Curbing Unequal Representation: The Impact of Labor Unions on Legislative Responsiveness in the US Congress[*]

Michael Becher[†]
Daniel Stegmueller[‡]

July 8, 2019

Abstract

While the tension between political equality and economic inequality is as old as democracy itself, a recent wave of scholarship has highlighted its acute relevance for democracy in America today. In contrast to the view that legislative responsiveness favoring the affluent is near to inevitable when income inequality is high, we argue that organized labor can be an effective source of political equality in the US House of Representatives. Our novel dataset combines income-specific estimates of constituency preferences based on 223,000 survey respondents matched to 27 roll-call votes with a measure of district-level union strength drawn from administrative records. We find that local unions significantly dampen unequal responsiveness to high incomes: a standard deviation increase in union membership increases legislative responsiveness towards the poor by about 6 to 8 percentage points. We rule out alternative explanations using district fixed effects, interactive and flexible controls accounting for policies and institutions, as well as a novel instrumental variable for unionization based on history and geography. We also show that the impact of unions operates via campaign contributions and partisan selection. Our findings underline calls to bring back organized labor into the analysis of political representation.

I. Introduction

Democratic theory holds that people's preferences should be equally represented in collective decision making regardless of their income and wealth. But it has long been recognized that there is a potentially fateful tension between the ideal of political equality and factual economic inequality. Considering the role of campaign contributions, for instance, it is easy to grasp that advantages in economic resources can lead to advantages in political representation even when elections are free and fair (Campante 2011). Over the last two decades, political science research assessing the degree of unequal representation in the US has boomed. It has highlighted worrying disparities in political responsiveness. However, there is much less evidence on what policies or institutions can dampen political inequality.

Larry Bartels' (2008) pioneering study revealed that US senators casting roll call votes respond much more to the views of the affluent than to middle income constituents. The preferences of the poor seem to be virtually ignored. Evidence of unequal representation has also been found for the House of Representatives, party platforms, and national and state policy (e.g., see Bartels 2016; Bhatti and Erikson 2011; Flavin 2012; Gilens 2012; Hertel-Fernandez et al. 2019; Rhodes and Schaffner 2017; Rigby and Wright 2013). Beyond the American context, recent scholarship documents similar patterns across a range of political systems in advanced industrialized democracies (Bartels 2017; Elsässer et al. 2018). Thus, it may appear that unequal democracy is a constant feature of capitalism. In contrast, we argue that organized labor can be an effective source of political equality in the US even in times of high economic inequality.

We contribute to this research agenda by analyzing the causal role labor unions play in enhancing legislative responsiveness to the poor. A large literature in comparative politics, economics, and sociology examines the effect of unions on economic inequality and redistributive politics (Ahlquist 2017). However, as noted by Kathleen Thelen, the study of labor unions typically "does not come up in the mainstream literature on American Politics" (Thelen 2019: 12).[1] Unions are one of the few membership organizations in national politics that advocate on the behalf of non-managerial workers (Schlozman et al. 2012) and sometimes even strike in the interest of others (Ahlquist and Levy 2013).

---

[1]There is of course a large literature on the role of unions in political life in general. It has documented that union membership is associated with lower income differentials in political participation (Leighley and Nagler 2007) or political knowledge (Macdonald 2019). Recently, the politics of teachers' unions has received increasing scholarly attention (Moe 2011). While providing important insight into the strategic actions of unions and their impact on individual political behavior, this literature does not address their impact on citizens' unequal representation in the actual decision making of lawmakers.

Yet few scholars have directly addressed the question whether stronger unions translate into less income-biased responsiveness by elected representatives. Studying the 110th House of Representatives, Ellis (2013) finds that district-level unionization is related to a smaller rich-poor gap for key legislative votes. Flavin (2018) conducts a cross-sectional state-level analysis and shows that American states with stronger unions exhibit less unequal representation. Gilens and Page (2014) conclude that mass-based interest groups, including unions, have little to no independent influence on national policy in the US.

However, the existing evidence is of correlational nature. Prior scholarship does not disentangle whether union strength merely correlates with or actually causes higher political equality. One major endogeneity threat stems from the potential of policy feedback. As John Ahlquist points out in a recent review, unions may influence "parties and policy, but policy and institutions also affect unionization rates" (Ahlquist 2017: 427). Hence, weaker unions may be more of a symptom rather than a cause of biased representation. While many workers express support for unionized jobs (Freeman and Rogers 2006), state and local politicians set varying rules that affect costs and benefits of union membership. Recent research highlights the political logic behind the expansion of public sector union laws in the 1960s and 1970s (Anzia and Moe 2016; Flavin and Hartney 2015) as well as the countervailing success of conservative groups to demobilize public sector unions since the early 2000s (Hertel-Fernandez 2018). In the private sector, 'right-to-work' laws hamper unionization efforts and can have profound political effects (Feigenbaum et al. 2018). Furthermore, any relationship between union strength and equal legislative responsiveness may be spuriously driven by the same underlying determinants. Following seminal research on the importance of social capital for making democracy work (Putnam 1993, 2000), workers in congressional districts with more social capital may be better represented in both their workplace and in Congress without a causal arrow running from the former to the latter.

In addition to concerns about causality, the literature faces the challenge of how to measure public preferences. When standard election surveys are sliced by income groups in subnational units (e.g., states or congressional districts) sampling noise becomes relevant and may produce "wobbly estimates" that hinder a decisive verdict (Bhatti and Erikson 2011: 232). Random measurement error generally leads to underestimation of the effect of public opinion on policy. While this concern can be mitigated by using the large Cooperative Congressional Election Study (CCES), it (and other large surveys) are not designed to be representative at the level of

congressional districts. One may thus worry that the seeming legislative underrepresentation of the poor may in part be an artifact of systematic sampling error.[2]

In this paper we address these problems using fine-grained data on local unions and corrected estimates of policy preferences. We account for local unobservables via district fixed effects and we conduct analyses where we introduce a new instrumental variable for union density. We examine the impact of unions on the preference-roll-call link in the contemporary US Congress, where unequal responsiveness by legislators and their staff has been well documented (Bartels 2008, 2016; Ellis 2013; Hertel-Fernandez et al. 2019; Rhodes and Schaffner 2017). We focus on members of the House of Representatives during the 109–112th Congress (2005-2012) since this setting enables us to capture within-state variation in union strength and within-district variation in preference polarization by income across a range of salient policy issues.

At its core, our dataset combines income-specific measures of constituency preferences based on 223,000 survey respondents matched to 27 roll-call votes with information on local unions extracted from more than 350,000 administrative records. We measure district-level preferences using multiple waves of the CCES and calculate preferences on 27 concrete policies for each income group in each congressional district. We employ small area estimation using micro-census data (this circumvents sampling issues with the CCES; but note that our findings are robust when using multilevel regression and poststratification [MRP] instead). Drawing on recent work by Becher et al. (2018), we measure the district-level strength of unions using mandatory reports filed by local unions to the Department of Labor.[3]

Our empirical analysis traces the legislative responsiveness of House members to the preferences of different income groups in their constituency conditional on district-level union strength. In line with previous research, we find that House members are significantly less responsive to the policy preferences of low-income than upper income constituents. However, this gap in responsiveness is smaller where unions are stronger, and it decreases significantly where union members are comparatively numerous. Our estimates suggest that a standard deviation increase in unionization increases responsiveness towards the poor by about 6 to 8 percentage points. The impact of unions is large enough to swing key votes.

---

[2]Other empirical issues have been raised with respect to research on policy adoption (Gilens 2012; Gilens and Page 2014). For instance, see Bashir (2015); Enns (2015); Erikson (2015); Soroka and Wlezien (2008).

[3]The study of Becher et al. (2018) examines the effect of union density and union concentration on roll call voting. It does not measure constituency preferences and thus cannot examine the extent of income-biased responsiveness. Furthermore, it lacks the exogenous source of variation in union density that we introduce in this paper.

To rule out alternative explanations, we start by controlling for district fixed effects. Then we allow for additional interactions between income-group preferences and alternative moderators, such as state-level policies, state fixed effects, district socio-demographics and proxies for social capital. For more robust causal identification, we leverage plausibly exogenous variation in union strength that stems from geography and the history of union mobilization in the middle of the twentieth century, rather than policy feedback, tastes, or institutions. At the time, virtually all coal and metal mines were unionized throughout the country. The location of these industries is mainly determined by nature, and initial unionization lead to subsequent spillovers in union membership in other industries (Holmes 2006). This suggest we can use local employment in mining in the 1950s as a valid instrumental variables for union membership more than 5 decades later. To further probe robustness of the findings with respect omitted variable bias and functional form assumptions, we employ a general post-double-selection estimator, which combines machine learning variable selection with causal inference methods. These additional tests confirm the initial regression results. Hence, we conclude that unions curb unequal representation in the legislative arena.

These findings have important implication for the debate about inequality and democracy. Scholars often find it straightforward to explain why the poor are poorly represented. By definition, those in lower part of the income distribution have less resources, and many studies have shown that they are less likely to participate in politics. In contrast, our results show that this outcome is not hardwired into the fabric of American democracy. In the congressional arena, local unions are a countervailing force. Our results imply that political efforts to weaken unions, as illustrated in recent reforms in Michigan or Wisconsin, are likely to exacerbate unequal responsiveness in representation. They magnitude of the union effect may also explain why unions remain under attack by conservative groups (Hertel-Fernandez 2018).

Our focus is complementary to research on contributions, political participation, or partisanship as important determinants of unequal legislative responsiveness (Barber 2016; Leighley and Oser 2018; Rhodes and Schaffner 2017). From our institutional perspective, these factors are mechanisms through which unions affect representation, and we provide some evidence in line with this.[4]

---

[4]Also complementary, a large literature examines how formal political institutions affect how the poor are represented across countries (e.g., Iversen and Soskice 2006; Long Jusko 2017).

Theoretical Motivation

Various strands of scholarship in political science and related fields suggest that labor unions are one of the few mass-membership organization that provide collective voice to lower income individuals (Ahlquist 2017; Ellis 2013; Flavin 2018; Freeman and Medoff 1984; Schlozman et al. 2012).[5] Consistent with a central premise of the collective voice perspective, contemporary unions in the US tend to take positions favored by less affluent citizens.[6] However, shared preferences between the less well-off and organized labor are by no means sufficient to alter inequalities in political representation. This requires an effective political transmission mechanism. Several prominent scholars of representation are skeptical. They suggest that unions have become too weak, too narrow, or too fragmented to have a significant egalitarian political impact in national politics (e.g., Gilens 2012: 175; Hacker and Pierson 2010: 143). As discussed above, the endogeneity of union organization to politics means that scholars cannot take a correlation between union density and political equality as sufficient evidence of causality, but should investigate it using a more credible research design (Ahlquist 2017).

The ability of unions to increase the rate of political participation—including voting, contacting officials, attending rallies, or making donations—of low- and middle-income citizens is often considered to be their key channel of political influence. Importantly, unions may also increase participation among non-members with similar policy preferences through get-out-the-vote campaigns and social networks (Leighley and Nagler 2007; Rosenfeld 2014; Schlozman et al. 2012).[7] Making contributions to favored candidates and campaigns complements the ability of unions to communicate with and mobilize members or to provide campaign volunteers. Indeed, unions are among the leading contributors to political action committees (PAC), accounting for a quarter of total PAC spending in 2009 (Schlozman et al. 2012: ch. 14). In contrast to

---

[5]Labor unions are organizations formed to bargain collectively, on behalf of their members, with employers over wages and conditions. Once formed, unions may (and often do) enter the political arena (Freeman and Medoff 1984; Olson 1965).

[6]Gilens (2012: 154-161) compares public positions of national unions with mass policy preferences across several hundred policy issues and finds that unions' positions are most strongly correlated with the preferences of the less well-off. Similarly, Schlozman et al. (2012: 87) conclude that unions are one of the few organizations in national politics "that advocate on behalf of the economic interest of workers who are not professionals or managers." See also Hacker and Pierson (2010); Schlozman (2015); Schlozman et al. (2018).

[7]Unions may also increase the salience of policy preferences, provide information to reduce uncertainty about which candidate/party is closer to members' preferred position, or shape the content of policy preferences through leadership or social interactions (Ahlquist and Levy 2013; Iversen and Soskice 2015; Kim and Margalit 2017; Mosimann and Pontusson 2017). While most of the literature on unequal representation takes the content of preferences as given (this study included), union effects on salience and, to some extent, political information are consistent with a mobilization mechanism and evidence presented later on. We agree that a full understanding of representation requires endogenizing citizens' preferences.

corporates and business organizations, union contributions "represent the aggregation of a large number of small individual donations" (Schlozman et al. 2012: 428).[8]

The credible threat of political mobilization can affect policy decisions by representatives in two general ways. First, it may shape who is elected in a given electoral district. If politicians are not exchangeable (because they differ in their commitments and beliefs), political selection is important. In an age of elite polarization, the partisan identity of a representative is often crucial for determining legislative voting (Bartels 2016; Rhodes and Schaffner 2017; Lee et al. 2004). Since the New Deal era, unions and union members have largely allied with the Democratic Party, given its stronger support for many of their broader policy demands (Lichtenstein 2013; Schlozman 2015).[9]

Second, unions' mobilization potential shapes the incentives of elected representatives, beyond their partisan affiliation and personal traits. Policymakers' rational anticipation of public reactions plays a central role in theories of accountability and dynamic responsiveness (Arnold 1990; Stimson et al. 1995). While many individual legislative votes do not affect the reelection prospects of representatives, on potentially salient votes they can face hard choices between party ideology and competing constituency preferences. On international trade agreements, for instance, Democratic representatives have faced cross-pressures between a more skeptical stance taken by unions and low-income constituents versus that of their own party (Box-Steffensmeier et al. 1997). On the other side of the aisle, Republican legislators, in the wake of the financial crisis, found themselves torn between their own partisan views on stimulus spending and the pressure from less well-off constituents (Mian et al. 2010).

Politicians' incentives are also linked to information. Theories of representation emphasize that members of Congress, and especially the House, face numerous voting decisions in each term, and it would be unrealistic to assume that they have access to reliable, unbiased polling data on constituency preferences on all the issues they face (Arnold 1990; Miller and Stokes 1963). Instead, representatives—with the help of their staffers—rely on alternative methods to assess public opinion, including constituent correspondence, town halls, contacts with community leaders, or local organizations. In this limited information context, local unions can enhance the visibility and perception of constituent preferences (Hertel-Fernandez et al. 2019).

Taken together, this reasoning implies that the district-level strength of labor unions increases the responsiveness by members of Congress to the less affluent. While we know

---

[8]While evidence on the direct effect of contributions on legislative behavior is mixed, recent field-experimental results demonstrate that contributions help to provide access (Kalla and Broockman 2016) or sway congressional staffers (Hertel-Fernandez et al. 2019).

[9]Political selection might also shape other political characteristics of representatives, such as their class background or race (Butler 2014; Carnes 2013).

from previous work that politicians are considerably more responsive to the preferences of the affluent than those of the less well-off, this bias should be reduced in districts with relatively higher union membership. Substantively, it is crucial to assess how far the presence of unions can move responsiveness toward the ideal of political equality.

