Social Networks and Voter Information

Victoria Mooers

October 10, 2024

(Preliminary and Incomplete)

Click here for latest version.

Abstract

Informed voters are essential for government accountability, and social networks are an important avenue through which voters acquire political information. However, U.S. congressional districts do not need to align with social networks, potentially impacting how easily voters learn about their representatives. I study whether the alignment between district boundaries and social networks affects voter knowledge, turnout, and campaign contributions in congressional elections. Using Facebook's Social Connectedness Index and an event study design, I find that an increase in the share of friends living in the same district increases voters' knowledge about their representative. For example, a 10 percentage point increase in this share raises the probability that a voter knows their representative's party by 3.3 percentage points, a 5% increase over the mean. Additionally, a higher share of friends in the same district decreases voter abstention. I use a model of information diffusion to simulate the share of informed voters under counterfactual district maps, creating a framework to evaluate the informational effects of proposed maps. These findings suggest that aligning political boundaries with social networks can enhance democratic engagement.

1 Introduction

Every ten years, the United States redraws its congressional district boundaries, which determine the constituents for each member of the U.S. House of Representatives. While the U.S. Constitution mandates that these districts have roughly equal populations, it places few other restrictions on how these lines are drawn. Literatures spanning social sciences, mathematics, and computer science have debated how to measure the fairness of these districts (e.g., Stephanopoulos and McGhee 2015). But the proposed measurements typically assume that changes in the district boundaries do not change voter turnout: in a given spot on the map, the expected number of voters from each political party is fixed—regardless of who we put them in a district with. However, voters do not act in isolation—they learn about politics from their social networks. For example, studies have shown that one-on-one interactions (e.g., Nickerson 2008) and social media (e.g., Enikolopov, Makarin, et al. 2020) can shape political behaviors, from election turnout to protest participation.

How does the alignment between social networks and political boundaries impact voters' information and behavior? Voters who live in the same district as a larger share of their friends may be more likely to hear about their representative through their network; in the aggregate, this could lead to sizable differences in voter knowledge between areas where social networks are more or less reflected by district boundaries. Additionally, more informed constituents may be more likely to vote (e.g., Snyder and Strömberg 2010), suggesting that changing district borders could also change who turns out to vote. Without understanding this relationship, our measures of gerrymandering and electoral fairness remain incomplete, overlooking the role of social networks in shaping political knowledge and participation.

In this paper, I estimate the causal impact of the county-level share of same-district friends (which I refer to as "congruence") on voters' knowledge, turnout, and campaign contributions in U.S. House of Representatives elections. Using an event study design, I leverage changes in congruence resulting from redistricting, which I demonstrate are plausibly exogenous. Next, I develop a model of political information diffusion within districts, and I use my reduced-form estimates to estimate the model's unknown parameter. With this model, I then simulate county-level shares of informed voters under counterfactual congressional district maps. These simulations provide a framework for evaluating the informational consequences of alternative maps.

I find that voters are more informed when their social networks better align with their congressional districts. For example, I find that a 10 percentage point (slightly less than one standard deviation) increase in congruence raises the probability that a voter knows their representative's party by 3.3 percentage points, from a mean of 62% (a 5% increase).[1] The same increase in congruence raises the probability that a voter recognizes their representative's name by 0.7 percentage points, from a mean of 93%. By Snyder and Strömberg 2010's estimates, this is equivalent to the effect of 15 additional newspaper articles about the representative (from a mean of 110 articles per congressional term). I find no effect on knowledge of governors

[1] During the period I study, Americans had on average 245 Facebook friends (Keith N. Hampton et al. 2012) and were estimated to have on average 600 acquaintances overall (McCormick et al. 2010), though only 10-25 people they trust (DiPrete et al. 2011). A 10 percentage point increase in friends is then, on average, an increase of about 25 Facebook friends or 60 acquaintances.

and senators—placebo outcomes, since these statewide offices are unaffected by district borders—suggesting that the increase in House knowledge does not reduce attention to other elected officials.

I also find that congruence increases engagement in House elections, in terms of both turnout and campaign contributions. A 10 percentage point increase in congruence raises turnout in House elections (relative to turnout in the top-of-ticket election) by 0.4 percentage point. equivalent to 97 additional newspaper articles, by Snyder and Strömberg 2010's estimates. For the full voting age population, this implies that for every four additional people who hear about their representative through their network, one more votes in the House election. Additionally, a 10 percentage point increase in congruence raises the share of campaign contributions to in-district House candidates (as a share of donations to all House candidates) by 7.4 percentage points, from a mean of 50%.

This paper makes three primary contributions:

First, I provide the first causal estimates of how the match between social networks and political boundaries affects voters' political knowledge and behavior. A large literature on government accountability emphasizes the importance of informed voters in government oversight, with informed voters receiving more public spending. This literature has focused on traditional media – such as TV, newspapers, and radio – as key sources of political information (e.g., Strömberg 2004, Eisensee and Strömberg 2007, Ferraz and Finan 2008). A related literature examines how the internet and social media heighten responsiveness to government effectiveness by increasing voters' access to information and easing coordination (e.g., Manacorda and Tesei 2020, Guriev et al. 2021; but Falck et al. 2014). I build on this work by highlighting the role of social networks themselves as a key information source for voters. My findings demonstrate that the alignment between social networks and political boundaries can create significant differences in voter knowledge levels; while my results are specific to the U.S. context, this mechanism may have implications for any setting where political boundaries are drawn.

Second, my design quantifies the aggregate impact of peer effects on voter knowledge and behavior in the United States, providing a middle ground between existing approaches. Studies that measure one-on-one interactions (e.g., Nickerson 2008) capture first order peer effects, but do not account for how these effects propagate through social networks (as shown in other contexts, e.g., Dahl et al. 2014), limiting their applicability for predicting the effects of large-scale policies like redistricting. Conversely, recent work estimating the causal effects of place on political behavior (Cantoni and Pons 2022, Brown et al. 2023) finds that the state voters live in explains much of the variation in voter turnout, but these estimates combine aggregate peer effects with the effects of state institutions (such as same-day registration or voter

ID laws). My design isolates the role of peer effects, independent of state-level institutional factors, making it particularly useful for understanding policies like redistricting. Other studies capture aggregate peer effects on voter information and behavior, though in different contexts. For example, Fafchamps et al. 2019 and Arias et al. 2019 examine peer effects within villages or precincts in Mozambique and rural Mexico, respectively, while Bond et al. 2012 analyzes the national-level impact of a Facebook Election Day reminder on U.S. voter turnout. My estimates, however, account for the geography of the U.S. social network, making this method especially relevant for redistricting. Moreover, because it relies on publicly available data, this approach can be applied to other contexts where political boundaries are redrawn.

Third, I contribute to the literature on measuring gerrymandering by developing a model of information diffusion within districts and using it to simulate the informational consequences of alternative congressional district maps. Existing measures of gerrymandering (e.g., Polsby and Popper 1991, McCartan, Kenny, Simko, Ebowe, et al. 2024) or models of strategic gerrymandering (e.g., Friedman and Holden 2008, Gul and Pesendorfer 2010, Kolotilin and Wolitzky 2020) often assume that changes to district boundaries do not affect the distribution of partisans in a given area. (However, Bouton et al. 2023 develop a model of strategic gerrymandering that accounts for turnout differences, and allow for voters' turnout decisions to endogenously respond to the turnout decisions of others in their district.) My model and simulations offer a new way to evaluate district maps, as they consider changes in voter behavior driven by how well districts align with social networks. While my simulations are partial equilibrium—they do not, for example, account for candidates' responses—this approach offers a foundation for reassessing the fairness of district maps once informational consequences are considered.

The remainder of this paper is organized as follows. In section 2, the empirical strategy is presented, with a discussion of the construction of congruence, as well as the event study design and tests of identifying assumptions. Section 3 presents the outcomes data. Section 4 presents and discusses the findings on voters' information, voter turnout, and campaign contributions, while section 5 provides robustness checks and a discussion of the limitations of the results. Section 6 outlines the theoretical model of information diffusion within districts, and section 7 explains how the model is estimated. Section 8 presents the simulated outcomes under counterfactual maps and discusses the implications for measures of gerrymandering. Section 9 concludes with policy recommendations and suggestions for future research.

2 Empirical Strategy and Networks Data

2.1 congruence: A Measure of District and Network Alignment

I study the consequences of geographical mismatch of social networks and political boundaries. In particular, I focus on the alignment of county-level social networks and congressional districts. I capture this alignment by constructing "congruence": for a randomly chosen person from the county, and a randomly chosen one of their friends, the probability that both individuals live in the same congressional district.

In order to construct congruence, I use data representing the Facebook friendship graph. I construct congruence for counties in the 48 contiguous U.S. states, and I show that congruence varies substantially across the U.S. I demonstrate that congruence is (not surprisingly) correlated with many determinants of social networks and district borders, but *changes* in congruence due to redistricting largely are not. Consequently, plausibly causal identification of the impacts of congruence can leverage these changes over time, as I discuss further in Section 2.2.

2.1.1 Proxy for Social Networks: Facebook Social Connectedness Index

For data on social networks, I use the Facebook Social Connectedness Index (SCI) – one of the best existing proxies for real-world social networks. Importantly, the SCI is publicly available, allowing researchers to apply my framework (developed in Sections 6-7) to simulate outcomes under alternative maps of their choice.

In essence, the SCI aggregates the Facebook friendship graph to provide a measure of the strength of social connection between two locations (such as counties) (Bailey, Cao, et al. 2018). For each pair of counties, $SCI_{c,k}$ is constructed as the relative probability of a friendship link between users in county c and county k:

$$\text{SCI}_{c,k} = \frac{\text{Friendship Links}_{c,k}}{\text{Facebook Users}_c \times \text{Facebook Users}_k}$$

That is, the SCI is the number of friendship links between the two locations, normalized by the total number of possible connections between them.[2]

I use the SCI for U.S. county-county pairs from the October 2021 snapshot.[3] The SCI is also available for

[2] The SCI is scaled from 1 to 1,000,000,000, areas with particularly small populations are removed, and noise is added to preserve privacy.

[3] There are 3,136 counties in the data, and each county appears in a pair with every other county (including itself). As such, there are 9,834,496 county-county pairs. The SCI only includes users who have interacted with (including simply logging into) any of Meta's apps (Facebook, Instagram, WhatsApp) in the 30 days prior to the snapshot. Locations are assigned based on users' provided information (such as their stated city) and their device connection information.

U.S. zip code-zip code pairs. I focus on county-county pairs to facilitate matching to county-level outcomes data for vote counts and campaign contributions. However, I also construct congruence using the zip code pairs (see Section 5), and I show that the results on voter information and self-reported turnout are robust to this alternative construction.

The SCI is an effective proxy for real-world social networks because it captures social ties that might not be revealed in geography-based proxies, like commuting flows, that rely on physical proximity. The SCI has been demonstrated to closely reflect offline networks (Bailey, Cao, et al. 2018; Bailey, Gupta, et al. 2021; Kuchler et al. 2022). Two features of the Facebook friendship graph aid this: First, it is very persistent over time, because Facebook friendships accumulate throughout a lifetime.[4] Second, Facebook usage rates across counties are uncorrelated with demographics like income (Chetty et al. 2022).[5] Nonetheless, in Section 5.2, I demonstrate that results are qualitatively similar if I use commuting flows as an alternative proxy for social networks.

2.1.2 Construction of congruence from SCI

Whereas the SCI gives the relative probability of a friendship between two counties, congruence is the share of a county's friends that live in the same congressional district as the people in the county. Accordingly, to construct congruence I need to appropriately aggregate the SCI. I do this by using the SCI to construct a matrix of county-county friendship shares, and then for each county summing friendship shares across same-district counties (adjusting for counties that intersect multiple districts).

First, the matrix of county friendship shares can be represented as

$$\Pi' = \begin{pmatrix} \pi_{1,1} & \cdots & \pi_{1,C} \\ \vdots & \ddots & \vdots \\ \pi_{C,1} & \cdots & \pi_{C,C} \end{pmatrix}$$

[4]For example, Enke et al. 2023 report that according to their correspondence with the SCI authors, the correlation between years of the SCI is above 0.99. Additionally, Bailey, Gupta, et al. 2021 finds that countries with higher social connectedness trade more, and that this relationship is similar in every year back to 1980; Kuchler et al. 2022 find that institutional investors are more likely to invest in firms in regions with higher social connectedness to the investor's region, and that this relationship remains at least back to 2007 (the start of their data).

[5]Similarly, at the time of the snapshot, survey-reported Facebook usage rates were also relatively even across demographic groups nationwide (Auxier and Anderson 2021). In particular, in 2021, Facebook usage rates (defined as whether you ever use the platform) among American adults varied slightly between urban and suburban (70%) and rural areas (67%); when sliced by race, income, and education, usage rates varied between 61% and 74%. The largest gaps emerge by age, with the lowest usage rates among 65+ year-olds (50%) and the highest usage rates among 30-49 year-olds (almost 80%); use among 18-29 year-olds reflected the national average at 70%. There are also minimal differences in Facebook usage rates by political party (Vogels et al. 2021). Facebook usage rates rose until 2016, and remained stable at around 70% of U.S. adults from then until at least 2021. As of 2021, Facebook was the social media platform with the least heterogeneity in usage rates by age. While 18-29 year-olds were the heaviest users of all other platforms, their Facebook use was only exceeded by their use of YouTube (95%) and Instagram (71%) (Auxier and Anderson 2021).

where $\mathcal{C} = [1, ..., C]$ is the set of all counties, and $\pi_{c,k}$ is the share of county c's friends that live in county k. $\pi_{c,k}$ is constructed as

$$\pi_{c,k} = \frac{\text{Links}_{c,k}}{\sum_{j \in C} \text{Links}_{c,j}}$$

(from here on referring to "Friendship Links" as "Links" for brevity). We can re-write the equation for $\text{SCI}_{c,k}$ as

$$\text{Links}_{c,k} = \text{SCI}_{c,k} \times \text{Facebook Users}_c \times \text{Facebook Users}_k$$

However, the number of Facebook users in each county is not made available, so this is not possible to directly construct. Bailey, Gupta, et al. 2020 argue that we can substitute the population of an area for the number of Facebook users. This requires the assumption that Facebook usage rates are the same across counties. This assumption is likely benign: as mentioned above, Chetty et al. 2022 demonstrate that while there is some variation in Facebook usage rates across counties, this variation cannot be predicted by demographics.

Replacing the number of Facebook users in a given county with the county's population and re-arranging, we get

$$\pi_{c,k} = \frac{\text{SCI}_{c,k} \times \text{Pop}_k}{\sum_{j \in C} \left(\text{SCI}_{c,j} \times \text{Pop}_j\right)}$$

which is feasible to calculate. I get each county's population from the 2020 Decennial Census.

Second, I adjust for the fact that district borders do not necessarily follow county borders by taking population-weighted averages. Counties can be fully contained within a single district, but they can also intersect multiple districts. For a given county c and district d, I use counts from the 2020 Decennial Census to construct the share of county c's population that lives in district d, represented by $q_{(c,d)}$. The probability that a randomly chosen individual from county c lives in district d is then $q_{(c,d)}$, and if we choose at random one of their friends in county k, the probability that friend is from district d is $q_{(k,d)}$. In turn, for a person living in county c and district d, the share of their friends that live in county k and district d is $\pi_{c,k} \times q_{(k,d)}$; the overall share of their friends that live in district d (regardless of county) is $\sum_{k \in C} \left(\pi_{c,k} \times q_{(k,d)}\right)$. Taking the population-weighted average across all districts county c intersects, represented by $D(c)$, gives congruence:

$$\bar{\pi}_c = \sum_{d \in D(c)} \sum_{k \in \mathcal{C}} \left(\pi_{c,k} \times q_{(c,d)} \times q_{(k,d)}\right)$$

Finally, the SCI is only available for one snapshot of the social network, in 2021. In order to derive congruence in each year over the period, I hold the social network fixed and I re-calculate congruence with each congressional border change – consequently, all changes in congruence are due solely to changes in the location of the district border.[6] As discussed above, social networks are slow-changing, so this assumption is not unreasonable. Further, holding the social network fixed at its 2021 structure and projecting it back in time will primarily introduce measurement error. I construct congruence for each year from 2002 to 2022, i.e. the 107th-117th Congresses.

2.1.3 Examples of congruence

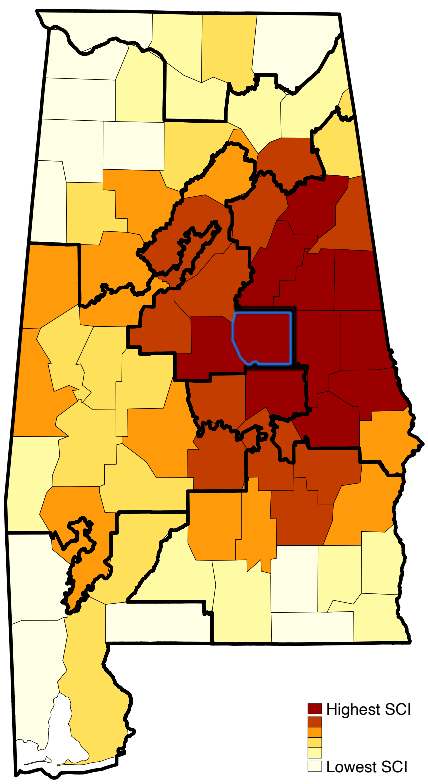
To illustrate the relationship between the SCI and congruence, as well a how congruence can change due to redistricting, consider Coosa County, Alabama, which experienced the biggest increase in congruence of any county following the most recent redistricting, which occurred in 2022 based on population counts in the 2020 Decennial Census.

Figure 1 shows the value of the SCI between Coosa County and each other county in Alabama. Coosa County is highlighted with a blue border. The counties that Coosa has the strongest social connections with are in dark red, while the counties that Coosa is most weakly connected to are in light yellow; there is an equal number of counties in each color bin. The maps only reflect Coosa County's connections to the other counties – they do not reflect how any other two counties are connected to each other. The map on the left displays the congressional district borders in Alabama immediately prior to redistricting (the borders used in the 2020 election), while the map on the right displays the borders immediately following redistricting (the borders used in the 2022 election).

Coosa County is most strongly connected to other counties to its east, while the strength of its connections drops off more quickly going west. Under the 117th Congress borders, Coosa County lies in the southeastern corner of its congressional district, with the district border following the north, east, and south borders of the county; Coosa County only shares a border with another county in its district on its western side, and only one of the counties it is most strongly connected to (darkest red) lies in the same district. Under the 118th Congress borders, Coosa County is moved into the district that had been east of it. Coosa County is still in the corner of the district, but its district border is reflected to the opposite corner, and Coosa County is now grouped into a district with all but two of the counties it is most strongly connected to.

Thus, in the left map, Coosa County is cut off from much of its social network, while in the right map Coosa County is grouped in with much of its social network. This is reflected in Coosa County's congruence

[6] Specifically, I assume that the share of a county's friends in each other county remains the same. To do this, I also assume that the populations are the same as Decennial Census 2020 populations (otherwise changes in population of one county would affect friendship shares for many other counties).



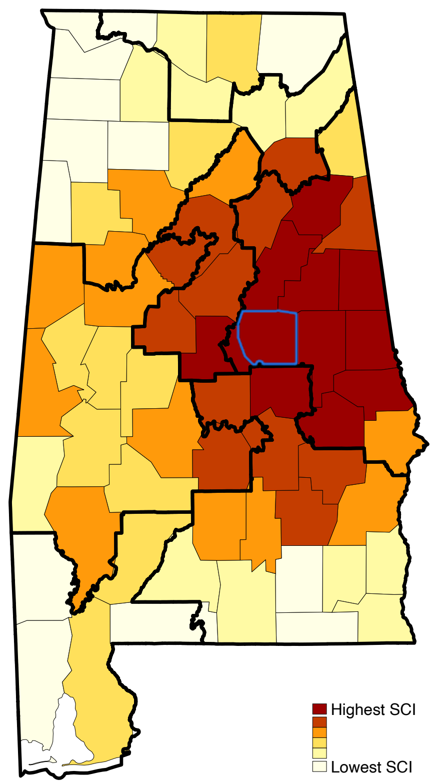


Figure 1: SCI of Coosa County, Alabama. Left: 117th Congress boundaries; Right: 118th Congress boundaries.

