Who Runs for Congress? A Study of State Legislators and Congressional Polarization

Connor Halloran Phillips[*] *Harvard University*

James M. Snyder, Jr.[†] *Harvard University and NBER*

Andrew B. Hall[‡§] *Stanford University*

November 16, 2022

Abstract

Why have the people running for Congress become more ideologically extreme over time, causing Congress to polarize? In this paper, we study how the pool of state legislators shapes the ideological distribution of congressional candidates. We geographically match state legislators to the US House districts that they could plausibly seek to represent. Although the national pool of state legislators has polarized over time, we find no evidence that they have polarized more where opportunities to win House seats are higher, or that more-extreme state legislators have become more likely to run for Congress relative to other state legislators in the same congressional districts. We conclude that the nationwide polarization of state legislators is particularly important in explaining the polarization of congressional candidates. The results emphasize the need to understand ideological selection into running for lower-level political offices in order to understand congressional polarization.

---

[*]Ph.D. Student, Harvard University.

[†]Leroy B. Williams Professor of History and Political Science, Harvard University.

[‡]Professor of Political Science, Stanford University. andrewbhall@stanford.edu, http://www.andrewbenjaminhall.com.

[§]We are grateful to *QJPS* editors Scott Ashworth and Joshua D. Clinton and to two anonymous reviewers for comments that helped to improve the manuscript.

Congress has polarized dramatically in recent decades, with legislators of the two parties more ideologically distant than ever before (McCarty, Poole and Rosenthal 2016). Understanding why is a major focus of the American politics literature today. From this research, we know that the set of people who run for Congress has itself polarized over this period, and that this development has likely caused a significant share of the overall growth in congressional polarization (Hall 2019; Theriault 2006; Thomsen 2017). Although this overall pattern of ideologically polarized candidates is apparent, we are only just starting to understand its underlying causes.

Existing research on congressional candidacy is largely limited to observing only the people who actually run for office. But to understand the roots of candidate-based polarization, it is necessary to observe the latent group of plausible candidates who might have run for each party in each congressional district—what we call the candidate pool. It could be that the candidate pool as a whole has polarized over time the way the set of people who actually run has, or it could be that the candidate pool as a whole has not changed, but more-extreme members of the candidate pool have become more likely than more-moderate members of the candidate pool to seek office over time. It is also possible that the extreme candidates in the pool have become better matched, geographically, to safe congressional districts where their party is likely to succeed.

Understanding these different mechanisms is important because it helps us identify the different potential incentives that affect who runs for office and polarization. If the entire candidate pool has polarized, then reducing polarization in Congress would require changing the candidate pool by, for example, encouraging the entry of more-moderate candidates in races for state and local office, which produce many plausible congressional candidates. If, on the other hand, polarization is mainly coming from differential selection of candidate-pool members into running for Congress, then reducing polarization requires encouraging existing, more-moderate members of the candidate pool to run. Finally, if the candidate

pool is better matched, geographically, to safe districts than it used to be, then this would point toward reforms to the districting of state and federal legislatures.

We introduce new data that allows us to observe an important sample from the latent candidate pool. In particular, we geographically match state senators and state house members, 1996–2016, to the congressional districts that contain their state legislative districts, so that we can track the part of the candidate pool that is comprised of sitting state legislators. This is in keeping with a small but important literature on national political ambitions among state legislators and the recruitment behavior of party leaders (e.g., Broockman et al. 2021; Maestas et al. 2006; Maestas, Maisel and Stone 2005; Thomsen 2014). State legislators constitute a particularly important set of candidates for Congress. Overall, during our period of study 12.3% of non-incumbent candidates were sitting or former state legislators, as were 15.8% of primary winners and 48.5% of general-election winners.

Using NP-Scores (Shor 2018) to measure state legislators' ideological positions, we offer a number of descriptive patterns concerning the candidate pool that help make sense of the changing ideologies of candidates for Congress.

First, after documenting that the overall candidate pool—and not just the state legislators who actually run for Congress—has polarized substantially, we show that the geographical match between candidate pools and congressional districts does not explain the rise in polarization since 1996. Not surprisingly, the number of Democratic state legislators is higher, and their ideology on average more extreme, in safe Democratic congressional districts, and vice-versa for Republicans. However, we see no evidence that this match has strengthened over time. The leftward shift of Democratic candidate pools and the rightward shift of Republican candidate pools have been roughly equal in safe, partisan districts and in more-competitive districts.

Second, we find that the ideological gap between candidates who do and do not run, from among the pool, has been largely fixed over time. Among the candidate pool, more-extreme state legislators disproportionately select into running for the House—but this selection is

only present in the Republican party, and it is of similar magnitude across the time period studied. Conversely, Democrats display very little differential selection into running for Congress. This finding contrasts with the differences in the overall national means, which indicate that the Democratic state legislators who run for Congress are significantly more liberal than those who do not. Existing work argues that Republicans are asymmetrically driving the rise in congressional polarization (Carmines 2011; Hacker and Pierson 2005, 2014, 2015); our results are consistent with this literature, and suggest that a key mechanism in this overall asymmetric polarization is the differential decision of more-extreme Republican state legislators to seek congressional seats. However, because this differential selection has not changed over time, it is unlikely to explain the increase in polarization over time by itself—though in conjunction with a polarizing candidate pool it can contribute importantly.

Finally, when we restrict our focus to state legislators who win congressional primary and general elections, we find that the ideological gap between these legislators and others in their pools is even larger than the average gap across all candidates, indicating that more-extreme state legislators are not punished electorally on average. This gap is especially pronounced in safe seats—defined as those with a heavily lopsided presidential vote—particularly among Republicans. Because the number of safe seats has risen over time, this difference helps contribute to the increase in polarization. However, as we show, the overall polarization of the pools still accounts for the bulk of the trend.

As a result, we conclude that the overall, countrywide polarization of the state legislative candidate pool is a key driver of candidate-based polarization. State legislators have polarized across all types of congressional districts, leading to a more-polarized set of congressional candidates finding electoral success. These results suggest that the main existing accounts of candidate-based polarization, Thomsen (2017) and Hall (2019), are incomplete: theories of candidate-based polarization should next examine the decision to run for state legislature and other lower offices in order to further explain the rise of polarized congressional candidates.

Background

The majority of polarization in Congress has occurred because of polarization among congressional candidates rather than changes in the outcomes of elections: Hall (2019) estimates that about 80% of the increase in polarization in the US House since 1980 would still have happened even if voters chose the most moderate candidate available in each election. Accordingly, understanding why Congress is polarized requires understanding why congressional candidates have polarized. Logically, this could have occurred in three (not mutually exclusive) ways:

(H1) Potential congressional candidates as a group have polarized, yielding more-extreme candidates as a byproduct.

(H2) Extremists have become relatively more likely to seek office than in the past.

(H3) Potential congressional candidates have become better sorted over time, matching extremists to safe districts where they are more likely to be successful.

Surprisingly little work is directed to addressing candidate polarization, with the notable exceptions of Thomsen (2014, 2017) and Hall (2019), both of whom provide evidence in support of (H2). Thomsen (2014, 2017) proposes the theory of "candidate fit," arguing that eligible moderates see less appeal to serving in a more partisan, polarized chamber than ideologues do and accordingly are less likely to run for Congress. Using data from the 2000–10 elections, Thomsen demonstrates that relatively moderate state legislators run for Congress less often than otherwise similar extremists in comparable districts, a difference that is especially pronounced among Republicans. Hall (2019) adds to this analysis by noting that extremists value winning a congressional seat more than moderates do because they bear a higher ideological cost from opposite-party victories. In combination with a lack of candidate fit, these differing incentives have led moderates to increasingly opt not to become candidates as the costs of candidacy (in terms of fundraising and enduring personal and negative campaigning) have risen and the benefits of serving in Congress have declined.