The strength of local unions underpins a credible mobilization *threat* that impacts the action of candidates and legislators. Anticipating mobilizing efforts by unions, a potential candidate may not even enter into the race; an elected, career-oriented, politician might be pressured to alter his or her vote even without a full mobilization effort as long as unions' mobilization capacity is visible. Thus, both campaign contributions and candidate selection should matter as a channel linking local union strength and representation.

Data and Measurement

Any effort to test the relevance of unions for unequal representation confronts major challenges of measurement and causal interpretation. Our empirical approach allows us to address these issues to an extent previously impossible. For the House of Representatives in the 109th to the 112th Congress, we created a panel of legislators' roll call votes matched to income-specific policy preferences at the district level, and district-level measures of union membership, based on digitized administrative records from the Department of Labor.[10] After matching information on roll call items for 223,000 CCES respondents to actual roll call votes, we estimate policy preferences for low and high income constituents in each district for 27 roll calls, adjusting for the fact that that the CCES is not a representative sample of district populations.

Measuring district preferences by income group

The CCES is an ideal starting point for our analysis, since it is a nationally representative study, includes a considerable number of roll call questions, and provides us with a large enough sample size to decompose income-group preferences by district. It largely addresses the problem that the number of observation in each subgroup defined by income and congressional district is too low (Bhatti and Erikson 2011). The roll calls included in the CCES concern key votes as identified by *Congressional Quarterly* and the *Washington Post* and cover a broad range

---

[10]Our analysis focuses on one apportionment period, which generally holds district boundaries constant (we show that the results are robust to cases of mid-period redistricting).

of issues (Ansolabehere and Jones 2010). Respondents are presented with the key wording of the bill (as used on the floor and in media reports) and are then asked to cast their own vote: "What about you? If you were faced with this decision would you vote for, against, or not sure?" Contrary to widely usual agree–disagree survey measures of issue preferences, matched roll call votes provide us with unequivocal evidence of policy congruence between respondent and legislator (Jessee 2009, Ansolabehere and Jones 2010: 585). We match 27 roll call items in the CCES to roll call votes cast in the House of the 109th to 112th Congress. These cover important legislative decisions, such as Dodd-Frank, the Affordable Care Act (and attempts to repeal it), the minimum wage increase, the ratification of the Central America Free Trade Agreement, or the Lilly Ledbetter Fair Pay Act. Table A.2 in the Appendix lists all matched CCES items and House bills included in our estimation sample.

The CCES provides a comparatively large sample size per district. However, it is not designed to be representative for congressional district populations. Thus, individuals with certain characteristics, such as particular combinations of income, race and education, may be underrepresented in the CCES sample of a given district. If this is the case, unadjusted policy preferences from the CCES will not reflect the target population and using them can lead to biased estimates of unequal representation in Congress.

Our solution to this problem is a small area estimation procedure that rebalances the survey sample to represent the district population using fine-grained Census data and flexible machine learning tools. The basic idea is to combine the survey data on voter preferences from the CCES, which may not be representatives of districts, with accurate data on the distribution of population characteristics in a given district from the Census Bureau's American Community Survey (ACS). The machine learning solution we propose is relatively new to the representation literature in political science. It has attractive properties that merit its application to this topic. However, our findings do not depend on this particular approach. We obtain comparable results when using the MRP approach widely used by political scientists (Lax and Phillips 2009).[11]

Concretely, we use about 3 million individual-level records from a synthetic sample of the ACS from 2006 to 2011, which provides an accurate representation of the district population. We stack both datasets, creating a structure where we have common district identifiers and individual covariates while responses to policy preference questions are missing in the Census portion of the data. As common covariates bridging CCES and Census we use the

---

[11]Online Appendix B.1 provides a more detailed description of our procedure and Table B.1 reports model results based on two alternative approaches of measuring preferences: MRP and using raw CCES means. The results are qualitatively the same as those based on the one presented in the main text. Compared to MRP, our approach yields somewhat more conservative estimates of the union effect on representation. Not surprisingly, given concerns about measurement error, effects based on raw means are somewhat smaller.

following demographic characteristics: gender, race (3 categories), education (5 categories), age (continuous) and family income (continuous). The latter is of particular relevance as we are interested in producing district–income group specific preferences. In the next step we fill missing preferences in the Census with matching data from CCES respondents. Since this is essentially a prediction problem, we can use powerful tools developed in the machine learning literature to achieve this task. We use an algorithm proposed by Stekhoven and Bühlmann (2011) to impute missing cells.[12] Compared to commonly used multivariate normal or regression imputation techniques, this strategy has the advantage that it is fully nonparametric, allowing for complex interactions between covariates, and deals with both continuous and categorical data. Our completed data-set now contains preferences for 27 roll call items of synthetic 'Census individuals', which are a representative sample of each House district.

With these data in hand, we assign individuals to income groups and calculate group-specific preferences for each roll call in each district. Following previous work in the representation literature (Bartels 2008, 2016), we delineate low- and high-income respondents using the 33th and 67th percentile of the distribution of family incomes. Following theories of constituency representation in Congress and methodological recommendations of Bhatti and Erikson (2011), we specify these income thresholds separately by congressional district. This accounts for the substantial differences in both average income and income inequality between US districts. It also ensures that within each district, income groups are of equal size. Online Appendix Table A.1 shows the distribution of income-group cutoffs. On average, our chosen cutoffs are close to those used in previous studies. The mean of our district-specific low-income cutoffs is around $39,000, while Bartels uses $40,000 (Bartels 2016: 240); our mean high-income cutoff is around $81,000, where Bartels employs a threshold of $80,000. However, beyond these averages lies considerable variation. In some districts, the 33rd percentile cutoff is as low as $16,500, while the 67th percentile reaches almost $160,000 in others.[13]

For each roll call, we then estimate district-level preferences of low- and high-income constituents, which we denote by (θ^l, θ^h), as the proportion of individuals voting 'yea'. Since preference estimates are in $[0, 1]$ they can be directly related to legislators' probability of voting 'yea' on a given roll call. Our data shows considerable variation in the distance of the policy preferences of those at the top and those at the bottom as illustrated in Figure I. It plots histograms of the difference between low-income and high-income preferences $(\theta^h - \theta^l)$ in congressional districts for six selected roll calls. For salient bills, such as increasing the

---

[12]Honaker and Plutzer (2016) use a similar approach (but rely on multivariate normal imputations) and further discuss its empirical performance in estimating small area attitudes and preferences.

[13]Results are relatively invariant to using alternative income thresholds (see Table C.1).
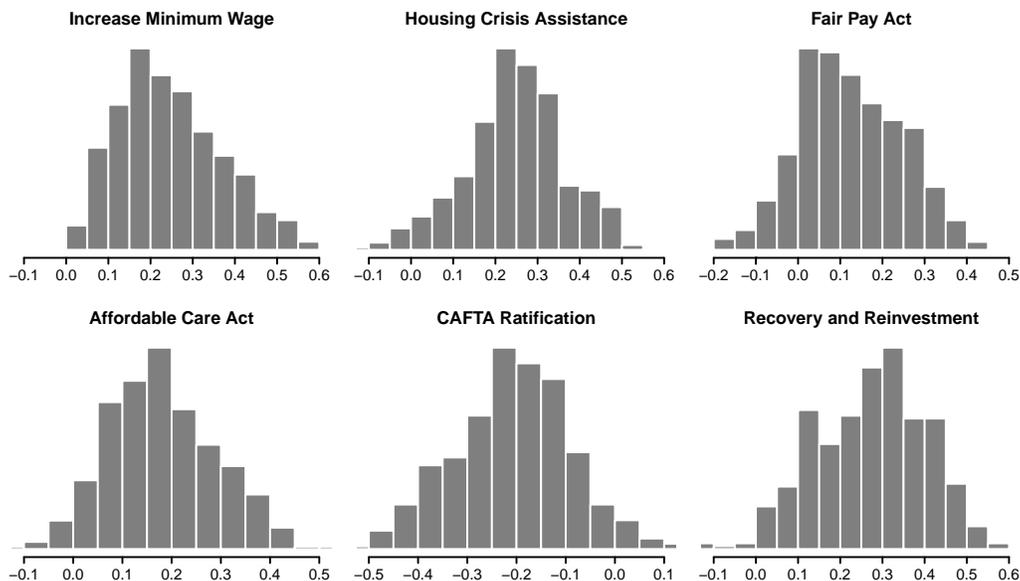
FIGURE I

District-level income gap in public support for 6 selected policies

Note: Each histogram plots the difference in support for a matched roll-call vote question between people in lower third and people in upper third of their district's income distribution for all House districts.

minimum wage (the Fair Minimum Wage Act), housing crisis assistance (the Housing and Economic Recovery Act), or Affordable Care Act, the vast majority of low-income constituents are more supportive than their high-income counterparts in each and every district. On other issues, such as the ratification of the Central America Free Trade Agreement, high income constituents are clearly in favor. In all examples, we find considerable across-district variation in the preference gap between low- and high-income constituents.[14] We leverage this variation over both roll calls and districts to estimate legislators' differential responsiveness to changes in policy preferences of different income groups, and how it might be moderated by union strength.

District-level union membership

To measure district-level union membership we draw on fine-grained administrative data compiled by Becher et al. (2018). Based on the Labor-Management Reporting and Disclosure Act

---

[14]Averaged over all districts and roll calls, there is a statistically significant gap between the preferences of the bottom third and the top. The mean of the (absolute) preference difference is 17 percentage points; the 10th percentile is 3 points while the 90th percentile is 32 percentage points.

(LMRDA) of 1959, unions have to file mandatory yearly reports (called LM forms) with Office of Labor-Management Standards (OLMS). The Civil Service Reform Act of 1978 introduced a similarly comprehensive system of reporting for federal employees (see Budd 2018). A mandatory part of each report is the number of members a union has. Failure to report, or reporting falsified information, is made a criminal offense under the LMRDA, and reports filed by unions are audited by the OLMS. The resulting database contains almost 30,000 local unions. It is based on 358,051 digitized individual reports that were cleaned, validated, geocoded, and matched to congressional districts. The number of union members in each congressional district can then be readily obtained as the sum of all reported union members. Figure II shows the distribution of union membership in House districts. It demonstrates that there is substantial variation in unionization between electoral districts even within states, which would be ignored by a state-level analysis. Note that while individual union locales exhibit significant variation in membership over time, aggregate district-level membership moves little during the period we study, in line with the national trend. Hence our analysis focuses on the between-district variation in unionization.
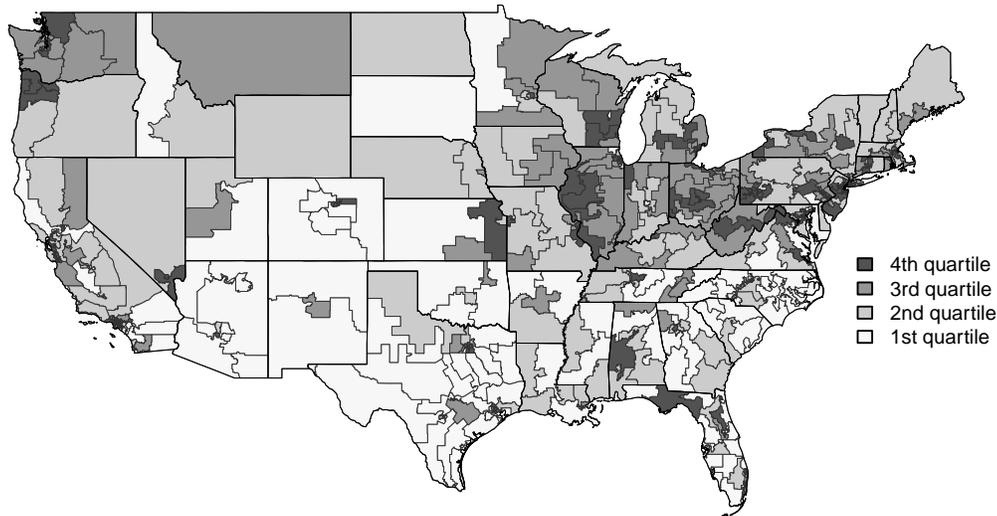
FIGURE II
Union membership in House districts, 109th-112th Congress.

Using LM forms provides important advantages over using measures derived from surveys. First, mandatory administrative filings are likely more reliable than population surveys, which often suffer from over-reporting and unit-nonresponse (Southworth and Stepan-Norris 2009: 311). Second, they allow us to estimate union membership numbers for smaller geographical units, which are usually unavailable in population surveys or only covered with insufficient

11

sample sizes.[15] Another advantage for the study of politics is that the presence of union locales is observable to politicians on the ground even in the absence of survey data.

A potential drawback of using LM forms is that some public unions (those that *exclusively* represent state, county, or municipal government employees) are exempt from filing. However, any union that covers at least one private sector employee is required to file. In practice, this leads to almost complete coverage, because unions are now increasingly organizing workers across different sectors and occupations (Lichtenstein 2013: 249).[16]

Basic Empirical Models

The fact that we observe several roll calls within a given congressional district allows us to specify a statistical model with district fixed effects, which capture unobservable characteristics of districts (and states) that are constant over roll-calls. To provide for a stricter test of the moderating effect of unions, we also allow a rich set of other district characteristics to moderate the link between income groups and legislators' voting behavior. This amounts to estimating models including interactions between observed district characteristics and group preferences. Later, we introduce an instrumental variable for union membership and a more general statistical model that makes less restrictive assumptions.

For each roll call vote j ($j = 1, \ldots, J$), we have measured preferences of low and high income citizens in a given congressional district d ($d = 1, \ldots, D$) denoted by ($\theta_{jd}^l, \theta_{jd}^h$). The level of (logged) union membership in a district is denoted by U_d. We specify relevant confounders in X_d. Depending on the particular specification (discussed in the next section) these will include (i) socio-economic district characteristics, (ii) measures of historical state union policies and state fixed effects, (iii) proxies for district-level social capital, (iv) as well as non-linear transformations of these. For ease of interpretation, we have scaled all inputs to have mean zero and unit standard deviation. Our model for the voting behavior of House members is the

---

[15]The most prominent data set on union membership, compiled by Hirsch et al. (2001), provides CPS-based estimates for states and metropolitan statistical areas; district identifiers are not available.

[16]National aggregates based on LM forms are in close agreement with measures from the CPS (Hirsch et al. 2001): the former estimates 13.21 million union members (excluding Washington, D.C.) while the latter yields 15.22 million. This difference is consistent with some degree of over-reporting in the (survey-based) CPS (Southworth and Stepan-Norris 2009: 311). It can also be interpreted as an upper bound for the non-coverage of some public sector unions. A more detailed analysis (Becher et al. 2018) finds that state-level aggregates from LM forms and the CPS are strongly correlated (r = 0.86).

following linear probability specification:

$$y_{ijd} = \mu^l \theta_{jd}^l + \mu^h \theta_{jd}^h + \eta^l (U_d \times \theta_{jd}^l) + \eta^h (U_d \times \theta_{jd}^h) +$$
$$\beta^l (X_d \times \theta_{jd}^l) + \beta^h (X_d \times \theta_{jd}^h) + \alpha_d + \epsilon_{ijd}$$

The key terms here are the interactions between union membership and the respective preferences of the affluent and the poor, $U_d \theta_{jd}^h$ and $U_d \theta_{jd}^l$. Thus, when η^l and η^h are zero the group-specific preference coefficients μ^l and μ^h indicate the change in the probability of legislators casting a supportive vote induced by a standard deviation change in the respective preferences of the poor and the affluent. The coefficient η^l indicates the conditional effect of a standard deviation change in logged union membership on the responsiveness of legislators' votes to the preferences of the poor. The corresponding effect for the affluent is given by η^h. Our theoretical expectation is that $\eta^l > 0$ and $\eta^h \leq 0$.

