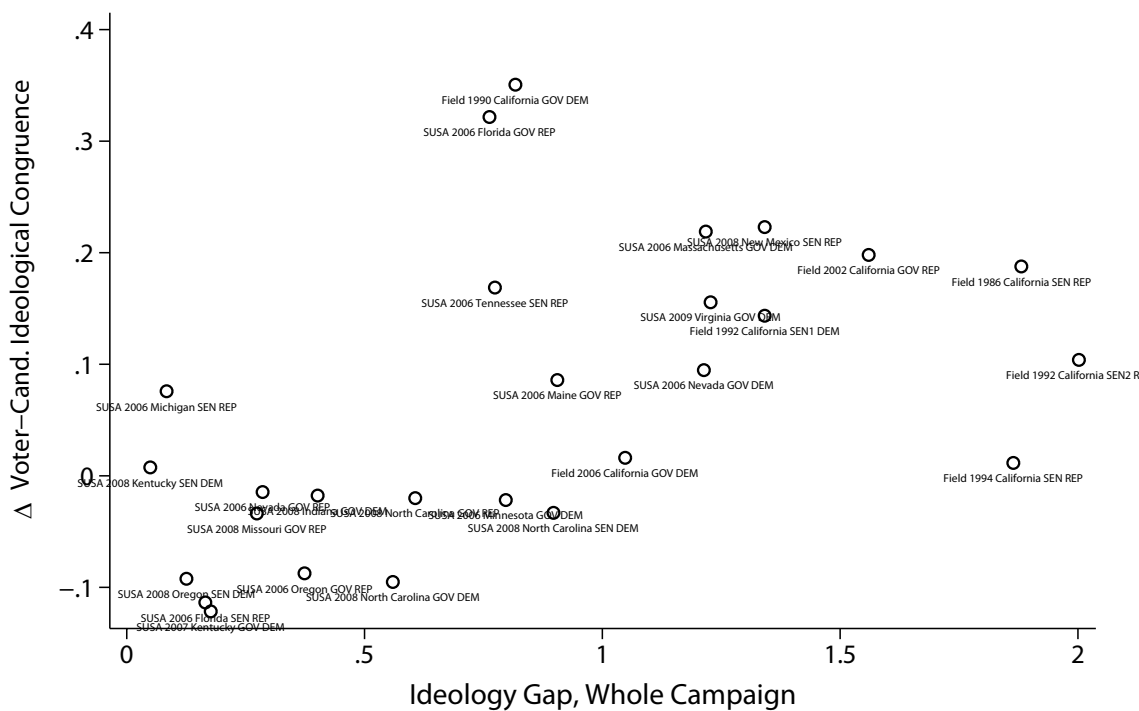
If these effects indeed reflect learning, then we might expect voters to learn more when the ideological gap is relatively large. In particular, we expect a positive association between the magnitude of the ideological gap between the candidates and the increase in congruence between voters' and candidates' ideological positions. This is precisely what we find.

This association is evident in a simple scatterplot of our coefficient estimates on the interaction between late poll and conservative voter ideology and our measures of ideological gap between the candidates. In Figure 3, we observe that the races with a relatively pronounced ideological gap between the candidates are also those where we find more "learning." One exception is the 1994 California Republican senate primary. In this case, the voters seemed to have identified the ideological position of candidates early in the primary. As is discussed in section 3.3, we find less learning in these types of cases.

voter learning in general elections. The incumbent may be substantially more influential in his state than any of the challengers and may induce people to modify their ideology to match his own in the reverse-causation model, which would upwardly bias the interaction effect.

Figure 3

Scatterplot of Δ Voter-Candidate Ideological Congruence
vs. Ideology Gap Between Candidates



To parametrically test whether this relationship exists, we add a triple interaction between the ideological gap measure, the late poll indicator, and the conservative respondent indicator variables. In other words, we run a triple difference specification of the following form, in which we compare conservative and liberal voters, early and late polls, and “placebo” and non-placebo races:

$$V_{ijt} = \alpha_j + \beta_{jC}C_{ijt} + \beta_{jL}L_{ijt} + (\gamma_{\text{int}} + g_j\gamma_{\text{gap}})C_{ijt}L_{ijt} + \epsilon_{ijt}. \quad (4)$$

The only modification is now to parameterize the interaction coefficient by restricting it to be affine in the ideology gap measure g_j . This is a full triple-difference specification since the coefficients on all terms depend on j , and hence on g_j . Our hypothesis of interest now has two parts: first, the interaction effect grows with the ideological gap, i.e. $\gamma_{\text{gap}} > 0$, and second, if the ideological gap is zero then there is no interaction effect, i.e. $\gamma_{\text{int}} \approx 0$. If these are both true – as they appeared to be in Figure 3 – then the differential trends in behavior among conservatives and liberals can be more plausibly attributed to the existence of ideological gaps between the candidates, and thus, that voters are learning about ideology and using what they learn to shape their vote.

We estimate equation 4 using various methods of constructing g_j . One straightforward way to obtain g_j is to use the ideology gap itself. However, one may be skeptical about the cardinalization assumptions implicit in using the gap directly. Therefore, we also consider setting g_j to be a dummy variable equal to 1 when the ideology gap is above a chosen cutoff. In Table 2 we present the results using a cutoff of 0.5, which is the same cutoff used to identify the “baseline” and placebo samples in Table 1.¹⁹

¹⁹We use 0.5 because Figure 3 shows that it is a natural break in the data. We also tried a cutoff of 0.7, which allows for more “placebo” primaries. The results do not substantively differ from the results using the 0.5 cutoff.

Table 2: Triple Differences

Gap Computation Period Gap Measure	Whole Period		First Third	
	Continuous	Cutoff 0.5	Continuous	Cutoff 0.5
Δ Voter-Cand. Ideological Congruence (γ_{int})	-0.016 (0.033)	-0.024 (0.032)	-0.043 (0.059)	-0.004 (0.051)
Δ Voter-Cand. Ideological Congruence x Ideological Gap (γ_{gap})	0.095** (0.032)	0.133** (0.039)	0.165** (0.055)	0.163** (0.059)
Observations	12279	12279	6367	6367
Number of Primaries	27	27	14	14
Number of Placebos	0	9	0	4

Robust standard errors in parentheses. Δ Voter-Candidate Ideological Congruence is the average interaction coefficient γ and estimates the amount by which conservative voters increase their support for the conservative candidate over the course of the campaign *over and above* the amount by which liberal voters increase their support for the conservative candidate. ** p<0.01, * p<0.05, + p<0.1.

We also varied the period over which we obtain the information from newspapers needed to construct g_j . If candidates manipulate their ideological stances or emphasize different issues over the course of the primary campaign, then sentences taken from newspaper stories and editorials later in the campaign might reflect these changes, and thus g_j might be endogenous. Thus, in addition to constructing g_j using all the sentences describing primary j (as described in the previous section), we also construct a measure of g_j using only the sentences from the first third of our sample for primary j . Focusing on the ideological gap between candidates in the early part of campaign reduces the potential endogeneity problem.²⁰

Table 3 presents the estimates of γ_{int} and γ_{gap} . Using the candidate ideological gap from either the whole period or the first third of the sample, our hypothesis is confirmed: $\gamma_{\text{gap}} > 0$ and $\gamma_{\text{int}} \approx 0$. In fact, γ_{gap} is very close to our “baseline” estimate in section 3 (it is 10 percentage points), and γ_{int} is always statistically insignificant and rather small. This increases our confidence that the results reflect voter learning. For these estimates, we only use races where we were able to code the ideological difference between the candidates using newspaper sentences for this analysis (see Table A.2 for a list of races for which we were able to quantify the ideological gap between the candidates).