Although informative, these analyses cannot rule out the potential confounding effect suggested by (H3)—that favorable electoral opportunities for extremists have increased—or the possibility that the differential propensity of extremists to seek office has been less important than the overall polarization of the candidate pool posited by (H1). To successfully discriminate among the effects of these three hypotheses, we need to introduce the concept of a candidate pool.

Defining the Candidate Pool

In every congressional race, only a fraction of the potential candidates actually decide to enter. Understanding selection into running for Congress therefore requires us to compare declared candidates to the unobserved, latent pool of potential candidates (Fowler and McClure 1989). We define the candidate pool for each party in a given district and election year as the set of individuals who could seriously contend for that party's nomination in the district during that cycle. This definition includes lower-level elected officeholders as well as non-elected officials who nevertheless have the connections, resources, and influence to potentially mount a credible campaign, often figures like attorneys or business executives.

With this concept, we can restate the three hypotheses above in terms of candidate pools:

(H1) Candidate pools have grown more extreme on average, mechanically yielding more-extreme candidates.

(H2) Within candidate pools, extremists have become relatively more likely to seek office than in the past.

(H3) Relatively extreme candidate pools in relatively safe districts have become a greater proportion of candidate pools. (This may have occurred from an increase in the number of safe districts, if candidate pools in such districts were already more extreme, or via sorting that rendered candidate pools in safe districts more extreme and those in marginal districts more moderate.)

The fundamental problem with studying candidate pools is that only the potential candidates who actually run are observed. We address this issue by focusing on a key group of potential candidates—sitting state senators and state house members. Political observers generally regard state legislators as high-quality potential congressional candidates, and approximately half of Congress is in fact comprised of former state legislators (Ramsdell 2018). The advantage of analyzing state legislators is that we can compare those who run for Congress with similarly situated peers who do not. Further, ideological scores based on state legislative roll-call voting records are available for both congressional candidates and noncandidates. One concern with limiting our focus to state legislators is that they might differ systematically from non-state legislators in ways that make them an unrepresentative subset of the pool. Later in the paper, however, we show that non-state legislators who run for and win seats in Congress are ideologically quite similar to the state legislators who do so, indicating that state legislators are an acceptable proxy for the overall candidate pool.

Therefore, by matching state legislators with their corresponding congressional districts, we can assess how the candidate pool has changed ideologically, and in turn by what mechanism these changes have driven polarization in Congress. We examine elections from 1996 to 2016, a longer window than in Thomsen's analysis and one that should allow us to track changes over time in the candidate pools and the patterns of candidacy. To preview our results, we find the strongest support for (H1), at least over the period we study. State legislators of each party within each congressional district have been growing more extreme on average. Conversely, although conservative Republicans are much more likely than moderate Republicans to run for Congress (a similar effect does not seem to exist among Democrats), this difference has been constant over the period studied, which implies that any contributions of (H2) to the *growth* of candidate polarization must have come before our period of study. This does not mean that the differential selection of candidates from the pool is not an important part of polarization—after all, as the pool polarizes, a fixed rate of differential selection further exacerbates polarization—but it suggests that understanding the incentives

of those who run for state legislature in the first place is important, too. Finally, the increasing number of safe districts with extreme pools has contributed to some polarization, particularly among Republicans, but this change only accounts for a small amount of the overall increase in congressional polarization.

Matching State Legislators to Congressional Districts

We use Shor and McCarty's scores (Shor 2018) to measure the ideology of state legislators. These scores are based on roll-call voting records and use responses to the National Political Awareness Test (NPAT) battery of position items to bridge across different state legislative chambers. Following Shor and McCarty, we call these *NP-Scores*. See Shor and McCarty (2011) for details. We study the years 1996–2016 because these NP-Scores are only available for most states starting in 1996.

We match state legislative districts to congressional districts based on the degree of overlap of their (voting) populations, using a variety of different sources. Most mappings are constructed from precinct-level election results, which we obtain for 1996–2012 from the Harvard Election Data Archive, for 2000 from American University and the University of Kentucky's Federal Elections Project, and for various years from files provided by Secretaries of State or State Boards of Elections. For states and years in the 2000s and 2010s where precinct-level results are not available, we use L2 voter file data from 2016 that reports district lines for both decades.

For each state senate and house district, we estimate the share of that district's voting population that lives within the boundaries of each congressional district in the state. We calculate these shares on the basis of registered voters (when the number of registered voters per precinct is available or when using the L2 data) or ballots cast in the top-of-the-ticket race (in other cases). We then average these shares within each redistricting cycle. Legislators are judged to be in their party's pool for a given congressional district if at least one-third

of their voters are located in that district.[1] Our findings are qualitatively similar when we set that cutoff at 25% and 50% (see Appendix A in the Supporting Information).[2]

In our basic setup, we link sitting state legislators to their relevant congressional districts.[3] However, former state legislators could also be thought of as being in the pool for Congress. In Appendix D in the Supporting Information, we explore a method for incorporating these former state legislators into the pool as well, arriving at similar conclusions.

We can use the precinct-level results and voter file data to match legislative and congressional districts that are contested in the same election cycle.[4] After each congressional redistricting, however, legislators representing the old state legislative districts choose whether or not to run in the new congressional districts. For these post-redistricting years, we match census data at the block level to the old state legislative districts and new congressional districts and calculate the share of the population in each state legislative district that lies within each new congressional district.[5]

---

[1] A simpler option would be simply to assign each state legislator to the congressional district that most overlaps their district. However, we observe a non-trivial number of cases (57, or 7.8% of all state legislators in our sample who run for Congress) in which a state legislator runs in a congressional district that is not the district with the largest overlap with their state legislative district.

[2] With the 33% and 25% cutoffs, it is possible for us to place a state legislator in multiple pools, which occurs for 12.1% of the observations in our sample. Accounts of state legislators' decisions to run for Congress often mention that at least some legislators have a legitimate prospect of seeking office in multiple districts (see, e.g., Fowler and McClure 1989, chap. 3).

[3] For our sample, 6.8% of congressional candidates, 8.9% of primary winners, and 31.1% of general election winners are sitting state legislators.

[4] Unfortunately, precinct names and boundaries often change across years, so it is difficult to use them to match districts across elections.

[5] We also employed an analogous procedure to match state legislative and congressional districts in several states where election returns and voter file data were not available but applicable census data was. In doing so, we used the Missouri Census Data Center's Geographic Correspondence Engine for years in which it has the relevant districts available and constructed the mappings ourselves for years when it did not. State legislative redistrictings that occur between congressional redistrictings (usually in response to a court order)

So far as we know, we are the first to construct and analyze candidate pools in this fashion. The closest analogue comes from Thomsen (2014, 2017) and Aldrich and Thomsen (2017), who match each state legislator to the congressional district having the greatest overlap with their district based on census population data.[6] These analyses, however, control for district-level characteristics rather than performing within-pool comparisons. In addition, our method of constructing pools incorporates the degree of overlap between state legislative and congressional constituencies, which allows us to place legislators from fragmented districts into multiple candidate pools or exclude them from pools altogether if they have no natural congressional constituency, which provides a more accurate portrait of the political opportunities available to them.