In order to mitigate the influence of unobserved confounders affecting legislators' voting behavior, we account for time-constant district-level unobservables by including district fixed effects, α_d.[17] Despite this, one may be worried that changes in responsiveness attributed to unions are spurious. As a conceptually straightforward approach to rule out further alternative explanations, we include the interactions between controls (both on the district- and state-level) and group preferences $X_d \theta_{jd}^l$ and $X_d \theta_{jd}^h$. They use within-district variation over roll-calls and preferences to estimate the conditional marginal effect of group preferences, making it less likely that our estimated effect of union membership is simply due to omitted confounders. Finally, ϵ_{ijd} are white-noise errors assumed independent of covariates. We account for heteroscedasticity and arbitrary within-district correlations when calculating standard errors.

Results

Before presenting evidence on the moderating effect of unions, we want to give a sense of the overall picture of legislators' responsiveness emerging from our data. Estimating a model as described above with district fixed effects but without accounting for local union organization (setting β^l, β^h and η^l, η^h to zero) or any other moderators, we find a clear gap in the responsiveness of legislators to the preferences of low- versus high-income individuals. A standard deviation (SD) increase in the preferences of the affluent is linked to an increase in

---

[17]Note that non-interacted effects of district-level union membership and covariates (which vary between districts, but are constant over roll calls) are absorbed in α_d.

the probability of legislators to cast a corresponding vote of 13.6 (±1.2) percentage points. In contrast, a SD increase in the preferences of the less well-off induces a much smaller change in legislators' behavior of 1.6 (±1.4) points. With a confidence interval ranging from −1.1 to 4.4, we cannot reject the null hypothesis that legislators do not respond to the preferences of low-income constituents in the average electoral district. The responsiveness gap between the two groups is sizable (at 11.9 (±2.5) percentage points) and significantly different from zero. This is consistent with with previous work on unequal representation in Congress (Bartels 2008, 2016; Bhatti and Erikson 2011; Ellis 2013). Next, we show that the extent of legislators' non-responsiveness depends crucially on the strength of local unions.

Turning to the effect of unions on legislative responsiveness, we start by summarizing our key finding graphically and then discuss more extensive model specifications. Figure III plots marginal effects of low- and high-income constituency preferences on representatives' roll-call votes at varying levels of union membership with 95% confidence intervals.[18] It shows that legislators' responsiveness to the policy preferences and low-income and high-income constituents depends on district-level union membership: as unionization increases, legislators' responsiveness to low-income constituents increases, while their responsiveness to high-income constituents declines by a similar amount. For example, moving from a district with median levels of union membership to one at the 75th percentile increases the responsiveness of legislators to low-income preferences by 8 percentage points, while it decreases responsiveness to high-income preferences by about 5 points. Given the initial responsiveness gap, this change is substantial enough to substantially level the playing field between affluent and poor.

Are these findings robust to confounding factors? Table I presents parameter estimates from a number of increasingly rich specifications designed to capture potential confounds. In specification (1), we begin with a baseline model (also plotted in Figure III) that includes district fixed effects but no further preferences-confounder interactions (setting β^l and β^h to zero). We find that a SD increase in district union membership increases legislators' responsiveness to the poor by about 11 (±1) percentage points, while at the same time decreasing the advantage in responsiveness enjoyed by the affluent by about 6 (±1) points.

Even after accounting for district fixed effects, alternative explanations for this result may be based on omitted variables that interact with group preferences. Reflecting policy feedback (Ahlquist 2017), the moderating effect we have ascribed to unions mostly reflects the fact that state governments have chosen policies that strengthen or weaken the ability of unions to organize. While collective action problems dilute incentives of politicians to make

---

[18]Calculated from a LPM of vote choice on preferences and union membership. It includes district fixed effects with district-clustered standard errors. See also specification (1) in Table I below.
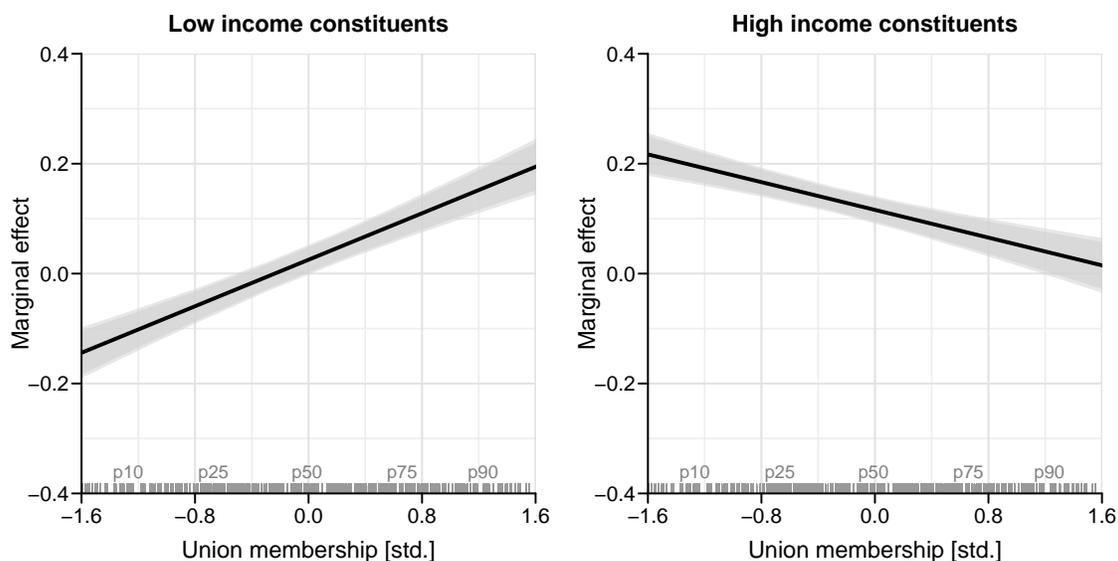
FIGURE III

District-level union membership as moderator of unequal representation.

Note: This figure plots changes in marginal effects of low- and high-income constituency preferences on representatives' roll-call votes conditional on district-level union membership. Shaded areas are 95% confidence intervals based on district-clustered standard errors. The sample distribution of (*z*-standardized) union membership is indicated above the x-axis.

politics using policies, in some circumstances they are overcome. Most relevant for unions are right-to-work and public sector bargaining laws (Anzia and Moe 2016; Feigenbaum et al. 2018; Flavin and Hartney 2015; Hacker and Pierson 2010; Hertel-Fernandez 2018). In specification (2), we therefore add two measures of historical state union policy: the share of years with right-to-work legislation and the share of years with mandatory collective bargaining laws for teachers since 1955, taken from Flavin and Hartney (2015). These enter X_d and are interacted with income group preferences θ^l and θ^h. In specification (3) we go one step further and allow for *any* state-level characteristic (such as institutions or historically-rooted popular anti-union sentiments) to moderate the marginal effect of income group preferences on legislators vote choice by including state-specific constants in X_d which are interacted with group preferences. The results from both extended specifications show that accounting for state-level policies and institutions as potential moderators does not change our core picture of the role of local union organization: where local unions are stronger the responsiveness gap between the affluent and the poor is reduced.

A more subtle problem concerns a form of simultaneity bias at the district level. There may be district-level factors shaping both the propensity to be a union member and to be politically

TABLE I

Union strength and representation. Effect of union membership on legislators'
responsiveness to income groups.

	(1)	(2)	(3)	(4)	(5)	(6)
Low income preferences	0.106	0.082	0.098	0.108	0.068	0.045
	(0.013)	(0.015)	(0.016)	(0.013)	(0.014)	(0.014)
High income preferences	−0.063	−0.036	−0.053	−0.066	−0.050	−0.028
	(0.012)	(0.013)	(0.014)	(0.012)	(0.013)	(0.013)
District fixed effects	✓	✓	✓	✓	✓	✓
Group preferences						
× union policy		✓				✓
× state constants			✓			
× district social capital				✓		✓
× district covariates					✓	✓

Note: N=15,780. N_d = 435. 27 roll call votes, 109th to 112th Congress. Entries are parameter estimates for η^l and η^h from a linear probability model with district-clustered standard errors in parentheses. All models include district fixed effects. Specifications (2) to (5) include coefficients for interaction (β^l, β^h) of income group preferences with state- or district-level confounders. Specification (2) includes two measures of historical state union policymaking, the share of years with right-to-work legislation and collective bargaining agreements. (3) interacts preferences with state fixed effects. (4) includes a social capital index (Rupasingha and Goetz 2008). (5) includes a large set of district-level characteristics (population size, degree of urbanization, shares of female, Black, Hispanic, BA degrees, employed in manufacturing, as well as median household income). Specification (6) includes all of the previously described measured variables.

active. If less affluent individuals with a higher capacity to organize and solve collective action problems cluster in specific districts, our estimates of the marginal impact of district union membership on responsiveness will be overly optimistic. Such a propensity may reflect critical historical junctures in labor organizations (Ahlquist and Levy 2013) or social capital (Putnam 1993, 2000).

As a first cut to address this problem using control variables, we add proxies tapping into social capital. The social capital index developed by Rupasingha and Goetz (2008) captures broader patterns of organizational and community life. It combines information on membership in voluntary associations, voter turnout, the Census response rate, and the number of non-profit organizations. The index is available at the county level and was spatially reweighted to congressional districts. The components of the index are measured contemporaneously with unionization and at least some of them (i.e., turnout) are potentially a consequence of union strength. Hence we interpret specification (4), which interacts voter preferences and district-level social capital, as a conservative test for the impact of unions. It makes clear that net of social capital, which may itself reflect prior union strength, district unionization shapes responsiveness: a SD increase in union membership still increases legislators' responsiveness

to the preferences of the poor by 6 (±1) percentage points, and lowers their responsiveness to the preferences of the affluent. Moreover, using alternative measures of social capital, the number bowling alleys and the number of congregations per 1000 inhabitants, yields the same result (see Appendix E).

In specification (5) we measure a large number of districts' socio-economic characteristics and allow them to interact with constituency preferences: population size, race (share of African Americans and Hispanics), education (share with BA or higher), the share of the working population employed in manufacturing, median household income, and the degree of urbanization (for descriptive statistics, see Table A.3). This set of covariates excludes obviously "bad controls" (Samii 2016) such as partisanship or campaign contributions that are a mechanism through which unions influence representation.[19] Again, our results point towards the existence of a clear moderating effect of unions. Our final specification, column (6) of Table I, includes *all* previous covariates and, again, confirms our core finding.[20]

Assessing Threats to Inference

So far our analysis has relied on district fixed effects and covariates interacted with preferences to rule out alternative explanations for the pro-poor, egalitarian impact of unions. Now we move to identification of the union effect based on an instrumental variable (IV) approach and a general selection model. They enable us to rule out alternative explanations based on unobservable omitted variables and relax functional form assumptions.

We instrument contemporary union density by exploiting the history and geography of the post-war labor movement in the United States. As proposed by Becher and Stegmueller (2019), the basic idea is to leverage the history of spillovers from unionization in a few capital-intensive industries, whose location was mainly determined by geographical factors, such as natural endowments, as opposed to local tastes or institutions. In the early 1950s coal and metal mines as well as steel plants were fully unionized across the country, irrespective of local political leanings (Holmes 2006: 2). In the environment of the time, extractive industries were easy targets for unionization due to their capital intensity, high sunk costs, and difficult working

---

[19]Unionization may also lead to sorting based on sociodemographic characteristics if people vote with their feet. Hence, our basic model excludes these factors, and the next section specifies an instrumental variable model.

[20]Additional robustness checks reported in Appendix E confirm these results (e.g., excluding states with redistricting or districts with extreme preferences, using alternative measures of social capital, accounting for errors-in-variables). Becher et al. (2018) find that union concentration can affect legislative voting. While we have not source of exogenous variation for their concentration variable, our results obtain when controlling for it (see Table E.1).

conditions. Thus, they were unionized even in the otherwise much more anti-union South. Importantly, the location of mining and even steel production is largely determined by nature, whether due to the availability of coal in the ground or the access to raw materials for steel mills. This means that the post-war level unionization in these industries is largely orthogonal to local factors.

Over time, positive spillovers from unionized mining and steel industries induced high levels of union membership across industries. Unions used their resources to organize new establishments, and workers switching industries brought with them tastes and skills for unionization. Using detailed local data, Holmes (2006) documents a positive relationship between the unionization of establishments (in health care and wholesale) in the 1990s and their proximity to mining and metal industries in the 1950s.

The combination of geographic location determined by nature and spillovers implies that part of the variation in union membership in congressional districts at the beginning of the twenty-first century is exogenous to local factors conditional on the intensity of mining and steel employment in the middle of the twentieth century. We need to assume that historical unionization in these capital-intensive sectors only affects contemporary political representation through contemporary unionization. We think this is a plausible approximation. Today, for instance, neither mining nor steel are uniformly unionized.[21]

We thus construct our instrument for current levels of union membership from data on district-level employment shares in mining and steel industries in the 1950s. We construct the latter by calculating them from 1950 Census samples spatially interpolated to current congressional districts.[22] Figure IV conveys that historical mining employment is strongly related to contemporary levels of union membership. It plots logged district-level employment in mining and steel industries (as share of working population) in the 1950 Census against logged contemporary district-level union membership in the 109th Congress. A simple linear model with state fixed effects suggests that a one percent increase in the instrument induces a 0.34 (±0.04) percent increase in the district-level share of union members. State fixed effects

---

[21]One possible violation of the exclusion restriction is that initial unionization in capital-intensive industries also fostered other institutions or organizations that represent low income groups. While we cannot rule this out completely, recall that our results are robust to adjusting for various alternative moderators, including social capital, churches, policies, and state fixed effects (see Table I and Table E.1) and to using a general selection model (below), which does not require this assumption.

[22]We use IPUMS 1 percent 1950 Census samples which use 'state economic areas' (SEA) as geographic identifiers. We create area-weighted crosswalks from SEAs to current congressional districts via spatial polygon intersection (using the GEOS GIS library) and use it to calculate historical stocks of mining/steel employment for each congressional district.

FIGURE IV

Instrumenting union density with post-war mining/steel industry employment

This Figure plots logged district-level employment in mining and steel industries (as share of working population) in 1950 against logged contemporary district-level union density. 109th Congress. Linear fit with robust 95% confidence interval superimposed.

capture variation in pro-business orientation of state in the 1950s, which may shape the intensity of mining or the location of steel industries.

Column (1) of Table II shows IV estimates for the effect of union strength on the link between preferences and legislative voting.[23] A SD increase in unionization increases legislators' responsiveness to the poor by about 8 (±1) percentage points and decreases responsiveness to the upper third by about 5 (±1) points. Our IV estimate lies within the set of estimates reported in Table I. In line with the discussion above, the F-statistic of 639.6 provides evidence of a strong first stage relationship. These results considerably strengthen the causal interpretation of our findings.