Interestingly, using the ideological gap from the newspaper coverage during first third of the sample yields somewhat larger coefficients. The differences are not substantively very remarkable.

²⁰In fact, we do not find much evidence that candidates change ideological positions during the primary campaign. We should also note that the substantive findings are not particularly sensitive to the cutoff used to identify the ideological gap. The estimate of the coefficient on the triple interaction variable is about the same if we use the ideological gap from the first half, as opposed to the first third, of the sample.

Table 3: Downballot Races

	Baseline	Placebo	Lt Gov & Atty Gen
Δ Voter-Cand. Ideological (γ_{avg}) Congruence	-0.006 (0.029)	-0.089 (0.059)	-0.041 (0.040)
Observations	5952	1702	3209
Number of Primaries	26	8	10
P-value	0.474	0.492	.344

Robust standard errors in parentheses. P-value is for the F-statistic for testing that all interaction coefficients are zero. Δ Voter-Candidate Ideological Congruence is the average interaction coefficient γ and estimates the amount by which conservative voters increase their support for the conservative candidate over the course of the campaign *over and above* the amount by which liberal voters increase their support for the conservative candidate.

** p<0.01, * p<0.05, + p<0.1.

3.3. More Evidence That It's Learning: Error correction

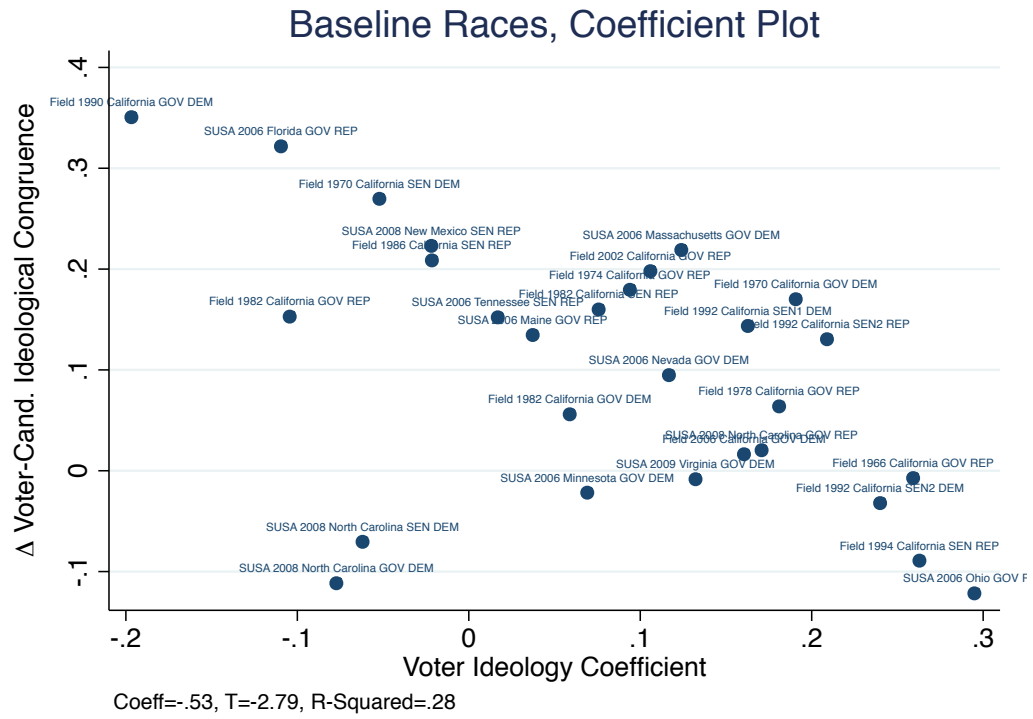
The results in Table 1 where we find no significant “learning” in primary races with an incumbent candidate may reflect the amount of information voters have about the candidates at the outset of the primary campaigns. We might expect that voters are more likely to “learn” about candidates’ ideological positions over the course of the campaign when they start with no perceptions or incorrect perceptions of these positions. Thus, another prediction consistent with learning concerns the relationship between the coefficient on ideology (β_{jC}) and the interaction coefficient (γ_j) in each primary. If conservatives are initially insufficiently in favor of the conservative candidate (low β_{jC}), then we should observe a greater increase in their support for the conservative candidate relative to liberals over the course of the campaign (higher γ_j), and if conservatives are overly in favor of the conservative candidate (high β_{jC}), then we should observe a smaller increase in their support for her relative to liberals over the course of the campaign (lower γ_j).

In Figure 4 we plot our estimates of γ_j against our estimates of β_{jC} for each race. The pattern in this figure is consistent with this prediction. There is a strong negative association between our estimates of these two coefficients. This negative association may be interpreted in the context of an error-correcting model of voter learning, in which voters have a prior perception of candidate ideology and then correct it over the course of the campaign.²¹

²¹The two outlier cases are for the 2008 Democratic primaries in North Carolina. The 2008 Democratic primaries may have been unusual since the Obama-Clinton presidential primary drew so much attention that year, especially for states – including North Carolina – that held their state primaries on the same day as the presidential primary. Also, the ideological gaps between the candidates in the North Carolina races are close to the 0.5 cutoff to be considered placebo cases, so it is not particularly surprising that the interaction coefficient is not very large.

Figure 4

Scatterplot of Δ Voter-Candidate Ideological Congruence
vs. Ideology Coefficient



Note: Baseline sample only.

3.4 Checking for Changes in the Composition of the Respondents

One potential concern with the above analysis is that the respondents in the Field Polls and SUSA were not the same across the polls within races. Consequently, the composition of the survey respondents in a primary might change substantially over the course of the campaign in a way that could lead to differential trends in vote intentions without there actually being learning. While we address this concern more directly in our internet panel surveys by polling the same respondents over time, we can also examine whether the results are sensitive to including various covariates in the Field Poll and whether there are any noticeable differences in observed covariates across the samples. When we include covariates – such as age, income, religion – our effect becomes slightly larger and more significant than the average interaction effect. We describe the specifications and show the results in Appendix C.

Besides including the covariates in our regressions, we can also check their stability in the primary electorate or (for the Field Poll) among conservatives and liberals separately. Finding stability would reassure us about the main finding, but finding instability does not necessarily invalidate it. Even if changes in the composition of the primary electorate take place, they might offset each other or even work against our hypothesis, and so not explain the differential trends we observe between conservative and liberal voters. We find few changes in the distributions of voter characteristics over the course of the primary campaigns, and that those that might have taken place were scattered and unsystematic (see Appendix C for details).

4. Downballot Races Using Polling Data

The above results suggest that voters do learn about candidates' ideological positions during primary election campaigns that generate relatively large amounts of information, that is, in gubernatorial and senate races. However, the campaigns for many offices do not receive the same amount of attention as these races do. Do voters in relatively low-information races, such as lieutenant governor or attorney general, also learn about their candidates' ideological positions?

In this section, we investigate whether we observe a similar pattern of learning about the ideological positions of candidates for downballot offices. Primaries for downballot offices often involve substantial ideological conflict – e.g., in the 1990 Republican primary for lieutenant governor in California one of the contestants accused the other of being insufficiently conservative during the Republican state convention. However, the amount of information about the candidates made available to voters during the primary campaigns tends to be relatively sparse, as the media often fails to cover these races and the candidates themselves have limited campaign resources. For example, the *Los Angeles Times* published, on average, 295 articles for each gubernatorial primary and 115 articles for each senatorial primary, but only published 44 articles, on average, for an attorney general primary, the downballot primary with the most media exposure. There were even fewer articles for other downballot primaries. Between 2002 to 2007, downballot candidate spending was on average 20% of gubernatorial candidate spending. We might therefore expect that voters would not learn as much about downballot candidates’ ideological positions as senate or gubernatorial candidates.