Our chosen period potentially contains 9748 distinct pools (435 congressional districts times 11 general elections plus 89 special elections, times 2 parties). Due to the inability to match state legislative districts to congressional districts in some states and years, or due to missing NP-Scores, our dataset has only 8484 of these pools (4242 district-years). Specifically, NP-Scores are unavailable in states and years accounting for 458 pools, and we exclude an additional 740 pools because of a lack of mappings. The main source of missingness in the mappings consists of redistrictings where sufficiently detailed information on either the old or new districts was unavailable, along with a handful of states where precinct-level results were incomplete. When at least one pool in a state cannot be constructed for a given year, we drop all districts in the state in that year from the analysis. This choice prevents included pools from, for example, skewing rural if urban precincts were more likely to provide incomplete results or skewing white if court-ordered redistrictings affected relatively nonwhite areas more

---

are easier to accommodate, as the mapping from the first post-redistricting elections gives the appropriate congressional district shares for each of the new districts in subsequent elections.

[6]Two other approaches also deserve mention. Maestas, Maisel and Stone (2005) study state legislators seeking national office by surveying all state legislators whose constituencies overlap a random sample of congressional districts. Fraga, Juenke and Shah (2020) use census block data to determine the proportion of a state legislative district represented by a non-white member of Congress.

often. We also omit the state of Nebraska, which has a nonpartisan legislature, accounting for the remaining 66 missing pools. Ultimately, we are able to incorporate pools from each of the other 49 states in at least some years.[7] In terms of state legislators, the dataset contains 20686 unique individuals. Many of these state legislators appear in multiple pools.

Another key variable in our analyses is the congressional district-level presidential vote, which we use as a proxy for a district's partisanship. These data are from the *Almanac of American Politics* for the 1990s and the Daily Kos's presidential returns by congressional district dataset for the 2000s and 2010s. We use the Democratic share of the two-party presidential vote in that year's presidential election (if a presidential election year) or the most recent presidential election (if not). In all years, we normalize this share by subtracting the national average and adding 0.5; thus, it is centered at 0.5 in a district reflective of the national vote. For convenience, we refer to this variable as *Democratic Presidential Vote* below. In many analyses, we use this variable to define safe and competitive districts. Competitive districts are those for which the normalized presidential vote is between 0.45 and 0.55, i.e., where the margin in the district is within 5% of the national margin. When the vote in a district falls outside this interval, we define the district as safe for the advantaged party. In our sample, 25.3% of the districts are competitive. This number falls sharply over time, from 37.8% in 1996 to 15.9% in 2016.

Additionally, we use data from Hirano and Snyder (2019) to determine which state legislators run for Congress and whether they win or lose their primary and general elections. In one analysis below, we study ideology scores constructed from Project Vote Smart's National Political Courage Test (formerly the NPAT); in another, we use common space

[7] In Appendix F in the Supporting Information, we verify that the missing pools do not appear to meaningfully differ from the included ones in ways that might impact our results.

W-NOMINATE scores; and in a third, we use CFScores from the DIME database (Bonica 2018).[8]

Descriptive Statistics on Candidate Pools

We begin with some simple summary statistics about the pools themselves. First, the pools vary greatly in size, from 0 to over 100 potential candidates (in Vermont for the Democrats, North Dakota for the Republicans, and New Hampshire for both parties). Pools tend to be larger in states where the number of state legislators per congressional district is larger, and in congressional districts where a party is favored to win in the general election based on the district's partisanship.

Figure 1 – Size of candidate pools across presidential vote share.

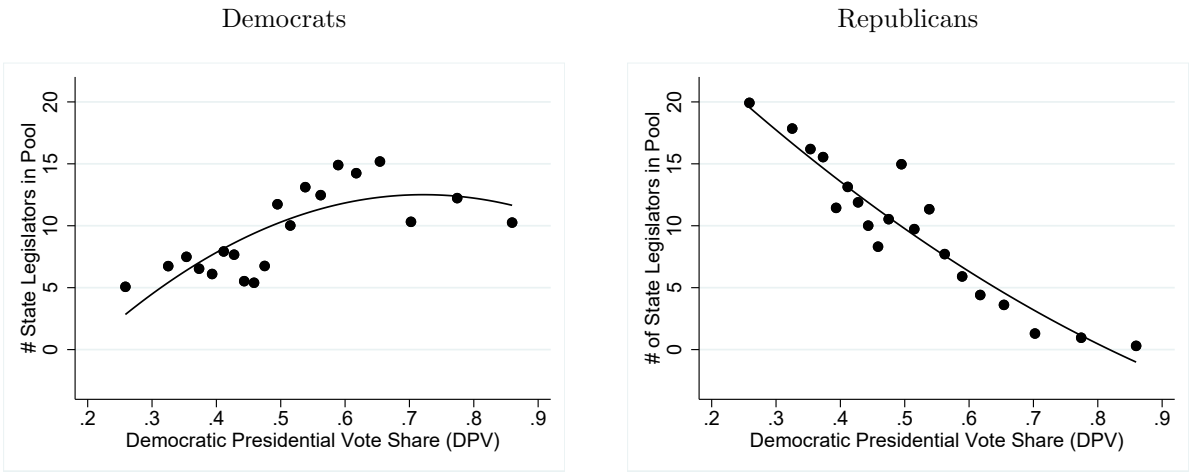


Figure 1 plots the number of state legislators in each party's pools against the congressional district's *Democratic Presidential Vote*. Each point represents a binned average. Note that we include pools with zero legislators in constructing these averages. These empty

[8]See Moskowitz, Rogowski and Snyder (2022) for more information about NPAT scores, see voteview.com (Lewis et al. 2019) for more information about NOMINATE, and see Bonica (2014) for more information about CFScores.

pools occur because we define candidate pools separately in each district for each party.[9] In general, a pool's size is an increasing function of a party's strength in the corresponding congressional district. However, very Democratic areas—those with *Democratic Presidential Vote* above 70%—tend to have somewhat smaller pools because many of these congressional districts are clustered in a few heavily urbanized states with small legislatures relative to the size of their congressional delegation (e.g., California).[10]

Table 1 presents summary statistics of NP-Scores for the pools (more-positive values indicate greater conservatism). Notably, the average Democratic pool is more liberal than the average ideology of Democratic legislators nationwide, but the pool average is more similar to the national average for Republicans. Pools in competitive districts tend to be somewhat more moderate, and pools in safe districts much more extreme, than the average pool. In addition, pools that produce congressional candidates are ideologically slightly more extreme than the average pool and pools that do not produce candidates.

The bottom panel of Table 1 presents standard deviations. Overall, Democratic state legislators in our sample are more ideologically heterogeneous than Republicans. Unsurprisingly, for both parties, within-state heterogeneity is noticeably lower than it is nationwide. Similarly, heterogeneity is even smaller within pools. This pattern implies that there is substantial heterogeneity in the mean ideology across pools, which is what we would expect given the geographic ideological sorting that is believed to exist in the US today.[11]

Figure 2 shows that this heterogeneity is strongly correlated with congressional district-level partisanship. In both parties, the average NP-Score falls sharply as *Democratic Presi-*

[9]The Democratic pool is empty for 396 district-years (9.3% of all 4242 district-years included in the dataset), and the Republican pool is for 621 (14.6%). In every congressional district, at least one party's pool was nonempty.

[10]When we normalize the number of state legislators in each pool by dividing by the ratio of the size of the state legislature to the number of congressional districts in the state, the relationship becomes much more linear, especially for the Democrats.

[11]See, e.g., Bishop (2008); Hopkins (2017).