Another approach to address the problem of union endogeneity is to use the post-double-selection estimator recently developed in econometrics (Belloni et al. 2014; Chernozhukov et al. 2015). It makes different assumptions than the IV approach but aims to recover the same causal effect, and thus ensures that our causal interpretation does not solely rely on the validity of a single approach.

---

[23]Formally, the 2SLS model instruments the two preference-union interaction terms in the estimation equation, $U_d \times \theta_{jd}^l$ and $U_d \times \theta_{jd}^h$, using $M_{d,1950} \times \theta_{jd}^l$ and $M_{d,1950} \times \theta_{jd}^h$, where $M_{d,1950}$ is the historical stock of mining and steel employment from the 1950 census in contemporary district d.

TABLE II

Probing the Causal Effect of Union Membership on
Legislators' Responsiveness. Instrumental variable and
post-double-selection estimates.

	(1) IV	(2) pDS	(3) pDS
Low income preferences	0.076	0.063	0.058
	(0.012)	(0.014)	(0.017)
High Income preferences	−0.052	−0.054	−0.038
	(0.012)	(0.013)	(0.015)
Semi-parametric terms		144	264
Post-LASSO terms		18	31
First-stage F statistic	639.6		
N	15674	7884	7813

Note: Table entries are estimates for η^L and η^H (robust s.e. in parentheses).
a Column 1 reports instrumental variable (2SLS) estimates. Union density instrumented with district mining employment share in 1950. First-stage F statistic is Kleibergen-Paap weak IV statistic.
b Post-double-selection estimates. See Appendix F for details. N=7,884 (50% validation sample, first half used for selection). Column (2) includes district characteristics in both linear and quadratic form and all their pairwise interactions; (3) adds union policy and social capital terms.

The post-double-selection estimator constructs a high-dimensional vector of controls from functional transforms of observables and their higher order interactions. It thus creates a partially linear model (Robinson 1988) using controls without the functional form restrictions usually imposed in regression models. It then estimates equations for both roll call votes and the interaction of union membership and preferences (including the set of high dimensional controls in all of them). Out of the large number of terms it selects confounders that predict both preferences and roll call votes using the LASSO. The selected set of covariates is used in a post-LASSO estimation step to account for the most relevant confounders. This estimator has low bias and yields accurate confidence intervals even under moderate selection mistakes (Belloni et al. 2014). Responsible for its robustness property is the step selecting the control set from all model equations. It finds controls whose omission leads to "large" omitted variable bias and includes them in the model. Any variables that are not included are therefore at most mildly associated with union density, district preferences, and the outcome, which decidedly limits the scope of omitted variable bias (Chernozhukov et al. 2015). Appendix F provides more details.

The results are in Table II. In specification (2) we include all district variables, their pairwise interactions and their interactions with district preferences, all in both linear and quadratic form. This leads to a vector of 144 covariate terms. In specification (3) we extend the set of possible controls and additionally include union policy variables and our measure of social capital leaving us with 264 controls. As the last line of Table II shows, the estimator selects a subset of these producing more flexible model specifications with the number of included controls being 18 and 31. Again, we find that increasing unionization positively affects the representation of low-income constituents. A SD increase in union membership increases legislators' responsiveness to low-income preferences by about 6 percentage points, while decreasing the responsiveness to the preferences of the affluent by about 4 points. The magnitude of the post-double selection estimates is in line with the IV results and the estimates we obtained in the richer specifications of our previous linear model (compare specifications (4) and (5) in Table I).[24]

Exploring Mechanisms

In this final empirical section, we explore two mechanisms of union influence: campaign contributions and partisan selection.[25] If *contributions* are a channel of union influence, we expect to observe that (i) in districts where local unions are stronger, unions and their members contribute more to sitting members of Congress; and (ii) that these contributions are positively linked to legislative responsiveness. If *partisan selection* is a relevant channel, (i) district-level union strength should have a positive impact on electing Democratic legislators, (ii) who in turn are more likely to vote in line with low-income preferences.

In a first step, we analyze the impact of union membership on the amount of (logged) labor contributions and selection of Democratic legislators. Our measure of contributions is calculated from raw campaign finance contribution data obtained from the Center for Responsive Politics. We sum contributions reported to the Federal Election Commission to candidates from the "labor" sector (excluding single-issue donations). Our count includes both individuals and PACs (but using either alone does not change our results). To guard against omitted variables, we

---

[24]We also investigated if our findings are dependent on the linear functional form imposed on the union-preference interaction terms. In a recent replication survey, Hainmueller et al. (2019) warn that "a large portion of published findings based on multiplicative interaction models are artifacts of misspecification." In Appendix G we show evidence that the moderating effect of unions is evident even in a nonparametric KRLS model without any *a priori* restrictions.

[25]Turnout is a related mechanism analyzed in the literature (e.g. Bartels 2008; Franko et al. 2016; Gilens 2012; Leighley and Oser 2018). While evidently important, for our purposes we prefer campaign contributions because they are clearly a choice variable related to maintaining a credible mobilization threat. Realized aggregate turnout, in contrast, depends more on idiosyncratic factors.

apply the instrumental variable strategy from above. Table III shows the results. In columns (1) - (3), we find that under OLS and IV specifications (with and without district controls) an increase in union membership systematically increases the amount of contributions from labor in that district. Consistent with some degree of endogeneity, the estimate from the IV models is about twice as large as the one from the OLS model. According to model (2) and converted to Dollar amounts, a standard deviation increase in union membership increases contributions from Labor by about $178,500. Turning to partisan selection, columns (4) - (6) support the argument that higher union membership entails a higher probability of a Democratic candidate being elected.

<div align="center">

TABLE III

Exploring Mechanisms, first step. The impact of union membership on the amount of labor contributions and selection of Democratic legislators.

</div>

	A: Contributions			B: Selection		
	(1) OLS	(2) IV	(3) IV	(4) OLS	(5) IV	(6) IV
Union membership	0.072	0.120	0.147	0.158	0.123	0.204
	(0.010)	(0.029)	(0.030)	(0.019)	(0.055)	(0.053)
First-stage F statistic		87.6	28.2		87.6	28.2
District-level controls			✓			✓

Note: Panel A show district-level regression of (log) labor contributions on (log) union membership. Panel B shows district-level LPMs of presence of Democratic representative on (log) union membership. Columns (2) and (5) instrument union density with district mining employment share in 1950. Kleibergen-Paap robust first-stage F statistic. Robust standard errors are in parentheses.

In a second step, we analyze the link between campaign contributions and unequal responsiveness as well as partisanship and unequal responsiveness. Table IV shows the results. Following the specification used in Table I, we estimate linear probability models regressing roll call votes on contributions interacted with constituency preferences, district fixed effects, and, in column (2), district covariates interacted with preferences. We find that in districts where labor contributions are higher, the marginal effect capturing a legislator's responsiveness to the preferences of low income constituents is significantly higher. Consistent with previous research (Bartels 2016; Rhodes and Schaffner 2017), the selection of Democratic legislators is also associated with higher responsiveness to the preferences of low income constituents compared to their Republican counterparts.

In sum, the pro-poor impact of unions rests in part on their ability to mobilize campaign contributions and getting Democratic candidates elected. This is consistent with arguments

based on mobilization threats and rational politicians. While intuitive, these results are by no means mechanical. National union organizations play an important role in lobbying Congress and their political resources need not be directed where unions' grassroots are strongest (Dark 1999). The evidence that district-level union membership nonetheless matters for legislative responsiveness is consistent with the argument that local union strength underpins a credible threat of mobilization that shapes political equality through political selection and post-electoral incentives. The importance of electoral selection visible in our results is in line with a larger body of research on elections and representation (Bartels 2016; Lee et al. 2004; Miller and Stokes 1963). Mobilization efforts by unions remain strongly linked to available human resources on the ground. Recent evidence also shows that the presence of local unions is linked to the perceptions of constituent preferences by congressional staffers (Hertel-Fernandez et al. 2019).

TABLE IV

Exploring Mechanisms, second step. Roll call votes as function of preferences moderated by labor contributions and Democratic legislators.

	(1)	(2)
A: Labor contributions		
× low income preferences	0.946	0.933
	(0.036)	(0.032)
× high income preferences	-0.735	-0.754
	(0.029)	(0.028)
B:Democratic representative		
× low income preferences	0.576	0.561
	(0.012)	(0.013)
× high income preferences	-0.411	-0.431
	(0.013)	(0.013)
District-level controls		✓

Note: Entries are coefficients from LPMs of legislators' vote as function of interaction between district preferences and (log) labor contributions (panel A) and Democratic representative (panel B). Column (2) adds district-level controls interacted with preferences. Robust standard errors are in parentheses.

Analyzing the heterogeneity of the impact of unions also sheds light on the mechanisms. For reasons of space, we relegate the complete discussion to Online Appendix H. We find that our results do not vary much by the type of the union (public or private) or whether the vote concerns a liberal or conservative position. However, the impact of union membership on legislators' responsiveness to low-income preferences is significantly larger for roll call votes

on which the AFL-CIO has taken an a prior position. This bolsters the interpretation that the egalitarian effect of unions is driven by their capacity for political action.

Conclusion

Dahl (1961) famously asked: who governs in a polity where political rights are equally distributed, but where large inequalities in income and wealth may bias representation? In the wake of rising income inequality in the US, scholars have identified the question of political inequality as one of the central challenges facing democracy in the twenty-first century (APSA 2004). While the scientific debate is ongoing, a growing number of studies has documented striking patterns of unequal responsiveness by income. When policy preferences diverge across income groups, legislators are biased toward the affluent at the expense of the middle-class and—especially—the poor. Many recent works conclude by asking what factors may improve political representation of the economically disadvantaged.

We contribute to this agenda by analyzing whether labor unions can serve as an effective collective voice institution that limits unequal representation in Congress. Going beyond existing work on this topic, we develop a research design to account for the fact that union strength is endogenous to politics. It is based on district fixed effects, rich data on alternative moderators of income-group preferences, and a instrumental variable approach. Against the view that unions are not relevant causal forces for political equality, we find that the district-level strength curbs unequal representation. While legislators are, on average, more responsive to the preferences of the affluent than to the preferences of the poor, this representation gap is highly variable. It is much less pronounced in districts where union membership is relatively higher.

Our findings cast a somewhat less pessimistic light on democratic representation in Congress. Despite high income inequality, polarization, expensive campaigns, and a legislature dominated by affluent politicians (Carnes 2013; Gilens 2012; Hacker and Pierson 2010; McCarty et al. 2006), they suggest that legislative underrepresentation of the poor is not an unavoidable feature of democracy in America. These results encourage research on public opinion and representation in the US to pay more attention to unions. Our research echoes the call to bring back unions into the mainstream of American politics (Ahlquist 2017; Thelen 2019).

References

Ahlquist, J. (2017). Labor unions, political representation, and economic inequality. *Annual Review of Political Science 17*, 409–432.

Ahlquist, J. S. and M. Levy (2013). *In the Interests of Others*. Princeton: Princeton University Press.

American Political Science Association, Task Force on Inequality and American Democracy (2004). American democracy in an age of rising inequality. *Perspectives on Politics 2*(4), 651–666.

Ansolabehere, S. and P. E. Jones (2010). Constituents' responses to congressional roll-call voting. *American Journal of Political Science 54*(3), 583–597.

Anzia, S. F. and T. M. Moe (2016). Do politicians use policy to make politics? the case of public-sector labor laws. *American Political Science Review 110*(4), 763–777.

Arnold, D. R. (1990). *The Logic of Congressional Action*. New Haven: Yale University Press.

Barber, M. J. (2016). Representing the preferences of voters, partisans, and voters in the us senate. *Public Opinion Quarterly 80*(1), 225–249.

Bartels, L. (2008). *Unequal Democracy: The Political Economy of the New Gilded Age* (1st ed.). Princeton: Princeton University Press.

Bartels, L. (2016). *Unequal Democracy: The Political Economy of the New Gilded Age* (2nd ed.). Princeton: Princeton University Press.

Bartels, L. M. (2017). Political inequality in affluent democracies: The social welfare deficit. Vanderbilt University, CSDI Working Paper 5-2017. [www.vanderbilt.edu/csdi/includes/Working_Paper_5_2017.pdf].

Bashir, O. S. (2015). Testing inferences about american politics: A review of the "oligarchy" result. *Research & Politics 2*(4), 1–7.

Becher, M. and D. Stegmueller (2019). Global economic shocks, local economic institutions, and legislative responses. Paper prepared for the 77th Annual Conference the Midwest Political Science Association, Chicago, April 4-7, 2019.

Becher, M., D. Stegmueller, and K. Kaeppner (2018). Local union organization and law making in the us congress. *Journal of Politics 80*(2), 39–554.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection amongst high-dimensional controls. *Review of Economic Studies 81*, 608–650.

Bhatti, Y. and R. S. Erikson (2011). How poorly are the poor represented in the u.s. senate? In P. K. Enns and C. Wlezien (Eds.), *Who Gets Represented?*, pp. 223–246. New York: Russel Sage Foundation.

Box-Steffensmeier, J. M., L. W. Arnold, and C. J. W. Zorn (1997). The strategic timing of position taking in congress: A study of the north american free trade agreement. *American Political Science Review 91*(2), 324–338.

Budd, J. W. (2018). *Labor Relations: Striking a Balance* (5 ed.). New York, NY: McGraw-Hill Education.

Butler, D. M. (2014). *Representing the Advantaged.* New York: Cambridge University Press.

Campante, F. R. (2011). Redistribution in a model of voting and campaign contributions. *Journal of Public Economics 95*(7-8), 646 – 656.

Carnes, N. (2013). *White-Collar Government: The Hidden Role of Class in Economic Policy Making.* Chicago, IL: University of Chicago Press.

Chernozhukov, V., C. Hansen, and M. Spindler (2015). Valid post-selection and post-regularization inference: An elementary, general approach. *Annual Review of Economics 7*(1), 649–688.

Dahl, R. A. (1961). *Who Governs?* New Haven: Yale University Press.

Dark, T. E. (1999). *The Unions and the Democrats.* Ithaca: Cornell University Press.

Ellis, C. (2013). Social context and economic biases in representation. *Journal of Politics 75*(3), 773–786.

Elsässer, L., S. Hense, and A. Schäfer (2018). Government of the people, by the elite, for the rich: Unequal responsiveness in an unlikely case. MPIfG Discussion Paper 18/5.

Enns, P. K. (2015). Relative policy support and coincidental representation. *Perspectives on Politics 13*(4), 1053–1064.

Erikson, R. S. (2015). Income inequality and policy responsiveness. *Annual Review of Political Science 18*, 11–29.

Feigenbaum, J., A. Hertel-Fernandez, and V. Williamson (2018). From the bargaining table to the ballot box: Political effects of right to work laws. NBER Working Paper 24259, [www.nber.org/papers/w22637].

Flavin, A. (2012). Inequality and policy representation in the american states. *American Politics Research 40*(1), 29–59.

Flavin, P. (2018). Labor union strength and the equality of political representation. *British Journal of Political Science 48*(4), 1075–1091.