The California field polls included surveys of voter preferences for a number of downballot offices (see Table A.2). Thus, we can estimate whether voters learn about the ideological positions of the downballot candidates using the same saturated model as in the above section. Again, we are mainly interested in the coefficients estimates on the interaction between the late polls and voter ideology, $\hat{\gamma}_j$. Since voters are not exposed to as much information about downballot candidates, we would expect the magnitude of $\hat{\gamma}_{\text{avg}}$ to be smaller for these races as compared to our estimate of this coefficient for the “baseline” gubernatorial and senate races in Table 1.

Because of the scarcity of newspaper coverage of downballot candidates, we could not have independent raters evaluate candidate ideology by coding newspaper sentences, as we did with top-of-the-ticket primaries. Instead, we used a variety of sources including roll-call voting scores, endorsements, newspaper articles, and various other media sources to classify the downballot candidates in 36 races.²² For several races we were unable to identify the

²²When we were unable to use roll-call voting scores or endorsements, we searched various media sources, primarily on-line newspapers, for information to identify the relative ideological positions of the candidates.

relative ideological positions of the candidates. In some cases, this was due to the candidates having similar ideological positions. More often we were unable to identify the positions due to lack of information about the candidates’ issue positions. Thus, similar to the “baseline” estimate for governors and senators, our “baseline” estimate of γ_{avg} for downballot races excludes the cases where we could not determine the relative ideological positions. We also estimate γ_{avg} for the two types of races that tend to receive slightly more media coverage and campaign resources: lieutenant governor and attorney general.

In the first column of Table 3, we present our “baseline” estimates of γ_{avg} . We find little evidence that voters were learning about the ideological positions of the candidates. The coefficient estimates are close to zero. The estimates are also small in magnitude and not statistically significant when we focus on the “placebo” cases where we had difficulty determining the ideological positions of the candidates.

In the third column of Table 3, we present our estimates of γ_{avg} for the primary campaigns for lieutenant governor and attorney general. Even in these downballot races that receive slightly more media coverage, $\hat{\gamma}_{\text{avg}}$ is still relatively small and not statistically significant. Moreover, the F-test fails to reject the null hypothesis that all the interaction coefficients are zero for each sample of races.²³

In contrast with our top-of-the-ticket findings, where voters appear to learn about candidates’ ideological positions and act on that knowledge, these results raise concerns about voters’ ability to make informed judgments in downballot primaries.²⁴ We return to this point in the conclusion.

5. Panel Survey Confirmation

Our analysis above relied on two assumptions that we investigate further by conducting

When the roll-call voting records disagreed with information from other media sources, we based our coding on the roll-call voting records. Thus, while Bowen is often portrayed as having a more liberal position than Ortiz in the media, she had a more moderate roll-call voting position relative to Ortiz.

²³In separate analyses, we find little evidence that the lack of a significant interaction effect for downballot offices reflects substantial learning about the candidates’ ideologies prior to the campaign period. Across all down ballot races, our estimates of the conservative main effects, $\hat{\beta}_{jC}$, are distributed around zero. We also fail to find the error correction we observe in senate and gubernatorial races (see Figure 4). Instead, we mostly observe regression to the mean – i.e. races with large main effects on voter conservative ideology tend to be associated with large negative coefficients on the conservative ideology and late poll interactions terms.

²⁴Of course, voters may be learning about nonideological aspects of these candidates.

internet panel surveys. First, we assumed that learning the candidates' ideologies lies behind the increased associations we found between ideology and vote choice. Second, we assumed that reverse causation – people adjusting their own ideology to match that of their preferred candidate – is absent. Both assumptions seem plausible in part because of the kinds of races we study: primary races without incumbents. In these contests, voters presumably had much to learn and lacked strong views about the candidates and thus seem unlikely to have adopted (or rejected) candidates' ideologies. Since these races are intra-party, voters were also not necessarily learning their parties' positions, only potential party nominees' positions, and so were unlikely to be learning and then following the ideologies of their parties.

To provide further assurance, however, we conducted our own panel surveys. We interviewed likely voters before and after seven gubernatorial and senate primaries in 2010 (see Table 4). With these surveys, we can test the learning assumption because we can determine who learned the candidates' ideologies and who did not. Since we also measured voters' ideologies earlier in the campaigns, we can be more certain that increased voter-candidate ideological congruence arises, not from voters changing their ideology, but from changing their votes.

Our panel surveys are not representative of voters. We recruited participants online through our own ads and through a consumer internet panel.²⁵ Although not representative, the panel nature of our data nevertheless allows us, to some extent, to vet further the cross-sectional findings.

In analyzing these data, we follow the earlier analyses as closely as possible. Most of the races we surveyed had a noticeable ideological gap between the major candidates, as shown in Table 4, which presents respondents' perceptions of the candidates on a seven-point ideology scale. In the Republican New Hampshire Senate primary, for example, our respondents placed the most moderate candidate, Bill Binnie, at 4.0 on a seven-point ideology scale and the conservative candidate, Ovide Lamontagne, at 6.4. Two races, however, had small gaps: the New Hampshire Republican gubernatorial primary and the Pennsylvania Democratic gu-

²⁵We recruited subjects through online ads shown on Facebook and through Decision Analysts (see appendix for details). We considered conducting a probability phone survey or a representative Internet panel through firms such as Knowledge Networks or Polimetrix, but the costs were prohibitive, into the hundreds of thousands of dollars, while ours cost only a few thousand dollars.

gubernatorial primary. We exclude these races and the race with an incumbent (Pennsylvania Senate Democratic race), leaving five primaries: the Michigan gubernatorial primaries for both parties, the New Hampshire Republican senate primary, and the Pennsylvania Republican gubernatorial and Senate primaries.

Table 4: Internet panel surveys and perceptions of ideology

Contest description	Ideological placements (7-point scale)		Vote coding	N
	Pre-election	Post-election		
Michigan Gov Dem				156
Virg Bernero	2.5	2.2	0	
Andy Dilon	3.9	4.6	1	
Michigan Gov Rep				195
Rick Snyder	5.5	5.2	0	
Mike Cox	5.5	5.7	1	
Mike Bouchard	5.7	5.8	1	
Peter Hoekstra	5.8	6.0	1	
New Hampshire Gov Rep				344
Karen Testerman	5.4	5.5	Placebo	
John Stephen	5.5	5.7		
Jack Kimball	5.7	5.8		
New Hampshire Sen Rep				365
Bill Binnie	4.3	4.0	0	
Jim Bender	5.2	5.2	0	
Kelly Ayotte	5.4	5.5	0	
Ovide Lamontagne	5.9	6.4	1	
Pennsylvania Gov Dem				93
Joe Hoeffel	2.3	2.3	Placebo	
Dan Onorato	3.7	3.7		
Jack Wagner	3.8	3.8		
Pennsylvania Gov Rep				96
Tom Cortbett	5.6	4.9	0	
Sam Rohrer	6.3	6.3	1	
Pennsylvania Sen Dem				163
Joe Sestak	2.1	2.1	Excluded	
Arlen Specter	4.2	4.2	bc. inc.	
Pennsylvania Sen Rep				161
Peg Luksik	5.9	5.9	0	
Pat Toomey	6.3	6.3	1	

Between the early and late interviews, many respondents learned about the candidates' positions. Measuring this learning, however, requires numerous coding decisions, especially in multi-candidate races. We adopt a simple approach and use a minimal definition of knowledge: we code respondents as knowing the positions if they placed at least one conservative candidate to the right of a liberal candidate.²⁶ Using this minimal test, about 42% passed early in the primary and 68% did so after the primary, for a net increase of 26%. This 26% net increase arises from 30% learning and 4% "unlearning."²⁷ The increase in minimal knowledge varies from race to race, as shown in Table 5, and is statistically significant at conventional levels in all races except the Pennsylvania Senate Republican primary.