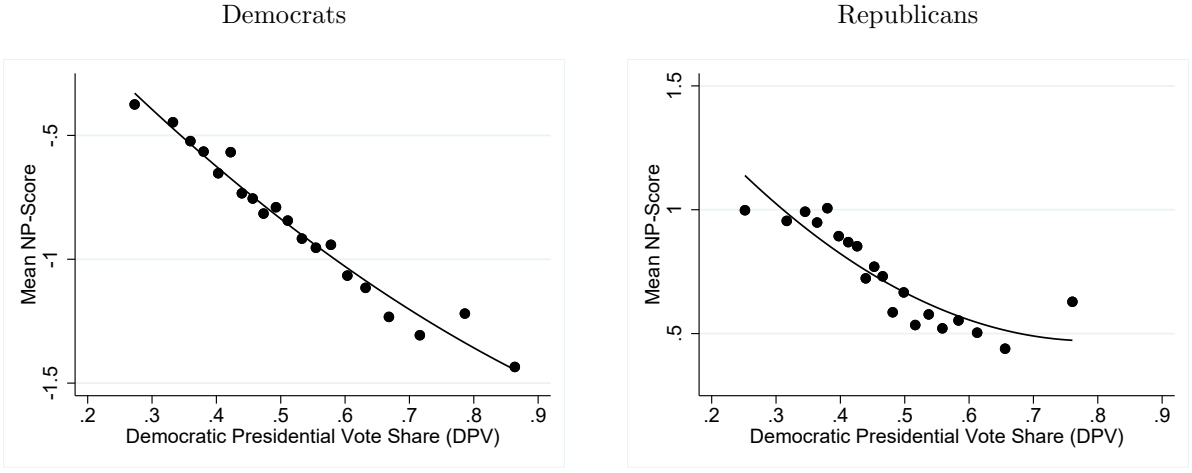
Table 1 – NP-Scores of state legislators, by party.

	Democrats			Republicans		
Item	Mean NP-Score	Median NP-Score	# Legs.	Mean NP-Score	Median NP-Score	# Legs.
National Average/Median	-0.79	-0.80	40,239	0.74	0.76	41,405
			# Pools			# Pools
All Pools	-0.86	-0.87	3,846	0.74	0.78	3,621
Pools in Safe Districts	-1.17	-1.11	1,476	0.91	0.90	1,690
Pools in Competitive Dists	-0.83	-0.86	1,001	0.63	0.68	1,053
Pools With 1+ Cong Cand	-0.89	-0.94	264	0.77	0.79	354
Pools With 0 Cong Cands	-0.86	-0.86	3,582	0.73	0.78	3,267
National Std. Dev.	0.53			0.39		
Avg. Within-State Std. Dev.	0.32			0.28		
Avg. Within-Pool Std. Dev.	0.27			0.24		

In the first row the unit of observation is the individual legislator. In rows 2-6 the unit of observation is the pool.

dential Vote increases—the pools tend to be more conservative in more-Republican districts—and in the case of Democratic pools the relationship is approximately linear.

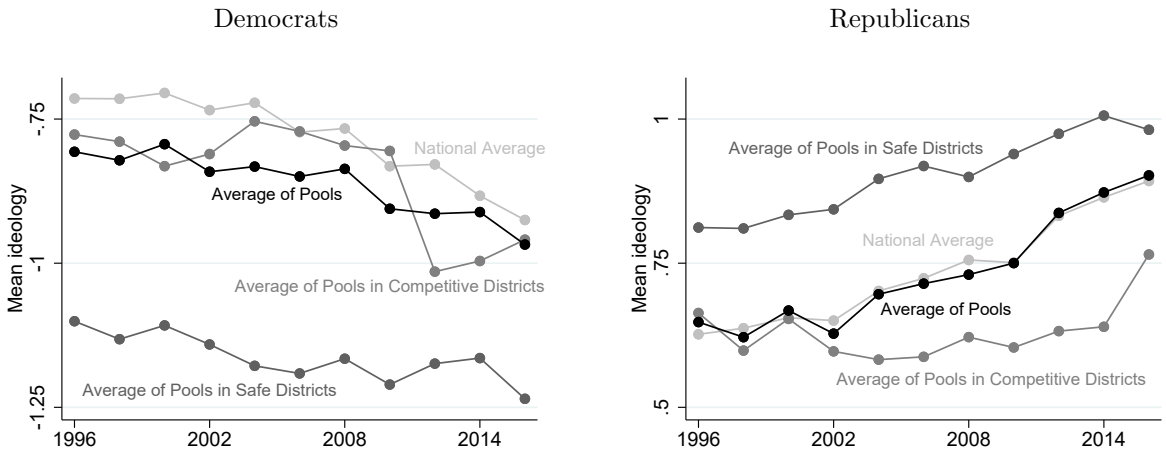
Figure 2 – Average NP-Score of pools across presidential vote share.



Similar Polarization Trends Across Districts

Figure 3 shows how ideology has trended over time. It presents the yearly averages of the first four rows of Table 1. Several patterns are noteworthy. First, for Republicans, the average ideology of the pools closely tracks the overall national average. For Democrats, both tend to move in the same direction, but the pool average is always more liberal than the national average. Overall, both parties' pools have polarized over the period of study.

Figure 3 – Average NP-Score of pools over time.



In both parties, the pools in competitive congressional districts tend to be more moderate than average. This is especially clear for Republicans starting in 2002 and for Democrats from 2002 through 2010. As is true in the overall sample, however, the most striking differences are for the pools in safe districts. For Democrats, the state legislators in these pools are on average much more liberal than they are in the average pool, and the gap has not changed much over time. Similarly, Republican pools in safe districts tend to be more conservative than average. Generally speaking, the pools in all types of districts have become increasingly extreme, with Democrats moving to the left and Republicans moving to the right, although the trend in competitive districts is a bit noisier. These findings suggest that the overall polarization of the candidate pool has been more important in driving congressional polarization than differential trends in different types of pools. To verify this impression, however,

we have to examine whether and how runners differ ideologically from their pools and if this pattern has varied over time.

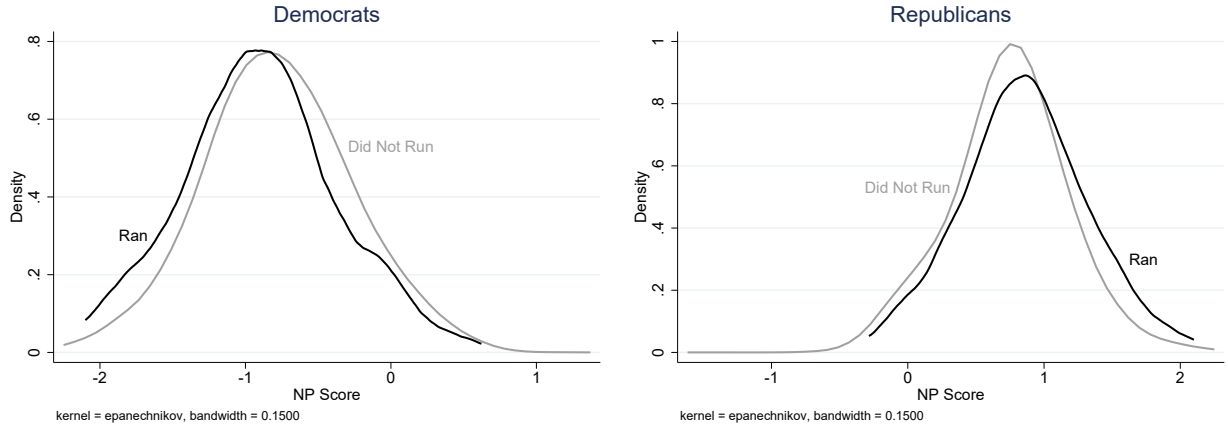
Selection into Running and Polarization

In this section, we analyze differences in NP-Scores between state legislators who run for Congress and those who do not. Table 2 presents some initial summary statistics. Democrats who run for Congress are noticeably more liberal than those state legislators who do not, and Republicans who run are similarly more conservative.[12]

Table 2 – NP-Score of state legislators who run vs. those who do not.