Flavin, P. and M. T. Hartney (2015). When government subsidizes its own: Collective bargaining laws as agents of political mobilization. *American Journal of Political Science 59*(4), 896–911.

Franko, W. W., N. J. Kelly, and C. Witko (2016). Class bias in voter turnout, representation, and income inequality. *Perspectives on Politics 14*(2), 351–368.

Freeman, R. B. and J. Medoff (1984). *What Do Unions Do?* New York: Basic Books.

Freeman, R. B. and J. Rogers (2006). *What Workers Want* (Updated ed.). New York: Russell Sage Foundation.

Gilens, M. (2012). *Affluence and Influence: Economic Inequality and Political Power in America.* Princeton: Princeton University Press and Russel Sage Foundation.

Gilens, M. and B. I. Page (2014). Testing theories of american politics: Elites, interest groups, and average citizens. *Perspectives on Politics 12*(3), 564–581.

Hacker, J. S. and P. Pierson (2010). *Winner-Take-All Politics.* New York, NY: Simon & Schuster.

Hainmueller, J., J. Mummolo, and Y. Xu (2019). How much should we trust estimates from multiplicative interaction models? simple tools to improve empirical practice. *Political Analysis 27*, 163–192.

Hertel-Fernandez, A. (2018). Policy feedback as political weapon: Conservative advocacy and the demobilization of the public sector labor movement. *Perspectives on Politics 16*(2), 364–379.

Hertel-Fernandez, A., M. Mildenberger, and L. Stokes (2019). Legislative staffers and representation in congress. *American Political Science Review 113*(2), 1–18.

Hirsch, B., D. Macpherson, and W. Vroman (2001). Estimates of union density by state. *Monthly Labor Review 124*(7), 51–55.

Holmes, T. J. (2006). Geographic spillover of unionism. NBER Working Paper 12025.

Honaker, J. and E. Plutzer (2016). Small area estimation with multiple overimputation. Manuscript. [http://hona.kr/papers/files/smallAreaEstimation.pdf].

Iversen, T. and D. Soskice (2006). Electoral institutions and the politics of coalitions: Why some democracies redistribute more than others. *American Political Science Review 100*(2), 165–181.

Iversen, T. and D. Soskice (2015). Information, inequality, and mass polarization: Ideology in advanced democracies information, inequality, and mass polarization ideology in advanced democracies. *Comparative Political Studies 48*(13), 1781 –1813.

Jessee, S. A. (2009). Spatial Voting in the 2004 Presidential Election. *American Political Science Review 103*(1), 59–81.

Kalla, J. L. and D. E. Broockman (2016). Campaign contributions facilitate access to congressional officials: A randomized field experiment. *American Journal of Political Science 60*(3), 545–558.

Kim, S. E. and Y. Margalit (2017). Informed preferences? the impact of unions on workers' policy views. *American Journal of Political Science 61*, 728–743.

Lax, J. R. and J. H. Phillips (2009). How should we estimate public opinion in the states? *American Journal of Political Science 53*(1), 107–121.

Lee, D. S., E. Moretti, and M. J. Butler (2004). Do voters affect or elect policies? evidence from the U. S. House. *Quarterly Journal of Economics 119*(3), 807–859.

Leighley, J. E. and J. Nagler (2007). Unions, voter turnout, and class bias in the U.S. electorate, 1964-2004. *Journal of Politics 69*(2), pp. 430–441.

Leighley, J. E. and J. Oser (2018). Representation in an era of political and economic inequality: How and when citizen engagement matters. *Perspectives on Politics 16*(2), 328–344.

Lichtenstein, N. (2013). *State of the Union: A Century of American Labor* (2nd ed.). Princeton: Princeton University Press.

Long Jusko, K. (2017). *Who Speaks for the Poor? Electoral Geography, Party Entry, and Representation.* Cambridge: Cambridge University Press.

Macdonald, D. (2019). How labor unions increase political knowledge: Evidence from the United States. *Political Behavior*, Forthcoming.

McCarty, N., K. T. Poole, and H. Rosenthal (2006). *Polarized America.* Cambridge, MA: MIT Press.

Mian, A., A. Sufi, and F. Trebbi (2010). The political economy of the us mortgage default crisis. *American Economic Review 100*(5), 1967–1998.

Miller, W. E. and D. E. Stokes (1963). Constituency influence in congress. *American Political Science Review 57*(1), 45–56.

Moe, T. M. (2011). *Special Interest: Teachers Unions and America's Public Schools.* Washington, D.C.: Brookings Institution.

Mosimann, N. and J. Pontusson (2017). Solidaristic unionism and support for redistribution in contemporary europe. *World Politics 69*(3), 448–492.

Olson, M. (1965). *The Logic of Collective Action.* Cambridge: Harvard University Press.

Putnam, R. (1993). *Making Democracy Work.* Princeton, NJ: Princeton University Press.

Putnam, R. (2000). *Bowling Alone. The collapse and revival of american community.* New York: Simon and Schuster.

Rhodes, J. H. and B. F. Schaffner (2017). Testing models of unequal representation: Democratic populists and republican oligarchs? *Quarterly Journal of Political Science 12*(s), 185–204.

Rigby, E. and G. C. Wright (2013). Political parties and representation of the poor in the american states. *American Journal of Political Science 57*(3), 552–565.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica 56*(4), 931–954.

Rosenfeld, J. (2014). *What Unions No Longer Do.* Cambridge: Harvard University Press.

Rupasingha, A. and S. J. Goetz (2008). US county-level social capital data, 1990-2005. The northeast regional center for rural development, Penn State University, University Park, PA.

Samii, C. (2016). Causal empiricism in quantitative research. *Journal of Politics 78*(3), 941–955.

Schlozman, D. (2015). *When Movements Anchor Parties.* Princeton: Princeton University Press.

Schlozman, K. L., S. Verba, and H. E. Brady (2012). *The Unheavenly Chorus. Unequal Political Voice and the Broken Promise of American Democracy.* Princeton: Princeton University Press.

Schlozman, K. L., S. Verba, and H. E. Brady (2018). *Unequal and Unrepresented: Political Inequality and the People's Voice in the New Gilded Age.* Princeton and Oxford: Princeton University Press.

Soroka, S. N. and C. Wlezien (2008). On the limits to inequality in representation. *PS: Political Science & Politics 41*(2), 319–327.

Southworth, C. and J. Stepan-Norris (2009). American trade unions and data limitations: A new agenda for labor studies. *Annual Review of Sociology 35*, 297–320.

Stekhoven, D. J. and P. Bühlmann (2011). Missforest. non-parametric missing value imputation for mixed-type data. *Bioinformatics 28*(1), 112–118.

Stimson, J. A., M. B. Mackuen, and R. S. Erikson (1995). Dynamic representation. *American Political Science Review 89*(3), 543–565.

Thelen, K. (2019). The american precariat: U.S. capitalism in comparative perspective. *Perspectives on Politics 17*(1), 1–23.

Appendix to Curbing Unequal Representation: The Impact of Labor Unions on Legislative Responsiveness in the US Congress

Contents

A. Data

In this appendix we present additional details on our dataset including details on the creation of some control variables and descriptive statistics.

Matched roll calls   Table A.2 displays Congressional roll calls matched to CCES items. We selected congressional roll calls based on content and, when several choices were available, based on their proximity to CCES fieldwork periods.

Income thresholds   Table A.1 presents an overview of the income thresholds we use to classify CCES respondents into income groups. We use two thresholds separating the lowest and highest income terciles. We calculate them from yearly American Community Survey files excluding individuals living in group quarters. For each congress, Table A.1 shows the average of all district-specific thresholds as well as the smallest and largest ones.

Public unions   Public unions captured (by name) in our data include the American Federation of State, County & Municipal Employees, National Education Association, American Federation of Teachers, American Federation of Government Employees, National Association of Government Employees, United Public Service Employees Union, National Treasury Employees Union, American Postal Workers Union, National Association of Letter Carriers, Rural Letter Carriers Association, National Postal Mail Handlers Union, National Alliance of Postal and Federal Employees, Patent Office Professional Association, National Labor Relations Board Union, International Association of Fire Fighters, Fraternal Order of Police, National Association of Police Organizations, various local police associations, and various local public school unions.

TABLE A.1

Distribution of district income-group reference points. Average threshold over all districts, smallest and largest value.

Congress	33th percentile			67th percentile		
	Mean	Min	Max	Mean	Min	Max
109	38123	16800	73675	77964	39612	146870
110	40127	18000	77000	83047	43600	155113
111	39021	17500	78262	82440	46000	160050
112	37381	16500	81000	79868	38500	158654

Note: Calculated from American Community Survey 1-year files. Household sample excluding group quarters. Missing income information imputed using Chained Random Forests.

Matched CCES–House roll calls included in our analysis.

Match	Bill	Date	Name	House Vote (Yea-Nay)	Bill Ideology†
(1)	HR 810	07/19/2006	Stem Cell Research Enhancement Act (Presidential Veto override)	235-193	L
(1)	HR 3	01/11/2007	Stem Cell Research Enhancement Act of 2007 (House)	253-174	L
(1)	S 5	06/07/2007	Stem Cell Research Enhancement Act of 2007	247-176	L
(2)	HR 2956	07/12/2007	Responsible Redeployment from Iraq Act	223-201	L
(3)	HR 2	01/10/2007	Fair Minimum Wage Act	315-116	L
(4)	HR 4297	12/08/2005	Tax Relief Extension Reconciliation Act (Passage)	234-197	C
(4)	HR 4297	05/10/2006	Tax Relief Extension Reconciliation Act (Agreeing to Conference Report)	244-185	C
(5)	HR 3045	07/28/2005	Dominican Republic-Central America-United States Free Trade Agreement Implementation Act	217-215	C
(6)	S 1927	08/04/2007	Protect America Act	227-183	C
(6)	HR 6304	06/20/2008	FISA Amendments Act of 2008	293-129	C
(7)	HR 3162	08/01/2007	Children's Health and Medicare Protection Act	225-204	L
(7)	HR 976	10/18/2007	Children's Health Insurance Program Reauthorization Act (Presidential Veto Override)	273-156	L
(7)	HR 3963	01/23/2008	Children's Health Insurance Program Reauthorization Act (Presidential Veto Override)	260-152	L
(7)	HR 2	02/04/2009	Children's Health Insurance Program Reauthorization Act	290-135	L
(8)	HR 3221	07/23/2008	Foreclosure Prevention Act of 2008	272-152	L
(9)	HR 3688	11/08/2007	United States-Peru Trade Promotion Agreement	285-132	C
(10)	HR 1424	10/03/2008	Emergency Economic Stabilization Act of 2008	263-171	L
(11)	HR 3080	10/12/2011	To implement the United States-Korea Trade Agreement	278-151	C
(12)	HR 3078	10/12/2011	To implement the United States-Colombia Trade Promotion Agreement	262-167	C
(13)	HR 2346	06/16/2009	Supplemental Appropriations, Fiscal Year 2009 (Agreeing to conference report)	226-202	L
(14)	HR 2831	07/31/2007	Lilly Ledbetter Fair Pay Act	225-199	L
(14)	HR 11	01/09/2009	Lilly Ledbetter Fair Pay Act of 2009 (House)	247-171	L
(14)	S 181	01/27/2009	Lilly Ledbetter Fair Pay Act of 2009	250-177	L
(15)	HR 1913	04/29/2009	Local Law Enforcement Hate Crimes Prevention Act	249-175	L
(16)	HR 1	02/13/2009	American Recovery and Reinvestment Act of 2009 (Agreeing to Conference Report)	246-183	L
(17)	HR 2454	06/26/2009	American Clean Energy and Security Act	219-212	L
(18)	HR 3590	03/21/2010	Patient Protection and Affordable Care Act	220-212	L
(19)	HR 3962	11/07/2009	Affordable Health Care for America Act	221-215	L
(20)	HR 4173	06/30/2010	Wall Street Reform and Consumer Protection Act of 2009	237-192	L
(21)	HR 2965	12/15/2010	Don't Ask, Don't Tell Repeal Act of 2010	250-175	L
(22)	S 365	08/01/2011	Budget Control Act of 2011	269-161	C
(23)	H CR 34	04/15/2011	House Budget Plan of 2011	235-193	C
(24)	H CR 112	03/28/2012	Simpson-Bowles/Copper Amendment to House Budget Plan	38-382	C
(25)	HR 8	08/01/2012	American Taxpayer Relief Act of 2012 (Levin Amendment)	170-257	L
(26)	HR 2	01/19/2011	Repealing the Job-Killing Health Care Law Act	245-189	C
(26)	HR 6079	07/11/2012	Repeal the Patient Protection and Affordable Care Act and […]	244-185	C
(27)	HR 1938	07/26/2011	North American-Made Energy Security Act	279-147	C

Note: The matching of roll calls to CCES items can be many-to-one.

† Coding of a bill's ideological character as (L)iberal or (C)onservative based on predominant support of bill by Democratic or Republican representatives, respectively.

TABLE A.3

Descriptive statistics of analysis sample

	Mean	SD	Min	Max	N
Roll-call vote: yea	0.568	0.495	0.000	1.000	15780
Constituent preferences					
Low income	0.593	0.220	0.047	0.979	15934
High income	0.555	0.198	0.037	0.967	15934
Low-High Gap	0.172	0.121	0.000	0.588	15934
Union membership [log]	9.705	1.046	6.094	13.619	15934
Population	7.022	0.723	4.697	9.980	15934
Share African American	0.124	0.146	0.004	0.680	15934
Share Hispanic	0.156	0.174	0.005	0.812	15934
Share BA or higher	0.275	0.097	0.073	0.645	15934
Median income [$10,000]	5.177	1.356	2.282	10.439	15934
Share female	0.508	0.010	0.462	0.543	15934
Manufacturing share	0.110	0.047	0.025	0.281	15934
Urbanization	0.790	0.199	0.213	1.000	15934
Social capital index	-0.675	0.758	-2.688	2.136	15789
Certification elections [log]	3.347	0.861	0.000	5.100	15934
Congregations [per 1000 persons]	0.765	1.147	0.062	6.453	15934

Note: Calculated from American Community Survey, 2006-2013. Note that when entered in models, variables are scaled to mean zero and unit SD. Preference gap is absolute difference in preferences between low and high income constituents in sample. Urbanization is calculated as the share of the district population living in an urban area based on the Census' definition of urban Census blocks (matched to congressional districts using the MABLE database). Congregations per 1000 inhabitants calculated from RCMS 2000 (spatially interpolated).

Descriptive statistics   Table A.3 shows descriptive statistics for all variables used in our analysis. Note that these are for the untransformed variables. In our empirical models, we standardize all inputs to have mean zero and unit standard deviation.

B. Estimation of District Preferences

In this section we describe how we estimate district-level preferences using three different strategies: (i) small area estimation using a matching approach based on random forests (which we use in the main text of our paper), (ii) estimation using multilevel regression and post-stratification (MRP), and (iii) unadjusted cell means. Each approach invokes different statistical and substantive assumptions. In the spirit of consilience, our aim here is to show that our substantive results do not depend on any particular choice.