Not only did these panel respondents learn, but they also appeared to act on this knowledge. Table 6 shows the findings for the panel respondents, presenting estimates from a series of models. As in the prior analysis, we use the Extremist Only classification of ideology (the results are similar for other classifications).²⁸

²⁶The results are similar with other codings.

²⁷These statistics exclude the two "placebo" races and the race with the incumbent.

²⁸For Republican primary voters, we code sixes and sevens as conservative and fours and lower as moderates. For Democratic primary voters, we code ones and twos as liberal and fours and higher as moderates. Other codings produced similar results.

Table 5: Learning about the candidates' positions

Contest description	% Knowing positions		Difference (net learning %)	p-value on difference
	Pre	Post		
Michigan Gov Dem	61.5	83.3	21.8	0.001
Michigan Gov Rep	33.3	71.2	37.9	0.001
New Hampshire Sen Rep	49.0	80.8	31.8	0.001
Pennsylvania Gov Gop	36.2	61.3	25.1	0.001
Pennsylvania Sen Rep	25.4	26.7	1.3	0.716

Since the sample of panel respondents is less representative than the cross-sectional polls we analyzed earlier, we start by replicating the previous analyses in this panel and assess whether the results are similar. We therefore first treat the panel as if it was two separate cross-sections and estimate a variant of equation 1. Since we have individual-level data, we estimate an individual-level model in a pooled data set, including indicator variables for primary and sample type, and clustering the standard errors at the individual. After excluding respondents because they failed to report an ideology or a vote preference, or because of the Extremist Only classification, the sample size in this pooled data set is 790 pre-election and 944 post-election, for a total of 1,734. As shown in Column 1 of Table 6, the cross-sectional and panel findings are reassuringly similar. The interaction between Conservative indicator variable and vote is positive and marginally statistically significant. On average, conservatives increased their support for the conservative candidate over the course of the campaign by 11 percentage points, slightly larger than the 10 percentage points we found in the cross-sectional surveys. This difference could arise because of differences between the samples, the shorter time between surveys, the races we sampled, or just by chance. The important point, however, is that the estimates are close despite these differences.

Table 6: Panel analyses

	(1)	(2)	(3)	(4)	(5)	(6)
Conservative indicator	0.19 (0.041)	0.22 (0.048)				
Post-election indicator	0.010 (0.031)	0.0018 (0.031)				
Conserv. \times Post	0.11** (0.036)	0.12** (0.036)				
Prior Conserv.			0.090* (0.047)	0.072 (0.068)	-0.10 (0.075)	-0.016 (0.066)
Prior Conserv. \times Learning					0.29** (0.11)	0.39** (0.092)
Learning indicator					0.072 (0.10)	-0.013 (0.084)
Prior vote			0.59 (0.034)	0.57 (0.057)	0.52 (0.080)	0.51 (0.075)
Prior vote \times Learning					-0.085 (0.087)	-0.11 (0.085)
Decision Analyst sample	-0.068 (0.046)	-0.080 (0.053)	-0.17 (0.054)	-0.17 (0.083)	-0.12 (0.080)	-0.088 (0.075)
Constant	0.32 (0.036)	0.30 (0.042)	0.23 (0.041)	0.24 (0.065)	0.26 (0.075)	0.18 (0.066)
Observations	1734	1436	718	389	389	507
R^2	0.246	0.254	0.473	0.385	0.451	0.401
SER	0.44	0.43	0.36	0.40	0.38	0.39

Notes: Ordinary least squares regression coefficients with robust standard errors in parentheses. In columns 1 and 2, the standard errors are clustered at the individual. The $N = 1734$ in column 1 is from 790 pre-election and 944 post-election. Column 6 adds the 118 respondents who failed to report a vote intent pre-election but who did report a vote choice post-election, coding their vote to 0.5. ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$ (shown only key coefficients).

To rule out reverse causation, we next present estimates based on measures of ideology only from the earlier interview. Since regression to the mean on the dependent variable can bias downwards coefficients on prior-wave variables (Finkel 1995), we modify the model for panel data:

$$V_{it} = \alpha_j + \gamma C_{it-1} + \rho V_{it-1} + \epsilon_{it} \quad (5)$$

Besides measuring conservativeness at $t - 1$, this model includes lagged vote, V_{t-1} , which addresses this potential bias and transforms the equation into a model of vote change. Since this model examines respondents from both waves, we retain only those who expressed a vote intent and an ideology in both interviews (718 of a possible 944 Extremist Only respondents). Column 2 shows that the cross-sectional estimate in Column 1 remains similar among this subset.

Column 3 presents the estimates of equation 5. The prior conservative coefficient is 0.09 and is statistically significant. It implies that a conservative became nine percentage points more likely to vote for the conservative candidate over the course of the campaign, a somewhat larger effect than in column 1. Since reverse causation cannot bias this estimate, we can therefore be more confident that it represents a causal effect of ideology.

We next examine whether learning lies behind this increase. To do so, we compare two groups: those who learned according to the minimal knowledge test (“learners”) and those who never passed the test (“never learners”). We focus on these two groups because they lend themselves to clear interpretations: the learners appear to know nothing early, but something late, while the never learners remained ignorant of even the most minimal knowledge. In contrast, drawing clear inferences about the other major group, those who passed the minimal test in both interviews, is hard because it may include individuals who learned more about the candidates’ positions between interviews. We therefore exclude respondents who passed the minimal test in both waves.²⁹ This leaves us with a sample of 389 Extremist Only respondents who learned or never learned. Column 4 shows that the

²⁹We also exclude respondents who passed the minimal test in their first interview, but failed to do so in their second.

effect of prior ideology is slightly smaller in this subsample to the overall estimate.

To estimate whether the tendency to gravitate towards like-minded candidates arises from learning, we add an indicator for the learners and the interaction between this indicator and prior conservative ideology.³⁰ The never learners are the comparison group (the omitted category). Column 5 shows the results. The interaction between prior conservative and learning is large, 0.29. Compared to conservatives who never learned, it suggests that conservatives who did learn increased their support for the conservative candidate over the course of the campaign by 29 percentage points more. As expected, conservatives who never learned did not shift towards the conservative candidate – if anything, they shifted away. Finally, column 6 adds the 118 respondents who failed to report a vote intent pre-election but who did report a vote choice post-election, coding their vote to 0.5. The estimate of the learning effect is larger: 0.39.

Taken together, these results support the cross-sectional findings, reassuring us that learning lies behind the increased associations between ideology and vote and further ruling out the reverse causation alternative. During primary campaigns, voters appear to learn about the candidates’ ideological positions and then act on this newly acquired knowledge, shifting their votes to the candidates who agree with them.

To provide added reassurance, we also directly test for reverse causation, that is, test whether individuals adopt the ideology of their preferred candidate. We estimate models such as those in columns 3-6, but with postelection ideology as the outcome variable. If people were following their preferred candidate, we should see prior vote influencing change in voters’ ideologies, especially among learners. These estimates, however, reveal no such findings. The coefficient on prior vote and its interaction with learning is always close to zero and precisely estimated.