	Democrats		Republicans	
Item	Mean NP-Score	# Obs.	Mean NP-Score	# Obs.
Ran for Congress	-0.905	306	0.853	404
Did Not Run	-0.777	35,171	0.743	36,199

Figure 4 – Differences in ideological distributions of NP-Scores, runners vs. non-runners.



[12]These differences are statistically significant at the .01 level using a simple t-test, even with the standard errors clustered by legislator.

Figure 4 explores these differences in more depth, presenting not just the difference in means for the two candidate pools but the differences in the entire distributions of NP-Scores. As the densities show, we see modest shifts away from the middle for runners as compared to non-runners in both parties.[13]

As we demonstrate below, however, the patterns change when we compare those who run only with other legislators in their pools rather than the entire set of co-partisan state legislators. For the rest of this section, we focus on within-pool comparisons. We begin with regressions of the following form:

$$NPScore_{id} = \alpha_d + \beta\, Ran_{id} + \epsilon_{id}$$

where i indexes state legislators and d indexes pools. Each pool corresponds to a party in a congressional district in a given year, and only legislators who are assigned to that pool are included. Also, we only include pools in which at least one state legislator ran for Congress and at least one did not, because all of the identifying variation comes from these pools. In all regressions, we cluster standard errors at the state level. Doing so is relatively conservative and allows ϵ_{id} to be correlated within legislators over time, within each pool, and even across pools within states.[14]

We focus our analyses on differences in mean NP-Scores, for simplicity. This choice could induce error in our analysis because ideological scalings tend to be less accurate for candidates with more-extreme positions, as well as because some distributional shifts might occur without having large effects on means. In Appendix C in the Supporting Information, we present parallel analyses in which we replicate all of the main tables below, but use two

[13] We can also examine differences in medians rather than means for these distributions. Among Republicans, the difference in median NP-Score for runners vs. non-runners is 0.099. For Democrats, this difference is -0.112.

[14] This last consideration might be relevant because some state legislators appear in multiple pools, even in the same year, and have the same ideology in all pools in which they appear.

ordinal rather than cardinal measures of the changes in the ideological positions of runners vs. non-runners—one based on ranking candidates in terms of their ideological positions with respect to their parties, and the other based on converting these ranks to a simple dummy variable indicating more or less extreme candidates. Because these alternative analyses yield similar conclusions, we do not present them in the body of the paper.

We do not give any causal interpretation to β: the purpose of this exercise is to isolate the proportion of polarization among congressional candidates that has come from selection into running within each party's pools. In other words, we are interested in decomposing the observed increase in polarization over these two decades rather than isolating the causal effect of any particular treatment, analogous to the approach taken by Hall (2019, chap. 2). In this regard, β simply gives the difference-in-means of ideology between runners and nonrunners that we could also obtain via t-tests, with a convenient way to adjust standard errors.[15]

Table 3 presents the results. The columns on the left are for Democrats, and the columns on the right are for Republicans. Each row presents the results of a separate regression for different types of pools. The columns labeled "Coeff." present the point estimates on β, the columns labeled "Std. Error" present the regression standard errors, and the number of observations for each regression is shown in square brackets.

The most dramatic patterns are for Republicans: those state legislators who run for Congress are significantly more conservative than those in their pools who do not. This relationship holds overall, as well as for all subsets of pool types, and the differences are substantively meaningful. For example, the first row shows that for all pools, the ideological difference is about 38% of the within-pool standard deviation among Republicans. Among

[15]In Appendix B in the Supporting Information, we present the reverse analyses, regressing running on NP-Score with pool fixed effects for pools that have both runners and non-runners. These regressions should not be thought of as causal either, because we do not control for other candidate-level characteristics affecting propensity to run for Congress that may covary with ideology.

Table 3 – Within-pool comparisons: NP-Score of state legislators who run vs. those who do not.

	Democrats			Republicans		
Sample	Coeff.	Std. Error	# Obs.	Coeff.	Std. Error	# Obs.
All pools	-0.008	(0.019)	[3898]	0.092	(0.021)	[5993]
Open seats	0.017	(0.021)	[2104]	0.092	(0.023)	[3001]
Safe districts	-0.015	(0.024)	[1718]	0.098	(0.029)	[3358]
Competitive districts	0.017	(0.035)	[1169]	0.106	(0.034)	[2066]

Democrats, the differences between those who run and those who do not are smaller than for Republicans, and the overall coefficient is very close to zero.[16]

When we compare the within-pool results in Table 3 and the full-sample summary statistics in Table 2, two differences are apparent. First, for Republicans, the within-pool difference in NP-Scores between those who run and those who do not is less than the overall difference between runners and non-runners. Second, for Democrats, the within-pool difference in NP-Scores between those who run and those who do not is negligible, but the overall difference is much larger and significant. That is, Democratic state legislators who run for Congress on average have approximately the same ideology as the other legislators within their pools, whereas overall, Democratic runners are more liberal than the non-runners. This pattern occurs because, as Table 1 shows, the average Democratic pool is more liberal than the average Democratic legislator.

One implication of these differences is that for many important research questions—for example, whether or not candidates are being strategic in their entry decisions—it is crucial to control for the characteristics of pools. Controlling for these characteristics, however, requires matching all state legislators who do not run to the congressional districts where they could plausibly run. Once a researcher has done the required matching, it is possible to

[16]The pattern of estimates is weakly consistent with strategic behavior—more-moderate candidates run in competitive areas. However, the differences between different types of districts are generally not statistically significant.

focus on within-pool comparisons and use pool-specific fixed effects (as we do above). This method is often superior to using a battery of control variables when attempting to deal with omitted-variable bias.

Variation Across Districts

Table 3 only shows small differences between safe and competitive districts. Figure 5 explores this finding graphically in more detail, via binned averages. To construct these averages, we first compute the difference between the NP-Score of each state legislator who runs for Congress and the average NP-Score of the others in their pool who do not run, which we call *Ideology Difference*. Next, we divide the congressional district pools into twenty equal-sized bins based on their *Democratic Presidential Vote*, and for each bin, we compute the average of *Ideology Difference*.[17] We plot these for Democrats on the left and Republicans on the right, along with the fitted values from quadratic regression curves.[18]