B.1. Small Area Estimation via Chained Random Forests

The core idea of our small area estimation strategy is based on the fact that we have access to two samples: one that is likely not representative of the population of all Congressional districts (the CCES), while the second one is representative of district populations by virtue of its sampling design (the Census or American Community Survey). By matching or imputing preferences from the former to the latter based on a common vector of observable individual characteristics, we can use the district-representative sample to estimate the preferences of individuals in a given district.[1]

Combining CCES and Census data using Random Forests   Figure B.1 illustrates this approach in more detail. We have data from m individuals in the CCES and n individuals in the Census (with $n \gg m$). Both sets of individuals share K common characteristics Z_k, such as age, race, or education. The first task at hand is then to match P roll call preferences Y_p that are only observed in the CCES to the census sample. This is a purely predictive task and it is thus well suited for machine learning approaches. We use random forests (Breiman 2001) to lean about $Y_p = f(Z_1, \ldots, Z_K)$ for $p = 1, \ldots, P$ using the algorithm proposed by Stekhoven and Bühlmann (2011). This approach has two key advantages. First, as is typical for approaches based on regression trees, it deals with both categorical and continuous data, allows for arbitrary functional forms, and can include higher order interactions between covariates (such as age×race×education). Second, we can assess the quality of the predictions based on our model before we deploy it to predict preferences in the Census. With the trained model in hand we can use $\widehat{f}(Z_1, \ldots, Z_K)$ in combination with observed Z in the Census sample to fill in preferences (i.e., completing the square in the lower right of Figure B.1). Using the completed Census data, we can estimate constituent district preferences as simple averages by district and income group since the Census sample is representative for each Congressional district's population.

Data details   Due to data confidentially constraints the Census Bureau does not provide district identifiers in its micro-data records. Instead, it identifies 630 Public Use Microdata areas. We create a synthetic Census sample for Congressional districts by sampling individuals from the full Census PUMA regions proportional to their relative share in a given districts. This information is based on a crosswalk from PUMA regions to Congressional districts created by recreating one from the other based on Census tract level population data in the MABLE Geocorr2K database. The 'donor pool' for this synthetic sample are the 1% extracts for the American Community Survey 2006-2011. We limit the sample to non-group quarter households and to individuals aged 17

---

[1]See Honaker and Plutzer (2016) for a more explicit exposition of this idea, evidence for its empirical reliability, and a comparison to MRP estimates.
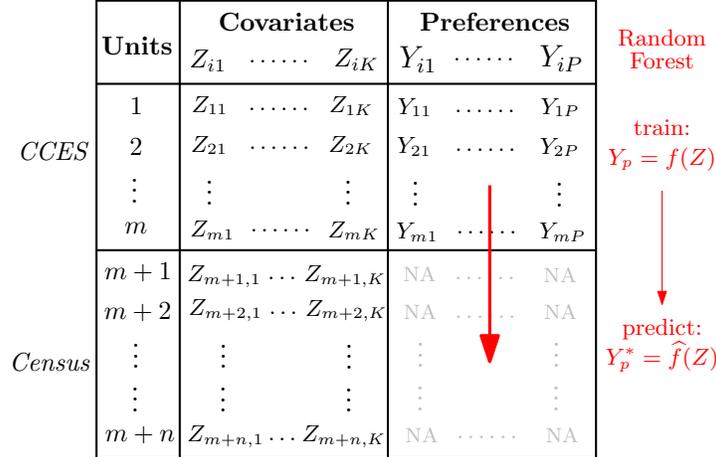
	Units	**Covariates** $Z_{i1} \quad \cdots\cdots \quad Z_{iK}$	**Preferences** $Y_{i1} \quad \cdots\cdots \quad Y_{iP}$	
				Random Forest
CCES	1	$Z_{11} \quad \cdots\cdots \quad Z_{1K}$	$Y_{11} \quad \cdots\cdots \quad Y_{1P}$	train:
	2	$Z_{21} \quad \cdots\cdots \quad Z_{2K}$	$Y_{21} \quad \cdots\cdots \quad Y_{2P}$	$Y_p = f(Z)$
	$\vdots$	$\vdots \qquad\qquad \vdots$	$\vdots \qquad\qquad \vdots$	
	m	$Z_{m1} \quad \cdots\cdots \quad Z_{mK}$	$Y_{m1} \quad \cdots\cdots \quad Y_{mP}$	
Census	$m+1$	$Z_{m+1,1} \cdots Z_{m+1,K}$	NA $\quad \cdots\cdots \quad$ NA	
	$m+2$	$Z_{m+2,1} \cdots Z_{m+2,K}$	NA $\quad \cdots\cdots \quad$ NA	predict:
	$\vdots$	$\vdots \qquad\qquad \vdots$	$\vdots \qquad\qquad \vdots$	$Y_p^* = \widehat{f}(Z)$
	$\vdots$	$\vdots \qquad\qquad \vdots$	$\vdots \qquad\qquad \vdots$	
	$m+n$	$Z_{m+n,1} \cdots Z_{m+n,K}$	NA $\quad \cdots\cdots \quad$ NA	

FIGURE B.1

Illustration of Small Area Estimation of District Preferences.

We use a sample of *m* individuals from the CCES that is not necessarily representative on the district-level, while a sample of *n* individuals from the Census is representative of district populations by design (Torrieri et al. 2014: Ch.4). We have access to bridging covariates Z_k that are common to both samples, while roll call preferences Y_p are only observed in the CCES. We train a flexible non-parametric model relating Y_p to Z and use it to predict preferences Y_p^* for Census individuals with characteristics Z. With preference values filled in, a district's income-group specific roll call preference can be estimated as the average of all units in that district.

and older providing us with data on 14 million (13,711,248) Americans. From this we create the synthetic district file which is comprised of 3,040,265 cases. This provides us with a Census sample including Congressional district identifiers. The sample for each district is representative of the district population (save for errors induced by the crosswalk). We thus use the distribution of important population characteristics (age, gender, education, race, income) to match data on policy preferences from the CCES.

We harmonize all covariates to be comparable between CCES and Census. For family income this entails an adjustment to the measure provided in the CCES. It asks respondents to place their family's total household income into 14 income bins.[2] We transform this discretized measure of income into a continuous one using a nonparametric midpoint Pareto estimator. It replaces each bin with its midpoint (e.g., the third category $20,000 to $29,999 gets assigned $25,000), while the value for the final, open-ended, bin is imputed from a Pareto distribution (e.g., Kopczuk et al. 2010). Using midpoints has been recognized for some time as an appropriate way to create scores for income categories (without making explicit distributional modeling assumptions). They have

---

[2]The exact question wording is: "Thinking back over the last year, what was your family's annual income?" The obvious issue here is that it is not clear which income concept this refers to (or, rather, which on the respondent employs). In line with the wording used in many other US surveys, we interpret it as referring to market income.

been used extensively, for example, in the American politics literature analyzing General Social Survey (GSS) data (Hout 2004).

Algorithm details   For easier exposition define a matrix D that contains both individual characteristics and roll call preferences. Let N be the number of rows of D. For any given variable v of D, D_v, with missing entries at locations $i_{mis}^{(v)} \subseteq \{1, \ldots, N\}$ we can separate out four parts:[3]

- Observed values of D_v: denoted as $y_{obs}^{(v)}$
- Missing values of D_v: $y_{mis}^{(v)}$
- Variables other than D_v with available observations $i_{obs}^{(v)} = \{1, \ldots, N\} \setminus i_{mis}^{(v)}$: $x_{obs}^{(v)}$
- Variables other than D_v with observations $i_{mis}^{(v)}$: $x_{mis}^{(v)}$

We now cycle through variables iteratively fitting random forest and filling in unobserved values until a stopping criterion c (indicating no further change in filled-in values) is met. Algorithmically, we proceed as follows:

Algorithm 1 Chained Random Forests

1: Start with initial guesses of missing values in D
2: $w \leftarrow$ vector of column indices sorted by increasing fraction of NA
3: **while** not c **do**
4:     $D_{old}^{imp} \leftarrow$ previously imputed D
5:     **for** v in w **do**
6:         Fit Random Forest: $y_{obs}^{(v)} \sim x_{obs}^{(v)}$
7:         Predict $y_{mis}^{(v)}$ using $x_{mis}^{(v)}$
8:         $D_{new}^{imp} \leftarrow$ updated imputed matrix using predicted $y_{mis}^{(v)}$
9:     Updated stopping criterion c
10: Return completed D^{imp}

To assess the quality of this scheme, we inspect the prediction error of the random forests using the out-of-bag (OOB) estimate (which can be obtaining during the bootstrap for each tree). We find it to be rather small in our application: most normalized root mean squared errors are around 0.11. This result is in line with simulations by Stekhoven and Bühlmann (2011) who compare it to other prediction schemes based on K nearest neighbors, EM-type LASSO algorithms,

---

[3]Note that this setup deals transparently with missing values in individual characteristics (such as missing education).

or multivariate normal schemes and find it to perform comparatively well with both continuous and categorical variables.[4]

B.2. *Multilevel Regression and Poststratification*

The approach described in the last section is closely related to MRP (Gelman and Little 1997; Park et al. 2006; Lax and Phillips 2013), which has become quite popular in political science. Both strategies involve fitting a model that is predictive of preferences given observed characteristics followed by a weighting step that re-balances observed characteristics to their distribution in the Census. What differentiates MRP from the previous approach is that it imposes more structure in the modeling step both in terms of functional form and distributional assumptions. By utilizing the advantages of hierarchical models with normally distributed random coefficients it produces preference estimates that are shrunken towards group means (Gelman et al. 2013: 116f.).[5] No such structural assumptions are made when using Random Forests. It will thus be instructive to compare how much our results depend on such modeling choices.

MRP implementation   For each roll call item in the CCES we estimate a separate model expressing the probability of supporting a proposal as a function of demographic characteristics. The demographic attributes included in our model broadly follow Lax and Phillips (2009, 2013) and are race, gender, education, age, and income.[6] Race is captured in three categories (white, black, other), education in five (high school or less, some college, 2-year college degree, 4-year college degree, graduate degree). Age is comprised of 6 categories (18-29, 30-39, 40-49, 50-59, 60-69, 70+) while income is comprised of 13 categories (with thresholds 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 100, 120, 150 [in $1,000]). Our model also includes district-specific intercepts. For each roll-call, we estimate the following hierarchical model using penalized maximum likelihood (Chung et al. 2013):

$$Pr(Y_i = 1) = \text{logit}^{-1}\left(\beta^0 + \alpha^{race}_{j[i]} + \alpha^{gender}_{k[i]} + \alpha^{age}_{l[i]} + \alpha^{educ}_{m[i]} + \alpha^{income}_{n[i]} + \alpha^{district}_{d[i]}\right) \qquad \text{(B.1)}$$

---

[4]See Tang and Ishwaran (2017) for further empirical validation of this strategy. See also Honaker and Plutzer (2016), who compare a similar matching strategy (but based on a multivariate normal model) with MRP estimated preferences using the CCES.

[5]This might be especially appropriate when some groups are small. The median number of respondents per district in the CCES is 506 and no district has fewer than 192 sampled respondents. But since we slice preferences further by income sub-groups, one may be worried that the sample size in some districts is small. MRP deals with this potential issue at the cost of making distributional assumptions.

[6]We also estimated a version of the model including a macro-level predictor, which has been found to improve the quality of the model. We use the demographically purged state predictor of Lax and Phillips (2013: 15), that is, the average liberal–conservative variation in state-level public opinion that is not due to variation demographic predictors. In our case this produces rather similar MRP estimates.

We employ the notation of Gelman and Hill (2007) and denote by $j[i]$ the category j to which individual i belongs. Here, β_0 is an intercept and the αs are hierarchically modeled effects for the various demographic groups. Each is drawn from a common normal distribution with mean zero and estimated variance σ^2:

$$\alpha_j^{race} \sim N\left(0, \sigma_{race}^2\right), \quad j = 1, \ldots, 3 \tag{B.2}$$

$$\alpha_k^{gender} \sim N\left(0, \sigma_{gender}^2\right), \quad k = 1, \ldots, 2 \tag{B.3}$$

$$\alpha_l^{age} \sim N\left(0, \sigma_{age}^2\right), \quad l = 1, \ldots, 6 \tag{B.4}$$

$$\alpha_m^{educ} \sim N\left(0, \sigma_{educ}^2\right), \quad m = 1, \ldots, 5 \tag{B.5}$$

$$\alpha_n^{income} \sim N\left(0, \sigma_{income}^2\right), \quad n = 1, \ldots, 13 \tag{B.6}$$

This setup induces shrinkage estimates for the same demographic categories in different districts. Note that using fixed effects for characteristics with few categories (Specifically, gender) does not impact our results. The district intercepts are drawn from a normal distribution with state-specific means $\alpha_{s[d]}$ and freely estimated variance:

$$\alpha_d \sim N\left(\alpha_{s[d]}^{state}, \sigma_{state}^2\right). \tag{B.7}$$

Our final preferences estimates for each income group on each roll call are obtained by using cell-specific predictions from the above hierarchical model, weighted by the population frequencies (obtained from our Census file) for each cell in each congressional district.

B.3. Model results under various preference estimation strategies

The estimates of district-level preferences obtained via our SAE approach and MRP are in broad agreement: The median difference in district preferences between SAE and MRP is 2.5 percentage points for low income and -0.1 percentage points for high income constituents. A large part of this difference is due to the heavier tails of the distribution of district preferences for each roll call estimated by our approach—perhaps not surprising given the shrinkage characteristics of MRP. To what extent do these differences in the distribution of preferences affect our estimated union effects?

Panel (A) of Table B.1 shows estimates for our six main specifications using MRP-based preferences. The results are unequivocal: using MRP estimated preferences leads to more pronounced estimates in all specifications. Using specification (6), which includes state policies, measures of district social capital and district covariates interacted with preferences, as well as district fixed

effects, we find that a unit increase in union membership increased responsiveness of legislators towards the preferences of low income constituents by about 8 (±2) percentage points (compared to only 5 points using our measurement strategy). Responsiveness estimated for high income preferences are similarly larger. Note that while larger, all estimates also carry increased confidence intervals.

TABLE B.1

Model results using different strategies to estimate district-level preferences.

	(1)	(2)	(3)	(4)	(5)	(6)
A: Multilevel Regression & Poststratification						
Low income preferences	0.182	0.158	0.181	0.185	0.115	0.081
	(0.021)	(0.024)	(0.026)	(0.022)	(0.022)	(0.022)
High income preferences	−0.136	−0.119	−0.139	−0.137	−0.091	−0.064
	(0.017)	(0.019)	(0.021)	(0.017)	(0.018)	(0.018)
B: Raw CCES means						
Low income preferences	0.080	0.061	0.063	0.080	0.043	0.028
	(0.010)	(0.011)	(0.012)	(0.010)	(0.011)	(0.011)
High income preferences	−0.027	−0.013	−0.010	−0.025	−0.018	−0.011
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)

Note: Specifications (1) to (6) are as in Table I in the main text but using different strategies to estimate district-level preferences of three income groups. Entries are effects of standard deviation increase in union membership on marginal effect of income group preferences on legislator vote.

As a further point of comparison, panel (B) shows preferences estimated via raw cell means in the CCES. Due to the issues discussed above, the raw data cannot be taken as gold standard, but it is nonetheless informative to see how much the results vary. Our core results even obtain when we simply use raw cell means without any statistical modeling to counter non-representative distributions of individual characteristics and small cell sizes. We find that in our strictest specification, a unit increase in union membership still increases responsiveness towards low income constituents by about 3 (±1) percentage points.