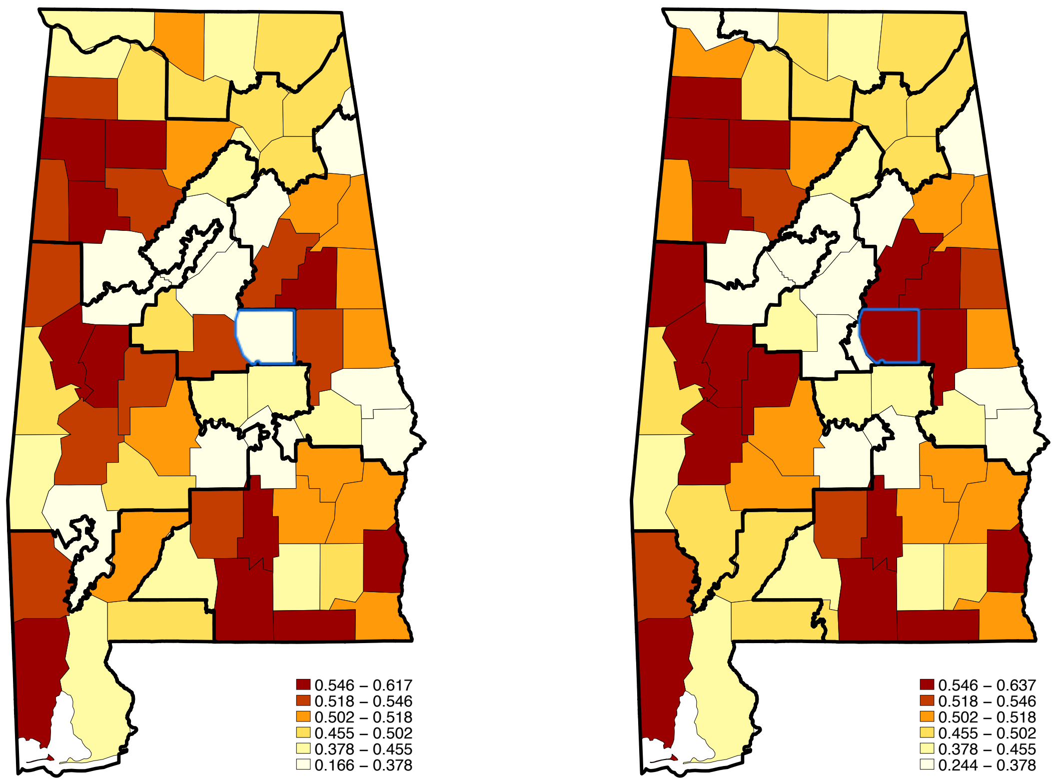


Figure 2: congruence of Alabama Counties. Left: 117th Congress boundaries; Right: 118th Congress boundaries.

before and after redistricting. In Figure 2, the left map represents the congruence of each county of Alabama before the 2022 redistricting, while the right map represents congruence after. Again, there is an equal number of counties in each bin, so congruence levels should be interpreted as congruence relative to other counties in Alabama. As we might predict from the SCI maps, Coosa County has among the lowest levels of congruence in Alabama under the 117th Congress borders. However, under the 118th Congress borders, when it is grouped in with the counties with larger shares of its friends, Coosa County has one of the highest levels of congruence in Alabama. In particular, Coosa County experiences a 39.3pp change in congruence, going from 16.6% under the old borders to 55.9% under the new borders.

2.1.4 Summary Statistics and Predictors of congruence

congruence varies substantially across the continental U.S. congruence is determined by both social networks and district borders; consequently, demographics and geographical features that are correlated with either of these determinants are highly correlated with congruence.

Among the continental 48 states over the full period, mean congruence is 41% with a standard deviation of 14pp; minimum congruence is 2% and maximum is 87%, while the 1st percentile is 8% and the 99th percentile is 67%. The middle 50% of counties have congruence between 32% and 51%, and the middle 80% of counties have congruence between 22% and 58%.[7]

Appendix Tables 6 and 7 summarize how various geographic and demographic features correlate with congruence, separately in 2010 and 2020. As might be expected given that social networks tend to follow state boundaries (Bailey, Cao, et al. 2018), counties in single district states[8] have higher congruence on average (53%). Additionally, due to the restriction that each congressional district within a state represent roughly the same population (across states averaging about 760,000 in 2020, see Eckman 2021; Whitaker 2017), counties with large populations (including most urban areas) are more likely to be split by a district boundary in order to accommodate this constraint; a one percent increase in county population is associated with a 0.05pp decrease in congruence.

The other determinant of congruence is the geography of social networks. The biggest predictor of social ties is distance (Bailey, Cao, et al. 2018), so counties that are further from a congressional district border will generally have higher congruence. Naturally, this is more likely to occur in geographically large districts, which are necessarily in areas with lower population density (again because each district is meant to have roughly the same population). This leads to higher congruence in rural areas. Simultaneously, urban areas have much more geographically dispersed social networks, because they have strong ties to other urban centers around the country (Bailey, Cao, et al. 2018); this further drives down congruence in urban areas.

2.1.5 Changes in congruence over Time

How is congruence changing over time? I examine changes in congruence due to redistricting following the 2010 and 2020 Decennial Censuses. Recall that I re-calculate congruence for each year by holding the social network fixed. I then calculate the change in congruence for a county following redistricting. In both years, the average change in congruence is nearly zero – 0.1pp in 2013, 0.3pp in 2023 – with a standard deviation of 6.5pp and 5.9pp respectively. In 2013, the biggest drop in congruence was by 36.6pp, while the biggest increase was by 32.7pp. In 2023, the biggest drop was by 31.0pp and the biggest increase was by 39.3pp. Following the 2010 Census, 500 counties (16%) experienced nearly zero change in congruence (specifically, an

[7] These statistics are roughly stable over the full period. Throughout this paper, I focus on the 48 contiguous states – i.e., in results I exclude Alaska, Hawaii, Washington, D.C., and territories. However, all counties as well as foreign friendships are included for calculating the scaled total number of friends (denominator) for each county.

[8] In 2022, these were Delaware, North Dakota, South Dakota, Vermont, and Wyoming (as well as Alaska, which is not included in my analysis). These five states contain 159 counties, or 5% of all counties in the data.

absolute change of less than 0.1pp), and following the 2020 Census, 467 counties (15%) experienced nearly no change in congruence. Thus, most counties experience some change in congruence, but very large changes are unusual.

In Appendix Tables 10 and 11, I show the predictors of changes in congruence. Most county characteristics are not correlated with changes in congruence, and most correlations that do exist disappear once media market and congressional district fixed effects are included (which I will do in my preferred specifications, discussed below). When media market and congressional district fixed effects are included, the share of the population that is white and non-Hispanic is positively correlated with changes in congruence and the share with income below the poverty line is negatively correlated. Accordingly, I control for these characteristics in my regressions.

2.2 Redistricting

Congruence measures variation in the match between social networks and congressional district boundaries, but it is not itself exogenous. Congruence is correlated with factors that determine district boundaries, factors that determine social networks, and sociodemographic characteristics. Accordingly, in order to have plausibly exogenous variation in congruence, I need to control for these factors, especially when they are possibly correlated with outcomes of interest.

In order to capture plausibly exogenous variation in congruence, I measure the impact on outcomes of a change in congruence due to congressional redistricting. I use an event study design, focusing on the redistricting that followed the 2010 Census. Focusing on a single redistricting event allows me to avoid concerns related to staggered treatment events, and also allows for a visual test of pre-trends in changes in congruence. The Census was conducted in April 2010, and states needed to draw new congressional district borders in time for the November 2012 elections. Accordingly, the congressional representatives first elected under the new borders assumed office in January 2013. As such, the last year before the treatment (i.e., a change in congruence) will consequently depend on the outcome. For outcomes that relate to the *current* representative, 2012 is the last year before treatment. For outcomes that relate to the upcoming election (therefore more related to the *next* representative), 2011 is the last year before treatment (or more commonly 2010, for outcomes only available in even years).

Assuming 2012 as the last year before treatment, the event studies accordingly take the following form:

$$y_{ict} = \lambda_t + \sum_{\tau=2006}^{\tau=2010} \beta_\tau \Delta \text{Congruence}_c \mathbb{I}(\tau = t) + \sum_{\tau=2014}^{\tau=2022} \beta_\tau \Delta \text{Congruence}_c \mathbb{I}(\tau = t) + X_{ct}\delta + Z_{ict}\gamma + \varepsilon_{ict} \quad (1)$$

where where y_{ict} is the outcome for a given individual i in county c in year t, $\Delta\text{Congruence}_c$ is the change in congruence experienced by county c between 2012 and 2013, λ_t are year fixed effects, X_{ct} is a vector of county-by-year controls (to further adjust for things like changing demographics over time), and Z_{ict} is a vector of individual controls. Errors ε_{ict} are clustered at the county level.

I can additionally include district-by-year fixed effects. This can be thought of as controlling for House election-specific factors that impact outcomes for all counties in the district. These can include characteristics of each of the candidates, scandals, national attention, levels of fundraising and campaign spending, etc.

Another concern may be that social networks may be highly correlated with media markets, and consequently congruence actually just reflects the impacts of TV and radio news or political advertisements bought at the media market level. To address this concern, I use the boundaries of the Nielsen Designated Market Areas and include DMA-by-year fixed effects.

Lastly, I control for partisan biases in network connections by constructing each county's exposure to Democrats. For each county, I multiply the share of the county's friends in each other county by the Democratic vote share in the county in the most recent presidential election; I then sum this across all counties the given county is connected to. In essence, this forms a rough approximation of the share of a county's friends that voted Democratic.

3 Outcomes Data

I study the impact of congruence on voters' knowledge and political behavior. I begin with survey data to study voters' knowledge of their representatives and their self-reported vote choices and candidate preferences. I then incorporate vote count data to reveal actual voting behavior, as well as data on campaign contributions to understand impacts on donation behavior.

3.1 Voters' Information

I test whether voters in more congruent counties are more informed by using responses in the Cooperative Election Study (CES) (formerly the Cooperative Congressional Election Study, or CCES; see for example

Schaffner, Ansolabehere, and Shih 2023) to measure voters' familiarity with their representatives.

The CES is a nationally representative survey that has run annually from 2006 to 2022 and ask about topics including demographics, political attitudes, political knowledge, and voting intentions and choices. In federal election years (i.e., all even years), a pre-election survey is conducted from late September to late October, and a post-election survey is conducted in November. In non-federal election years (i.e., all odd years) a single survey is conducted in the fall. I use the pre-election surveys (or single surveys in odd years) for 2006-2022. The CES sample consists of 50,000+ adults in every federal election year since 2010 (>30,000 in 2006 and 2008) and 10,000+ adults in every odd year. I use the CES's cumulative weights, which re-weight observations to make sample sizes comparable across years (see Kuriwaki 2018). The CES includes each respondent's county and congressional district, enabling me to link respondents to county-level congruence measures and to observe responses to questions about each respondent's own representative.

I construct three binary variables to assess how familiar respondents are with their current representative. Detailed descriptions of these variables are in Appendix Table 2. Respondents are asked to "Please indicate whether you've heard of this person and if so which party he or she is affiliated with...". They are asked this about their current House representative, both of their senators, and their governor. Respondents can answer "Never Heard of Person", "Republican", "Democrat", "Other Party/Independent", or "Not Sure". The first dummy variable, "Heard of Incumbent", is coded as 0 if the respondent answered "Never Heard of Person" and 1 otherwise.[9] This variable captures whether the respondent claims to have any familiarity with their representative at all: do they even recognize the name? The second dummy variable, "Selected Party", is coded as 0 if the respondent answered "Never Heard of Person" *or* "Not Sure", and 1 otherwise. This variable indicates whether, beyond recognizing the representative, the respondent claims to have some knowledge about them: they claim to know the party the representative belongs to (though they may just be guessing). Lastly, the third dummy variable, "Selected Correct Party", is coded as 1 if the respondent selected the correct party for the incumbent and 0 otherwise. While lucky guesses cannot be ruled out, this variable generally indicates that the respondent at least knows enough about their representative to know what party their representative belongs to. Appendix Table 3 shows that, as expected, fewer people select their representative's party (68.6%) than claim to have heard of them (93.2%), and fewer still select the correct party (61.6%—though, among those who select a party, the overwhelming majority select the correct party).

[9] In 2006, 2007, and 2009, respondents do not have the option to say "Never Heard of Person" and instead can only say "Not Sure." Consequently, I drop these years in regressions using the "Heard of Incumbent" variable. Note that 2008, 2010, and 2011 still provide observations of this variable prior to the redistricting that follows the 2010 Decennial Census, because the new districts first apply in 2012.

As when constructing the congruence measure, I only include respondents in the 48 contiguous states.[10] Additionally, not all counties are represented in every year; in even years, there is at least one respondent from 80-90% of counties, while in odd years about two-thirds of counties have at least one respondent. Because the weighted sample is representative of people living in the U.S. (rather than of U.S. counties) and more people live in urban areas (which tend to have lower congruence), the average respondent's county congruence is slightly lower at 37% (compared to 41% for the average county).[11]

3.2 Voter Turnout

I test impacts of congruence on voter turnout and vote shares using both survey responses in the CES as well as county-level vote count data from Dave Leip's Election Atlas.

I begin with CES survey responses in order to study impacts on voting within the same sample as the information outcomes. The pre-election surveys ask respondents questions about their voting intentions (e.g., who they prefer among candidates running), while the post-election survey asks respondents about who they ended up voting for. I use both the pre-election survey and the post-election survey: while the post-election survey asks about actual vote choices, outcomes from the pre-election survey utilize the same sample as the information outcomes (because there is some attrition between surveys).

Next, I use county-level vote counts from Dave Leip's Election Atlas to measure the impacts of congruence on actual voting outcomes. I use the period spanning 2002-2020; 2002 is the first election under the district boundaries that are in place through the 2010 election, and 2020 is the last election under the district boundaries that are first used in the 2012 election.

In order to include district-by-year fixed effects, I take two approaches. First, when using this county-level vote count data, I include only counties that are in a single congressional district, or I only link the county to the congressional district that a majority of its population is in. Second, I use other data sources to construct the same voting outcomes at the county-by-district-level, using precinct-level vote count data (see Appendix Section B.2.1 for details). I do not use the precinct counts for my main results as it does not fully span my time period. However, I find qualitatively similar results using this data (described in Section

[10] In these 48 states and across all 17 years of the CES, there are 612,085 respondents (552,307 excluding 2006, 2007, and 2009). I exclude missing responses to the candidate party recognition question (<2% of respondents in each year; for most of these cases, the House candidate name is missing in the survey). When including individual demographic controls, I similarly exclude respondents who did not answer the relevant demographic questions. I also exclude a small number of respondents in 2006 and 2007 that are assigned to counties that are not in their state of residence. Lastly, in the 2020 survey, 925 respondents in North Carolina were assigned to incorrect congressional districts, and consequently were shown the candidate names for the wrong district. I exclude these respondents, since they were not asked about their familiarity with their own representative. See Schaffner, Ansolabehere, and Luks 2021.

[11] The distribution is otherwise similar to the county-level distribution, with a standard deviation of 11pp, a minimum of 9% and a maximum of 74%.

5).

I focus on turnout in the House election relative to turnout in the top-of-ticket election, i.e., the election that is likely to receive the most attention and have the most force in driving voters to show up at the polling booth. I define the "top-of-ticket" election as the Presidential election when it occurs (every four years), and in midterm years as the Senate election (if occurring, which it does for about two-thirds of counties in midterm years), else the Governor election (if occurring, which it does for about a quarter of counties in midterm years).[12] Accounting for the top-of-ticket election helps to further control for factors unrelated to the House election that may drive differences in overall turnout, including differences in the cost of voting. In particular, I construct the difference between the number of votes cast in the top-of-ticket election and the number of votes cast in the House election, as a share of the top-of-ticket votes cast. This measure captures the share of voters who, despite having paid the cost of voting in order to vote in the top-of-ticket race, choose to abstain from the House election. (This is also referred to as "roll-off," the seemingly paradoxical phenomenon that Timothy J Feddersen and Pesendorfer 1996 seek to explain with their model of the "Swing Voter's Curse," and which Miller 2022; Snyder and Strömberg 2010 also study empirically.) Details and descriptive statistics are provided in Appendix Section B.2.2.

3.3 Campaign Contributions

I test impacts of congruence on donation behavior using data on campaign contributions to House candidates from Kuziemko et al. 2023, which is constructed from the Federal Election Commission campaign contribution data in Bonica 2014. For a given contribution, Kuziemko et al. 2023 use geocoding to identify whether the contributor lives in the same congressional district as the House candidate they are donating to. Then, for each Census tract, they construct the aggregate amount of donations to in-district candidates and to out-of-district candidates (both the dollar amount, and the number of contributors). With this data in hand, for each county I construct the share of contributions to in-district candidates, for 2002-2016.

4 Results

Section 4 presents my main results that congruence increases voters' knowledge about their representatives, and accordingly decreases abstention in House elections. I also find that congruence increases shifts donations to same-district House candidates and away from out-of-district candidates.

[12] In each midterm year, there are a few states where neither a Senate election nor a Governor election occurs, but this only applies to about 10% of counties on average.

4.1 Voters

4.1.1 Voters' Knowledge about Representatives

I estimate positive and significant coefficients for the impact of congruence on voters' knowledge about their representatives. Figure 3 presents the event studies for outcomes "Heard of Incumbent," "Selected Party," and "Selected Correct Party" constructed from the CES data. I find that a change in congruence has an immediate and persistent impact on voters' knowledge. I focus on even years of the CES survey until 2022 (the last year before the next national redistricting event).[13] The β and dashed line on the figures indicate the estimated aggregate effect, from the specification of the form:

$$y_{ict} = \alpha + \beta_1 \Delta\text{Congruence}_c + \beta_2 \mathbb{I}(t > 2012) + \beta_3 \Delta\text{Congruence}_c \times \mathbb{I}(t > 2012) + X_{ct}\delta + Z_{ict}\gamma + \varepsilon_{ict}$$

In the event studies shown, I include district-by-year fixed effects, DMA-by-year fixed effects, and individual demographic controls from the CES; however, adding the fixed effects and controls beyond the district-by-year fixed effects makes little difference. Further, results are similar when county fixed effects are included. Congruence is measured on a scale from 0 to 1, and outcome variables are binary. As such, reported estimates give the change in probability of the outcome (measured between 0 and 1) that would result from a 0 to 1 change in congruence.

The event studies show that the change in voter knowledge due to changes in congruence in redistricting most strongly takes effect in the first survey after redistricting (2014). Impacts are relatively stable over time. The otherwise stability of the estimates may be attributable to an attention story: only voters who have a high share of friends in their district are reminded by their friends about their representative often enough to actually remember their representative's name and political party when asked to fill out the survey.

Based on these estimates, if we assume linear impacts, an increase in a county's congruence by 10pp would increase the probability that a respondent in that county has heard of their representative by 0.7pp (recall from Table 3 that the mean is 93.2%). The same change in a county's congruence would increase the probability a respondent in that county selects a party by 3.2pp (from mean 68.6%) and selects the correct party by 3.3pp (from mean 61.7%).[14]

[13] The odd years have a sample about one-fifth the size of even years. As such, including odd years yields similar results with noisy estimates on the odd-year coefficients. Focusing on even years also gives consistency in interpretation: the event studies thus reflect voters' knowledge of their current representative shortly before the election that will replace or re-elect that representative. Further, voting outcomes (in the next section) are mostly only available in even years.

[14] Recall from Section 2.1.5 the a one standard deviation change in congruence following redistricting is roughly 5pp; one

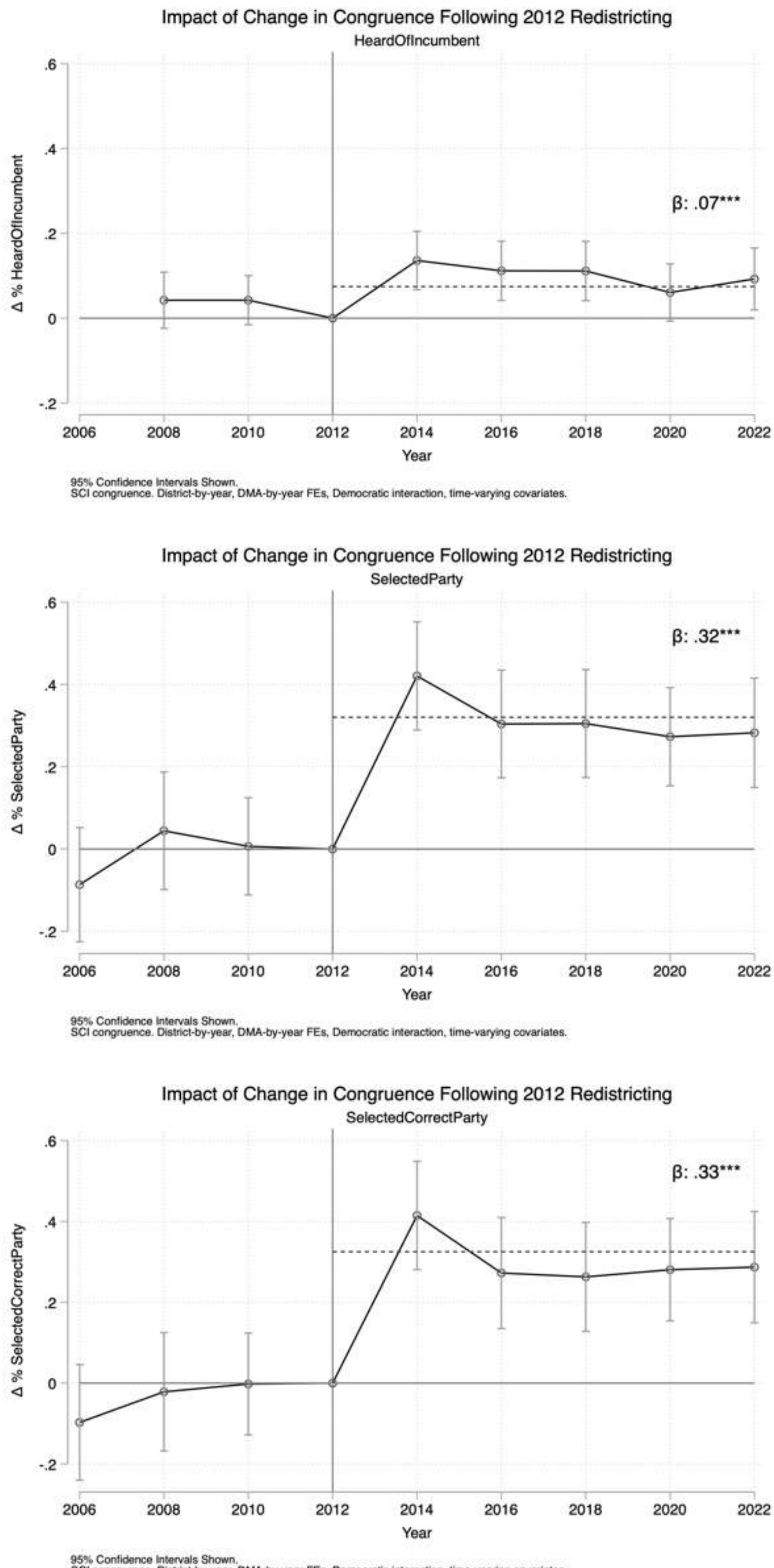


Figure 3: Effects of Increase in Congruence on Voter Familiarity with Representative

As discussed in more detail in Section 5, I find similar results when using commuting flows as an alternative proxy of social networks. Additionally, I do not find evidence that congruence increases voters' knowledge on placebo outcomes (i.e., the same three outcomes but for the respondent's governor and senators).