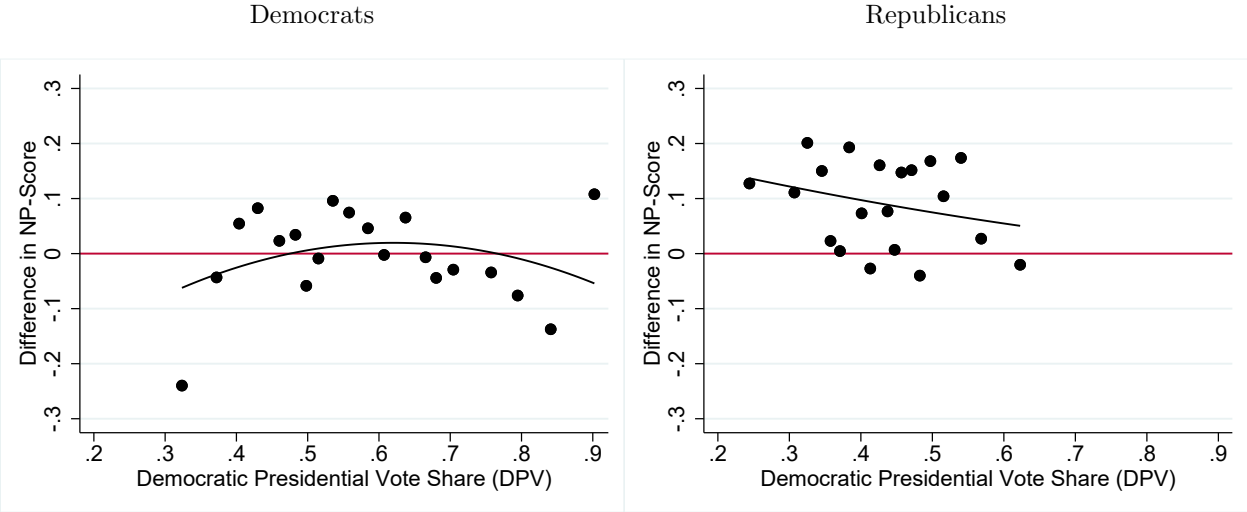
Democrats do not show a clear monotonic pattern across values of *Democratic Presidential Vote*. However, as Table 3 above suggests, the highest values of *Ideology Difference* tend to occur in competitive districts.

For Republicans, the values of *Ideology Difference* are more consistently positive. Also, the relationship between *Ideology Difference* and *Democratic Presidential Vote* is generally downward sloping, although it is noisy. If we compare the most Republican districts with

[17] *Ideology Difference* is non-missing only for pools in which at least one state legislator ran for Congress and at least one did not.

[18] Notice that pools exist for a wider range of the *Democratic Presidential Vote* variable for Democrats compared to Republicans. Many heavily Democratic congressional districts (i.e., those with a *Democratic Presidential Vote* of about 65% or more) do not have any Republican state legislators. The same phenomenon occurs on the opposite end—that is, some districts have a *Democratic Presidential Vote* below about 30% and no Democratic legislators—but in fewer instances. These patterns reflect a variety of geographic factors, including the urban concentration of Democratic voters, requirements regarding majority-minority districts, and possibly partisan gerrymandering.

Figure 5 – Within-pool difference in NP-Score between state legislators who ran and did not run across presidential vote share.



other districts, the tendency for Republican runners to be more conservative than their pools is even more pronounced. For example, define *very safe* districts as those with a *Democratic Presidential Vote* below 35%. When we run a regression similar to those above that also includes an interaction term, *Ran* × *Very Safe District*, the coefficient on this term is positive and statistically significant at the 0.05 level.[19] That is, in the very safe districts those legislators who run are significantly more conservative relative to their counterparts in other districts.

Safe districts are especially important to consider for two reasons. First, almost all of the candidates who win primaries in these districts go on to win the general election and become members of Congress. Second, as noted above, the number of competitive congressional districts has fallen over our period of study. By definition, this decline increases the number of safe districts, and more of these have become safe Republican districts than have become safe Democratic districts. This disproportionate growth, together with the

[19]Specifically, the coefficient on *Ran* is 0.078 (SE 0.022), and the coefficient on *Ran* × *Very Safe District* is 0.073 (SE 0.033). In the analogous analysis for Democrats, neither the main effect nor the interaction term is significant.

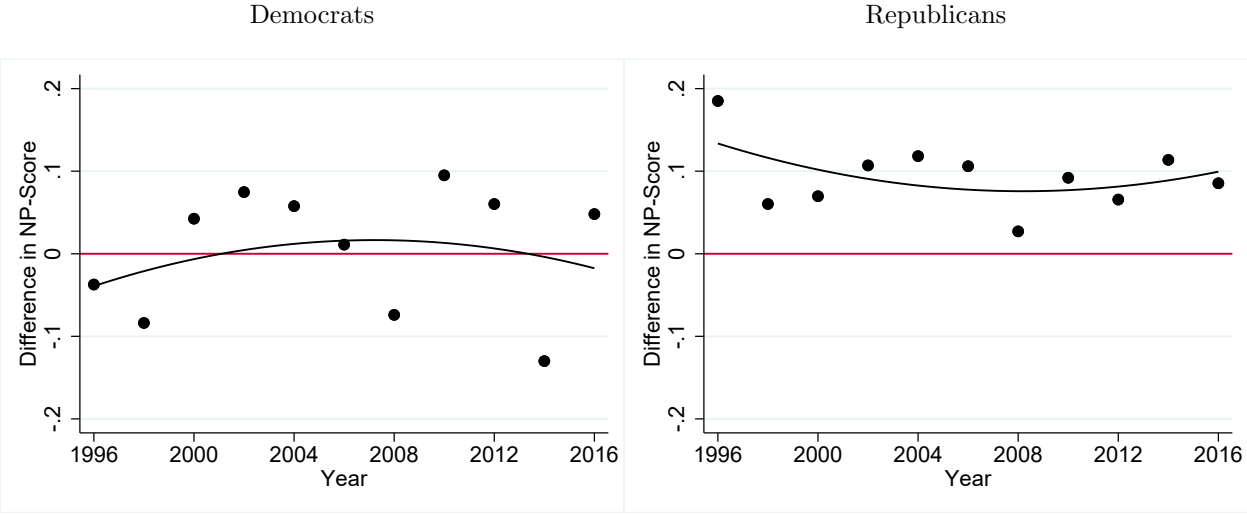
greater conservatism of Republican state legislators who run in these districts, could be one factor that has increased polarization, as detailed by our hypothesis (H3).

Separate from district-level partisanship, we also investigate whether the differences between runners and non-runners are greater in larger candidate pools, or in candidate pools with more ideological dispersion. These analyses are available in Appendix E in the Supporting Information, and offer tentative evidence that both of these factors are associated with larger differences, at least among Republicans.

Variation Over Time

Given that polarization in Congress increased sharply over our period of study, it is important to consider changes over time. For each year in our data, Figure 6 shows the average value of *Ideology Difference*. Again, we include quadratic regression curves and present results for the Democrats on the left and the Republicans on the right.

Figure 6 – Ideology of state legislators who ran and did not run over time.



The patterns are clear: in neither party is there a significant trend over time. This finding implies that, at least for the period 1996–2016, changes in who runs from the pools are not driving the increase in polarization (meaning we find no support for our hypothesis H2).

Who Wins from the Pool?

We now study who wins in primary and general elections, again focusing on within-pool comparisons of NP-Scores. For primaries, we run regressions similar to those above but with the independent variable *Won Primary* in place of the variable *Ran*. This variable is set equal to 0 for state legislators who either did not compete or who lost the primary. We restrict attention to pools that contain a primary winner and at least one candidate with *Won Primary* $= 0$.[20] For general elections, we run analogous regressions with the independent variable *Won General* in place of *Ran* and on the analogous subset of pools. In all regressions, we cluster standard errors at the state level.

Table 4 presents the results. The top panel analyzes primaries, and the bottom panel analyzes general elections.

Table 4 – Within-pool comparisons: NP-Score of state legislators who win primary or general elections vs. others.