In sum, all three approaches lead to the same qualitative conclusions about the moderating effect of unions on unequal representation in Congress. The two alternative approaches to deal with the problem that CCS surveys are not representative of congressional districts by design suggest that a larger effect of unions than the naive approach using the unadjusted survey data. Quantitatively, our preferred estimates are based on small area estimation via random forests as they are less reliant on normality assumptions and are systematically more conservative than those based on MRP.

C. Alternative Income Thresholds

This section discusses the impact of different income thresholds on our results. In Table I in the main text, preferences of income groups are based on a *district-specific* income thresholds splitting the population into three groups (at the 33rd and 66th percentile). Thus, voters are classified as 'low income' relative to other voters in their congressional district. For example, during the 111th Congress a voter with an income of $40,000 would be part of the low income group in most of Massachusetts' districts (where low income thresholds vary from about $40,000 to $50,000), but not in the 8th (where the threshold is about $30,000). If income threshold were *state-specific* instead, he or she would be considered low income everywhere in the state (as the state-specific low income threshold is now ≈$47,000).

TABLE C.1
Model results using different definitions of income groups.

	(1)	(2)	(3)	(4)	(5)	(6)
A: State-specific income thresholds						
Low income preferences	0.105	0.082	0.097	0.107	0.067	0.044
	(0.013)	(0.015)	(0.016)	(0.013)	(0.014)	(0.014)
High income preferences	−0.062	−0.036	−0.052	−0.065	−0.049	−0.027
	(0.012)	(0.013)	(0.014)	(0.013)	(0.013)	(0.013)
B: Shifted income thresholds: p20 - p80						
Low income preferences	0.098	0.077	0.09	0.100	0.063	0.042
	(0.012)	(0.013)	(0.014)	(0.012)	(0.013)	(0.012)
High income preferences	−0.054	−0.031	−0.046	−0.057	−0.044	−0.025
	(0.011)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)

Note: Specifications (1) to (6) are as in Table I in the main text but with income groups defined via different income thresholds. Entries are effects of standard deviation increase in union membership on marginal effect of income group preferences on legislator vote.

Not all states display as much variation in income-group thresholds. Thus, using state- instead of district-specific thresholds does not alter our core results in an appreciable way. As Panel (A) shows, the resulting marginal effects estimates for all six model specifications are remarkably similar when using preferences of income groups defined by state-specific thresholds. In panel (B) we no longer divide the population into three equally sized income groups. Instead, we restrict the low-income group to only those below the 20th percentile of the (district-specific) income distribution. Similarly, we classified as high income only those above the 80th percentile. Our

resulting estimates for the union-responsiveness marginal effects are slightly smaller, but still of a substantively relevant magnitude and statistically different from zero.

D. Measures of District Organizational Capacity

In the empirical analysis reported in this Appendix, we use the number of religious congregations as another proxy for associational life, complementing the social capital index used in the main text. In a previous version of the paper, we also use certification elections as a proxy for unions' mobilization capacity. Here we provide some background and explain in more detail how we calculate both variables.

Congregations   As one proxy for district level social capital we use the number of congregations per inhabitant. The number of congregations in a given district is not readily available for the years covered in our study. Therefore, we spatially aggregate county-level measures from the 2010 Religious Congregations and Membership Study to the congressional district level using areal interpolation techniques that take into account the population distribution between counties and districts. We use a geographic country-to-district equivalence file calculated from Census shapefiles. This is combined with population weights for each country-district intersection derived using the Master Area Block Level Equivalency database v1.3.3 (available from the Missouri Census Data Center), which calculates them based on about 5.3 million Census blocks. With these weights in hand we can interpolate county-level to district-level congregation counts using weighted means (for states with at-large districts, this reduces to a simple summation, as counties are perfectly nested within districts).

NLRB certification elections   In a previous version of the paper, we also used union certification elections as a proxy for workers' capability to organize for collective action. As has been pointed out by readers and discussants, one concern with this variable is that it may be driven to a significant extent by the existing stock of local unions, as unionization requires people and resources. While it may be useful to distinguish realized union membership from unionization effort and our results are robust to accounting for NLRB elections, in line with the suggestions we dropped this robustness test. For completeness and consistency, we document the construction of the measure.

The formation of unions is regulated by the National Labor Relations Act (NLRB) enacted in 1935 (see Budd 2018: ch. 6). A successful union organization process usually requires an absolute majority of employees voting for the proposed union in a certification election held under the guidelines of the NLRB. Getting the NLRB to conduct an election requires that there is sufficient

interest among employees in an appropriate bargaining unit to be represented by a union. For proof of sufficient interest, the NLRB requires that at least 30% of employees sign an authorization card stating they authorize a particular union to represent them for the purpose of collective bargaining. Building support and collecting the required signatures takes organizational effort. For workers, unionization has features of a public good. Everybody may gain through better conditions from collective bargaining, but contributing to the organizational drive is costly for each individual. Beyond mere opportunity costs, there also is a non-zero risk of being (illegally) fired by the employer for those especially active. If more than 50% of employees sign authorization cards, then the union can request voluntary recognition without a certification election. However, the employer has the right to deny this, in which case a certification election is held. In his labor relations textbook, Budd (2018: 199) notes that voluntary card check recognition is "the exception rather than the norm because employers typically refuse to recognize unions voluntarily."

We use the NLRB's database on election reports to extract all attempts to certify (or de-certify) a local union. They are available from `www.nlrb.gov`. Each database entry is a vote concerning a bargaining unit; the average unit size is 25 employees. There are about 2200 elections each year. Each individual case file usually provides address information on the employer and the site where the election was held. Using this information, we geocode each individual case report and locate it in a congressional district.

E. Additional Robustness Tests

In this section we describe several additional robustness tests.

Redistricting   First, we address the fact that several cases of court-ordered redistricting in Georgia and Texas lead to inter-Census changes in district boundaries. We exclude both states in specification (1) of Table E.1 and find our results unchanged.

Alternative measures of social capital   Next we consider two alternative measures of social capital. First, the number of bowling alleys in an area (Putnam 2000). As the social capital index used in the main text, this variables comes from Rupasingha and Goetz (2008) and was spatially reweighted to districts. Second, the number of congregations per inhabitant. The construction of this measure is explained above (Appendix D). These two measures are less likely than the social capital index to be the result of unionization. We find that these changes in measurement do not qualitatively alter our findings. Unsurprisingly, the estimated union effects are somewhat larger than in the specification adjusting for the social capital index.

1:1 mapping of CCES preferences to roll calls　We begin by limiting our sample by creating a unique mapping between preferences and roll call votes. Some of our CCEs preferences estimates are linked to more than one Congressional roll call. To investigate if this affects our results, specification (3) uses a 1:1 map dropping additionally available roll calls after the first match. This reduces the sample size to 11,104 respondents. We find that our results are not influenced by this change.

TABLE E.1
Additional robustness tests.

	Low income preferences		High income preferences		N
(1) Redistricting	0.067	(0.014)	−0.051	(0.013)	12,784
(2a) Social capital: bowling alleys	0.070	(0.015)	−0.051	(0.014)	14,153
(2b) Social capital: churches	0.072	(0.015)	−-0.051	(0.014)	14,282
(3) Injective preference roll call map	0.063	(0.013)	−0.041	(0.013)	11,104
(4) Extreme preferences excl.	0.074	(0.016)	−0.048	(0.015)	13,308
(5) New York excluded	0.070	(0.015)	−0.048	(0.014)	14,730
(6) Local Union Concentration	0.065	(0.014)	−0.047	(0.014)	15,780
(7) Trimmed LPM estimator	0.074	(0.015)	−0.055	(0.014)	15,426
(8) Errors-in-variables	0.062	(0.004)	−0.054	(0.004)	15,345

Note: Based on specification (5) of Table I. (7) uses trimmed estimator of Horrace and Oaxaca (2006). Specification (8) shows results from an errors-in-variables model implemented in a Bayesian framework (entries are posterior means and standard deviations). See text for details.

Extreme preferences excluded　In specification (4) we investigate if extreme district preferences on some roll calls drive our results. To do so, we trim the distribution of preferences at the bottom and the top. For each roll call we exclude districts with preference estimates below the 5th and above the 95th percentile. Using only trimmed preferences has no appreciable impact on our estimates.

New York excluded　Another test estimates our model with the state of New York excluded from the sample. While Becher et al. (2018) found that LM form estimates of union strength correlate highly with aggregated state-level estimates derived from the Current Population survey, they note that this correlation is lower for New York. In specification (5) we thus show that our results are not affected by its exclusion.

Union Concentration　Our data on local unions are from Becher et al. (2018), who also find that the local concentration of unions is an important dimension. While Becher et al. (2018) show that both dimensions (membership and concentration) vary independently, it is prudent to check if

our results on the impact of union membership on representation still obtain when accounting for the structure of union organization. In specification (6) we show this to be the case.

Trimmed LPM estimator A seventh, more technical, specification implements the trimmed estimator suggested by Horrace and Oaxaca (2006). It accounts for the fact that we estimate a linear probability model to a binary dependent variable, which entails the possibility that the model-implied linear predictor lies outside the unit interval. Our results in Table E.1 indicate that this change does not materially affect our core results (if anything, they become slightly larger).

Errors-in-variables Our final test accounts for the errors-in-variables problem caused by the fact that our district preference measures are based on estimates. While, in general, standard errors for our district-level estimates are quite small relative to the quantity being measured and one expects a downward bias in parameter estimates in a linear model with errors-in-variables, we estimate this specification to get a sense of the quantitative magnitude of the change in parameter estimates.[7] We find that adjusting for measurement error produces very little quantitative change; both estimates are within the confidence bounds of our non-corrected estimates.

F. Post-Double-Selection Estimator

The post-double-selection models in the main text provide a relaxation of the linearity and exogeneity assumptions made in our main model. To do so we use the double-post-selection estimator proposed by Belloni et al. (Belloni et al. 2013, 2017). Specifically, this model setup aims to reduce the possible impact of omitted variable bias by accounting for a large number of confounders in the most flexible way possible. This can be achieved by moving beyond restricting confounders to be linear and additive, and instead considering a flexible, unrestricted (non-parametric) function. This leads to the formulation of the following partially linear model (Robinson 1988) equation (for ease of exposition we omit district fixed effects in the notation and ignore i subscripts):

$$y_{jd} = \mu^l \theta_{jd}^l + \mu^h \theta_{jd}^h + \eta^l U_d \theta_{jd}^l + \eta^h U_d \theta_{jd}^h + g(Z_d) + \epsilon_{jd} \tag{F.1}$$

---

[7]We implement this model in a Bayesian framework, where we incorporate the measurement error model directly into the posterior distribution. To specify the variance of the measurement error for low and high income group preferences, we average the standard errors of the district-group means from the raw CCES data (pre-Census matching). Measurement error variance is slightly larger for low income preferences (0.029) than for high income preferences (0.025). We use the setup proposed in Richardson and Gilks (1993), implemented in Stan (v.2.17.0) and estimated (due to the size of our data set) using mean field variational inference. We use normal priors with mean zero and standard deviation (SD) of 100 for all regression coefficients, and inverse Gamma priors with shape and scale 0.01 for residuals. In the measurement error equation, we use normal priors with mean zero and SD of 10 for the mean of the measurement error and a student-t prior with 3 degrees of freedom and mean 1, SD 10 for the standard deviation of the measurement. The reported entries are posterior means and standard deviations.

with $E(\epsilon_{jd}|Z_s, U_d, \theta_{jd}) = 0$. Here, y is the vote of a representative in a given district, U_d is the level of union density. The function $g(Z_d)$ captures the possibly high-dimensional and nonlinear influence of confounders (interacted with income group preferences). The utility of this specification as a robustness tests stems from the fact that it imposes no a priori restriction on the functional form of confounding variables. A second key ingredient in a model capturing biases due to omitted variables is the relationship between the treatment (union density) and confounders. Therefore, we consider the following auxiliary treatment equation

$$U_d = m(Z_d) + v_i, \quad E(v_i|Z_d = 0), \tag{F.2}$$

which relates treatment to covariates Z_d. The function $m(Z_d)$ summarizes the confounding effect that potentially create omitted variable bias if $m \neq 0$, which is to be expected in an observational study such as ours.

The next step is to create approximations to both $g(\cdot)$ and $m(\cdot)$ by including a large number (p) of control terms $w_d = P(Z_d) \in \mathbb{R}^p$. These control terms can be spline transforms of covariates, higher order interaction terms, etc. Even with an initially limited set of variables, the number of control terms can grow large, say $p > 200$. To limit the number of estimated coefficients, we assume that g and m are approximately sparse (Belloni et al. 2013) and can be modeled using s non-zero coefficients (with $s \ll p$) selected using regularization techniques, such as the LASSO (see Tibshirani 1996; see Ratkovic and Tingley 2017 for a recent exposition in a political science context):

$$y_{jd} = \mu^l \theta_{jd}^l + \mu^h \theta_{jd}^h + \eta^l U_d \theta_{jd}^l + \eta^h U_d \theta_{jd}^h + w_d' \beta_{g0} + r_{gd} + \zeta_{jd} \tag{F.3}$$

$$U_d = w_d' \beta_{m0} + r_{mi} + v_d \tag{F.4}$$

Here, r_{gi} and r_{mi} are approximation errors.

However, before proceeding we need to consider the problem that variable selection techniques, such as the LASSO, are intended for prediction, not inference. In fact, a "naive" application of variable selection, where one keeps only the significant w variables in equation (F.3) fails. It relies on perfect model selection and can lead to biased inferences and misleading confidence intervals (see Leeb and Pötscher 2008). Thus, one can re-express the problem as one of prediction by substituting the auxiliary treatment equation (F.4) for D_d in (F.3) yielding a reduced form equation with a composite approximation error (cf. Belloni et al. 2013). Now both equations in the system represent predictive relationships and are thus amenable to high-dimensional selection techniques.

Note that using this dual equation setup is also necessary to guard against variable selection errors. To see this, consider the consequence of applying variable selection techniques to the

outcome equation only. In trying to predict y with w, an algorithm (such as LASSO) will favor variables with large coefficients in $\bar{\beta}_0$ but will ignore those of intermediate impact. However, omitted variables that are strongly related to the treatment, i.e., with large coefficients in β_{m0}, can lead to large omitted variable bias in the estimate of η even when the size of their coefficient in $\bar{\beta}_0$ is moderate. The Post-double selection estimator suggested by Belloni et al. (2013) addresses this problem, by basing selection on *both* reduced form equations. Let $\hat{I}_1$ be the control set selected by LASSO of y_{jd} on w_d in the first predictive equation, and let $\hat{I}_2$ be the control set selected by LASSO of U_d on w_d in the second equation. Then, parameter estimates for the effects of union density and the regularized control set are obtained by OLS estimation of equation (F.1) with the set $\hat{I} = \hat{I}_1 \cup \hat{I}_2$ included as controls (replacing $g(\cdot)$). In our implementation we employ the root-LASSO (Belloni et al. 2011) in each selection step.