4.1.2 Voters' Choices

How does information translate into vote choices? I examine the impact of congruence on voter turnout and on voters' candidate preferences.

Survey Responses I start with examining subjects self-reported voting preferences and choices in the CES survey. This allows me to look at voting outcomes using the same sample as the information outcomes. The CES asks voters about their voting intentions and preferred candidates (in the pre-survey, run in September or October) and later about their actual vote choices (in the post-survey, run following the November election).[15] I run event studies analogous to equation 1 to examine the impact of congruence on voting-related outcomes; however, here I treat 2010 as the base year, as 2012 elections occur under the new district boundaries, and accordingly congruence with the new district may begin to impact voter behavior in the 2012 election.

First, consider voters' House candidate preferences reported in the pre-survey. Subjects are asked "In the general election for U.S. House of Representatives in your area, who do you prefer?" and are shown a list of names of candidates running in the election for their district. Subjects can choose a name, or indicate no preference for any particular candidate with options like "No One" or "Not Sure." Accordingly, I construct indicators for whether the subject prefers the incumbent (i.e., the name of their preferred candidate matches the name of their current House representative), prefers an opponent (i.e., the subject chooses the name of a candidate that is not the incumbent), or prefers neither. In Figure 4, I show event studies for these three outcomes; I always restrict to cases in which an incumbent exists.[16]

Results are noisier, but an increase in congruence is associated with an increase in preference for the incumbent. However, this increased preference for the incumbent does not come at the cost of preference for the opponent (which remains unchanged) but rather comes from a reduction in subjects reporting that they prefer no candidate.

standard deviation of congruence itself is 11pp. The largest changes in congruence following redistricting are around 30pp.

[15] All subjects that complete the pre-survey are asked to participate in the post-survey, though there is some attrition. Weights do not account for this attrition.

[16] I define a candidate as an incumbent if they are an incumbent for *anyone* in the survey – i.e., a candidate is an incumbent if they are currently serving in the House. Consequently, in 2012 the definition of "incumbent" is somewhat spurious, as due to redistricting, there are many subjects for whom an incumbent exists, but that incumbent is not their current representative.

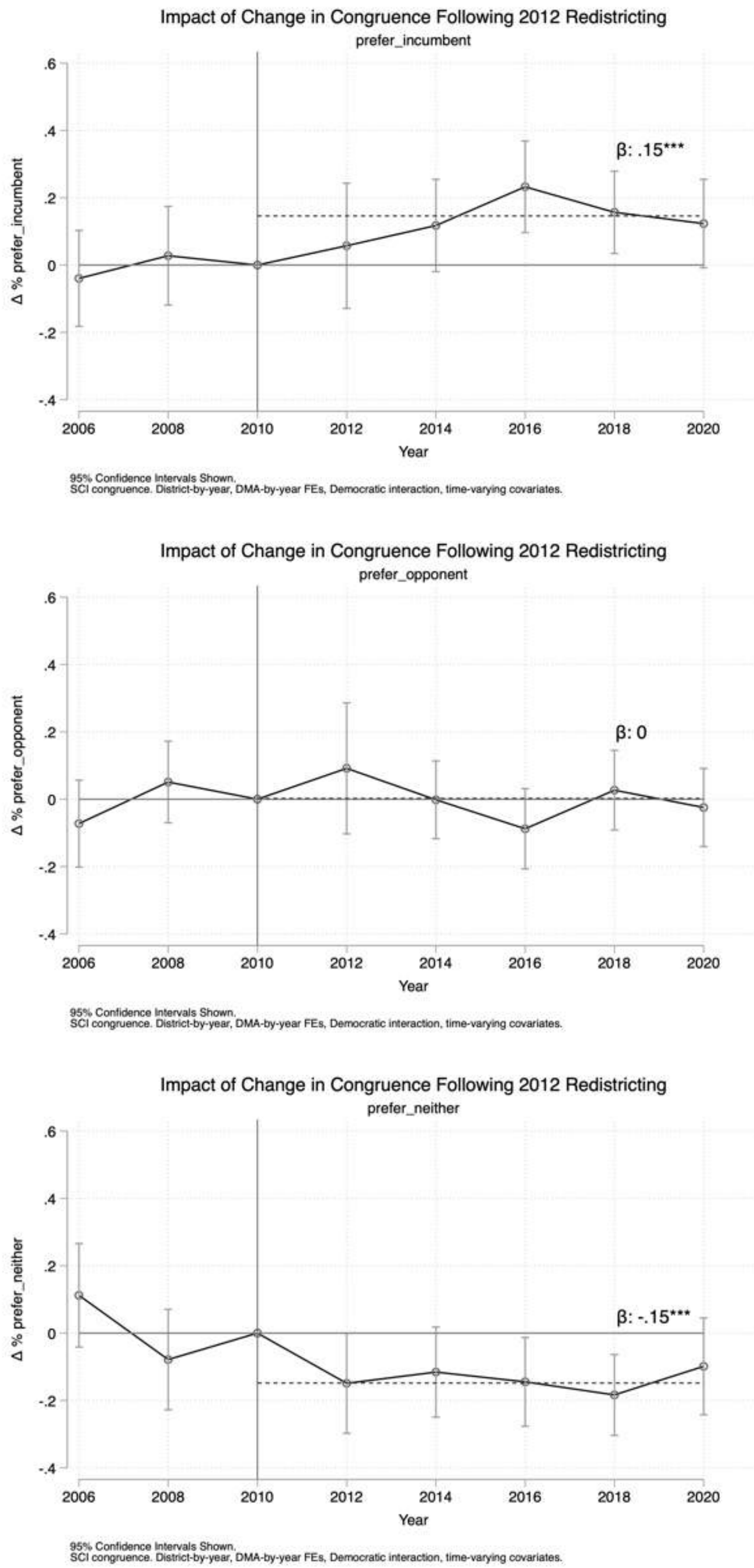


Figure 4: Effects of Increase in Congruence on Voter Preferences in CES

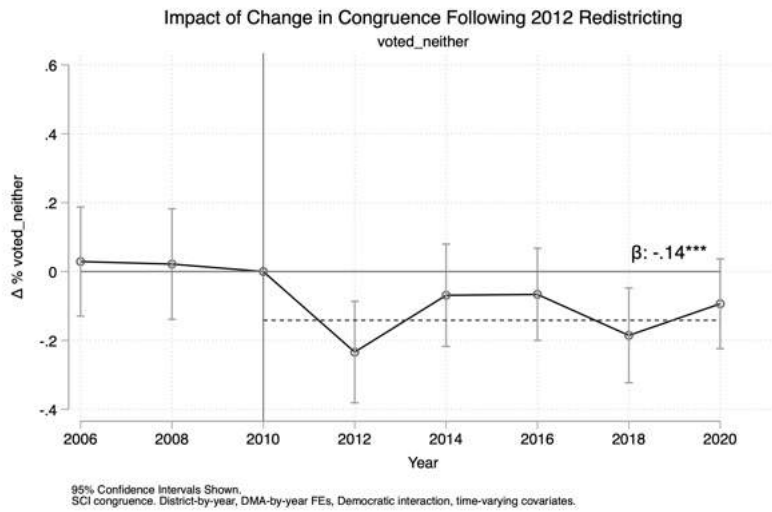


Figure 5: Effects of Increase in Congruence on Voter Choices Reported in CES - Roll-Off

Second, I examine whether these preferences translate into actual changes in votes. Here, in order to disentangle effects on vote choice from effects on turnout (which I address below), I restrict to subjects who voted in the general election.[17] Here again, I find the same pattern: I find no impacts on votes for the opponent but an increase in reporting voting for the incumbent (see Appendix Figure 14), driven by a decrease in subjects who report *not* voting in the House election, as seen in Figure 5. Because I have restricted the sample to general election voters, this outcome is equivalent to roll-off: turning out for the general election, but choosing not to vote in the House election.

Together, conditional on already turning out to vote, congruence may increase extensive margin participation in House elections. To test this, I turn next to actual vote count data.

Vote Count Data Figure 6 shows the impact of an increase in congruence on turnout in the House election, relative to top-of-ticket turnout; specifically, the outcome is the share of top-of-ticket voters who abstain in the House election, or "roll-off." Consequently, a decrease in roll-off corresponds to an increase in turnout. The specification used includes district-by-year fixed effects (for the district the majority of a county's population is in – counties for which no district has a majority of the county's population are dropped), DMA-by-year fixed effects, and county-by-year demographic controls. Results are similar when

Note, however, that we can exclude 2012 and the results are similar.

[17] The CES links survey respondents to state voter rolls and constructs indicators of whether respondents are active registered voters and of which elections there is a record of the respondent voting in. I include both subjects who are in this manner validated as turning out in the general election, as well as subjects who self-reported turnout when asked whether they voted in the November election.

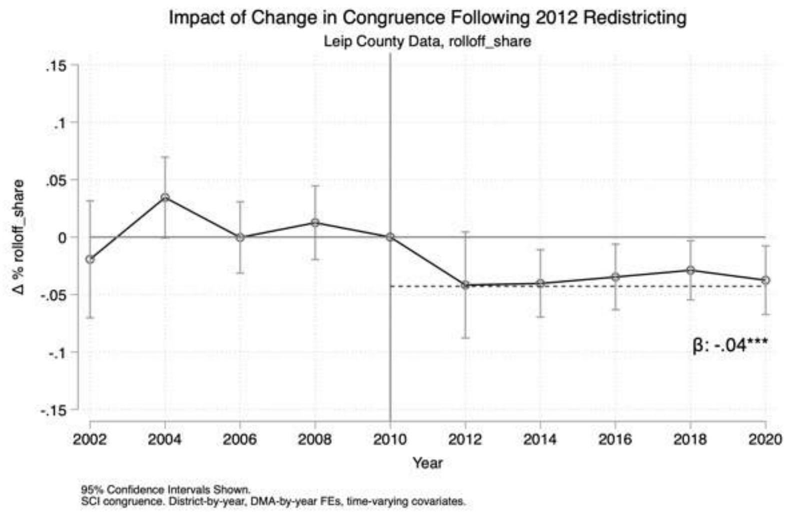


Figure 6: Effects of Increase in Congruence on Roll-Off

restricting to counties fully within one congressional district.

The negative impact indicates that congruence reduces roll-off: if a county becomes more congruent, its voters become more likely to vote in the House election *conditional* on turning out to vote in the top-of-ticket election. Recalling that mean rolloff is about 4pp, the estimate indicates that a 10pp increase in congruence reduces rolloff by 0.04pp, or by 10%.

4.2 Campaign Contributions

Figure 7 shows the impact of an increase in congruent on the share of dollars contributed to in-district candidates, as share of all county donations to House candidates. In particular, a 10pp increase in congruence is associated with a 7.4pp increase in the share of contributions to in-district candidates, from a mean of 50%.

I do not find any impact on total donations to House candidates (see Appendix Section C.3.3), indicating that this increase is driven by shifting donations away from out-of-district candidates and towards in-district candidates, rather than changing the overall amount that contributors are allocating towards House races.

5 Robustness

I find in my main specification that congruence has a positive effect on voters' knowledge of their representatives. I explore the robustness of this finding by testing whether congruence impacts placebo outcome and

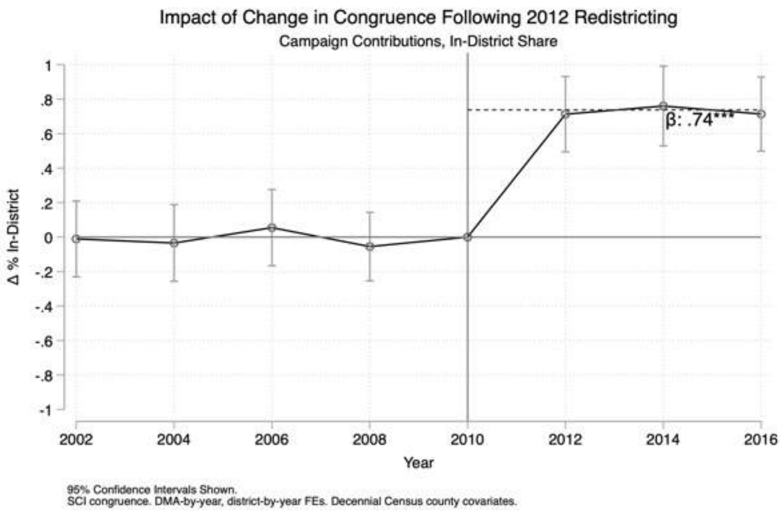


Figure 7: Effects of Increase in Congruence on Share of Contributions to In-District Candidates

by constructing an alternative measure of congruence using commuting flows.

5.1 Placebo Outcomes

I test whether congruence impacts voters' knowledge of their governor and senators: because these offices are elected through statewide elections, and consequently congressional district borders are not relevant for them, congruence should not impact them.

CES respondents answer similar questions about whether they have heard of and can identify the party of their governor and each of their senators. From these responses, I construct outcome variables analogous to the ones in the main analysis, and which measure whether voters have heard of, select a party for, and select the correct party for their governor and their senators. These variables are summarized in Appendix Table 14.

I find no significant impact of congruence on these nine outcomes. Results are reported in Appendix Section C.2.1.

5.2 Commuting Flows as an Alternative Network Measure

Commuting flows can be used as an alternative measure of social networks: the number of people that commute between two counties reflects patterns of who is regularly physically proximate to each other. Replicating the analysis using commuting flows can shed light on the extent to which the SCI captures "real world" offline networks. I use the 2016 5-Year ACS County-County Commuting Flows, which report the average number of people that commute between two counties, and I construct commuting congruence as the share of a county's commuters that stay within the county's district when commuting. For county i in district J (which contains counties j) and all US counties K (which contains counties k),

$$\text{Commuting Congruence}_i = \frac{\sum_{j \in J} \text{Commuters}_{i,j}}{\sum_{k \in K} \text{Commuters}_{i,k}}$$

In Appendix Table 15, I report results for the effect of commuting congruence on voters' familiarity with their representatives. *[Table is out-of-date, to be updated.]* Estimates are of smaller magnitudes but otherwise are similar: across specifications, commuting congruence has a positive effect on measures of voters' familiarity with their representatives, with significance at at least the 5% level for all but one estimate. I interpret the smaller estimates as reflecting the fact that commuting flows are a rougher approximation of social networks than the SCI. Additionally, the larger effects when using the SCI to construct congruence likely also reflect use of Facebook to share news about representatives.

5.3 Border Pairs Specification

An alternative identification strategy that does not rely so heavily on the 2012 redistricting event is to compare pairs of counties that lie across a district border from each other (Snyder and Strömberg 2010, Spenkuch and Toniatti 2018). The two counties in a pair should be largely similar, except for which district they are assigned to. In particular, because they are in different districts, they will likely have different congruence levels. Accordingly, we can identify the impact of congruence by comparing deviations from the county-pair's mean in one county to deviations from the county-pair's mean in the neighboring county. The specification for this design is

$$y_{ct} = \alpha_c + \mu_{pt} + \beta \text{Congruence}_{c,t} + X'_{ct}\delta + \varepsilon_{ct}$$

where y_{ct} is the outcome of interest for county c in year t, μ_{pt} is the pair-by-year fixed effect, β is the coefficient of interest, and X'_{ct} is a vector of time-varying county-level controls. I restrict to counties fully within one district. Because counties can border multiple other counties across a district border, I follow Spenkuch and Toniatti 2018 and collapse all outcomes to the county level then include one observation for every pair that a given county is in.

Because the sample becomes quite restricted when we focus only on border counties within one district, precision decreases substantially. I do not include district-by-year fixed effects because there is not enough data to accommodate them, so I instead include state-by-year fixed effects. I also restrict to only comparing pairs within the same state (though results are qualitatively similar when I include all county pairs).

Appendix Table 16 reports the results of the border pairs specification. *[Table is out-of-date, to be updated.]* With the border pairs design, I find very similar results as in the redistricting design, except estimates on "Selected Party" become insignificant after adding DMA-by-year fixed effects.

5.4 Zip Codes

To be added.

6 Model of Information Diffusion within Districts

The following model formalizes the connection between my primary measure (congruence) and the share of voters in a county that are informed about their representative.

Consider pieces of news about congressional representatives arising and spreading in a population. In a given area, what is the steady state share of people who have learned some relevant (i.e., sufficiently recent) news about their representative? I represent this process using a mean-field approximation, applying Jackson and López-Pintado (2013)'s model of diffusion with homophily and heterogenous types.

6.1 Types of individuals

The society consists of a continuum of agents $N = [0, 1]$. Within the society, each agent is assigned a type based on where they live.

In particular, let $\mathcal{D} = [1, ..., D]$ be the set of all congressional districts and $\mathcal{C} = [1, ..., C]$ be the set of counties. There is no ranking between districts and counties: counties can be fully within districts, districts can be fully within counties, or neither. Agents are characterized by the congressional district $d \in \mathcal{D}$ and county $c \in \mathcal{C}$ in which they reside. An agent is of type (c, d) if they live in the intersection of county c and district d. Accordingly, there $C \times D$ possible types.The society is partitioned by type, such that $n(c, d) \in [0, 1]$ is the fraction of agents of type (c, d).

6.2 Friendship shares between types

The share of friends each type has of each other type can be described by the matrix

$$\Pi = \begin{pmatrix} \pi_{(1,1),(1,1)} & \cdots & \pi_{(1,1),(C,D)} \\ \vdots & \ddots & \vdots \\ \pi_{(C,D),(1,1)} & \cdots & \pi_{(C,D),(C,D)} \end{pmatrix}$$

where $\pi_{(c,d)(c',d')} \geq 0$ is the share of type (c,d)'s friends that live in (c',d'). Equivalently, in any given encounter, this is the probability that an agent from county c and district d meets an agent from county c' and district d'. Accordingly, $\sum_{c'=1}^{C}\sum_{d'=1}^{D}\pi_{(c,d)(c',d')} = 1$. Assume that if $\pi_{(c,d)(c',d')} = 0$ then $\pi_{(c',d')(c,d)} = 0$, because friendships are mutual (but observe that otherwise $\pi_{(c,d)(c',d')}$ need not equal $\pi_{(c',d')(c,d)}$). If $c \cap d = \emptyset$, for any c' and d' $\pi_{(c,d)(c',d')} = 0$.

6.3 Information sharing process

Any individual can be informed (state 1) or uninformed (state 0) about their own congressional representative at any given point in time. Individuals only care about news about their own representative; consequently, they only become informed if they receive a piece of news about their *own* representative. A random set of agents are initially informed because they are seeded with a piece of news about their own representative. Once informed, agents forget the news and become uninformed at rate $\delta > 0$.

Uninformed agents can become informed if they receive news from an informed friend from the same district. In particular, each period, every agent meets with one friend to receive news. The meeting does not need to be reciprocal: one agent can receive news from the other without the reverse being true.[18] For brevity, I will say that an agent "meets" a friend to mean that an agent "receives news from" a friend, using the two terms interchangeably. While an agent may meet any friend, they are only interested in the news about their own representative, as this is the only information that matters for their decision at the ballot box. Accordingly, an uninformed individual becomes informed if (i) the friend they meet is from their same district, (ii) that friend is informed. However, assume that there are some frictions to communicating information, such that when an uninformed agent meets an informed same-district friend, the informed friend's news is communicated with probability $\alpha \in (0,1]$.[19]

[18] As Jackson and López-Pintado 2013 explain, assuming that the meetings are reciprocal requires adding the constraint that $n(i,j)\pi_{(i,j)(k,l)} = n(k,l)\pi_{(k,l)(i,j)}$ – that is, that the number of interactions from type (i,j) to type (k,l) in a period is the same as the number of interactions from type (k,l) to type (i,j) in a period.

[19] α captures frictions in communication from both ends of the interaction: both the probability that the recipient of the news does not pay attention to it, as well as the probability that the conveyer of the news fails to pass it on. α is the probability of transmission, conditional on an uninformed agent receiving news from an informed, same-district agent.

6.4 Timing

To summarize, consider the process as occurring in discrete periods that each proceed as follows:

1. Begin each period with some agents in state 1 (informed) and some agents in state 0 (uninformed).
2. News is shared. Each agent meets one friend.
3. Uninformed agents who meet an informed same-district friend become informed with probability α. (No change occurs if an informed agent meets another informed agent, or if agents from different districts meet.)
4. At the end of each period, a share δ of informed individuals become uninformed.

6.5 Individual transition probabilities

Per unit of time, what is the probability that an uninformed individual becomes informed? This will depend on the share of an individual's same-district friends that are informed.

Let $\rho_{(c,d)}(t)$ denote the probability a type (c, d) agent is informed at time t. Let $\tilde{\rho}_{(c,d)}(t)$ represent the probability that a type (c, d) agent meets a same-district informed friend at time t. That is, $\tilde{\rho}_{(c,d)}(t)$ is the share of type (c, d)'s friends that (i) live in the same district, and (ii) are informed at time t. $\tilde{\rho}_{(c,d)}(t)$ is constructed as the weighted average share of informed friends, with weights given by the friendship shares from Π, and with friends from other districts treated as if they are all uninformed:

$$\tilde{\rho}_{(c,d)}(t) = \sum_{c' \in \mathcal{C}} \sum_{d' \in \mathcal{D}} \left(\pi_{(c,d)(c',d')} \times \rho_{(c',d')}(t) \times \mathbb{I}\{d' = d\} \right) \tag{2}$$

$$= \sum_{c' \in \mathcal{C}} \left(\pi_{(c,d)(c',d)} \times \rho_{(c',d)}(t) \right) \tag{3}$$

Therefore, at time t, the rate that an uninformed type-(c, d) agent transitions to informed is $\alpha\tilde{\rho}_{(c,d)}(t)$: the frictions in sharing information multiplied by the probability of meeting a same-district informed friend. Representing time $t \in \mathbb{R}_+$ as continuous, the dynamics are

$$\frac{d\rho_{(c,d)}(t)}{dt} = \underbrace{(1 - \rho_{(c,d)}(t))}_{\text{Share in state 0}} \underbrace{(\alpha\tilde{\rho}_{(c,d)}(t))}_{\text{Rate } 0 \to 1} - \underbrace{\delta\rho_{(c,d)}(t)}_{\text{Rate } 1 \to 0 \times \text{Share in state 1}} \tag{4}$$

That is, the change in the probability a type (c, d) person is informed is given by the difference between the share of (c, d) people who become newly informed and the share of (c, d) people who become newly uninformed.