³⁰Specifically, we estimate

$$V_{it} = \alpha_j + \gamma_{nl}C_{it-1} + \omega L_i + \gamma_l C_{it-1}L_i + \rho_{nl}V_{it-1} + \rho_l V_{it-1}L_i + \epsilon_{it}, \quad (6)$$

where L_i is an indicator variable for learning, γ_{nl} captures the increase in voter-candidate congruence among never learners, γ_l captures it among learners, and ρ_{nl} and ρ_l do the same for prior vote.

6. Conclusions

Although the conventional wisdom is that U.S. voters are not well informed about politics, a number of studies find that many voters acquire a substantial amount of information about candidates' policy positions and attributes over the course of electoral campaigns, especially for the presidential candidates. What is less well known is whether the information voters gain during campaigns actually affects their voting decisions. Our results suggest that voters learn about candidates during state primary election campaigns and this information affects their vote choices. This pattern is most apparent for gubernatorial and senate races with no incumbent candidate.

In downballot races, however, this pattern appears absent – we see little evidence of voter learning about ideology. This could be due to the lack of easily available information for voters. For example, as noted above, many of these races receive little media coverage, and the candidates often have little money for campaign advertising. Voters may also pay little attention to ideology in these races, since it presumably matters less for the job. Does it matter if the state treasurer, auditor, or secretary of state is liberal or conservative? This issue deserves further study, since one possible implication is that the amount of information is too low to allow voters to make meaningful choices; so, perhaps, these offices should be filled by appointment rather than elections.

Our study focuses on the effect of learning about candidate ideology, but voters may act on what they learn about candidates' other attributes as well. Do voters in primaries learn about candidates' positions on salient and divisive issues – e.g., gay marriage inside the Democratic party, or immigration policy inside the Republican party – and vote on the basis of these positions? Voters might also learn about “electability.” If voters are strategic in the sense that they look ahead to the general election, then electability might interact in subtle and interesting ways with candidate ideology. For example, voters who are extremely strategic and well informed might even respond to the events in the primary campaign of the opposing party.

Finally, our study says nothing about how voters learn. Is it through media coverage? Campaign advertising? Debates? Word-of-mouth? Are there other cues – race, ethnicity,

and gender – that help voters identify or make educated guesses about candidates’ ideologies? At this point, we cannot say, but future research can address these important questions.³¹ For example, if campaign advertising by candidates and interest groups proves to be a key channel, this would have important implications for campaign finance reforms.

³¹Some previous studies find that in nonpartisan elections voters use proxies for party ID to help decide how to vote, when such proxies are available. See, e.g., Smith and Squire (1988) on judicial reconfirmations in California, and McDermott (2005) on candidate occupation designations for downballot California offices.

REFERENCES

- Abramowitz, Alan I. 1978. "Impact of a Presidential Debate on Voter Rationality." *American Journal of Political Science* 22(3): 680-90.
- Alvarez, R. Michael. 1997. "Why Study Only Presidential Campaigns? Statewide Races as a Laboratory for Campaign Analysis." Unpublished manuscript, California Institute of Technology.
- Alvarez, R. Michael. 2001. *Information and Elections*. Ann Arbor, MI: University of Michigan Press.
- Atkin, Charles and Gary Heald. 1976. "Effects of Political Advertising." *Public Opinion Quarterly* 40:2:216-228.
- Bartels, Larry M. 1993. "Messages Received: The Political Impact of Media Exposure." *American Political Science Review* 87: 267-285.
- Bartels, Larry M. 2002a. "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior* 24(2): 117-50.
- Bartels, Larry M. 2002b. "The Impact of Candidate Traits in American Presidential Elections." In *Leaders' Personalities and the Outcomes of Democratic Elections*, edited by Anthony King, 44-70. Oxford: Oxford University Press.
- Benoit, William L. and Glenn J. Hansen. 2004. "Presidential Debate Watching, Issue Knowledge, Character Evaluation, and Vote Choice." *Human Communication Research* 30:1:121-144.
- Berelson, Bernard R., Paul F. Lazarsfeld and William N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: University of Chicago Press.
- Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. Chicago: University Of Chicago Press.
- Carsey, Thomas M., and Geoffrey C. Layman. 2006. "Changing Sides or Changing Minds? Party Identification and Policy Preferences in the American Electorate." *American Journal of Political Science* 50(2): 464-77.
- Clarke, Peter, and Susan H. Evans. 1983. *Covering Campaigns: Journalism in Congressional Elections*. Stanford, Calif.: Stanford University Press.
- Craig, Stephen C., James G. Kane, and Jason Gainous. 2005. "Issue-Related Learning in a Gubernatorial Campaign: A Panel Study." *Political Communication* 22: 483-504.
- Druckman, James N. 2005 "Media Matter: How Newspapers and Television News Cover Campaigns and Influence Voters." *Political Communication* 22: 463-481.

- Fan, David P. and Albert Tims. 1989. "The Impact of the News Media on Public Opinion: American Presidential Elections, 1987-1988." *International Journal of Public Opinion Research* 1:2:151-63.
- Finkel, Steven E. 1993. "Reexamining the 'Minimal Effects' Model in Recent Presidential Campaigns." *The Journal of Politics* 55:1:1-21.
- Finkel, Steven E. 1995. *Causal Analysis with Panel Data*. Thousand Oaks, California: Sage Publications, Inc.
- Frank, Thomas. 2004. *What's the Matter with Kansas?*. New York: Henry Holt and Co.
- Gelman, Andrew, and Gary King. 1993. "Why are American Presidential Election Campaign Polls So Variable When Votes are So Predictable?" *British Journal of Political Science* 23: 409-451.
- Gronke, Paul. 2000. *The Electorate, the Campaign, and the Office: A Unified Approach to Senate and House Elections*. Ann Arbor, MI: University of Michigan Press.
- Henderson, Mike. 2009. "Campaign Learning and Issue Voting: Evidence from the 2000, 2004, and 2008 Elections." Unpublished manuscript.
- Highton, Benjamin. 2006. "Alternative Tests for the Effects of Campaigns and Candidates on Voting Behavior." In *Capturing Campaign Effects*, edited by Henry E. Brady and Richard Johnston. Ann Arbor, MI: University of Michigan Press.
- Hollbrook, Thomas M. 1996. *Do Campaigns Matter?* Thousand Oaks, CA: SAGE Publications.
- Iyengar, Shanto, and Donald Kinder. 1987. *News that Matters: Television and American Opinion*. Chicago: The University of Chicago Press.
- Iyengar, Shanto and Adam F. Simon. 2000. "New Perspectives and Evidence on Political Communication and Campaign Effects." *Annual Review of Psychology* 51:1:149-169.
- Johnston, Richard, André Blais, Henry E. Brady, and Jean Crête. 1992. *Letting the People Decide: Dynamics of a Canadian Election*. Stanford, Calif.: Stanford University Press.
- Krasno, Jonathan S. 1994. *Challengers, Competition, and Reelection: Comparing Senate and House Elections*. New Haven, CT: Yale University Press.
- Lazarsfeld, Paul F., Bernard R. Berelson, and Hazel Gaudet. 1948. *The People's Choice: How the Voter Makes up His Mind in a Presidential Campaign*. Second edition. New York: Columbia University Press.
- Lenz, Gabriel S. 2009. "Learning and Opinion Change, Not Priming: Reconsidering the Evidence for the Priming Hypothesis." *American Journal of Political Science* 53: 821-837.