This estimator has low bias and yields accurate confidence intervals even under moderate selection mistakes (Belloni and Chernozhukov 2009; Belloni et al. 2014).[8] Responsible for this robustness is the indirect LASSO step selecting the U_d-control set. It finds controls whose omission leads to "large" omitted variable bias and includes them in the model. Any variables that are not included ("omitted") are therefore at most mildly associated to U_d and y_{jd}, which decidedly limits the scope of omitted variable bias (Chernozhukov et al. 2015).

G. NONPARAMETRIC EVIDENCE FOR UNION-PREFERENCES INTERACTION

As discussed in the main text, we want to estimate a specification that makes as little *a priori* assumptions about functional form relationships between variables (including their interactions). Thus, we non-parametrically model $y_{ijd} = f(z)$ with $z = [\theta^l_{jd}, \theta^h_{jd}, U_d, X_d]$ by approximating it via Kernel Regularized Least Squares (Hainmueller and Hazlett 2014), $y = Kc$. Here, K is an $N \times N$ Gaussian Kernel matrix

$$K = \exp\left(\frac{-\|Z_d - z_j\|^2}{\sigma^2}\right) \tag{G.1}$$

with an associated vector of weights c. Intuitively, one can think of KRLS as a local regression method, which predicts the outcome at each covariate point by calculating an optimally weighted sum of locally fitted functions. The KRLS algorithm uses Gaussian kernels centered around an observation. The weights c are chosen to produce the best fit to the data. Since a possibly large number of c values provide (approximately) optimal weights it makes sense to prefer values of c that produce "smoother" function surfaces. This is achieved via regularization by adding a squared

---

[8]For a very general discussion see Belloni et al. (2017).

L2 penalty to the least squares criterion:

$$c^* = \underset{c \in \mathcal{R}^D}{\operatorname{argmin}} \left[(y - Kc)'(y - Kc) + \lambda c' Kc \right], \tag{G.2}$$

which yields an estimator for c as $c^* = (K + \lambda I)^{-1} y$ (see Hainmueller and Hazlett 2014, appendix). This leaves two parameters to be set, σ^2 and λ. Following Hainmueller and Hazlett (2014), we set $\sigma^2 = D$ the number of columns in z and let λ be chosen by minimizing leave-one-out loss.

The benefit of this approach is twofold. First, it allows for an approximation of highly nonlinear and non-additive functional forms (without having to construct non-linear terms as we do in the post-double selection LASSO). Second, it allows us to check if the marginal effects of group preferences changes with levels of union density *without* explicitly specifying this interaction term (and instead learning it from the data). To do the latter one can calculate pointwise partial derivatives of y with respect to a chosen covariate $z^{(d)}$ (Hainmueller and Hazlett 2014: 156). For any given observation j we calculate

$$\widehat{\frac{\partial y}{\partial z_j^{U_d}}} = \frac{-2}{\sigma^2} \sum_i c_i \exp \left(\frac{-\|Z_d - z_j\|^2}{\sigma^2} \right) \left(Z_d^{U_d} - z_j^{U_d} \right). \tag{G.3}$$

These yields as many partial derivatives as there are cases. We apply a thin plate smoother (with parameters chosen via cross-validation) to plot these against district-level union membership in Figure G.1. Perhaps unsurprisingly, we find that the assumption of an exactly linear interaction specification is too restrictive, especially in the case of the preferences of high income constituents.

However, the most noteworthy result clearly is the fact that, using a non-parametric model not including an *a priori* interaction between union membership and preferences, we find clear evidence that union membership moderates the relationship between preferences and legislative voting. For low income constituents, increasing district-level union membership steadily increases the marginal effect of their preferences on legislators' vote choice. Moving from low levels of union membership (at the 25th percentile) to median levels of union membership increase low-income preference responsiveness by about 5 percentage points. An equally sized increase from the median to the 75th percentile increases responsiveness by almost 8 percentage points. We also find similar (albeit weaker) evidence for an interaction between high income group preferences and union membership.
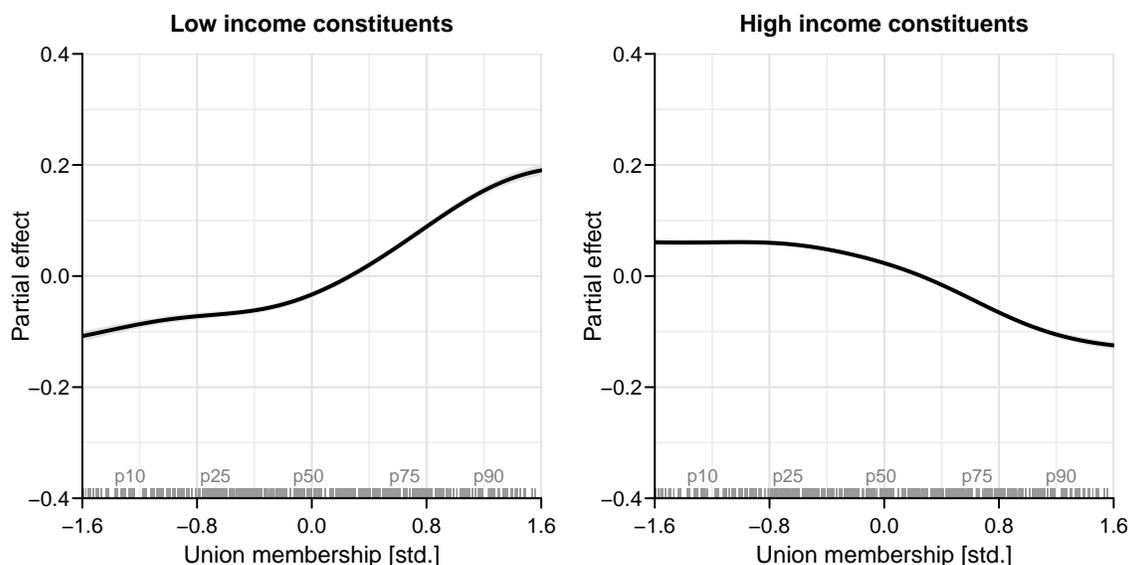
FIGURE G.1

Nonparametric estimate of interaction between union membership and preferences

Note: This figure plots partial effects (summarized using thin-plate spline smoothing) of preferences of low and high income constituents on legislative votes at levels of district union membership. Estimates obtained via KRLS.

H. HETEROGENEITY

Union type Is our finding driven by a particular type of union? A recent strand of research stresses the special characteristics of public unions and their political influence (e.g., Anzia and Moe 2016; Flavin and Hartney 2015). Hence, one may ask whether our findings mainly reflect the influence of private-sector unions since public sector unions are too narrow in their interests to mitigate unequal responsiveness. Panel (A) of Table H.1 provides some evidence on this question. The administrative forms used to measure union membership do not distinguish between private and public unions, and local unions may contain workers from both the private and the public sector. To calculate an approximate measure of district public union membership, we identify unions with public sector members (based on their name) and create separate union membership counts for "public" and the remaining "non-public" unions (see appendix A for details).

Our findings suggests that the coefficient for the impact of a districts' public union membership on the responsiveness of legislators to the preferences of the poor is sizable (at about 7 percentage points) and clearly statistically different from zero. At the same time, the coefficient for the remaining "non-public" unions is slightly reduced. The difference between the two estimates is not statistically distinguishable from zero. This finding does not support the hypothesis of a null-effect

of public sector unions. It also suggests that the changing private-public union composition will not necessarily lead to less collective voice in Congress.

Bill ideology Panel (B) explores whether the effect of unions varies with the ideological direction of the bill that is voted on. Based on the partisan vote margin of the roll call vote, we define an indicator variable for conservative roll calls and estimate separate coefficients for each bill type. We find that union effects are relevant (and significant) for both bill types, they are larger for conservative votes. A standard deviation increase in union membership increases responsiveness to the preferences of low-income constituents by about 9 ($\pm$2) percentage points for conservative bills compared to about 5 ($\pm$1) points for liberal bills. The difference is larger for the preferences of high income constituents. In both cases the difference in marginal effects between liberal and conservative bills is statistically significant. Our findings suggest that union influence is more relevant for bills that have (potentially) adverse consequences for low income constituents. We trace this issue further in the next specification.

TABLE H.1
Effect heterogeneity. Marginal effects of unionization on legislative
responsiveness to low and high income groups.

	Low income	High income
(A) Private vs. Public unions		
Public unions	0.074 (0.016)	−0.058 (0.015)
Non-public unions	0.054 (0.016)	−0.027 (0.016)
(B) Bill ideology		
Conservative bill	0.086 (0.017)	−0.086 (0.018)
Liberal bill	0.052 (0.014)	−0.028 (0.013)
(C) AFL-CIO endorsement		
No position	0.054 (0.014)	−0.054 (0.013)
Endorsement	0.077 (0.015)	−0.040 (0.014)

Note: Estimates for η^L and η^H with cluster-robust standard errors in parentheses. N=15,780. Panel (A) shows separate effects for district counts of union members for unions classified as public or non-public (see text). Statistical tests for the difference in union type yield $p = 0.172$ for low income preferences and $p = 0.027$ for high income ones. Panel (B) estimates separate effects for bills classified as conservative or liberal based on their predominant party vote. Tests for significance of difference: $p = 0.009$ for low and $p = 0.000$ for high income preferences. Panel (C) classifies bills with economic content where the AFLCIO has taken a public stand for or against it (depending on bill content). Tests for significance of difference: $p = 0.003$ for low income, $p = 0.049$ for high income preferences.

Union voting recommendations In panel (C) we consider bills with economic content and that have (or have not) been endorsed explicitly by the largest union confederation, the AFL-CIO. Our

definition of endorsement is based on voting recommendations made publicly by the AFL-CIO.[9] AFL-CIO recommendations signal the salience of the issue to unions, and they were made for more than half of the votes in the analysis. Panel (C) shows that the impact of union membership on legislators' responsiveness for bills especially relevant to low-income citizens is about 2 percentage points larger for votes on which the AFL-CIO had taken a prior position. This difference is statistically different from zero ($p = 0.003$).[10] The fact that districts with higher union membership see better representation of the less affluent more so when issues are salient to unions bolsters the interpretation that our main result is actually driven by unions' capacity for political action. This finding is also consistent with micro-level studies of the effects of union position-taking (Ahlquist et al. 2014; Kim and Margalit 2017).

REFERENCES

Ahlquist, J. S., A. B. Clayton, and M. Levi (2014). Provoking preferences: Unionization, trade policy, and the ilwu puzzle. *International Organization 68*(1), 33–75.

Anzia, S. F. and T. M. Moe (2016). Do politicians use policy to make politics? the case of public-sector labor laws. *American Political Science Review 110*(4), 763–777.

Becher, M., D. Stegmueller, and K. Kaeppner (2018). Local union organization and law making in the us congress. *Journal of Politics 80*(2), 39–554.

Belloni, A. and V. Chernozhukov (2009). Least squares after model selection in high-dimensional sparse models. *Bernoulli 19*(2), 521–547.

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica 85*(1), 233–298.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection amongst high-dimensional controls. *Review of Economic Studies 81*, 608–650.

Belloni, A., V. Chernozhukov, and C. B. Hansen (2013). Inference for high-dimensional sparse econometric models. In D. Acemoglu, M. Arellano, and E. Dekel (Eds.), *Advances in Economics and Econometrics: Tenth World Congress*, Volume 3, pp. 245–295. Cambridge: Cambridge University Press.

Belloni, A., V. Chernozhukov, and L. Wang (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika 98*(4), 791–806.

---

[9]Taken from the AFL-CIO "legislative scorecard", `https://aflcio.org/what-unions-do/social-economic-justice/advocacy/scorecard`.

[10]For high-income preferences the estimate for η^h is smaller for endorsed bills but still significantly different from zero.

Breiman, L. (2001). Random forests. *Machine Learning 45*(1), 5–32.

Budd, J. W. (2018). *Labor Relations: Striking a Balance* (5 ed.). New York, NY: McGraw-Hill Education.

Chernozhukov, V., C. Hansen, and M. Spindler (2015). Valid post-selection and post-regularization inference: An elementary, general approach. *Annual Review of Economics 7*(1), 649–688.

Chung, Y., S. Rabe-Hesketh, V. Dorie, A. Gelman, and J. Liu (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika 78*(4), 685–709.

Flavin, P. and M. T. Hartney (2015). When government subsidizes its own: Collective bargaining laws as agents of political mobilization. *American Journal of Political Science 59*(4), 896–911.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel / Hierarchical Models*. Cambridge University Press.

Gelman, A. and T. C. Little (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodologist 23*, 127–135.

Gelman, A., H. S. Stern, J. B. Carlin, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian data analysis* (Third ed.). Boca Raton: CRC Press.

Hainmueller, J. and C. Hazlett (2014). Kernel regularized least squares: Reducing misspecification bias with a flexible and interpretable machine learning approach. *Political Analysis 22*(2), 143–168.

Honaker, J. and E. Plutzer (2016). Small area estimation with multiple overimputation. Manuscript. [`http://hona.kr/papers/files/smallAreaEstimation.pdf`].

Horrace, W. C. and R. L. Oaxaca (2006). Results on the bias and inconsistency of ordinary least squares for the linear probability model. *Economics Letters 90*, 321–327.

Hout, M. (2004). Getting the most out of the GSS income measures. GSS Methodological Report 101.

Kim, S. E. and Y. Margalit (2017). Informed preferences? the impact of unions on workers' policy views. *American Journal of Political Science 61*, 728–743.

Kopczuk, W., E. Saez, and J. Song (2010). Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937. *Quarterly Journal of Economics 125*(1), 91–128.

Lax, J. R. and J. H. Phillips (2009). How should we estimate public opinion in the states? *American Journal of Political Science 53*(1), 107–121.

Lax, J. R. and J. H. Phillips (2013). How should we estimate sub-national opinion using mrp? preliminary findings and recommendations. Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago.

Leeb, H. and B. M. Pötscher (2008). Can one estimate the unconditional distribution of post-model-selection estimators? *Econometric Theory 24*(2), 338–376.

Park, D. K., A. Gelman, and J. Bafumi (2006). State-level opinions from national surveys: Poststratification using multilevel logistic regression. In J. E. Cohen (Ed.), *Public opinion in state politics*, pp. 209–28. Stanford: Stanford University Press.

Putnam, R. (2000). *Bowling Alone. The collapse and revival of american community*. New York: Simon and Schuster.

Ratkovic, M. and D. Tingley (2017). Sparse estimation and uncertainty with application to subgroup analysis. *Political Analysis 25*(1), 1–40.

Richardson, S. and W. R. Gilks (1993). A bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology 138*(6), 430–442.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica 56*(4), 931–954.

Rupasingha, A. and S. J. Goetz (2008). US county-level social capital data, 1990-2005. The northeast regional center for rural development, Penn State University, University Park, PA.

Stekhoven, D. J. and P. Bühlmann (2011). Missforest. non-parametric missing value imputation for mixed-type data. *Bioinformatics 28*(1), 112–118.

Tang, F. and H. Ishwaran (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal 10*, 363–377.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B 58*(1), 267–288.

Torrieri, N., ACSO, DSSD, and SEHSD Program Staff (2014). American community survey design and methodology. United States Census Bureau. [`www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html`].