6.6 Steady state

In the steady state, $\frac{d\rho_{(c,d)}(t)}{dt} = 0$ for all (c, d). Consequently, at the steady state, the probability that an individual of type (c, d) is informed is given by

$$\rho_{(c,d)} = \frac{\alpha\tilde{\rho}_{(c,d)}}{\alpha\tilde{\rho}_{(c,d)} + \delta} \tag{5}$$

6.7 Aggregating to county-level

When friendship shares are only known at a more aggregated level, such as counties, I can construct the aggregated steady state probabilities by taking population-weighted averages. In particular, let $\pi_{c,c'}$ represent the probability that an agent from county c meets an agent from county c' (regardless of district) in any given meeting. The county-county friendship matrix summarizes these probabilities:

$$\Pi' = \begin{pmatrix} \pi_{1,1} & \cdots & \pi_{1,C} \\ \vdots & \ddots & \vdots \\ \pi_{C,1} & \cdots & \pi_{C,C} \end{pmatrix}$$

Let $D(c)$ be the set of districts that county c intersects with. Recall that $n(c, d)$ is the fraction of agents of type (c, d). Then, the share of county c's population in each district d it intersects is $q_{(c,d)} = \frac{n(c,d)}{\sum_{d' \in D(c)} n(c,d')}$.

Assume that friendships are uniformly distributed within a county. Then, $\pi_{(c,d)(c',d')} = \pi_{c,c'} \times q_{(c',d')}$: the probability a type (c, d) agent meets a type (c', d') agent is approximated by the probability an agent from county c meets an agent from county c', multiplied by the probability that an agent living in c' also lives in d'.

The share of people in a given county that are informed about their own district at time t can then be constructed as $\rho_c(t) = \sum_{d \in D(c)} q_{(c,d)}\rho_{(c,d)}(t)$: the population-weighted average share of informed people, summing across each district the county intersects.

The probability that, in any given meeting, an agent from county c meets an informed agent who lives in the same district is

$$\tilde{\rho}_c(t) = \sum_{c' \in C} \sum_{d \in D(c)} \left(\pi_{c,c'} \times \rho_{c'}(t) \times q_{(c,d)} \times q_{(c',d)}\right) \tag{6}$$

where $\pi_{c,c'} \times \rho_{c'}(t)$ gives the probability an individual in c' is informed weighted by the probability of meeting an individual from c', and $q_{(c,d)} \times q_{(c',d)}$ represents the probability that the two agents live in the same district.[20]

Accordingly, the steady state probability that an individual from county c is informed is

$$\rho_c = \frac{\alpha\tilde{\rho}_c}{\alpha\tilde{\rho}_c + \delta} \tag{7}$$

6.8 Congruence

The joint probability of meeting a friend who lives in the same district *and* is informed can be broken down into the probability of meeting a same-district friend (i.e., *congruence*), multiplied by the probability that said friend is informed. In particular, let congruence for county c be represented by $\bar{\pi}_c = \sum_{c' \in \mathcal{C}} \sum_{d \in D(c)} \left(\pi_{c,c'} \times q_{(c,d)} \times q_{(c',d)}\right)$. Upon meeting a same-district friend, let $\tilde{r}_c(t)$ represent the probability that the friend is informed:

$$\tilde{r}_c(t) = \frac{1}{\bar{\pi}_c} \sum_{c' \in C} \sum_{d \in D(c)} \left(\pi_{c,c'} \times \rho_{c'}(t) \times q_{(c,d)} \times q_{(c',d)}\right)$$

(For more detail, Appendix Section A.1 provides a full derivation.)

The steady state probability that an individual from county c is informed can then be written as

$$\rho_c = \frac{\alpha\bar{\pi}_c\tilde{r}_c}{\alpha\bar{\pi}_c\tilde{r}_c + \delta} \tag{8}$$

For a given county c, it were possible to increase congruence $\bar{\pi}_c$ without decreasing the share of informed friends in other counties,[21] then such an increase in congruence would lead to an increase in the share of informed agents in c. It is not generally true that an increase in $\bar{\pi}_c$ for county c will not cause a decrease in $\rho_{c'}$ for some county c'; however, in order for an increase in $\bar{\pi}_c$ to be associated with a decrease in ρ_c, the decrease in $\rho_{c'}$ must be large enough that $\sum_{d \in D(c)} \pi_{c,c'} \times \rho_{c'}(t) \times q_{(c,d)} \times q_{(c',d)}$ offsets the first order increase

[20] To be clear, $q_{(c',d)} = 0$ whenever $c' \cap d = \emptyset$.

[21] Specifically, counties for which for which $D(c) \cap D(c') \neq \emptyset$.

in ρ_c driven by a larger share of connections same-district friends in counties that experience no change (or an increase) in the share informed.

6.9 Predicting Abstention Rates: Swing Voter's Curse

How does the share of informed voters affect voter turnout rates? Timothy J Feddersen and Pesendorfer 1996 model the "Swing Voter's Curse," which illustrates that even in an environment with costless voting[22] it can be rational for an uninformed voter to abstain. For intuition, consider an environment where voters are choosing between two alternatives (alternative 1 and alternative 0) and an unknown state of the world (state 1 or state 0), and all voters prefer the alternative that matches the state of the world. Some voters are informed and some voters are uninformed; assume there is at least one of each. Assuming all other uninformed voters behave similarly, then conditioning on the event in which she is pivotal, an uninformed voter is strictly better off abstaining: when her vote is pivotal, she is voting for the inferior option (against the choice of an informed voter). Timothy J. Feddersen and Pesendorfer 1999 and McMurray 2013 extend these insights to analyze rational abstention under different structures of information and preferences. Battaglini et al. 2009, Morton and Tyran 2011, Mengel and Rivas 2017, and Mooers et al. 2024 test models of information aggregation under abstention in laboratory experiments, and across different settings find that voters with lower qualities signals are more likely to abstain.

6.9.1 Set-up

I apply **feddersenSwingVoterCurse1996**'s model to this setting to predict how county-level turnout in House elections responds to changes in the share of informed voters in a county. Here, I sketch the set-up, relabeled for this setting. Consider the election occurring in a single district d. Assume that there are two possible states of the world $z \in \{0, 1\}$, and the state is chosen by nature before the election. Assume that ex ante each state is equally likely, which is common knowledge.[23] Informed voters know the state with certainty, while uninformed voters think either state is equally likely.

In each county c, assume that an exogenously given share of voters are partisans: a share $p_{0,c}$ are partisans for party 0, a share $p_{1,c}$ are partisans for party 1, and a share $p_{i,c} = 1 - p_{0,c} - p_{1,c}$ are independents. 0-partisans always prefer a candidate from party 0, regardless of the state, and analogously for 1-partisans. Independent voters want to elect the candidate from the party that matches the state of the world.

[22] For example, when the cost of voting has already been paid in order to turn out for a top-of-ticket race, and voters are deciding whether to cast a vote in a down-ballot race.

[23] Assume an equally likely state for expositional simplicity, but any common knowledge prior is permitted – in fact, the abstention rate is independent of the prior.

The exact number of voters is uncertain. In particular, uniformly from the district's population, nature chooses a set of voters by $M+1$ individual draws; in each draw, nature chooses an agent with probability $(1-p_\phi)$. Thus, if an agent is selected, she is an independent with probability $p_{i,c}/(1-p_\phi)$, a 0-partisan with probability $p_{0,c}/(1-p_\phi)$, and a 1-partisan with probability $p_{1,c}/(1-p_\phi)$. These probabilities are common knowledge.

The population of district d is represented by the mass $n_d = \sum_{c\in C(d)} n(c,d)$ where $C(d)$ is the set of all counties c such that $c \cap d \neq \emptyset$. Accordingly, the probability that a chosen agent is from county c is $\frac{n(c,d)}{n_d}$. District-wide, the partisan shares are $p_x^d = \sum_{c\in C(d)} \frac{n(c,d)}{n_d} p_{x,c}$ for each $x \in \{i, 0, 1\}$.

Each agent from a given county c is informed with probability ρ_c (the steady state share of informed voters in the county, as determined by the network structure). The probability of being informed is independent of partisanship, but partisans always vote for their preferred party regardless of their beliefs about the state of the world. Consequently, the probability of being informed only is relevant for the strategy of the independents. Across the full district, voters are informed with probability $\rho^d = \sum_{c\in C(d)} \frac{n(c,d)}{n_d} \rho_c$

6.9.2 Strategies

Agents choose an action $s \in \{\phi, 0, 1\}$ where ϕ represents abstaining, and 0 or 1 represents voting for the candidate from party 0 or 1. Partisans never abstain, and always vote for their preferred candidate. Informed independents always vote for the candidate they know to match the state of the world.

As Timothy J Feddersen and Pesendorfer 1996detail, uninformed independents play a mixed strategy in order to maximize the probability that the election chooses the candidate that matches the state of the world. In equilibrium, the uninformed independents mix between abstaining (with probability τ_ϕ^d) and voting for either candidate (with probabilities τ_0^d and τ_1^d, respectively, such that $1 = \tau_\phi^d + \tau_0^d + \tau_1^d$). In particular, the uninformed independents would like for the informed voters to decide the election outcome. Consequently, when they vote, they vote for the candidate with a smaller district-wide partisan advantage, so as to close the partisan gap and maximize the probability that the informed independents decide the election. Across the district, if $p_i^d(1-\rho^d) < |p_0^d - p_1^d|$ (i.e., if the share of uninformed independents is less than the partisan gap), then all uninformed independents vote for the candidate with a smaller partisan advantage. Otherwise, uninformed independents randomize between abstaining and voting for the candidate with a smaller partisan share, and they abstain with a probability such that in expectation they exactly cancel out the partisan advantage.

6.9.3 Abstention rates

District-wide In a district with a large enough population,[24] if $p_i^d(1-\rho^d) > |p_0^d - p_1^d|$ then the probability an uninformed independent abstains is

$$\tau_\phi^d = 1 - \frac{|p_0^d - p_1^d|}{p_i^d(1-\rho^d)}$$

The overall abstention rate $t^d \in [0, 1)$ in the district is then

$$t^d = p_i^d(1-\rho^d) - |p_0^d - p_1^d|$$

So long as the partisan difference in the district is not too large, the abstention rate is decreasing in the share of informed voters. The abstention rate increases as the difference in the partisan shares decreases.

County-wide Within the district, all uninformed independents abstain at the same rate. However, each county has a different share of informed voters, as well as different shares of independents. Accordingly, the overall abstention rate observed in county c in the district d election is

$$\begin{aligned} t_c^d &= \tau_\phi^d p_{i,c}(1-\rho_c) \\ &= p_{i,c}(1-\rho_c) - \frac{p_{i,c}(1-\rho_c)}{p_i^d(1-\rho^d)}|p_0^d - p_1^d| \end{aligned}$$

In a given county, the observed abstention rate is again decreasing in the share of informed voters in the county.

If a county intersects multiple districts, then the overall observed abstention rate in the county is the population-weighted average across all districts the county intersects:

$$t_c = \sum_{d \in D(c)} q_{(c,d)} t_c^d$$

7 Model Estimation and Counterfactuals

Can congressional district maps be drawn to increase the share of informed voters? If so, what are the consequences for turnout, and in turn the bias or competitiveness of districts?

To answer these questions, I estimate parameters governing the model of information diffusion, and I then use the model to simulate the share of informed voters and changes in turnout under different congressional

[24] Districts have an average population of 760,000.

maps.

7.1 Estimation

Recall that for a given county c, the steady-state share of informed voters is given by

$$\rho_c = \frac{\alpha \bar{\pi}_c \tilde{r}_c}{\alpha \bar{\pi}_c \tilde{r}_c + \delta}$$

Further, if we let $\lambda = \delta/\alpha$, we can write

$$\rho_c = \frac{\bar{\pi}_c \tilde{r}_c}{\bar{\pi}_c \tilde{r}_c + \lambda}$$

$\bar{\pi}_c$ can be calculated from the data, as can the friendship shares and population shares required to construct each $\tilde{r}_c$, given a vector of each ρ_c. As such, the only unknown parameter determining the steady state share of informed voters is λ. Consequently, I focus on estimating λ.

From the reduced form results, I have estimated the impact of a change in congruence on the change in share informed (i.e., $\hat{\beta} = \frac{\text{cov}(\Delta\pi, \Delta\rho)}{\text{var}(\Delta\pi)}$). Consequently, I can estimate λ by using indirect inference Gourieroux et al. 1993 to find the λ^* that gives a simulated $\tilde{\beta}$ that most closely matches $\hat{\beta}$. I do this as follows:

1. Draw a value of λ (say, from a discretized grid). Vectors $\bar{\pi}^{\text{before}}$ and $\bar{\pi}^{\text{after}}$ are observed. Draw a random initial vector of $\rho^{\text{before}}(0)$.

2. Given λ, $\bar{\pi}^{\text{before}}$, and $\rho^{\text{before}}(0)$, simulate $\rho^{\text{before}}(1), ..., \rho^{\text{before}}(t)$, until convergence to the steady state is reached. This gives the simulated steady-state value of the share informed before the change in congruence.

3. Repeat the process for $\bar{\pi}^{\text{after}}$ and $\rho^{\text{after}}(0)$.

4. Calculate $\tilde{\beta}(\lambda) = \frac{\text{cov}(\Delta\pi, \Delta\rho)}{\text{var}(\Delta\pi)}$, where $\Delta\pi = \bar{\pi}^{\text{after}} - \bar{\pi}^{\text{before}}$ and $\Delta\rho = \rho^{\text{after}} - \rho^{\text{before}}$.

5. Repeat for many values of λ and choose the λ^* that minimizes the difference between $\tilde{\beta}(\lambda^*)$ and the $\hat{\beta}$ estimated in my regressions on the CES data.

I estimate the λ that corresponds to a model where the “share informed” corresponds to the share who have heard of their incumbent (call it λ^{Heard}), and I also estimate the λ when the “share informed” is defined as the share who correctly select the incumbent’s party (call it λ^{Correct}).

Using a grid search, I find that $\lambda^{\text{Heard}} = 0.008$ and $\lambda^{\text{Correct}} = 0.039$. The below figures show the differences between the simulated $\tilde{\beta}(\lambda)$ and empirical $\hat{\beta}$ over a range of 800 values of λ. Observe that in each case,

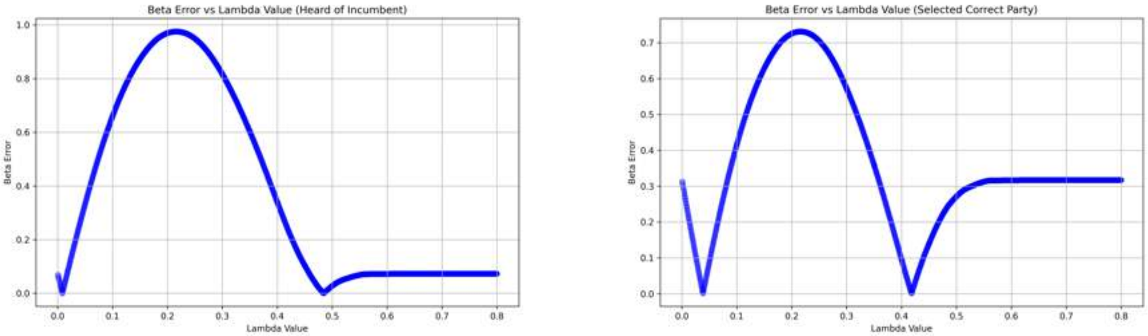


Figure 8: Difference Between $\hat{\beta}$ and $\tilde{\beta}$ for Given λ

there are two possible values of λ that minimize the error. This is because, when written in matrix form, the steady-state can be written as quadratic in the vector of ρ values. However, the larger value of λ can be ruled out in each case, as it corresponds to a steady state where the share informed is very near zero. Indeed, the difference between $\hat{\beta}$ and $\tilde{\beta}(\lambda)$ flattens above this because for large enough λ, all λs larger than it lead to convergence to the steady state where $\rho = 0$—and as such no change in congruence can lead to any change in ρ (because ρ is always 0).

7.2 Counterfactuals

With these estimates of λ in hand, I can then calculate the share of informed voters under counterfactual maps. I first consider an alternative map proposed by an Ohio citizens' group, which had the stated purpose of designing districts to better reflect communities. I next consider a large set of simulated feasible alternative congressional district boundaries, in order to understand the map characteristics congruence correlates with.

7.2.1 Map Proposed by Ohio Citizens Redistricting Commission

Ohio has a long history of court battles over gerrymandering, prompting a variety of new proposals for how districts should be drawn in the state. One such proposal is the creation of the Ohio Citizens Redistricting Commission (OCRC), which would be a citizen-led commission consisting of five Democrats, five Republicans, and five Independents chosen to be representative of the Ohio population. A measure creating this commission is on the ballot in Ohio in the upcoming November 2024 election.

As part of the push to make the commission part of law, a non-partisan independent citizens group by the same name developed a congressional district map, which they claim would be a more fair representation of Ohio communities by avoiding splitting communities across district borders and by improving minority

	Unweighted		Weighted	
Outcome (Means)	**Enacted**	**OCRC**	**Enacted**	**OCRC**
Congruence	42.6	43.9	34.1	35.5
Share Informed	83.8	84.2	79.8	80.5

Table 1: Comparison of Current Ohio Enacted Map and OCRC Map

representation. The current enacted map and OCRC's proposed map are in Appendix Figure 18.

Given the stated goal of making districts better reflect communities, the OCRC map makes for an interesting case study: I evaluate first whether it "succeeds" in the sense of increasing congruence, and then I examine whether this leads to a higher share of informed voters. Note also that Ohio's state constitution requires minimizing splitting counties when drawing congressional district maps, and Ohio is the largest state with such a requirement. As such, most counties in Ohio are fully within one district, reducing the extent to which population-weighted averages are required to construct county congruence.

Congruence and Information I examine the consequences of this proposed map by constructing congruence and then simulating the share of informed voters under the map. I compare the OCRC map to the current enacted congressional district map in Ohio. Throughout this section, I define "informed" as equivalently to correctly knowing the representative's party (i.e., I use λ^{Correct}). Appendix Figure 20 shows the differences in congruence and share informed between the two maps.

Overall, both average congruence and the share of informed voters are higher under the OCRC map, as shown in Table 1. The table shows the county means, either unweighted or weighted by the county's population. The difference between the maps is higher when considering weighted means, because the citizens' map increases congruence more for urban areas.

Abstention Rates I apply the Swing Voter's Curse model in order to estimate abstention rates. To determine the share of Democrats, Republicans, and Independents in each county, I use CES self-reported partisanship. Survey respondents state whether they are a "Strong Republican," "Not Very Strong Republican," or are an "Independent," and analogously for Democrats. I define "independents" as all of those who state they are an independent as well as those who are "not very strong" partisans. With this in hand, for a given county, I know every parameter required by the Swing Voter's Curse Model: the partisan shares as well as the share of informed voters.

I calculate the district-wide shares of partisans as well as the district-wide share informed, because these

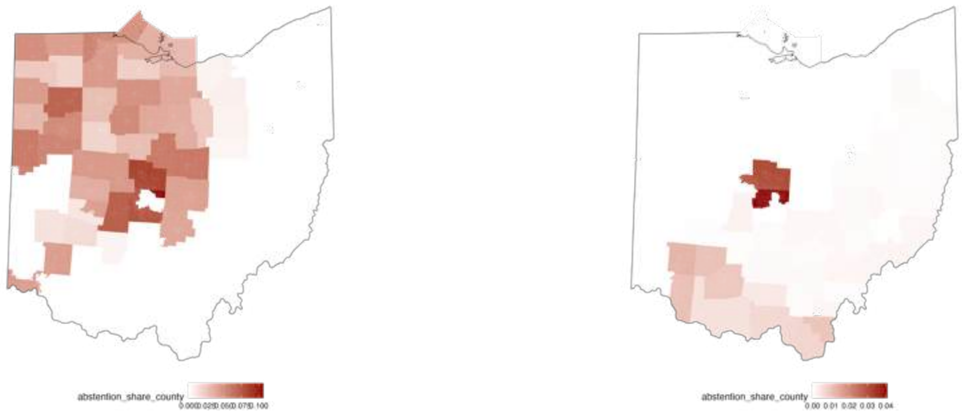


Figure 9: Simulated Abstention Rate in House Elections. Left: Current enacted map; Right: OCRC map.

are the parameters that would matter for a given uninformed independent's voting decision (as opposed to the county-wide shares). However, the county-level share informed is required to determine the abstention rate that would be observed in a given county.

Recall that voters only abstain if the district is sufficiently competitive – i.e., if the difference in the share of Democrats and the share of Republicans is sufficiently small. In Ohio, under both the current enacted map and the OCRC map, some districts are insufficiently competitive for abstention rates to be non-zero. Figure 9 shows the district-wide abstention rates under the current enacted map and the OCRC map; white areas are districts that are not competitive enough for the share of informed voters to matter. Notice that each map makes different parts of the state competitive or non-competitive.

Partisan Differences Largely because the OCRC map increases congruence in some urban areas, the share of informed voters increases more in Democratic-leaning counties. Figure 10 plots the difference in congruence between the OCRC map and the enacted map against the share of Democratic voters in the county (based on the average Democratic vote share in state-wide elections from 2016 to 2020). Ignoring county populations (i.e., considering the unweighted regression line), the OCRC map does not necessarily favor Democratic counties; in fact, higher increases in congruence are associated with higher Republican vote shares. However, once accounting for county populations, higher congruence under the OCRC map is associated with Democratic counties because some larger, urban counties have higher congruence under the OCRC map. Turning to the share of informed voters, a higher share of informed voters under the OCRC map is associated with a higher Democratic vote share.