- Lipsitz, Keena. 2004. "The Significance of Rich Information Environments: Voter Knowledge in Battleground States." Paper presented at the annual meeting of the Midwest Political Science Association, Chicago.
- Martinez, Michael D., Stephen C. Craig, and James G. Kane. 2001. "Information and Voting: A Panel Study." Paper presented at the annual meetings of the American Political Science Association, San Francisco.
- McDermott, Monika L. 2005. "Candidate Occupations and Voter Information Shortcuts." *The Journal of Politics* 67:1:201-219.
- Patterson, Thomas E. and Robert D. McClure. 1976. *The Unseeing Eye: The Myth of Television Power in National Politics*. New York: G.P. Putnam's Sons.
- Rice, Laurie. 2004. "Learning About Presidential Primary Candidates: Campaign Advertising, Learning, and Recall." Paper presented at the annual meeting of the American Political Science Association, Chicago.
- Sears, David O. and Steven H. Chaffee. 1979. "Uses and Effects of the 1976 Debates: An Overview of Empirical Studies." In *The Great Debates*, edited by Sidney Kraus. Bloomington IN: Indiana University Press.
- Shaw, Daron R. "The Effect of TV Ads and Candidate Appearances on Statewide Presidential Votes, 1988-96." *The American Political Science Review* 93:2:345-61.
- Squire, Peverill and Eric Smith. 1988. "The Effect of Partisan Information on Voters in Nonpartisan Elections." *The Journal of Politics* 50:1:168-79
- Tedin, Kent L. and Richard W. Murray. 1981. "Dynamics of Candidate Choice in a State Election." *The Journal of Politics* 43:2:435-455.
- Vavreck, Lynn, Constantine J. Spiliotes, and Linda L. Fowler. 2002. "The Effects of Retail Politics in the New Hampshire Primary." *American Journal of Political Science* 46(3): 595-610.
- Weaver, David H. 1996. "What Voters Learn from Media." *Annals of the American Academy of Political and Social Science* 546: 34-47.
- Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.
- Zaller, John. 1994. "Politics as Usual: The Rise and Fall of Candidate Perot." Manuscript. University of California, Los Angeles.
- Zhao, Xinshu and Glen L. Bleske. 1993. "The Effect of 'Horse Race' Reporting in Increasing Voters' Issue Knowledge." Paper presented at the annual meeting of AEJMC, Kansas City.

Appendix A

For all of the gubernatorial and senate primary campaigns in Table A.1, we gathered information about the candidates' ideological positions from interest group endorsements and various media reports. As we describe in the text, we vetted our coding with independent raters who read sentences from newspaper coverage of the campaign. More specifically, we perform standardized searches for newspaper articles about each primary election in our sample. We searched Factiva because this source allowed us to obtain the articles in a way that we could also select sentences from each article that contained particular key terms. In selected sentences that contained the name of at least one candidate, the office for which they are running, and at least one ideology related key-word.

The following key-words were used to identify sentences with potential ideological content: liberal, progressive, conservative, moderate, centrist, mainstream, abortion, pro-choice, pro-life, the unborn, unborn children, prayer, vouchers, family values, gay marriage, gay rights, same-sex marriage, death penalty, criminals, crime rate, law-and-order, gun control, gun rights, handgun, assault weapons, concealed weapons, illegal immigrants, illegal aliens, immigration, border security, tax cut, tax increase, flat tax, income tax, anti tax, taxes, tax and spend, government spending, congressional spending, state spending, wasteful spending, cut spending, big spender, health care, labor, welfare, minimum wage, the environment, environmental, endangered species, regulation, drilling, global warming, climate change, Iraq, terror, defense spending, national security.

We then assigned between two and five coders to read each sentence and score the ideology of each candidate mentioned in the sentence on a 1-5 scale on the basis of the sentence, with 1 corresponding to very liberal and 5 corresponding to very conservative. In practice, Democratic candidates fall within the 1-3 segment of the scale, and Republican candidates fall within the 3-5 segment. Occasionally sentence would be repeated as news sources carried reports from other sources. We attempted to limit these repeats and dropped them whenever they were identified.

For each primary, we consider the top two vote getters and compute the ideology scores for each candidate for each coder by averaging the scores for each relevant sentence that the coders found useful in assessing the given candidate's ideology. We then compute the difference in ideology scores between the top two candidates for each coder, and average them to obtain the *ideology gap* for the entire primary. We assign ideology gaps only to primaries with a sufficient number of useful newspaper sentences about the candidates. Primaries with low ideology gaps have only slight ideological differentiation, while primaries with high ideology gaps have severe ideological differentiation, with a moderate in the party running against an extremist. We assign the major candidate with the higher candidate ideology score to be the conservative candidate in the election, and the major candidate with the lower score to be the liberal candidate.

The ideology gap in our sample of races ranges from 0 to 2. Figure A.1 shows a histogram of the ideological gaps. Thus, our data contains primaries with trivial ideological differences as well as primaries with very significant ideological differences, i.e., with an extremist running against a moderate. For our baseline sample, in which we expect to find voter learning

about ideology, we select all primaries with an ideology score of 0.5 or greater. For the remaining primaries, we do not expect to find large interaction effects. We will refer to the set of primaries with an ideology gap less than 0.5 as the “placebo” races.

We should note that ideological classification based on the independent sentence codings matched our own assessments in all but one case. That one case was the 2006 Michigan Republican Senator race, between Bouchard and Butler. We code Bouchard as conservative, while the coders code Butler as conservative. The ideology gap in this race is 0.08, which is one of the smallest ideology gaps for all coded races (it is at the 5th percentile of the ideology gaps).

For the downballot races we were unable to use the same sentence coding method because of the relative scarcity of newspaper reports about the primary campaigns. To identify the liberal and conservative candidates in these races we used three sources of information: 1) DW-nominate scores for former state legislators, 2) newspaper and interest group endorsements, and 3) various media reports. Most of the races are coded from various media reports (mainly newspaper articles) that were gathered by two research assistants. In addition to there being relatively few newspaper articles covering these races, many of the articles that did cover these races did not provide information relevant for assessing the candidates’ ideological positions. We attempted to identify the primary races where the ideological division between the candidates were relatively more pronounced.

Table A.1: Governor and Senate Races Studied

Contest Description	More Conservative Cands	More Liberal Cands
Field 1966 California Gov Rep	Reagan	Christopher
Field 1970 California Gov Dem	Yorty	Unruh
Field 1970 California Sen Dem	Tunney	Brown
Field 1974 California Gov Rep	Reinecke	Flournoy
Field 1978 California Gov Rep	Briggs, Davis	Maddy, Wilson
Field 1982 California Gov Dem	Garamendi	Bradley, Obledo
Field 1982 California Gov Rep	Curb	Deukmejian
Field 1982 California Sen Rep	Wilson	McCloskey
Field 1986 California Sen Rep ⁺	Herschensohn	Zschau
Field 1990 California Gov Dem ⁺	Feinstein	Vandekamp
Field 1992 California Sen1 Dem ⁺	Levine, McCarthy	Boxer
Field 1992 California Sen2 Dem	Davis	Feinstein
Field 1992 California Sen2 Rep ⁺	Herschensohn	Campbell
Field 1994 California Sen Rep ⁺	Dannemeyer	Huffington, Squires
Field 2002 California Gov Rep ⁺	Simon	Jones, Riordan
Field 2006 California Gov Dem ⁺	Westly	Angelides
Susa 2006 Florida Gov Rep ⁺	Gallagher	Crist
Susa 2006 Florida Sen Rep ^{*+}	Harris, McBride	Collins, Monroe
Susa 2006 Maine Gov Rep ⁺	Woodcock	Mills
Susa 2006 Massachusetts Gov Dem ⁺	Gabrieli, Reilly	Patrick
Susa 2006 Michigan Sen Rep ^{*+}	Butler	Bouchard
Susa 2006 Minnesota Gov Dem ⁺	Hatch	Lourey
Susa 2006 Nevada Gov Dem ⁺	Gibson	Titus
Susa 2006 Nevada Gov Rep ^{*+}	Beers	Gibbons
Susa 2006 Ohio Gov Rep	Blackwell	Petro
Susa 2006 Oregon Gov Rep ^{*+}	Atkinson, Mannix	Saxton
Susa 2006 Tennessee Sen Rep ⁺	Bryant	Corker
Susa 2007 Kentucky Gov Dem ^{*+}	Lunsford, Richards	Beshear, Henry
Susa 2008 Indiana Gov Dem ^{*+}	Thompson	Schellinger
Susa 2008 Kentucky Sen Dem ^{*+}	Lunsford	Fischer
Susa 2008 Missouri Gov Rep ^{*+}	Steelman	Hulshof
Susa 2008 New Mexico Sen Rep ⁺	Pearce	Wilson
Susa 2008 North Carolina Gov Dem ⁺	Perdue	Moore
Susa 2008 North Carolina Gov Rep ⁺	Smith	McCrorry
Susa 2008 North Carolina Sen Dem ⁺	Hagan	Neal
Susa 2008 Oregon Sen Dem ^{*+}	Merkley	Novick
Susa 2009 Virginia Gov Dem ⁺	Deeds	Moran