	Democrats			Republicans		
Sample	Coeff.	Std. Error	# Obs.	Coeff.	Std. Error	# Obs.
Primary Election Winners vs. Others in Pool						
All pools	-0.032	(0.029)	[2379]	0.108	(0.023)	[3488]
Safe districts	-0.043	(0.032)	[904]	0.164	(0.034)	[2014]
Competitive districts	-0.022	(0.047)	[789]	0.065	(0.031)	[1104]
General Election Winners vs. Others in Pool						
All pools	-0.039	(0.029)	[1184]	0.136	(0.034)	[1921]
Safe districts	-0.045	(0.032)	[878]	0.170	(0.039)	[1467]
Competitive districts	-0.034	(0.074)	[252]	0.043	(0.051)	[448]

For Republicans, the results for all pools (first row of each panel) mirror those in Table 3 for those who run. That is, both primary and general election winners have significantly higher NP-Scores than the average state legislator in their pools. In safe districts, the patterns are even more striking. For both primaries and general elections, the winners are much more conservative than the pool averages. For competitive districts, on the other hand,

[20] Again, all of the identifying variation comes from these cases.

the winners are still more conservative than their pools, but the differences are smaller and insignificant.[21]

For Democrats, both primary and general election winners are overall ideologically somewhat to the left of their pools (see first row of each panel), a pattern that holds in both safe and competitive districts, although the differences are not statistically significant. Note that the relationship is stronger than that found in Table 3 above for those who run. This pattern suggests that, at least among the subset of candidates and potential candidates who are state legislators, Democratic primaries may select more-liberal candidates. The differences between winners and others in their pools are similar for primary and general election winners, suggesting that general elections select less strongly on ideology among state legislators.[22] The estimates are generally noisy, however, so we must treat them with some caution.[23]

As with the differences between runners and their pools depicted in Figure 6, these relationships do not vary significantly over time: a time trend interacted with *Won Primary* and *Won General* is insignificant for both Democrats and Republicans.

[21] When we compare safe districts with all others or with competitive districts, the differences in the estimated β are statistically significant at the 0.05 level for both primary elections and general elections. The difference between primary winners and others in their pools for competitive districts is marginally statistically significant, but as Appendices A and D show, this result is not robust to alternative definitions of the pools.

[22] This pattern is consistent with the ideas that state legislators tend to be stronger general election candidates regardless of their ideology and that primary winners are virtually assured of election in safe districts.

[23] In particular, only 24 Democratic pools in our sample contain general election winners in competitive districts.

Are the Findings Specific to State Legislators?

We are only able to measure one portion of the overall pool of people who might plausibly run for the House, namely, sitting state legislators who have cast roll-call votes and so can be scaled ideologically using NP-Scores. But are state legislators more ideologically extreme, or moderate, than other plausible candidates for Congress? The answer is unclear because we lack information about the overall pool of non-state legislative candidates. We can, however, assess whether any significant differences exist between the state legislators and non-state legislators who choose to run, and between the state legislators and non-state legislators who win seats in Congress. We find no significant difference in both cases after we control for pool characteristics, suggesting that our analyses of state legislators provide helpful evidence about overall trends in who runs for the House.

To compare state legislators with non-state legislators, we use the NPAT scores. These are available for 26.5% of the (non-incumbent) candidates who run for Congress between 1996 and 2016, including 15.6% of the state legislators who run and 27.6% of the non-state legislators who run. Ideally we would focus on within-pool or at least within-district comparisons, asking whether state legislators who run are significantly different from non-state legislators who run in the same pool or district. However, the number of cases becomes quite small and possibly quite unrepresentative. Thus, in this analysis we simply control for one of the most important distinguishing characteristics of pools, *Democratic Presidential Vote*.

We also compare state legislators who win and go on to serve in Congress with non-state legislators who win, using their W-NOMINATE scores. Although winners are of course a non-random subsample of all candidates, we can at least make a universal comparison within this group. Again, the comparisons are not within-pool, so we control for *Democratic Presidential Vote*.

The results are shown in Table 5. As the table indicates, we do not find noticeable differences between state legislators and non-state legislators for either sample or measure.

Table 5 – Ideology of state legislators compared to non-state legislators.

	Democrats	Republicans
NPAT Scores		
State Legislator	-0.009	-0.000
	(0.033)	(0.029)
Dem Pres Vote Share	-0.349	-0.512
	(0.072)	(0.054)
Observations	1235	1246
W-NOMINATE Scores		
State Legislator	-0.000	0.002
	(0.013)	(0.015)
Dem Pres Vote Share	-0.911	-1.013
	(0.053)	(0.091)
Observations	283	359

This suggests that the state legislators who run are not ideologically much different than the non-state legislators who run. It should also be noted that by definition, only candidates who actually win can contribute to congressional polarization. Therefore, even if the subset of candidates in the NPAT scores analysis is unrepresentative, the W-NOMINATE analysis should reassure us that state legislators are ideologically similar to the set of non-state legislators among those who make it to Congress, the relevant individuals for an analysis of congressional polarization.

Finally, we can also use Bonica (2018)'s CFScores, which scale candidates and office-holders ideologically based on the pattern of donations they receive (see Bonica 2014 for details), to compare state legislators who run for Congress with non-state legislators who do so. These scores are available for a greater number of candidates than NPAT scores and provide a helpful sense of candidates' positions, though recent work raises concerns that they may not reflect within-party differences in ideology, as measured by roll-call votes, very accurately (Barber 2022; Hill and Huber 2017; Tausanovitch and Warshaw 2017). Nevertheless, conducting the analysis with these independently derived scores is a useful robustness check. In addition, enough candidates have CFScores to enable us to conduct our preferred analy-

sis, using pool fixed effects to compare the ideologies of state legislators and non-legislators within the same pool. The first two columns of Table 6 accordingly replicate the analyses in Table 5 using CFScores, and the second two columns show the effect of including pool fixed effects (the coefficient on *Democratic Presidential Vote* drops out in the third and fourth models because *Democratic Presidential Vote* is constant within a pool). Interestingly, the analysis using *Democratic Presidential Vote* seems to indicate that state legislators of both parties who run for Congress are more moderate than non-state legislators who do so, but controlling for pool dramatically reduces the size of these coefficients and drops them below conventional levels of significance.[24] These results both serve as confirmation of the findings in Table 5 that state legislators who run for Congress do not meaningfully differ from non-state legislators who do so and demonstrate the importance of conducting within-pool analysis: in this case, a typical regression including district-level covariates finds results that differ from those estimated using pool fixed effects.

Table 6 – Ideology of state legislators compared to non-state legislators, CFScores.

	Democrats	Republicans	Democrats	Republicans
State Legislator	0.130	-0.096	0.077	-0.057
	(0.041)	(0.021)	(0.067)	(0.038)
Dem Pres Vote Share	0.363	-0.427	–	–
	(0.141)	(0.240)		
Observations	2186	2447	2186	2447
Pool Fixed Effects	No	No	Yes	Yes

[24]The observed difference between state legislators and non-state legislators in the first two columns does not change over time, with the coefficient on an interaction between year and the dummy variable indicating a state legislator very close to zero and insignificant for both parties. Thomsen (2017) also finds, using CFScores, that the overall ideological distribution of congressional candidates who were state legislators is very similar to the overall ideological distribution of candidates who were not state legislators.

Conclusion

Polarization in the U.S. House has risen sharply in recent decades, and one key driver appears to be increasing polarization in the types of candidates who run. However, existing work cannot distinguish among several competing reasons for why the people running for the House might be polarizing. It could be that the entire set of citizens who seriously consider running for Congress has polarized; or that those who are more ideologically extreme differentially choose to run more over time; or that strategic considerations lead more-extreme candidates to run in safe districts, and the number of these districts has increased. Distinguishing among these accounts is valuable in part because they point to different potential policies for reducing polarization, but doing so requires us to measure not only the ideology of people who run for the House, but also the ideology of citizens who plausibly could have run but chose not to.