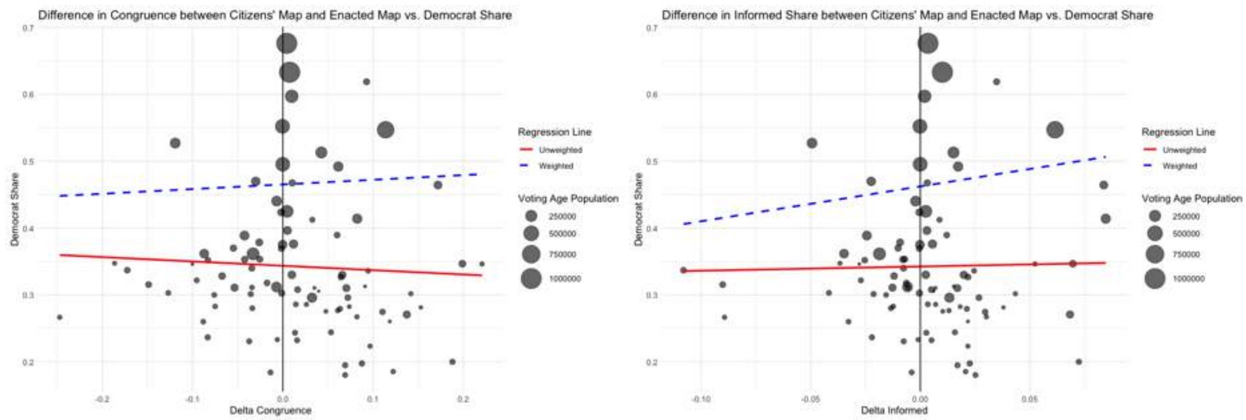


Figure 10: Partisanship vs. Difference between OCRC map and Enacted Map

7.2.2 Comparisons Across Many Simulated Boundaries

McCartan, Kenny, Simko, Kuriwaki, et al. 2021 simulate 5,000 congressional district maps for each of the 50 states. They use Monte Carlo simulation, and the maps are constrained to follow the given state's redistricting laws. While in general drawing an optimal map is known to be very difficult *[add citations...]*, their dataset instead enables looking across a distribution of feasible alternative maps, and comparing a feature of interest of a given map against the distribution of the feature across the sampled maps. They calculate commonly used measures of gerrymandering for each map. I use maps from their database to calculate congruence, shares informed, and abstention shares under counterfactual maps. In particular, for a given map, I first calculate congruence. I then know the vector $\bar{\pi}$ for the map; λ^{Correct} has been estimated; and so I can simulate ρ under the map.

As an example, in Appendix Section D.3 I consider two congressional district maps for Texas: the current district boundaries, and the district boundaries among the 5,000 Texas maps in the McCartan, Kenny, Simko, Kuriwaki, et al. 2021 database that minimize the efficiency gap (Stephanopoulos and McGhee 2015). A lower efficiency gap implies more competitive districts; under the Swing Voter's Curse model, the share of informed voters only matters for the abstention rate when districts are sufficiently competitive.

Appendix Figure 21 shows the differences between the current map and the efficiency gap minimizing map, with the differences in congruence on the left and the differences in the share informed (in particular, the share who can correctly select their representative's party) on the right. Most notably, under the efficiency gap minimizing map, counties in urban areas such as Austin, Dallas, and Houston would become more congruent. In turn, they would also experience increases in the share of informed voters. Under the current map, the simulated Texas-wide mean share informed is 79.98%; this increases the 81.03% under the efficiency

Outcome	Congruence (Unweighted)	Congruence (Weighted)
Compactness (Edge)	0.093***	0.132***
Compactness (Polsby-Popper)	0.497***	-0.013
County Splits	-38.841***	-82.224***
Municipal Splits	-62.116***	-395.18***
Efficiency Gap	0.497***	0.290*
Partisan Bias	0.389***	1.237***
E(Dem. Seats)	-27.313***	-3.457
Dissimilarity Index (Democrat vs Republican)	-0.075***	0.665***
Dissimilarity Index (Black vs Other)	-0.036	0.047
Dissimilarity Index (Hispanic vs Other)	-0.049	1.076***

gap minimizing map.

I also simulate abstention rates under the current district map in Texas, in Appendix Figure 22. Darker red indicates higher abstention rates. Grey counties do not have any respondents in the CES over the period 2010-2020 (as they are very low population counties, and it is not guaranteed that there will be a respondent from every county in every year). White counties are generally in districts where the partisan gap is large enough that abstention is predicted to be zero.

I simulate congruence across all 5,000 Texas maps in the McCartan, Kenny, Simko, Kuriwaki, et al. 2021 dataset. The table below reports the correlation between congruence and several common measures of gerrymandering, across these maps. Some of these correlations are likely to be unique to the geography and culture of Texas, such as higher congruence being associated with a stronger Republican bias (as indicated by the positive coefficients for partisan bias and the efficiency gap, and negative coefficients for the expected number of Democratic seats). Other features may be more common across states, such as congruence being positively correlated with increased compactness and fewer county or municipal splits.

8 Conclusion

Counties across the U.S. vary substantially in their social cohesiveness with their congressional district – their congruence. While people living in the average county share a representative with about half of the county's friends, this varies from 2% to 87%. I show that congruence increases voters' familiarity with their representative: when a county becomes more congruent due to redistricting, voters are more likely to recognize the name and know the party of their representative. I find similar results for voter information regardless of whether I construct the network using the SCI or commuting flows, which strengthens the case that these impacts are not unique to Facebook users. I also find that congruence decrease rates of abstaining in the House election. The impacts of congruence on electoral outcomes will inform policy by providing new dimensions on which to assess proposed district boundaries. This evidence is especially important as detailed social network data, like the SCI, has become publicly available for the first time in recent years – both enabling its use by policymakers to draw fairer districts, but also gerrymanderers who may seek to exploit it.

References

Angelucci, C., Cagé, J., & Sinkinson, M. (2020). *Media Competition and News Diets* (Working Paper No. 26782). National Bureau of Economic Research. https://doi.org/10.3386/w26782

Arias, E., Balán, P., Larreguy, H., Marshall, J., & Querubín, P. (2019). Information Provision, Voter Coordination, and Electoral Accountability: Evidence from Mexican Social Networks. *American Political Science Review*, *113*(2), 475–498. https://doi.org/10.1017/S0003055419000091

Auxier, B., & Anderson, M. (2021). Social Media Use in 2021.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social Connectedness: Measurement, Determinants, and Effects. *Journal of Economic Perspectives*, *32*(3), 259–280. https://doi.org/10.1257/jep.32.3.259

Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R., & Stroebel, J. (2020). International Trade and Social Connectedness, 55.

Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R., & Stroebel, J. (2021). International trade and social connectedness. *Journal of International Economics*, *129*, 103418. https://doi.org/10.1016/j.jinteco.2020.103418

Battaglini, M., Morton, R. B., & Palfrey, T. R. (2009). The Swing Voter's Curse in the Laboratory. *Review of Economic Studies*, *77*(1), 61–89. https://doi.org/10.1111/j.1467-937X.2009.00569.x

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298. https://doi.org/10.1038/nature11421

Bonica, A. (2014). Mapping the Ideological Marketplace. *American Journal of Political Science*, *58*(2), 367–386. https://doi.org/10.1111/ajps.12062

Bouton, L., Genicot, G., Castanheira, M., & Stashko, A. (2023). Pack-Crack-Pack: Gerrymandering with Differential Turnout.

Brown, J. R., Cantoni, E., Chinoy, S., Koenen, M., & Pons, V. (2023). The Effect of Childhood Environment on Political Behavior: Evidence from Young U.S. Movers, 1992–2021. *NBER Working Paper*. https://doi.org/http://www.nber.org/papers/w31759

Cantoni, E., & Pons, V. (2022). Does Context Outweigh Individual Characteristics in Driving Voting Behavior? Evidence from Relocations within the United States. *American Economic Review*, *112*(4), 1226–1272. https://doi.org/10.1257/aer.20201660

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., ... Wernerfelt, N. (2022). Social capital I: Measurement and associations with economic mobility. *Nature*, *608*(7921), 108–121. https://doi.org/10.1038/s41586-022-04996-4

Dahl, G. B., Løken, K. V., & Mogstad, M. (2014). Peer Effects in Program Participation. *American Economic Review*, *104*(7), 2049–2074. https://doi.org/10.1257/aer.104.7.2049

DiPrete, T. A., Gelman, A., McCormick, T., Teitler, J., & Zheng, T. (2011). Segregation in Social Networks Based on Acquaintanceship and Trust. *American Journal of Sociology*, *116*(4), 1234–1283. https://doi.org/10.1086/659100

Eckman, S. J. (2021). *Apportionment and Redistricting Process for the U.S. House of Representatives* (tech. rep. R45951). Congressional Research Service.

Eisensee, T., & Strömberg, D. (2007). News Droughts, News Floods, and U. S. Disaster Relief. *The Quarterly Journal of Economics*, *122*(2), 693–728. https://doi.org/10.1162/qjec.122.2.693

Enikolopov, R., Makarin, A., & Petrova, M. (2020). Social Media and Protest Participation: Evidence From Russia. *Econometrica*, *88*(4), 1479–1514. https://doi.org/10.3982/ECTA14281

Enikolopov, R., Petrova, M., & Zhuravskaya, E. (2011). Media and Political Persuasion: Evidence from Russia. *American Economic Review*, *101*(7), 3253–3285. https://doi.org/10.1257/aer.101.7.3253

Enke, B., Fisman, R., Freitas, L. M., & Sun, S. (2023). *Universalism and Political Representation: Evidence from the Field* (tech. rep. w31265). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w31265

Fafchamps, M., Vaz, A., & Vicente, P. C. (2019). Voting and Peer Effects: Experimental Evidence from Mozambique. *Economic Development and Cultural Change*.

Falck, O., Gold, R., & Heblich, S. (2014). E-lections: Voting Behavior and the Internet. *American Economic Review*, *104*(7), 2238–2265. https://doi.org/10.1257/aer.104.7.2238

Feddersen, T. J. [Timothy J], & Pesendorfer, W. (1996). The Swing Voter's Curse. *The American Economic Review*, *86*(3), 408–424.

Feddersen, T. J. [Timothy J.], & Pesendorfer, W. (1999). Abstention in Elections with Asymmetric Information and Diverse Preferences. *American Political Science Review*, *93*(2), 381–398. https://doi.org/10.2307/2585402

Ferraz, C., & Finan, F. (2008). Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes. *The Quarterly Journal of Economics*, *123*(2), 703–745. https://doi.org/10.1162/qjec.2008.123.2.703

Friedman, J. N., & Holden, R. T. (2008). Optimal Gerrymandering: Sometimes Pack, But Never Crack. *American Economic Review*, *98*(1), 113–144. https://doi.org/10.1257/aer.98.1.113

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, *8*(S1), S85–S118. https://doi.org/10.1002/jae.3950080507

Gul, F., & Pesendorfer, W. (2010). Strategic Redistricting. *American Economic Review*, *100*(4), 1616–1641. https://doi.org/10.1257/aer.100.4.1616

Guriev, S., Melnikov, N., & Zhuravskaya, E. (2021). 3G Internet and Confidence in Government. *The Quarterly Journal of Economics*, *136*(4), 2533–2613. https://doi.org/10.1093/qje/qjaa040

Jackson, M. O., & López-Pintado, D. (2013). Diffusion and contagion in networks with heterogeneous agents and homophily. *Network Science*, *1*(1), 49–67. https://doi.org/10.1017/nws.2012.7

Keith N. Hampton, Lauren Sessions Goulet, & Cameron Marlow. (2012). *Why most Facebook users get more than they give* (tech. rep.). Pew Research Center.

Kolotilin, A., & Wolitzky, A. (2020). The Economics of Partisan Gerrymandering. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3698690

Kuchler, T., Li, Y., Peng, L., Stroebel, J., & Zhou, D. (2022). Social Proximity to Capital: Implications for Investors and Firms (I. Goldstein, Ed.). *The Review of Financial Studies*, *35*(6), 2743–2789. https://doi.org/10.1093/rfs/hhab111

Kuriwaki, S. (2018). Cumulative CES Common Content. https://doi.org/10.7910/DVN/II2DB6

Kuziemko, I., Longuet-Marx, N., & Naidu, S. (2023). *"Compensate the Losers?" Economic Policy and Partisan Realignment in the US* (tech. rep. w31794). National Bureau of Economic Research. Cambridge, MA. https://doi.org/10.3386/w31794

Manacorda, M., & Tesei, A. (2020). Liberation Technology: Mobile Phones and Political Mobilization in Africa. *Econometrica*, *88*(2), 533–567. https://doi.org/10.3982/ECTA14392

McCartan, C., Kenny, C. T., Simko, T., Ebowe, E., Zhao, M. Y., & Imai, K. (2024). Redistricting Reforms Reduce Gerrymandering by Constraining Partisan Actors. https://doi.org/10.48550/ARXIV.2407.11336

McCartan, C., Kenny, C. T., Simko, T., Kuriwaki, S., Garcia, G., Wang, K., Wu, M., & Imai, K. (2021). 50-State Redistricting Simulations. https://doi.org/10.7910/DVN/SLCD3E

McCormick, T. H., Salganik, M. J., & Zheng, T. (2010). How Many People Do You Know?: Efficiently Estimating Personal Network Size. *Journal of the American Statistical Association*, *105*(489), 59–70. https://doi.org/10.1198/jasa.2009.ap08518

McMurray, J. C. (2013). Aggregating Information by Voting: The Wisdom of the Experts versus the Wisdom of the Masses. *The Review of Economic Studies*, *80*(1), 277–312. https://doi.org/10.1093/restud/rds026

Mengel, F., & Rivas, J. (2017). Common value elections with private information and informative priors: Theory and experiments. *Games and Economic Behavior*, *104*, 190–221. https://doi.org/10.1016/j.geb.2017.03.009

Miller, M. G. (2022). Candidate Extremism and Voter Roll-Off in US House Elections. *Legislative Studies Quarterly*, *47*(4), 791–821. https://doi.org/10.1111/lsq.12356

MIT Election Data and Science Lab. (2021). *County presidential election returns 2000-2020*. https://doi.org/10.7910/DVN/VOQCHQ

Mooers, V., Campbell, J., Casella, A., de Lara, L., & Ravindran, D. (2024). Liquid Democracy. Two Experiments on Delegation in Voting.

Morton, R. B., & Tyran, J.-R. (2011). Let the experts decide? Asymmetric information, abstention, and coordination in standing committees. *Games and Economic Behavior*, *72*(2), 485–509. https://doi.org/10.1016/j.geb.2010.09.007

Nickerson, D. W. (2008). Is Voting Contagious? Evidence from Two Field Experiments. *American Political Science Review*, *102*(1), 49–57. https://doi.org/10.1017/S0003055408080039

Polsby, D. D., & Popper, R. (1991). The Third Criterion: Compactness as a Procedural Safeguard Against Partisan Gerrymandering. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2936284

Prat, A., & Strömberg, D. (2005). *Commercial Television and Voter Information* (CEPR Discussion Paper No. 4989).

Schaffner, B., Ansolabehere, S., & Luks, S. (2021). Cooperative Election Study Common Content, 2020. https://doi.org/10.7910/DVN/E9N6PH

Schaffner, B., Ansolabehere, S., & Shih, M. (2023). Cooperative Election Study Common Content, 2022. https://doi.org/10.7910/DVN/PR4L8P

Snyder, J. M., & Strömberg, D. (2010). Press Coverage and Political Accountability. *Journal of Political Economy*, *118*(2), 355–408. https://doi.org/10.1086/652903

Spenkuch, J. L., & Toniatti, D. (2018). Political Advertising and Election Results*. *The Quarterly Journal of Economics*, *133*(4), 1981–2036. https://doi.org/10.1093/qje/qjy010

Stephanopoulos, N. O., & McGhee, E. M. (2015). Partisan Gerrymandering and the Efficiency Gap. *University of Chicago Law Review*.

Strömberg, D. (2004). Radio's Impact on Public Spending. *The Quarterly Journal of Economics*, *119*(1), 189–221. https://doi.org/10.1162/003355304772839560

Vogels, E., Auxier, B., & Anderson, M. (2021). Partisan differences in social media use show up for some platforms, but not Facebook.

Whitaker, L. P. (2017). *Congressional Redistricting Law: Background and Recent Court Rulings* (tech. rep. R44798). Congressional Research Service.

Appendix

A Model Derivations

A.1 Constructing Weighted Average Congruence

For any person in district d, there are two general types of people they can be friends with: people who are also in district d, and people who are outside of the district (call them members of $nd \equiv \mathcal{D} \setminus d$).[25] To take a shortcut with notation, say that any person who is in county c and in district d has (on average) a share $\pi^{d}_{(c,d)}$ of friends in district d, and they have a share $\pi^{nd}_{(c,d)} = 1 - \pi^{d}_{(c,d)}$ of friends outside of their district, where $\pi^{d}_{(c,d)} = \sum_{c' \in \mathcal{C}} \pi_{(c,d)(c',d)}$, and $\pi^{nd}_{(c,d)} = \sum_{c' \in \mathcal{C}} \sum_{f \in \mathcal{D} \setminus d} \pi_{(c,d)(c',f)}$.

In order to construct $\pi^{d}_{(c,d)}$ (the share of type (c,d)'s friends that also live in d), I sum as follows:

$$\begin{aligned}
\pi^{d}_{(c,d)} &= \sum_{c' \in \mathcal{C}} \pi_{(c,d)(c',d)} \\
&= \sum_{c' \in \mathcal{C}} \sum_{d' \in D(c)} \left(\pi_{(c,d)(c',d')} \times \mathbb{I}\left\{d' = d\right\}\right) \\
&= \sum_{c' \in \mathcal{C}} \sum_{d' \in D(c)} \left(\pi_{c,c'} \times q_{(c',d')} \times \mathbb{I}\left\{d' = d\right\}\right) \\
&= \sum_{c' \in \mathcal{C}} \left(\pi_{c,c'} \times q_{(c',d)}\right)
\end{aligned}$$

Next, define $\bar{\pi}_c$ as the share of county c's friendships that are between people in the same district. Put differently, $\bar{\pi}_c$ represents the probability that a randomly chosen person from county c interacts with a person from their own district (without conditioning on which district $d \in D(c)$ the county c person is from). We can construct $\bar{\pi}_c$ as follows:

[25] Both people in or outside the district may be in the same or different counties: for example, my county may be split across two districts (such that there are other people in my same county but in a different district), or my district may contain multiple counties (such that there are people in my same district but in a different county).

$$\begin{aligned}
\bar{\pi}_c &= \sum_{d \in D(c)} \left(\pi^d_{(c,d)} \times q_{(c,d)} \right) \\
&= \sum_{d \in D(c)} \sum_{c' \in \mathcal{C}} \sum_{d' \in D(c')} \left(\pi_{c,c'} \times q_{(c,d)} \times q_{(c',d')} \times \mathbb{I}\{d' = d\} \right) \\
&= \sum_{d \in D(c)} \sum_{c' \in \mathcal{C}} \left(\pi_{c,c'} \times q_{(c,d)} \times q_{(c',d)} \right) \\
&= \sum_{c' \in \mathcal{C}} \left[(QQ^T) \circ \Pi' \right]_{c,c'}
\end{aligned}$$

where Q represents the matrix of the population shares $q_{(c,d)}$, and observing that $q_{(k,d)} = 0$ whenever $k \cap d = \emptyset$.

$\bar{\pi}_c$ is congruence, or the share of friends that live in the same district. $\pi^d_{(c,d)}$ is an analogous to congruence, but specifically for people living in county c and district d (i.e., for people in county c and district d, the share of their friends that live in district d).

A.1.1 Constructing $\tilde{r}_c$

$\tilde{r}_c$ represents the probability that a randomly chosen friend is informed, conditional on that friend being in the same district as the chosen person from county c (but not conditioning on the person from c being from any particular district within c).

For each county, we know the share of that county that is in general informed about their own district, ρ_c.[26] Accordingly, conditional on a person in county c meeting a person in the same district as them, the probability that that person is informed is:

$$\tilde{r}_c = \sum_{k \in C} P(\text{in county } k | \text{in same district}) \times \rho_k$$

Using Bayes' Rule, we know that

$$P(\text{in county } k | \text{in same district}) = \frac{P(\text{in same district} | \text{in county } k) P(\text{in county } k)}{P(\text{in same district})}$$

[26] I omit time subscripts for brevity; the same relations apply in the steady state.

which in turn gives

$$P(\text{in county } k | \text{in same district}) = \frac{\sum_{d \in D(c)} \left(q_{(c,d)} \times q_{(k,d)}\right) \times \pi_{c,k}}{\bar{\pi}_c}$$

Accordingly,

$$\begin{aligned}
\tilde{r}_c &= \sum_{k \in C} \frac{\sum_{d \in D(c)} \left(q_{(c,d)} \times q_{(k,d)}\right) \times \pi_{c,k}}{\bar{\pi}_c} \times \rho_k \\
&= \frac{1}{\bar{\pi}_c} \sum_{k \in C} \sum_{d \in D(c)} \left(q_{(c,d)} \times q_{(k,d)} \times \pi_{c,k} \times \rho_k\right)
\end{aligned}$$

B Data Descriptions

B.1 Variables from CES

B.2 Construction of Vote Count Measures

B.2.1 County-by-congressional district measures

I construct the voting outcomes at the county-by-CD-level by using precinct-level vote count data from the Harvard Election Data Archive (for 2000-2010) and the MIT Election Data and Science Lab (for 2016-2020), combined with county-by-congressional-district vote count data from Dave Leip's Election Atlas (for House elections) and Daily Kos (for President, Senator, and Governor elections).