* indicates a placebo race. + indicates a race where we have a measure of the ideological gap between the candidates.

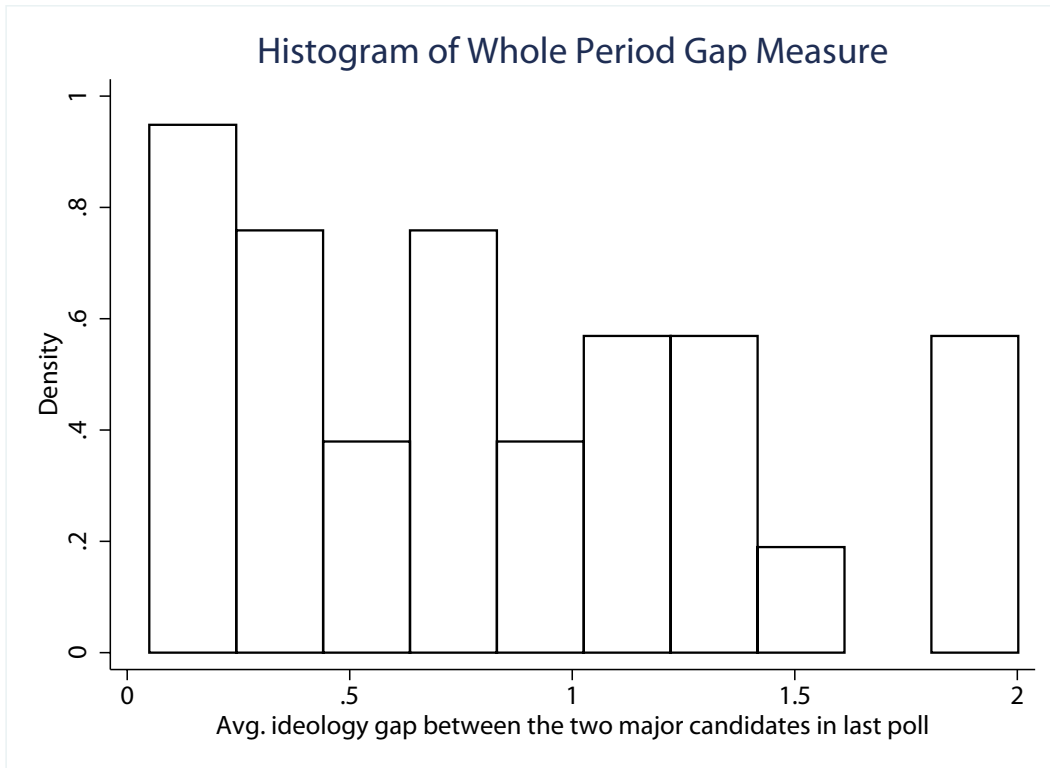
Table A.2: Downballot Races Studied

Contest Description	More Conservative Cand	More Liberal Cand
Field 1966 California Lt Gov Dem	Anderson	Braden
Field 1970 California Lt Gov Dem	Wenke	Alquist
Field 1974 California Atty Gen Dem	Bugliosi	Norris
Field 1978 California Lt Gov Rep	Reinecke	Flournoy
Field 1978 California Atty Gen Dem	Pines	Burke
Field 1978 California Atty Gen Rep*	Deukmejian	Browning
Field 1986 California Lt Gov Rep	Richardson	Curb
Field 1986 California Cont Dem*	Garamendi	Davis
Field 1986 California Cont Rep	Sebastiani	Campbell
Field 1990 California Lt Gov Rep	Bergeson	Seymour
Field 1990 California Atty Gen Dem*	Reiner	Smith
Field 1990 California Sec State Rep*	Levy	Flores
Field 1990 California Treasurer Rep	Buchanan	Hayes
Field 1990 California Insurance Com Dem	Garamendi	Press
Field 1994 California Lt Gov Rep	Wright	Statham
Field 1994 California Sec State Dem	Miller	Woo
Field 1994 California Treasurer Dem	Roberti	Angelides
Field 1994 California Controller Dem	Areias	Perata
Field 1994 California Controller Rep	Mclintock	Morris
Field 1994 California Insurance Com Dem	Torres	Margolin
Field 1994 California Insurance Com Rep	Bannister	Quackenbush
Field 2002 California Sec State Dem*	Alioto	Shelley
Field 2002 California Sec State Rep	Olberg	Schaefer
Field 2002 California Treasurer*	Conlon	Toman
Field 2002 California Controller Dem	Westly	Klehs
Field 2002 California Controller Rep	Mclintock	Andal
Field 2002 California Insurance Com Dem	Garamendi	Calderon
Field 2002 California Insurance Com Rep	Mendoza	Stitch
Field 2006 California Lt Gov Dem	Garamendi	Speier
Field 2006 California Atty Gen Dem	Delgadillo	Brown
Field 2006 California Sec State* Dem	Bowen	Ortiz
Field 2006 California Treasurer Rep	Parrish	Richman
Field 2006 California Controller* Dem	Chiang	Dunn
Field 2006 California Controller Rep	Strickland	Malconado

* indicates a placebo race.

Figure A.1

Histogram of Ideological Gaps Between Candidates



Appendix B

We also examine the robustness of our findings to broadening the classification of liberal and conservative voters as well as the types of polls considered early and late polls. The specific changes are discussed in the main text. Here we present main findings.

In Table B.1 we see that the magnitude of $\hat{\gamma}_{\text{avg}}$ attenuates as we broaden the criteria for classifying polls as being or early and late. This is perhaps not surprising as we would expect that the polls after the initial poll but before the 45 day cutoff may already start incorporating some learning about ideological positions by the respondents. The polls after the 45 day cutoff but before the last poll would include respondents who may still have not learned the candidates' ideologies but who will eventually learn this information by the last poll.

Similarly we see that the magnitude of $\hat{\gamma}_{\text{avg}}$ also attenuates as we begin to incorporate more ideologically moderate respondents into our non-liberal and non-conservative respondent classifications. This perhaps not surprising as incorporating more moderates increases the potential for measurement error as many of the moderates may actually be leaning liberal or leaning conservative. However, even in the most exhaustive, “minimally separated” specification – where we use the *Exhaustive* and *45-Day-Cutoff* procedures – the average interaction coefficient is 6.8 percentage points.