We collect new data to match state legislators to U.S. House districts and create the pools of legislators who could plausibly run in each district. Putting our analyses together, we see two important factors for explaining the rise of polarization in the U.S. House.

First, the overall polarization of state legislatures has led to more-polarized state legislators running for the House (H1). Our analyses show that Republicans who run for the House tend on average to be more extreme than their pools, whereas Democrats are representative of their pools. As the underlying pools themselves polarize, they consequently produce more extreme candidates, with the relatively constant relationship between the ideology of Republican candidates and that of their pools in particular exacerbating polarization among Republican congressional candidates. Also, although sorting occurs across pools—the ideology of the pool of state legislators in a congressional district tends to track the partisanship of the district in the expected direction—the average increase in polarization among the pools essentially tracks the increasing polarization among state legislators nationally.

Second, the increasing number of safe districts is also a factor, though less important quantitatively. Safe districts are particularly responsible for sending more-extreme state

legislators to Congress, especially in the Republican party. More extreme Republican state legislators have not become more likely to run for the House than other co-partisans in the same candidate pool, but they are enjoying increasing electoral success over time as their House districts become increasingly safe (H3). A reduction in the number of safe districts might not affect who runs, but by affecting who wins it might reduce polarization further.

One way to assess the relative contributions of these two factors is to decompose the sources of the overall increase in polarization among state legislators elected to Congress over the period. In all, our dataset contains ten election years. Taking the first three (1996-2000) and the last three (2012-2016) as representative of the beginning and end of the period we study, polarization (i.e., the difference between the average Republican state legislator elected to Congress and the average Democratic state legislator elected to Congress) increases by 0.483 points on the NP-Score scale. The increase in the average gap between Republican and Democratic pools between the same two periods is 0.323, or 66.9% of the overall increase. Thus, about two-thirds of the overall increase in polarization comes from the polarization of pools over time along with the constant gap between Republican runners and their pools, and only one-third occurs as a result of pools producing more-polarized winners on average (of which the increase coming from the growing number of safe districts comprises a part). Conversely, multiplying the average difference in the winners-pool gap between safe and competitive districts with the increase in the proportion of safe districts between 1996-2000 and 2012-2016 yields 0.013, or only 2.7% of the overall increase.

Thus, the polarization of state legislatures—and the resulting polarization of pools of eligible congressional candidates—has accounted for the vast majority of congressional polarization coming from former state legislators during the 1996–2016 period. Although our study does not focus on congressional candidates who are not state legislators, the level of polarization does not appear to differ meaningfully between these two groups, in particular among the subset who are successful in getting elected to Congress, the ultimate category of interest. To understand why the House is polarizing, and how we might reduce polarization

in the future, we therefore must understand why state legislators are polarizing. This will require careful study of the conditions and incentives that local politicians and others face in considering whether to become state legislators.

References

Aldrich, John H., and Danielle M. Thomsen. 2017. "Party, Policy, and the Ambition to Run for Higher Office." *Legislative Studies Quarterly* 42 (2): 321–343.

Barber, Michael. 2022. "Comparing Campaign Finance and Vote-Based Measures of Ideology." *The Journal of Politics* 84 (1): 613–619.

Bishop, Bill. 2008. *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart*. Boston: Houghton Mifflin Harcourt.

Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58 (2): 367–386.

Bonica, Adam. 2018. "DIME Scores for Congressional Candidates for 1980-2018 Election Cycles." https://web.stanford.edu/∼bonica/data.html.

Broockman, David E., Nicholas Carnes, Melody Crowder-Meyer, and Christopher Skovron. 2021. "Why Local Party Leaders Don't Support Nominating Centrists." *British Journal of Political Science* 51 (2): 724–749.

Carmines, Edward G. 2011. "Review Symposium: Class Politics, American-Style." *Perspectives on Politics* 9 (3): 645–647.

Fowler, Linda L., and Robert D. McClure. 1989. *Political Ambition: Who Decides to Run for Congress*. New Haven, CT: Yale University Press.

Fraga, Bernard L., Eric Gonzalez Juenke, and Paru Shah. 2020. "One Run Leads to Another: Minority Incumbents and the Emergence of Lower Ticket Minority Candidates." *The Journal of Politics* 82 (2): 771–775.

Hacker, Jacob S., and Paul Pierson. 2005. *Off Center: The Republican Revolution and the Erosion of American Democracy*. New Haven, CT: Yale University Press.

Hacker, Jacob S., and Paul Pierson. 2014. "After the "Master Theory": Downs, Schattschneider, and the Rebirth of Policy-Focused Analysis." *Perspectives on Politics* 12 (3): 643–662.

Hacker, Jacob S., and Paul Pierson. 2015. "Confronting Asymmetric Polarization." In Nathaniel Persily, ed., *Solutions to Political Polarization in America*. New York: Cambridge University Press, 59–70.

Hall, Andrew B. 2019. *Who Wants to Run?: How the Devaluing of Political Office Drives Polarization*. Chicago: University of Chicago Press.

Hill, Seth J., and Gregory A. Huber. 2017. "Representativeness and Motivations of the Contemporary Donorate: Results from Merged Survey and Administrative Records." *Political Behavior* 39 (1): 3–29.

Hirano, Shigeo, and James M. Snyder, Jr. 2019. *Primary Elections in the United States*. Cambridge, UK: Cambridge University Press.

Hopkins, David A. 2017. *Red Fighting Blue: How Geography and Electoral Rules Polarize American Politics*. Cambridge, UK: Cambridge University Press.

Lewis, Jeffrey B., Keith T. Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet. 2019. "Voteview: Congressional Roll-Call Votes Database." https://voteview.com.

Maestas, Cherie D., L. Sandy Maisel, and Walter J. Stone. 2005. "National Party Efforts to Recruit State Legislators to Run for the U.S. House." *Legislative Studies Quarterly* 30 (2): 277–300.

Maestas, Cherie D., Sarah Fulton, L. Sandy Maisel, and Walter J. Stone. 2006. "When to Risk It? Institutions, Ambitions, and the Decision to Run for the U.S. House." *American Political Science Review* 100 (2): 195–208.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2016. *Polarized America: The Dance of Ideology and Unequal Riches*. Second ed. Cambridge, MA: The MIT Press.

Moskowitz, Daniel J., Jon C. Rogowski, and James M. Snyder, Jr. 2022. "Parsing Party Polarization in Congress." Working Paper. https://www.dropbox.com/s/fydc74lrgfm9y52/npat-paper.pdf?raw=1.

Ramsdell, Molly. 2018. "How Many Former State Legislators Serve in Congress?" *The NSCL Blog*, November 2. https://www.ncsl.org/blog/2018/11/02/how-many-former-state-legislators-serve-in-congress.aspx.

Shor, Boris. 2018. "Individual State Legislator Shor-McCarty Ideology Data, May 2018 Update." https://doi.org/10.7910/DVN/6QWX7Q.

Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105 (3): 530–551.

Tausanovitch, Chris, and Christopher Warshaw. 2017. "Estimating Candidates' Political Orientation in a Polarized Congress." *Political Analysis* 25 (2): 167–187.

Theriault, Sean M. 2006. "Party Polarization in the US Congress: Member Replacement and Member Adaptation." *Party Politics* 12 (4): 483–503.

Thomsen, Danielle M. 2014. "Ideological Moderates Won't Run: How Party Fit Matters for Partisan Polarization in Congress." *The Journal of Politics* 76 (3): 786–797.

Thomsen, Danielle M. 2017. *Opting Out of Congress: Partisan Polarization and the Decline of Moderate Candidates*. Cambridge, UK: Cambridge University Press.