Variable	Description
Heard of Representative	When shown the name of their current House representative and asked to indicate the party their representative is affiliated with, respondent did not indicate they had “Never Heard of Person” , and they instead chose “Republican”, “Democrat”, “Other Party/Independent”, or “Not Sure”. Binary. From pre-survey.
Selected Party	When shown the name of their current House representative and asked to indicate the party their representative is affiliated with, respondent did not indicate they had “Never Heard of Person” or “Not Sure”, and they instead chose “Republican”, “Democrat”, or “Other Party/Independent”. Binary. From pre-survey.
Selected Correct Party	When shown the name of their current House representative and asked to indicate the party their representative is affiliated with, respondent chose the correct party. Binary. From pre-survey.
Prefer Incumbent	When asked “In the general election for U.S. House of Representatives in your area, who do you prefer?”, respondent chose the name of their current House representative. Binary. From pre-survey. Missing if there is no incumbent running.
Prefer Opponent	When asked “In the general election for U.S. House of Representatives in your area, who do you prefer?”, respondent chose the name of someone other than their current House representative. Binary. From pre-survey. Missing if there is no incumbent running.
Prefer Neither	When asked “In the general election for U.S. House of Representatives in your area, who do you prefer?”, respondent did not choose the name of any candidate. Binary. From pre-survey. Missing if there is no incumbent running.
Voted for Incumbent	When asked “For whom did you vote for U.S. House?”, respondent chose the name of their current House representative. Binary. From post-survey. Missing if there is no incumbent running. Missing if both “Voted in General Election” variables are missing.
Voted for Opponent	When asked “For whom did you vote for U.S. House?”, respondent chose the name of someone other than their current House representative. Binary. From post-survey. Missing if there is no incumbent running. Missing if both “Voted in General Election” variables are missing.
Voted for Neither	When asked “For whom did you vote for U.S. House?”, respondent did not choose the name of any candidate. Binary. From post-survey. Missing if there is no incumbent running. Missing if both “Voted in General Election” variables are missing.
Voted in General Election (Validated)	Respondent can be linked to state voter rolls, and there is a record of the respondent voting in the general election. Binary. From post-survey.
Voted in Primary Election (Validated)	Respondent can be linked to state voter rolls, and there is a record of the respondent voting in the primary election. Binary. From post-survey.
Voted in General Election (Self-Report)	Respondent answered that they voted in the general election. Binary. From post-survey.

Table 2: Descriptions for CES Outcome Variables

Variable	**Observations**	**Mean (%)**	**SD (pp)**
Heard of Representative	545,185	93.2	25.2
Selected Party	604,254	68.6	46.4
Selected Correct Party	604,254	61.7	48.6
Prefer Incumbent	419,545	40.14	49.0
Prefer Opponent	419,545	26.7	44.3
Prefer Neither	419,545	33.1	47.1
Voted for Incumbent	385,212	41.0	49.2
Voted for Opponent	385,212	29.1	45.4
Voted for Neither	385,212	29.9	45.8
Voted in General Election (Validated)	417,421	57.5	49.4
Voted in Primary Election (Validated)	381,277	31.8	46.6
Voted in General Election (Self-Report)	388,262	87.8	32.8

Table 3: CES Data: Summary Statistics

Variable	**Description**
House Turnout, Relative to Top-of-Ticket (“Roll-Off”)	$\frac{\text{\# Votes in Top-of-Ticket Race} - \text{\# Votes in House Race}}{\text{\# Votes in Top-of-Ticket Race}}$ For main analysis, from Dave Leip’s Election Atlas (county-level). For robustness, from Harvard Election Data Archive, Daily Kos, Dave Leip’s Election Atlas, and MIT Election Data and Science Lab (for county-by-congressional district-level). Elections where there is no top-of-ticket race are excluded.
Turnout in Top-of-Ticket Election	Turnout in the top-of-ticket election, as a share of the Voting Age Population (VAP), i.e. the population over age 18. Vote counts from Dave Leip’s Election Atlas, VAP from Census. Elections where turnout exceeds the VAP are excluded; identical to House turnout when the House election is top-of-ticket.
Turnout in House Election	Turnout in the House election, as a share of the Voting Age Population (VAP). Vote counts from Dave Leip’s Election Atlas, VAP from Census. Elections where turnout exceeds the VAP are excluded.

Table 4: Descriptions for Voting Outcome Variables

Variable	**Observations**	**Mean (%)**	**SD (pp)**
Roll-Off	29,133	4.42	12.22
Turnout in Top-of-Ticket Election	30,206	51.34	13.44
Turnout in House Election	30,308	49.07	13.49

Table 5: Voting Outcomes: Summary Statistics

	(1) No FE	(2) DI	(3) DMA	(4) State	(5) Dist	(6) DMA+State	(7) DMA+Dist	(8) DMA+State+DI	(9) DMA+Dist+DI
single_district	0.131***	0.134***	0.097***	0.000	0.000	0.000	0.000	0.000	0.000
	(0.008)	(0.008)	(0.018)	(.)	(.)	(.)	(.)	(.)	(.)
log_population	-0.047***	-0.047***	-0.036***	-0.039***	-0.023***	-0.035***	-0.024***	-0.033***	-0.024**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
log_pop_den	-0.035***	-0.034***	-0.036***	-0.038***	-0.024***	-0.035***	-0.027***	-0.034***	-0.026**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Interpolation of poppct_rural on year	0.202***	0.195***	0.129***	0.141***	0.080***	0.125***	0.089***	0.120***	0.086**
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
share_foreign	-1.276***	-1.314***	-0.914***	-0.923***	-0.613***	-0.868***	-0.673***	-0.905***	-0.703**
	(0.056)	(0.055)	(0.064)	(0.059)	(0.069)	(0.064)	(0.079)	(0.065)	(0.081)
share_moved	-0.775***	-0.809***	-0.677***	-0.620***	-0.550***	-0.671***	-0.574***	-0.649***	-0.562**
	(0.051)	(0.050)	(0.046)	(0.046)	(0.045)	(0.046)	(0.047)	(0.046)	(0.047)
(mean) share_white_not_hispanic	0.175***	0.161***	0.134***	0.144***	0.045***	0.135***	0.055***	0.109***	0.030
	(0.014)	(0.015)	(0.019)	(0.017)	(0.017)	(0.018)	(0.019)	(0.019)	(0.020)
(mean) share_ed_nohs	0.755***	0.674***	1.079***	1.111***	0.929***	1.056***	0.965***	1.024***	0.949**
	(0.054)	(0.054)	(0.064)	(0.066)	(0.066)	(0.063)	(0.069)	(0.064)	(0.069)
share_collegep	-1.188***	-1.155***	-0.977***	-1.064***	-0.829***	-0.948***	-0.829***	-0.928***	-0.821**
	(0.033)	(0.033)	(0.033)	(0.034)	(0.038)	(0.034)	(0.040)	(0.034)	(0.040)
(mean) share_below_pov	0.486***	0.504***	0.384***	0.437***	0.299***	0.338***	0.290***	0.429***	0.348**
	(0.039)	(0.038)	(0.043)	(0.044)	(0.045)	(0.044)	(0.046)	(0.044)	(0.047)
log_median_income	-0.260***	-0.251***	-0.216***	-0.242***	-0.149***	-0.206***	-0.150***	-0.211***	-0.156**
	(0.012)	(0.012)	(0.013)	(0.013)	(0.015)	(0.014)	(0.016)	(0.014)	(0.016)
dem_share2008	-0.275***	-0.508***	-0.232***	-0.233***	-0.098***	-0.222***	-0.122***	-0.315***	-0.151**
	(0.019)	(0.040)	(0.021)	(0.021)	(0.022)	(0.021)	(0.023)	(0.034)	(0.041)
Observations	3107	3107	3101	3107	2985	3101	2979	3101	2979

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

Notes: Each cell represents a regression of congruence on the variable indicated on the left. Demographics from 2010 Decennial Census and 2010-2014 ACS. "Single District State" is a dummy variable indicating the state has only one congressional district. "Pop. Density" is the population per square mile. "% Foreign Born" is the share of non-citizens and naturalized citizens. "% Moved Last Year" is the share of people who moved within the last year, including those that moved within the same county. "% No HS" is the share of individuals 25 or over who do not hold a high school degree. "% College/Grad" is the share of individuals 25 or older who hold at least a college degree. "Democratic Share 2008" is the share of votes for Obama in the 2008 presidential election. Column "No FE" includes no fixed effects, "DI" includes a control for the county's share of Democratic friends, and "DMA", "State" and "Dist" each represent whether DMA, state, or district fixed effects were included.

Table 6: Correlates of Congruence, 2010

	(1) No FE	(2) DI	(3) DMA	(4) State	(5) Dist	(6) DMA+State	(7) DMA+Dist	(8) DMA+State+DI	(9) DMA+Dist+DI
single_district	0.127***	0.140***	0.094***	0.000	0.000	0.000	0.000	0.000	0.000
	(0.008)	(0.008)	(0.018)	(.)	(.)	(.)	(.)	(.)	(.)
log_population	-0.047***	-0.043***	-0.036***	-0.040***	-0.026***	-0.035***	-0.026***	-0.024***	-0.016**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
log_pop_den	-0.035***	-0.032***	-0.037***	-0.039***	-0.027***	-0.036***	-0.028***	-0.025***	-0.018**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Interpolation of poppct_rural on year	0.202***	0.184***	0.135***	0.152***	0.096***	0.132***	0.100***	0.088***	0.067**
	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
share_foreign	-1.283***	-1.206***	-0.892***	-0.949***	-0.632***	-0.849***	-0.668***	-0.721***	-0.584**
	(0.053)	(0.051)	(0.058)	(0.056)	(0.066)	(0.059)	(0.072)	(0.051)	(0.065)
share_moved	-0.826***	-0.776***	-0.693***	-0.628***	-0.569***	-0.677***	-0.592***	-0.450***	-0.426**
	(0.055)	(0.054)	(0.048)	(0.052)	(0.049)	(0.048)	(0.050)	(0.049)	(0.050)
(mean) share_white_not_hispanic	0.223***	0.203***	0.172***	0.188***	0.079***	0.171***	0.093***	0.028	-0.036
	(0.014)	(0.015)	(0.018)	(0.017)	(0.019)	(0.018)	(0.020)	(0.019)	(0.022)
(mean) share_ed_nohs	0.642***	0.428***	0.938***	1.000***	0.771***	0.920***	0.769***	0.560***	0.492**
	(0.065)	(0.064)	(0.073)	(0.081)	(0.077)	(0.073)	(0.078)	(0.069)	(0.075)
share_collegep	-1.097***	-1.019***	-0.915***	-1.018***	-0.777***	-0.890***	-0.767***	-0.675***	-0.587**
	(0.030)	(0.033)	(0.030)	(0.031)	(0.035)	(0.031)	(0.037)	(0.036)	(0.042)
(mean) share_below_pov	0.580***	0.517***	0.446***	0.546***	0.347***	0.404***	0.333***	0.449***	0.388**
	(0.039)	(0.039)	(0.042)	(0.046)	(0.045)	(0.044)	(0.045)	(0.040)	(0.042)
log_median_income	-0.260***	-0.228***	-0.227***	-0.267***	-0.164***	-0.218***	-0.161***	-0.172***	-0.134**
	(0.012)	(0.012)	(0.014)	(0.013)	(0.014)	(0.014)	(0.015)	(0.013)	(0.014)
dem_share2016	-0.383***	-0.342***	-0.314***	-0.326***	-0.193***	-0.306***	-0.219***	-0.064**	0.066**
	(0.018)	(0.020)	(0.019)	(0.019)	(0.022)	(0.019)	(0.023)	(0.028)	(0.033)
dem_share2020	-0.273***	-0.402***	-0.337***	-0.399***	-0.255***	-0.370***	-0.282***	-0.088**	0.054
	(0.015)	(0.039)	(0.020)	(0.020)	(0.024)	(0.021)	(0.025)	(0.037)	(0.045)
Observations	3107	3107	3101	3107	2988	3101	2981	3101	2981

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

Notes: Each cell represents a regression of congruence on the variable indicated on the left. Demographics from 2020 Decennial Census and 2015-2019 ACS. "Single District State" is a dummy variable indicating the state has only one congressional district. "Pop. Density" is the population per square mile. "% Foreign Born" is the share of non-citizens and naturalized citizens. "% Moved Last Year" is the share of people who moved within the last year, including those that moved within the same county. "% No HS" is the share of individuals 25 or over who do not hold a high school degree. "% College/Grad" is the share of individuals 25 or older who hold at least a college degree. "Democratic Share 2016" is the share of votes for Clinton in the 2016 presidential election; "Democratic Share 2020" is the share of votes for Biden in the 2020 presidential election. Column "No FE" includes no fixed effects, "DI" includes a control for the county's share of Democratic friends, and "DMA", "State" and "Dist" each represent whether DMA, state, or district fixed effects were included.

Table 7: Correlates of Congruence, 2020

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
single_district	0.112***	0.000	0.000	0.114***	0.000	0.000	0.113***	0.000
	(0.007)	(.)	(.)	(0.007)	(.)	(.)	(0.007)	(.)
log_population	-0.021***	-0.004	0.004	-0.021***	-0.004	0.003	-0.019***	-0.004
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
log_pop_den	-0.002	-0.010***	-0.009***	-0.001	-0.008***	-0.009**	-0.001	-0.008***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)
Interpolation of poppct_rural on year	-0.060***	-0.042***	-0.030***	-0.059***	-0.040***	-0.029***	-0.057***	-0.041***
	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)
share_foreign	-0.619***	-0.436***	-0.525***	-0.623***	-0.471***	-0.547***	-0.712***	-0.475***
	(0.050)	(0.064)	(0.084)	(0.050)	(0.067)	(0.088)	(0.061)	(0.082)
share_moved	-0.271***	-0.268***	-0.281***	-0.287***	-0.266***	-0.279***	-0.310***	-0.266***
	(0.050)	(0.056)	(0.057)	(0.050)	(0.056)	(0.057)	(0.050)	(0.056)
(mean) share_white_not_hispanic	0.151***	0.131***	0.086***	0.140***	0.097***	0.066**	0.138***	0.096***
	(0.014)	(0.021)	(0.023)	(0.015)	(0.026)	(0.028)	(0.015)	(0.026)
(mean) share_ed_nohs	0.470***	0.502***	0.592***	0.436***	0.507***	0.595***	0.416***	0.506***
	(0.071)	(0.085)	(0.093)	(0.073)	(0.085)	(0.093)	(0.073)	(0.085)
share_collegep	-0.563***	-0.480***	-0.420***	-0.544***	-0.434***	-0.394***	-0.566***	-0.435***
	(0.047)	(0.051)	(0.057)	(0.048)	(0.054)	(0.061)	(0.048)	(0.054)
(mean) share_below_pov	0.234***	0.124**	0.111*	0.244***	0.137**	0.117*	0.223***	0.136**
	(0.060)	(0.060)	(0.062)	(0.060)	(0.061)	(0.063)	(0.061)	(0.061)
log_median_income	-0.018	-0.027	-0.017	-0.023	-0.032	-0.021	-0.017	-0.032
	(0.021)	(0.021)	(0.022)	(0.021)	(0.021)	(0.022)	(0.021)	(0.021)
dem_share2008				-0.037**	-0.061**	-0.038	0.083**	-0.058
				(0.016)	(0.025)	(0.029)	(0.036)	(0.040)
Fixed effects	No	DMA, State	DMA, Dist	No	DMA, State	DMA, Dist	No, DI	DMA+State+DI

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

Notes: Joint tests. Demographics from 2010 Decennial Census and 2010-2014 ACS. "Single District State" is a dummy variable indicating the state has only one congressional district. "Pop. Density" is the population per square mile. "% Foreign Born" is the share of non-citizens and naturalized citizens. "% Moved Last Year" is the share of people who moved within the last year, including those that moved within the same county. "% No HS" is the share of individuals 25 or over who do not hold a high school degree. "% College/Grad" is the share of individuals 25 or older who hold at least a college degree. "Democratic Share 2008" is the share of votes for Obama in the 2008 presidential election. Column "No FE" includes no fixed effects, "DI" includes a control for the county's share of Democratic friends, and "DMA", "State" and "Dist" each represent whether DMA, state, or district fixed effects were included.

Table 8: Correlates of Congruence - Joint Tests, 2010

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
single_district	0.110***	0.000	0.000	0.110***	0.000	0.000	0.111***	0.000
	(0.007)	(.)	(.)	(0.007)	(.)	(.)	(0.007)	(.)
log_population	-0.021***	-0.003	0.005	-0.021***	-0.003	0.005	-0.019***	0.000
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
log_pop_den	0.004**	-0.007**	-0.008**	0.005**	-0.006*	-0.008**	0.002	-0.007**
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
Interpolation of poppct_rural on year	-0.027**	-0.010	0.002	-0.026**	-0.009	0.002	-0.028***	-0.011
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)
share_foreign	-0.496***	-0.325***	-0.357***	-0.501***	-0.354***	-0.368***	-0.498***	-0.525***
	(0.051)	(0.065)	(0.079)	(0.051)	(0.068)	(0.081)	(0.051)	(0.072)
share_moved	-0.219***	-0.195***	-0.263***	-0.225***	-0.195***	-0.262***	-0.238***	-0.205***
	(0.056)	(0.058)	(0.061)	(0.056)	(0.058)	(0.061)	(0.056)	(0.057)
(mean) share_white_not_hispanic	0.178***	0.142***	0.094***	0.166***	0.095***	0.075**	0.179***	0.079***
	(0.013)	(0.021)	(0.024)	(0.016)	(0.030)	(0.033)	(0.017)	(0.030)
(mean) share_ed_nohs	0.214***	0.229**	0.214**	0.201**	0.235**	0.217**	0.188**	0.190*
	(0.081)	(0.104)	(0.107)	(0.081)	(0.103)	(0.106)	(0.080)	(0.100)
share_collegep	-0.627***	-0.545***	-0.476***	-0.601***	-0.470***	-0.444***	-0.606***	-0.405***
	(0.044)	(0.048)	(0.054)	(0.049)	(0.060)	(0.066)	(0.049)	(0.059)
(mean) share_below_pov	0.381***	0.253***	0.186***	0.388***	0.259***	0.189***	0.388***	0.218***
	(0.057)	(0.058)	(0.058)	(0.058)	(0.058)	(0.058)	(0.057)	(0.057)
log_median_income	0.003	-0.003	-0.021	-0.000	-0.013	-0.025	0.002	-0.002
	(0.018)	(0.018)	(0.019)	(0.018)	(0.019)	(0.020)	(0.018)	(0.019)
dem_share2016				-0.026	-0.071**	-0.032	0.010	0.099***
				(0.019)	(0.030)	(0.033)	(0.022)	(0.034)
Fixed effects	No	DMA, State	DMA, Dist	No	DMA, State	DMA, Dist	No	DMA+State+DI

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Each cell represents a regression of congruence on the variable indicated on the left. Demographics from 2020 Decennial Census and 2015-2019 ACS. "Single District State" is a dummy variable indicating the state has only one congressional district. "Pop. Density" is the population per square mile. "% Foreign Born" is the share of non-citizens and naturalized citizens. "% Moved Last Year" is the share of people who moved within the last year, including those that moved within the same county. "% No HS" is the share of individuals 25 or over who do not hold a high school degree. "% College/Grad" is the share of individuals 25 or older who hold at least a college degree.
"Democratic Share 2016" is the share of votes for Clinton in the 2016 presidential election; "Democratic Share 2020" is the share of votes for Biden in the 2020 presidential election. Column "No FE" includes no fixed effects, "DI" includes a control for the county's share of Democratic friends, and "DMA", "State" and "Dist" each represent whether DMA, state, or district fixed effects were included.