We also examine the robustness of our findings to the different types of polls and also across political parties. In particular, we examine whether the effect varied between the Field Poll surveys and the SUSA surveys and whether it varied between the polls for the Republican and Democratic primaries.

Table B.2 presents the interaction coefficients – how much conservative voters shift towards conservative candidates more than do liberal voters – for the various subsamples of the baseline sample. The average interaction coefficients for all the subsamples are positive, significant, and large. The magnitudes are also similar, except that the Field Poll races have a larger interaction coefficient than do the SUSA races. This is partly because the Field Poll races go back to the 1960s, during which both the Republican and the Democratic parties both had distinct moderate and extremist wings. The Republican party had a relatively liberal “Rockefeller” wing and a conservative wing, while the Democratic party had a pro-war and anti-war wing. In contrast, the SUSA races only cover recent primary elections, and currently there is probably less polarization among Democrats than among Republicans.³²

³²See, for example, Frank (2004) for a discussion of the ideological cleavage inside the Republican party.

Table B.1: Baseline Estimates Including Moderates and Intermediate Polls

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
Average Interaction Coefficient	0.100** (0.0209)	0.0829** (0.0163)	0.0957** (0.0189)	0.0777** (0.0153)	0.0827** (0.0162)	0.0649** (0.0129)
Observations	12,099	22,804	13,507	25,550	17,835	34,742
R-squared	0.551	0.536	0.556	0.541	0.563	0.548
Ideology	Extremes	Extremes	Intermediate	Intermediate	Exhaustive	Exhaustive
Polls	First + Last	45 Day Cutoff	First + Last	45 Day Cutoff	First + Last	45 Day Cutoff
Number of Primaries	28	26	28	26	28	26
P-value	3.39e-08	2.43e-09	2.57e-09	2.22e-09	3.06e-09	4.99e-09

Robust standard errors in parentheses.

P-value is the P-value of the F-statistic for testing that all interaction coefficients are zero.

** p<0.01, * p<0.05, + p<0.1

Table B.2: Effects by Data Source and Party

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Avg Interaction Coefficient	0.067** (0.018)	0.123** (0.030)	0.069* (0.029)	0.115** (0.029)	0.083** (0.030)	0.111* (0.043)	0.139** (0.038)	0.122** (0.034)	0.017 (0.046)
Observations	16822	5881	6218	6921	5178	2992	2889	3929	2289
R-squared	0.551	0.563	0.540	0.539	0.566	0.600	0.514	0.484	0.613
Sample	All	Field	SUSA	Rep	Dem	Field R	Field D	SUSA R	SUSA D
Number of Primaries	37	16	12	15	13	9	7	6	6
P-value	7.03e-07	8.64e-05	3.12e-05	2.12e-06	0.00123	0.0234	0.000317	3.69e-06	0.337

Robust standard errors in parentheses.

P-value is the P-value of the F-statistic for testing that all interaction coefficients are zero.

** p<0.01, * p<0.05, + p<0.1

Appendix C

Since the Field Polls allow us to view unit records for all the survey respondents, we can include socio-demographic information about the respondents directly into our specification. We include covariates for age, gender, race, education, income, homeownership, religion and location in California (all discretized into categories). Our coefficient of interest is the interaction between the conservative respondent and late poll indicator variables. We present average interaction effects for the following specifications:

$$V_{ijt} = \alpha_j + \beta_{jC}C_{ijt} + \beta_{jL}L_{ijt} + \gamma_j C_{ijt}L_{ijt} + \varepsilon_{ijt} \quad (7)$$

$$V_{ijt} = \alpha_j + \sum_{k=1}^K \delta_k X_{ijt}^k + \beta_{jC}C_{ijt} + \beta_{jL}L_{ijt} + \gamma_j C_{ijt}L_{ijt} + \varepsilon_{ijt} \quad (8)$$

$$V_{ijt} = \alpha_j + \sum_{k=1}^K \delta_{jk} X_{ijt}^k + \beta_{jC}C_{ijt} + \beta_{jL}L_{ijt} + \gamma_j C_{ijt}L_{ijt} + \varepsilon_{ijt} \quad (9)$$

The first specification is just the baseline specification. The second included covariates in the intercept term with coefficients that are not allowed to vary across primary elections. The last allows the coefficients on the covariates δ_{jk} to vary with the primary election j under consideration. Our covariates are age, gender, race, education, income, homeownership, religion and location in California (all discretized into categories).

Table C.1 presents the estimated average interaction effects after including the additional covariates. The estimates are slightly higher and more significant than the average interaction effect for the Field Poll subsample of the “baseline” specification without the additional covariates, though not implausibly higher.

Besides including the covariates in our regressions, we can also check their stability in the primary electorate or (for the Field Poll) among conservatives and liberals separately. We find few changes in the distributions of voter characteristics over the course of the primary campaigns, and that those that might have taken place were scattered and unsystematic.

To conduct this analysis, we perform χ^2 tests for the null hypothesis that the distribution of a given covariate is the same in the last poll of a primary as in the first poll of a primary for all available covariates and all primaries, both within each ideology and for the electorate as a whole. We find few changes in the distributions of voter characteristics over the course of the primary campaigns, and that those that might have taken place were scattered and unsystematic. Among conservatives in the Field Poll, we have 13 primaries and 8 covariates, leading to 104 tests, of which only 6 reject at the 5% level, which is very close to what we expect by chance. Among liberals in the Field Poll, we also have 104 tests with 8 rejections at the 5% level, also close to nominal size. For the electorate as a whole, for the Field Poll, we have 13 primaries and 9 covariates (ideology being one of them, with the ideologies being conservative, liberal and excluded in order to account for the ideology distribution among voters polarizing); hence 112 tests. We obtain 17 rejections at the 5% level, which is more than twice nominal size, but nevertheless not too worrisome. Finally, for the SUSA data, we have 21 primaries and 9 covariates (age, Bush approval, gender, race, education, income,

ideology, urban status and region of state), but we do not have all covariates for all primaries, so we have 146 tests. Of these, we reject at the 5% level in 12 cases, which is close to twice nominal size. Importantly, there is no clear pattern to the rejections. In particular, no single variable is predominantly responsible for a large number of the rejections. A table with the results from these χ^2 tests is available from the authors upon request.

Table C.1 Robustness to Covariates in the Field Poll Races

VARIABLES	Baseline	Covariates Levels	Covariates Interactions
Δ Voter-Cand. Ideological Convergence	0.128** (0.033)	0.166** (0.032)	0.141** (0.031)
Observations	5369	4492	4492
R-squared	0.574	0.573	0.616
Number of Primaries	15	15	15

Robust standard errors in parentheses. ** p<0.01, * p<0.05, + p<0.1.

Appendix D: Additional details on Internet sample recruitment

We recruited respondents for the Internet panel through two sources. First, we placed ads on Facebook inviting respondents to take part in the surveys, targeting them at individuals 20 years or older, who lived in the state holding the primary, and who reported in their profile a partisan or ideological identification consistent with the primary. We titled the ads with the state's abbreviation, e.g., "MI Elections Survey" and included a subtitle about the survey, e.g., "Researchers at [omitted for anonymity] want your opinion on the upcoming Michigan elections, take our survey and test your knowledge!" At the end of the survey, we asked respondents if we could contact them after the election for a follow-up survey and, if so, for their e-mail addresses (about two-thirds provided them). Second, we recruited respondents from a large consumer Internet panel, Decision Analyst, for the New Hampshire primaries.