Table 9: Correlates of Congruence - Joint Tests, 2020

	(1) No FE	(2) DI	(3) DMA	(4) State	(5) Dist	(6) DMA+State	(7) DMA+Dist	(8) DMA+State+DI	(9) DMA+Dist+DI
single_district	-0.001	-0.001	-0.001	0.000	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.005)	(.)	(.)	(.)	(.)	(.)	(.)
log_population	-0.001	-0.002**	-0.001	-0.002***	-0.001	-0.001	-0.001	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log_pop_den	-0.000	-0.001	-0.001	-0.002***	-0.001	-0.001	-0.000	-0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Interpolation of poppct_rural on year	0.009***	0.010***	0.007*	0.013***	0.009**	0.008**	0.006	0.008**	0.004
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
share_foreign	-0.041**	-0.039**	-0.003	-0.058***	-0.022	-0.015	-0.015	-0.016	-0.023
	(0.016)	(0.016)	(0.025)	(0.021)	(0.031)	(0.025)	(0.033)	(0.025)	(0.033)
share_moved	-0.028	-0.026	-0.027	-0.021	-0.017	-0.025	-0.021	-0.024	-0.015
	(0.023)	(0.023)	(0.026)	(0.025)	(0.026)	(0.026)	(0.027)	(0.026)	(0.027)
(mean) share_white_not_hispanic	0.023***	0.024***	0.009	0.016**	0.022**	0.008	0.022**	0.007	0.014
	(0.006)	(0.006)	(0.008)	(0.007)	(0.009)	(0.008)	(0.010)	(0.009)	(0.011)
(mean) share_ed_nohs	-0.033	-0.027	-0.001	0.040	0.001	0.024	-0.016	0.022	-0.026
	(0.027)	(0.027)	(0.035)	(0.031)	(0.038)	(0.035)	(0.040)	(0.036)	(0.040)
share_collegep	-0.010	-0.017	0.007	-0.034*	-0.007	-0.006	-0.002	-0.004	0.007
	(0.016)	(0.017)	(0.020)	(0.018)	(0.023)	(0.020)	(0.024)	(0.020)	(0.024)
(mean) share_below_pov	-0.036*	-0.037**	-0.053**	0.002	-0.057**	-0.036	-0.075***	-0.035	-0.066**
	(0.018)	(0.018)	(0.022)	(0.021)	(0.024)	(0.023)	(0.025)	(0.023)	(0.025)
log_median_income	0.001	-0.000	0.007	-0.012**	0.004	0.001	0.012	0.001	0.011
	(0.005)	(0.005)	(0.006)	(0.005)	(0.007)	(0.006)	(0.008)	(0.006)	(0.008)
dem_share2008	0.008	-0.011	0.002	-0.012	-0.019	-0.001	-0.024*	0.013	0.000
	(0.007)	(0.014)	(0.010)	(0.010)	(0.013)	(0.011)	(0.014)	(0.016)	(0.022)
Observations	3107	3107	3101	3107	2985	3101	2979	3101	2979

Table 10: Predictors of Changes in Congruence - 2010

	(1) No FE	(2) DI	(3) DMA	(4) State	(5) Dist	(6) DMA+State	(7) DMA+Dist	(8) DMA+State+DI	(9) DMA+Dist+DI
single_district	-0.001	-0.001	-0.001	0.000	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.005)	(.)	(.)	(.)	(.)	(.)	(.)
log_population	-0.001*	-0.001**	-0.001	-0.002***	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log_pop_den	-0.001	-0.001	-0.001	-0.002***	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Interpolation of poppct_rural on year	0.007**	0.009***	0.005	0.011***	0.005	0.006*	0.006	0.005	0.005
	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
share_foreign	-0.033**	-0.036**	0.010	-0.047**	0.025	-0.000	0.018	0.004	0.022
	(0.015)	(0.015)	(0.022)	(0.018)	(0.029)	(0.022)	(0.030)	(0.023)	(0.030)
share_moved	-0.032	-0.033	-0.034	-0.028	-0.036	-0.034	-0.044	-0.030	-0.038
	(0.024)	(0.024)	(0.026)	(0.025)	(0.026)	(0.026)	(0.027)	(0.026)	(0.028)
(mean) share_white_not_hispanic	0.024***	0.024***	0.010	0.018**	0.006	0.009	0.008	0.007	0.004
	(0.006)	(0.006)	(0.008)	(0.007)	(0.009)	(0.008)	(0.009)	(0.009)	(0.010)
(mean) share_ed_nohs	-0.050*	-0.050*	-0.027	0.008	-0.012	-0.009	-0.001	-0.022	-0.017
	(0.029)	(0.030)	(0.036)	(0.032)	(0.039)	(0.036)	(0.041)	(0.038)	(0.042)
share_collegep	-0.005	-0.009	0.012	-0.026	0.006	0.001	0.002	0.013	0.019
	(0.015)	(0.016)	(0.018)	(0.016)	(0.022)	(0.019)	(0.023)	(0.021)	(0.025)
(mean) share_below_pov	-0.010	-0.009	-0.012	0.036*	0.003	0.005	-0.003	0.006	-0.001
	(0.017)	(0.017)	(0.022)	(0.020)	(0.023)	(0.022)	(0.023)	(0.022)	(0.023)
log_median_income	-0.001	-0.002	0.004	-0.015***	-0.004	-0.001	0.001	0.000	0.002
	(0.005)	(0.005)	(0.006)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
dem_share2016	-0.012*	-0.019**	-0.005	-0.021**	-0.005	-0.007	-0.009	0.001	0.004
	(0.007)	(0.008)	(0.009)	(0.008)	(0.012)	(0.010)	(0.012)	(0.014)	(0.019)
dem_share2020	0.003	-0.000	0.008	-0.024***	-0.009	-0.006	-0.010	0.007	0.009
	(0.006)	(0.013)	(0.009)	(0.009)	(0.012)	(0.010)	(0.012)	(0.017)	(0.022)
Observations	3107	3107	3101	3107	2988	3101	2981	3101	2981

Table 11: Predictors of Changes in Congruence - 2020

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(
single_district	-0.001	0.000	0.000	-0.004*	0.000	0.000	-0.004*	0.000	0.0
	(0.002)	(.)	(.)	(0.002)	(.)	(.)	(0.002)	(.)	(
log_population	-0.000	0.001	-0.002	-0.000	0.001	-0.002	-0.000	0.001	-0.
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.0
log_pop_den	0.002	0.000	0.004*	0.000	-0.000	0.004*	0.000	-0.000	0.0
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.0
Interpolation of poppct_rural on year	0.011	0.012	0.012	0.010	0.011	0.012	0.010	0.011	0.0
	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.0
share_foreign	0.025	-0.008	0.018	0.029	0.004	0.017	0.023	-0.025	-0.
	(0.029)	(0.042)	(0.050)	(0.029)	(0.042)	(0.050)	(0.036)	(0.049)	(0.0
share_moved	-0.000	0.012	0.008	0.014	0.011	0.008	0.013	0.008	0.0
	(0.029)	(0.035)	(0.036)	(0.030)	(0.035)	(0.036)	(0.030)	(0.035)	(0.0
(mean) share_white_not_hispanic	0.016*	-0.008	0.006	0.026***	0.004	0.006	0.026***	0.001	0.0
	(0.009)	(0.013)	(0.015)	(0.010)	(0.016)	(0.017)	(0.010)	(0.016)	(0.0
(mean) share_ed_nohs	-0.041	0.046	-0.004	-0.010	0.044	-0.004	-0.011	0.038	-0.
	(0.044)	(0.060)	(0.068)	(0.045)	(0.061)	(0.068)	(0.044)	(0.060)	(0.0
share_collegep	-0.018	0.022	-0.034	-0.036	0.006	-0.033	-0.038	0.004	-0.
	(0.026)	(0.032)	(0.037)	(0.027)	(0.035)	(0.040)	(0.027)	(0.035)	(0.0
(mean) share_below_pov	-0.044	-0.073*	-0.088**	-0.053	-0.077*	-0.088**	-0.054	-0.080*	-0.0
	(0.039)	(0.042)	(0.042)	(0.039)	(0.042)	(0.042)	(0.040)	(0.042)	(0.0
log_median_income	-0.014	-0.008	-0.004	-0.010	-0.006	-0.004	-0.009	-0.004	-0.
	(0.011)	(0.012)	(0.013)	(0.011)	(0.012)	(0.013)	(0.011)	(0.013)	(0.0
dem_share2008				0.033***	0.021	-0.001	0.043*	0.048*	0.0
				(0.010)	(0.018)	(0.020)	(0.022)	(0.026)	(0.0
Fixed effects	No	DMA, State	DMA, Dist	No	DMA, State	DMA, Dist	No, DI	DMA+State+DI	DMA+

Table 12: Predictors of Changes in Congruence - 2010 - Joint Tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
single_district	-0.003	0.000	0.000	-0.003	0.000	0.000	-0.003	0.000	0
	(0.002)	(.)	(.)	(0.002)	(.)	(.)	(0.002)	(.)	
log_population	-0.001	0.000	0.001	-0.001	0.000	0.001	-0.001	0.001	0
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0
log_pop_den	0.001	-0.001	-0.002	0.001	-0.001	-0.002	0.001	-0.001	-0
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0
Interpolation of poppct_rural on year	0.008	0.005	0.005	0.008	0.005	0.005	0.008	0.005	0
	(0.006)	(0.007)	(0.008)	(0.006)	(0.007)	(0.008)	(0.006)	(0.007)	(0
share_foreign	0.071***	0.058	0.078*	0.073***	0.061*	0.079*	0.073***	0.055	0
	(0.027)	(0.036)	(0.045)	(0.027)	(0.035)	(0.044)	(0.027)	(0.038)	(0
share_moved	-0.021	-0.025	-0.045	-0.018	-0.025	-0.045	-0.016	-0.025	-0
	(0.029)	(0.033)	(0.034)	(0.029)	(0.033)	(0.034)	(0.030)	(0.033)	(0
(mean) share_white_not_hispanic	0.029***	0.012	0.010	0.035***	0.016	0.011	0.033***	0.016	0
	(0.009)	(0.012)	(0.014)	(0.010)	(0.017)	(0.018)	(0.011)	(0.017)	(0
(mean) share_ed_nohs	-0.108**	-0.059	-0.056	-0.100**	-0.060	-0.056	-0.098**	-0.061	-0
	(0.048)	(0.058)	(0.065)	(0.048)	(0.058)	(0.065)	(0.048)	(0.058)	(0
share_collegep	-0.019	0.009	0.012	-0.033	0.002	0.011	-0.032	0.004	0
	(0.025)	(0.030)	(0.034)	(0.028)	(0.037)	(0.041)	(0.028)	(0.037)	(0
(mean) share_below_pov	0.045	0.034	0.015	0.041	0.034	0.014	0.041	0.033	0
	(0.038)	(0.039)	(0.042)	(0.038)	(0.039)	(0.042)	(0.038)	(0.040)	(0
log_median_income	-0.007	-0.000	-0.005	-0.005	0.001	-0.005	-0.006	0.001	-0
	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.014)	(0.012)	(0.013)	(0
dem_share2016				0.014	0.007	0.001	0.008	0.012	0
				(0.012)	(0.020)	(0.023)	(0.015)	(0.024)	(0
Fixed effects	No	DMA, State	DMA, Dist	No	DMA, State	DMA, Dist	No	DMA+State+DI	DMA-

Table 13: Predictors of Changes in Congruence - 2020 - Joint Tests

B.2.2 Variable Descriptions

C Additional Empirical Results

C.1 Congruence Descriptive Statistics

C.1.1 Correlates of Congruence

Tables 6 and 7 summarize how congruence varies with predictors of district boundaries and social networks. Every cell of the table represents a separate regression. These tables also show how congruence is associated with socioeconomic, demographic, and political characteristics of counties' populations, as these characteristics are likely to be correlated with the outcomes of interest. County population, population density, and the share of the population that is rural comes from the 2010 or 2020 Decennial Census, accordingly. Democratic presidential vote shares are from MIT Election Data and Science Lab 2021. The remaining demographics come from the 2010-2014 ACS for 2010 and the 2015-2019 5-Year ACS for 2020.[27] The determinants of district boundaries – whether the state has a single district, and the county's population – significantly predict congruence. In particular, as expected, counties in single district states are more congruent. Population is negatively correlated with congruence. Congruence is increasing in the share of the county that is rural. Congruence is decreasing in population density. Congruence is negatively correlated with the share of the population that was born outside of the U.S. and the share that moved within the last year: these are both indicators of more mobile populations, which intuitively are more likely to have more geographically dispersed networks. Congruence is positively correlated with the share of the population that is white as well as the share of the population without a high school degree and the share of the population below the poverty line. Analogously, the share of the population with at least a college degree is negatively associated with congruence. Concordantly, congruence is negatively correlated with the Democratic presidential vote share. Tables 8 and 9 show the analogous joint tests. **C.1.2 Changes in Congruence**

C.2 Voters' Knowledge

C.2.1 Placebo Tests for Voter Information

Table 14 provides summary statistics for the nine outcome variables used for placebo tests. These distributions are generally similar to those for House representatives (Table 3), though respondents are generally more likely to recognize and select the correct party for their Senators and Governor. The following nine figures show the results of the placebo tests. In general, congruence does not significantly predict the placebo outcomes. **C.2.2 Commuting Congruence**

Table 15 shows results from specifications that construct congruence using commuting flows. *[Table results are out of date, to be updated/use event studies instead.]* **C.2.3 Border Pairs**

Variable	**Observations**	**Mean (%)**	**SD (pp)**
Heard of Governor	549,740	96.3	18.8
Selected Governor Party	608,985	81.3	39.0
Selected Correct Gov. Party	608,985	74.7	43.5
Heard of Senator 1	549,244	94.9	22.0
Selected Senator 1 Party	608,414	75.0	43.3
Selected Correct Sen. 1 Party	608,414	67.7	46.7
Heard of Senator 2	549,246	94.9	22.0
Selected Senator 2 Party	608,402	74.3	43.7
Selected Correct Sen. 2 Party	608,402	66.9	47.0

Table 14: CES Data: Summary Statistics for Placebo Outcomes

	(1) County & Year FEs Only	(2) Add District x Year FEs	(3) Add Individual Demographic Controls	(4) Add County-Year Controls
	Heard of Incumbent			
Commuting Congruence	0.050**	0.048*	0.049**	0.065***
	(0.023)	(0.025)	(0.024)	(0.025)
	[0.032]	[0.050]	[0.042]	[0.008]
Obs	544,910	544,910	544,910	401,874
R^2	0.032	0.078	0.130	0.134
	Selected Party			
Commuting Congruence	0.149***	0.112***	0.132***	0.131***
	(0.039)	(0.037)	(0.033)	(0.036)
	[0.000]	[0.002]	[0.000]	[0.000]
Obs	603,967	603,966	603,966	460,937
R^2	0.042	0.098	0.224	0.225
	Selected Correct Party			
Commuting Congruence	0.135***	0.117***	0.139***	0.150***
	(0.041)	(0.039)	(0.035)	(0.037)
	[0.001]	[0.003]	[0.000]	[0.000]
Obs	603,967	603,966	603,966	460,937
R^2	0.044	0.108	0.251	0.251
Ind. Controls			X	X
County x Year Controls				X
FEs	County, Year	County, District x Year	County, District x Year	County, District x Year

*** p<0.01, **p<0.05, * p<0.1
Standard errors clustered at the county level in parentheses. P-values in square brackets.
"Heard of Incumbent" not available in 2006, 2007, or 2009. Individual controls include gender, race, education, age categories, and whether the respondent is affiliated with the same party as their representative. County-by-year controls include population and shares by race, age categories, gender, and county urban population share.

Table 15: Effect of Commuting Congruence on Voter Familiarity with Representative

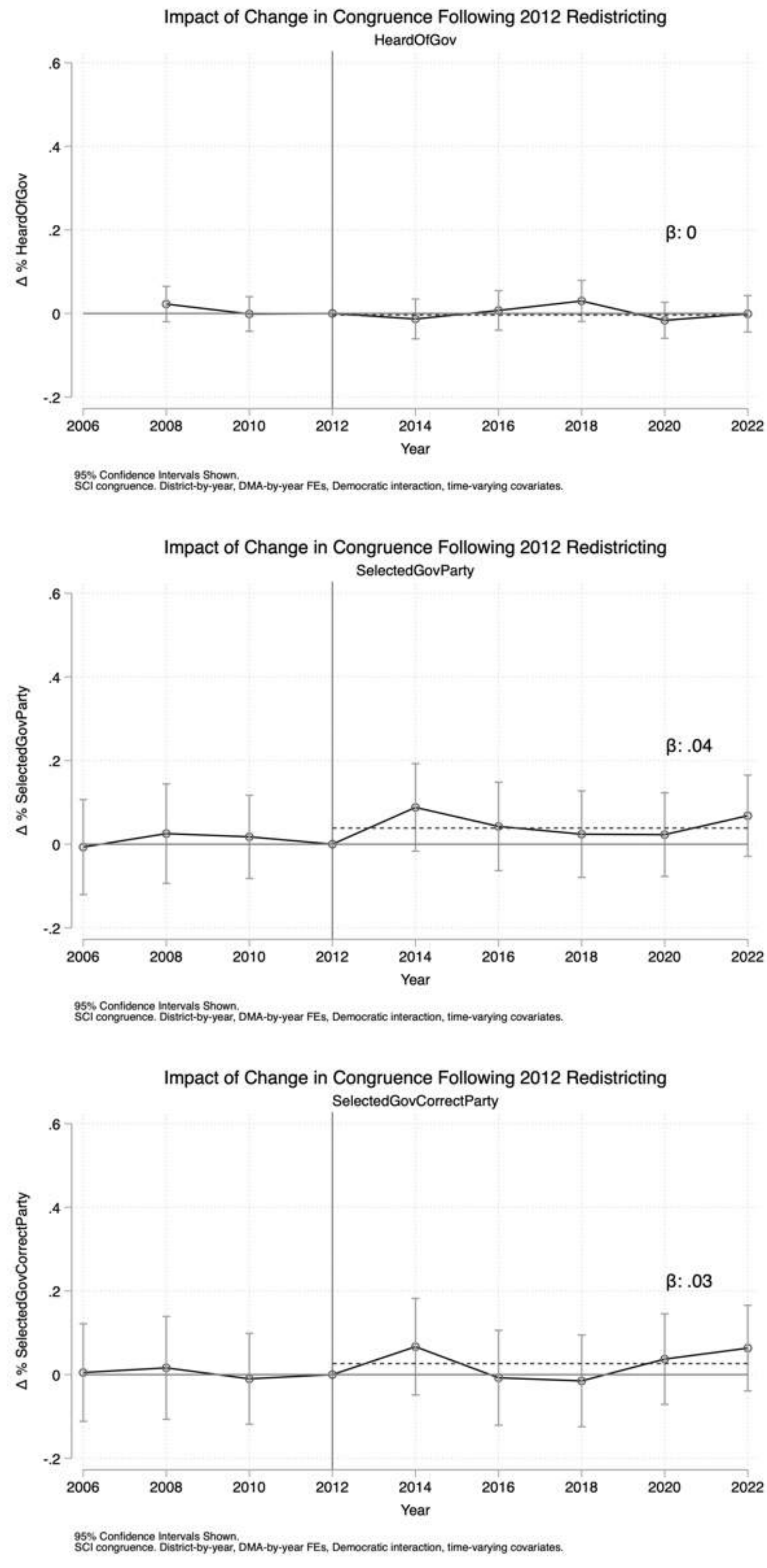


Figure 11: Effect of Increase in Congruence on Knowledge of Governor

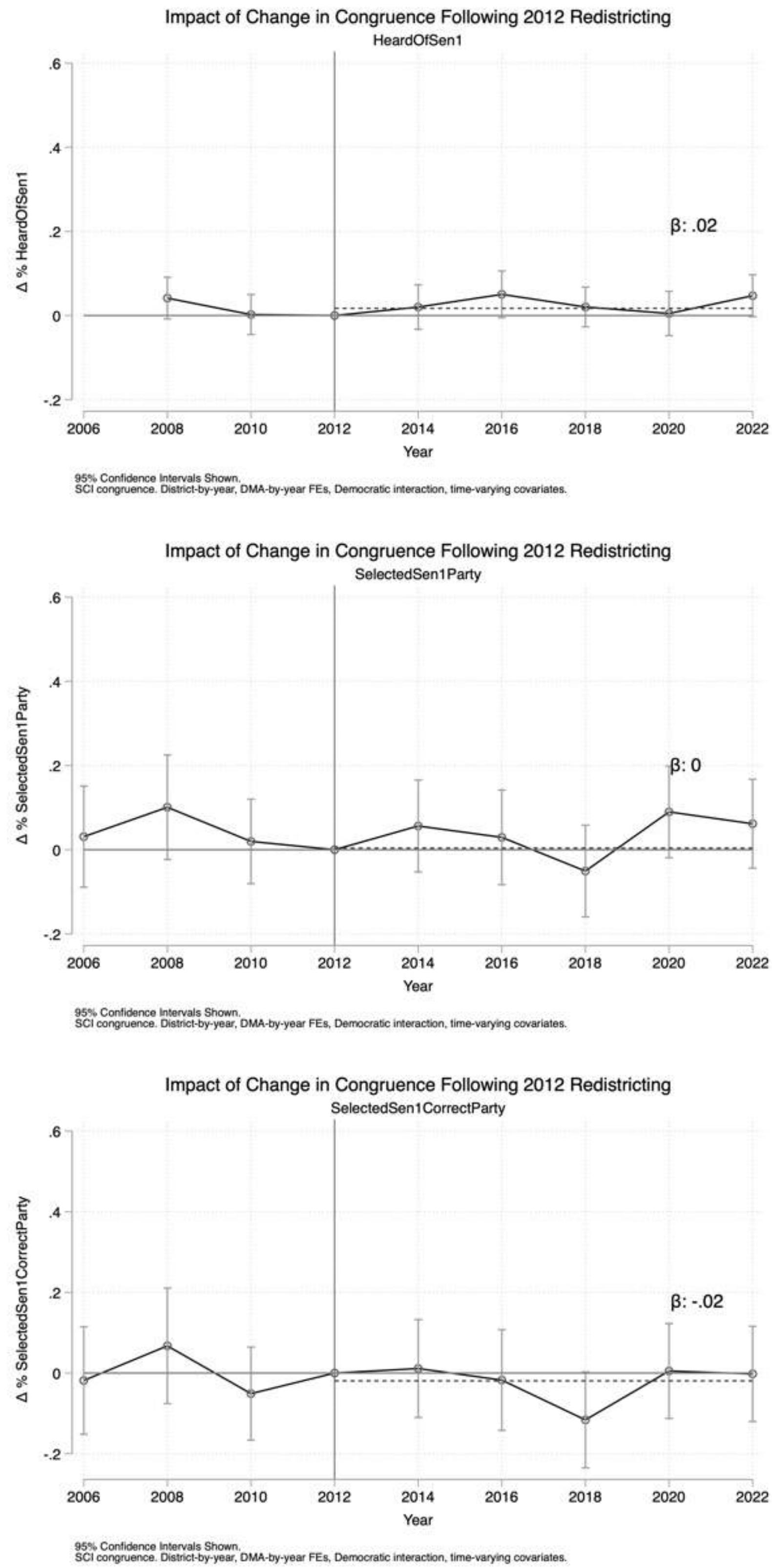


Figure 12: Effect of Increase in Congruence on Knowledge of Senator 1

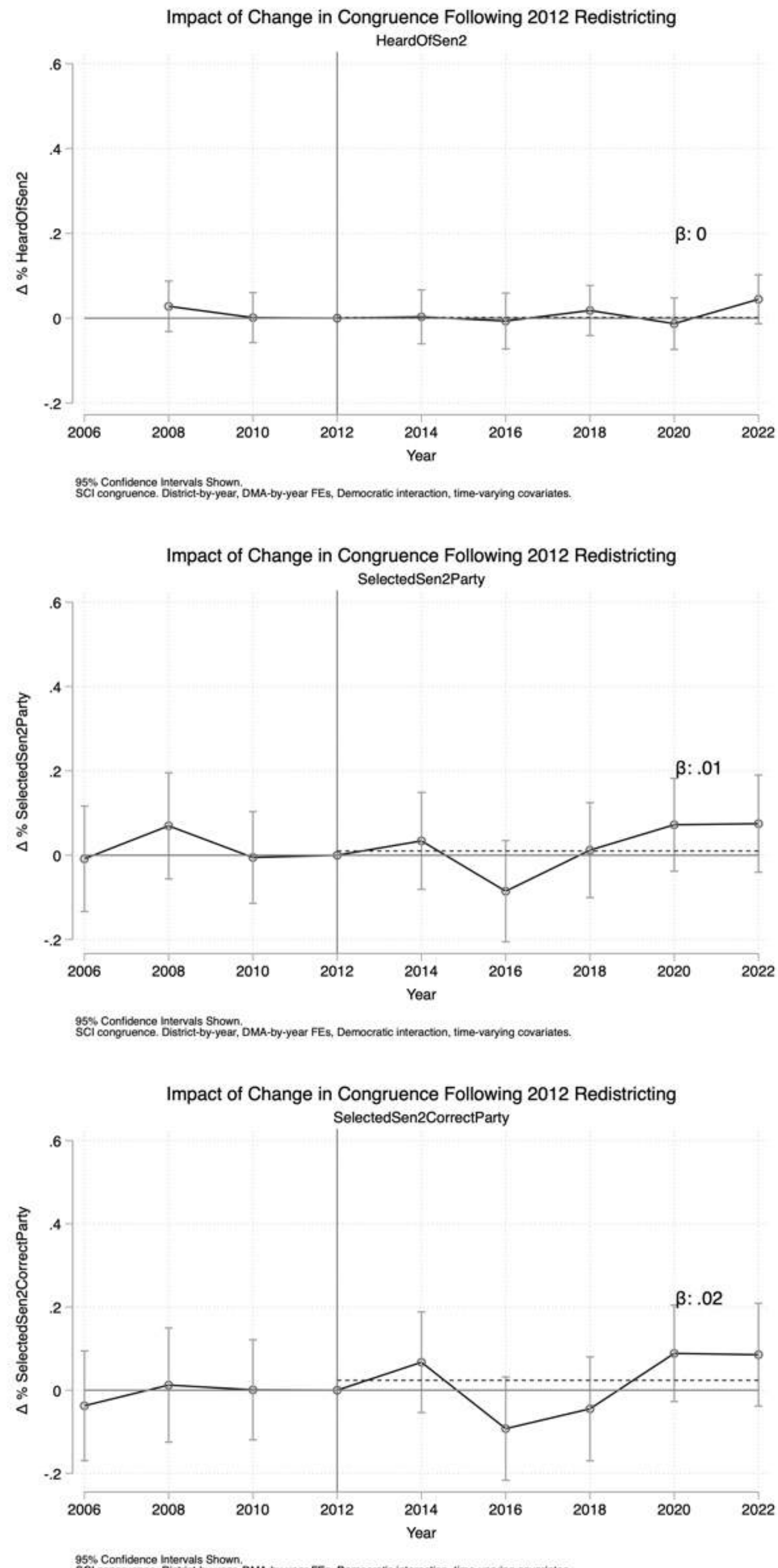


Figure 13: Effect of Increase in Congruence on Knowledge of Senator 1

	(1) County & Pair x Year FEs Only	(2) Add State x Year FEs	(3) Add DMA x Year FEs	(4) Add Dem. Exposure	(5) Add Individual Demographic Controls	(6) Add County-Year Controls
	Heard of Incumbent					
Congruence	0.254***	0.254***	0.296***	0.294***	0.253***	0.256***
	(0.080)	(0.080)	(0.091)	(0.091)	(0.093)	(0.091)
	[0.002]	[0.002]	[0.001]	[0.001]	[0.007]	[0.005]
Obs	22,094	22,094	21,508	21,508	21,508	21,508
R^2	0.606	0.606	0.699	0.699	0.718	0.720
	Selected Party					
Congruence	0.368***	0.368***	0.214	0.217	0.162	0.136
	(0.141)	(0.141)	(0.155)	(0.156)	(0.151)	(0.151)
	[0.009]	[0.009]	[0.168]	[0.165]	[0.284]	[0.367]
Obs	25,798	25,798	25,126	25,126	25,126	25,126
R^2	0.620	0.620	0.709	0.709	0.740	0.742
	Selected Correct Party					
Congruence	0.608***	0.608***	0.399**	0.406**	0.346**	0.343**
	(0.161)	(0.161)	(0.176)	(0.176)	(0.165)	(0.160)
	[0.000]	[0.000]	[0.023]	[0.022]	[0.036]	[0.032]
Obs	25,798	25,798	25,126	25,126	25,126	25,126
R^2	0.631	0.631	0.716	0.717	0.749	0.752
Dem. Exposure				X	X	X
Ind. Controls					X	X
County x Year Controls						X
FEs	County, Pair x Year	County, Pair x Year, District x Year	County, Pair x Year, District x Year, DMA x Year	County, Pair x Year, District x Year, DMA x Year	County, Pair x Year, District x Year, DMA x Year	County, Pair x Year, District x Year, DMA x Year

*** p<0.01, **p<0.05, * p<0.1
Standard errors clustered at the county level in parentheses. P-values in square brackets.
"Heard of Incumbent" not available in 2006, 2007, or 2009. Individual controls include gender, race, education, age categories, and whether the respondent is affiliated with the same party as their representative. County-by-year controls include population and shares by race, age categories, gender, and county urban population share.

Table 16: Effect of Congruence on Voter Familiarity with Representative, within Border Pairs

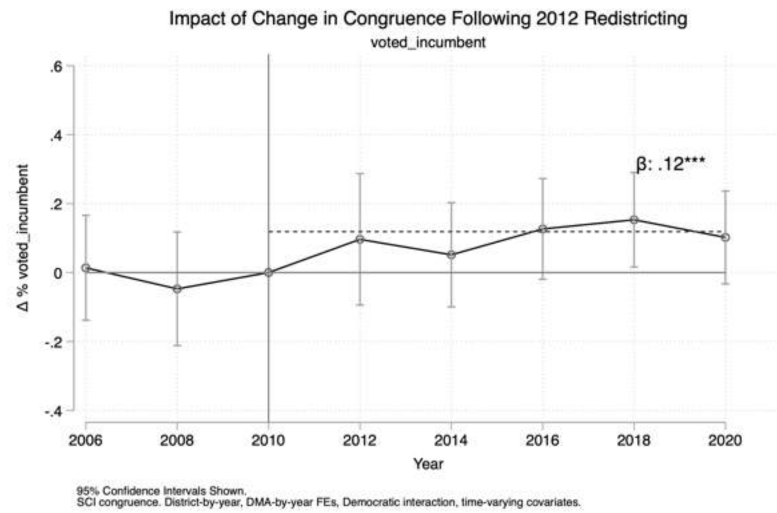
explored in prior literature (Snyder and Strömberg 2010, Prat and Strömberg 2005, Enikolopov, Petrova, et al. 2011, Angelucci et al. 2020, Eisensee and Strömberg 2007, Strömberg 2004). In the main results, I use Nielsen Designated Market Area (DMA) regions, which reflect TV and radio markets, in order to include DMA-by-year fixed effects: advertisements are purchased at the DMA level, so all counties DMA-wide receive the same advertisements. Consequently, DMA-by-year fixed effects allow me to capture the impact of congruence *holding fixed* TV and radio news and advertisements. An alternative approach, that could address concerns about social networks driving a county's influence on the purchase of advertising in the media market, is to construct a measure of "DMA congruence." Such a concern might be based on the logic in Snyder and Strömberg 2010 that a media market that is better aligned with a congressional district will produce more content about the district's representative, thereby making voters in that district more informed about their representative.[28] Conveniently, DMAs contain many counties, but DMA borders follow county borders, so no county is in multiple DMAs. Accordingly, for county i in district J, which contains counties j, and in DMA M, which contains counties m, DMA congruence is defined as

$$\text{DMA Congruence}_i = \frac{\sum_{j\in(J\cap M)} \text{Population}_j}{\sum_{m\in M} \text{Population}_m}$$

	(1) County & Year FEs Only	(2) Add District x Year FEs	(3) Add Individual Demographic Controls	(4) Add County-Year Controls
		Heard of Incumbent		
Congruence	0.067	0.118***	0.133***	0.145***
	(0.041)	(0.046)	(0.045)	(0.047)
	[0.100]	[0.010]	[0.003]	[0.002]
DMA Congruence	0.054***	0.030*	0.028	0.035*
	(0.016)	(0.018)	(0.017)	(0.018)
	[0.001]	[0.096]	[0.102]	[0.054]
Obs	460,951	460,951	460,951	402,066
R^2	0.035	0.083	0.132	0.133
		Selected Party		
Congruence	0.191**	0.214***	0.269***	0.278***
	(0.086)	(0.078)	(0.074)	(0.077)
	[0.027]	[0.006]	[0.000]	[0.000]
DMA Congruence	0.122***	0.093***	0.088***	0.098***
	(0.035)	(0.032)	(0.030)	(0.032)
	[0.000]	[0.003]	[0.004]	[0.002]
Obs	520,034	520,033	520,033	461,154
R^2	0.046	0.103	0.211	0.210
		Selected Correct Party		
Congruence	0.123	0.194**	0.253***	0.280***
	(0.093)	(0.082)	(0.077)	(0.079)
	[0.189]	[0.019]	[0.001]	[0.000]
DMA Congruence	0.131***	0.115***	0.108***	0.121***
	(0.038)	(0.032)	(0.030)	(0.032)
	[0.001]	[0.000]	[0.000]	[0.000]
Obs	520,034	520,033	520,033	461,154
R^2	0.047	0.113	0.236	0.235
Ind. Controls			X	X
County x Year Controls				X
FEs	County, Year	County, District x Year	County, District x Year	County, District x Year

*** $p<0.01$, ** $p<0.05$, * $p<0.1$
Standard errors clustered at the county level in parentheses. P-values in square brackets.
"Heard of Incumbent" not available in 2006, 2007, or 2009. Individual controls include gender, race, education, age categories, and whether the respondent is affiliated with the same party as their representative. County-by-year controls include population and shares by race, age categories, gender, and county urban population share.

Table 17: Effect of Congruence and DMA Congruence on Voter Familiarity with Representative



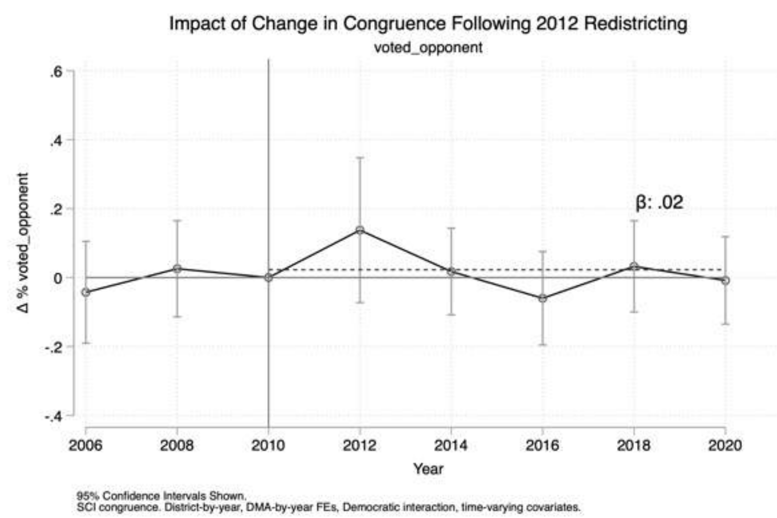


Figure 14: Dynamic Effects of Change in Congruence on Voter Choices Reported in CES - Incumbent vs Challenger

I can then conduct a horse race and test whether I estimate a significant effect of my original congruence measure after controlling for DMA congruence.[29] Results are shown in Appendix Table 17. *[Table results are out-of-date, to be updated/use event studies instead.]* In general, results are similar to the original specification.

C.3 Voters' Choices

C.3.1 CES Survey Responses

In the post-survey, respondents are asked who they voted for. Based on these self-reports, I find that congruence increases the share of general election voters reporting voting for the incumbent, without affecting the share who vote for the challenger. Instead, as described in the main text, the share of general election voters who report skipping the House election decreases.

C.3.2 Vote Count Data

Using county-level vote count data, I look at the impact of a change in congruence on actual turnout in the top-of-ticket (i.e., President, Senate, or Governor) election and turnout in the House election. I measure this as total votes in the respective election divided by the share of the county's population that is age 18 or older (i.e., the voting age population). While the voting age population does not account for those who are ineligible to vote (e.g., non-citizens or people with a felony record in some states), I do control for the share of the population that are non-citizens in the set of county-by-year covariates. The sample includes only counties that have at least 50% of population in one district, and district-by-year fixed effects reflect that district. (Results are qualitatively similar when restricting to counties 100% in a single district.) Here we find no impacts of congruence on turnout in the top-of-ticket race and in the House race. However, this does not necessarily contradict the finding of congruence decreasing roll-off: When I construct roll-off as $\frac{\text{Votes in Top-of-Ticket Election} - \text{Votes in House Election}}{\text{Voting Age Population}}$, I find a significant 2pp decrease in roll-off. *[The spike in 2004 I am investigating.]* This is because the estimates of House turnout are too imprecise to detect an impact of this scale, and also have a slight negative pre-trend; in the roll-off estimates, controlling for the top-of-ticket election both reduces the noise in the estimates and eliminates much of the pre-trend.

C.3.3 Campaign Contributions

Congruence has no impact on total contributions to House candidates; instead, it increases contributions to in-district candidates at the cost of contributions to out-of-district candidates. I also find similar results

[28] Of course, Snyder and Strömberg 2010 actually find no effect of TV and radio market congruence in their paper (though it is constructed somewhat differently), and instead find effects of newspaper market congruence, which I have not yet included

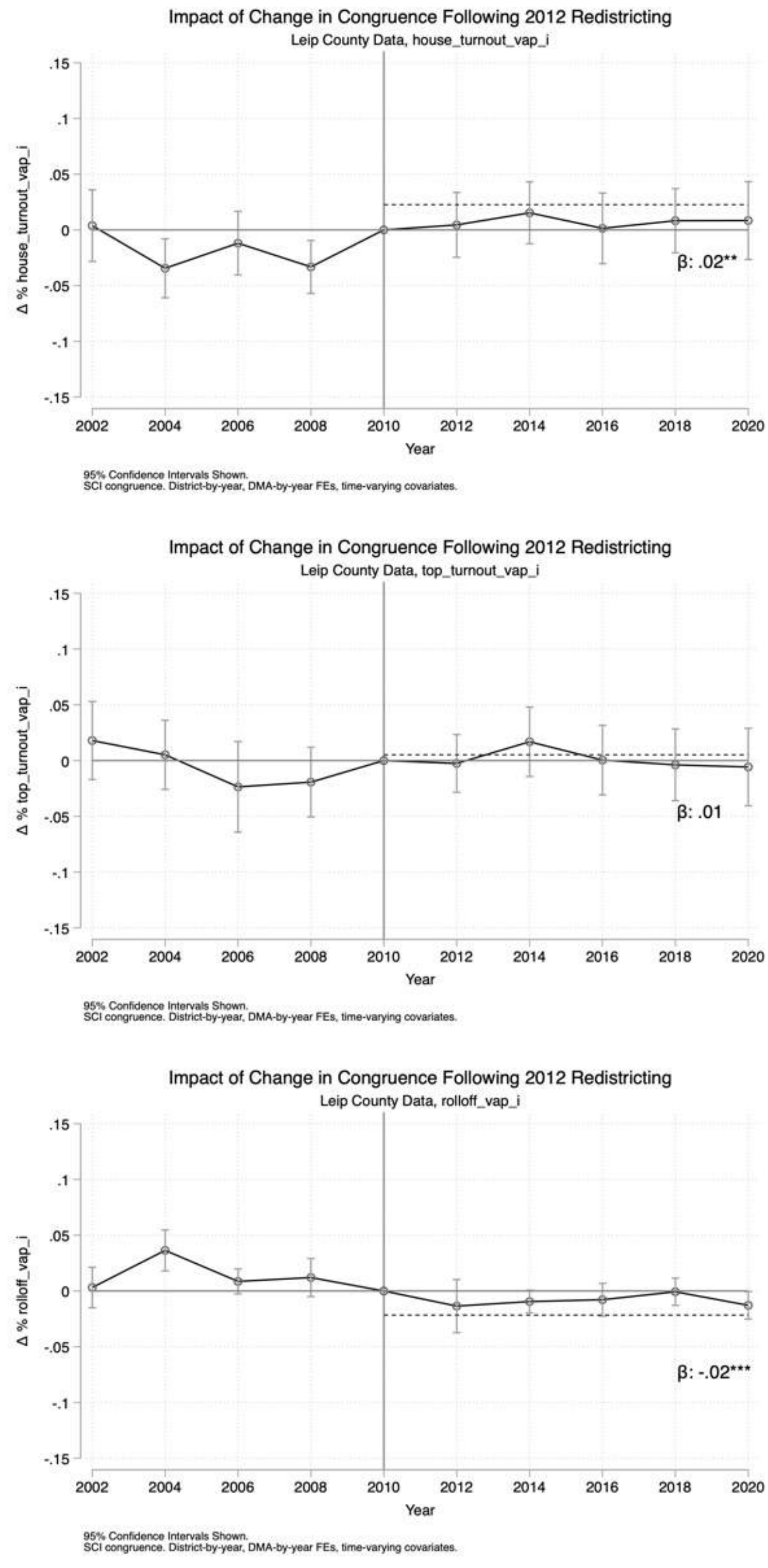


Figure 15: Effects of Increase in Congruence on Turnout

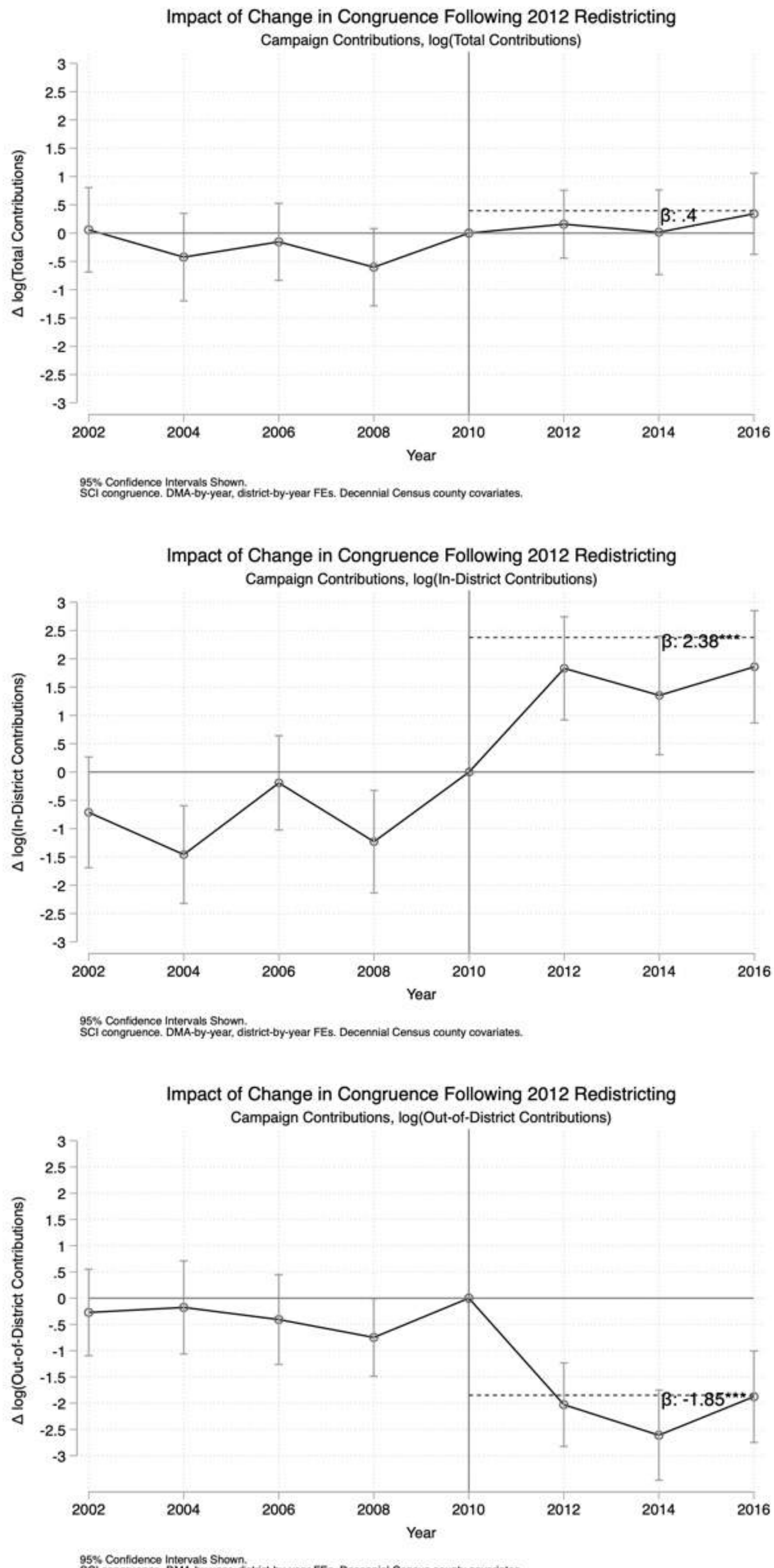


Figure 16: Effects of Increase in Congruence on Log of Campaign Contributions

when restricting to only primary elections; only general elections; or excluding large donations (i.e., excluding Census tracts where the average donation per contributor exceeds $1,000).

D Simulations

D.1 Additional Background on Ohio

D.2 Simulated Ohio Maps

[Add maps and table for Democrat map vs Republican map]

D.3 Simulated Texas Maps

here.

[29]DMA congruence and the originally defined county congruence are about 25% correlated.

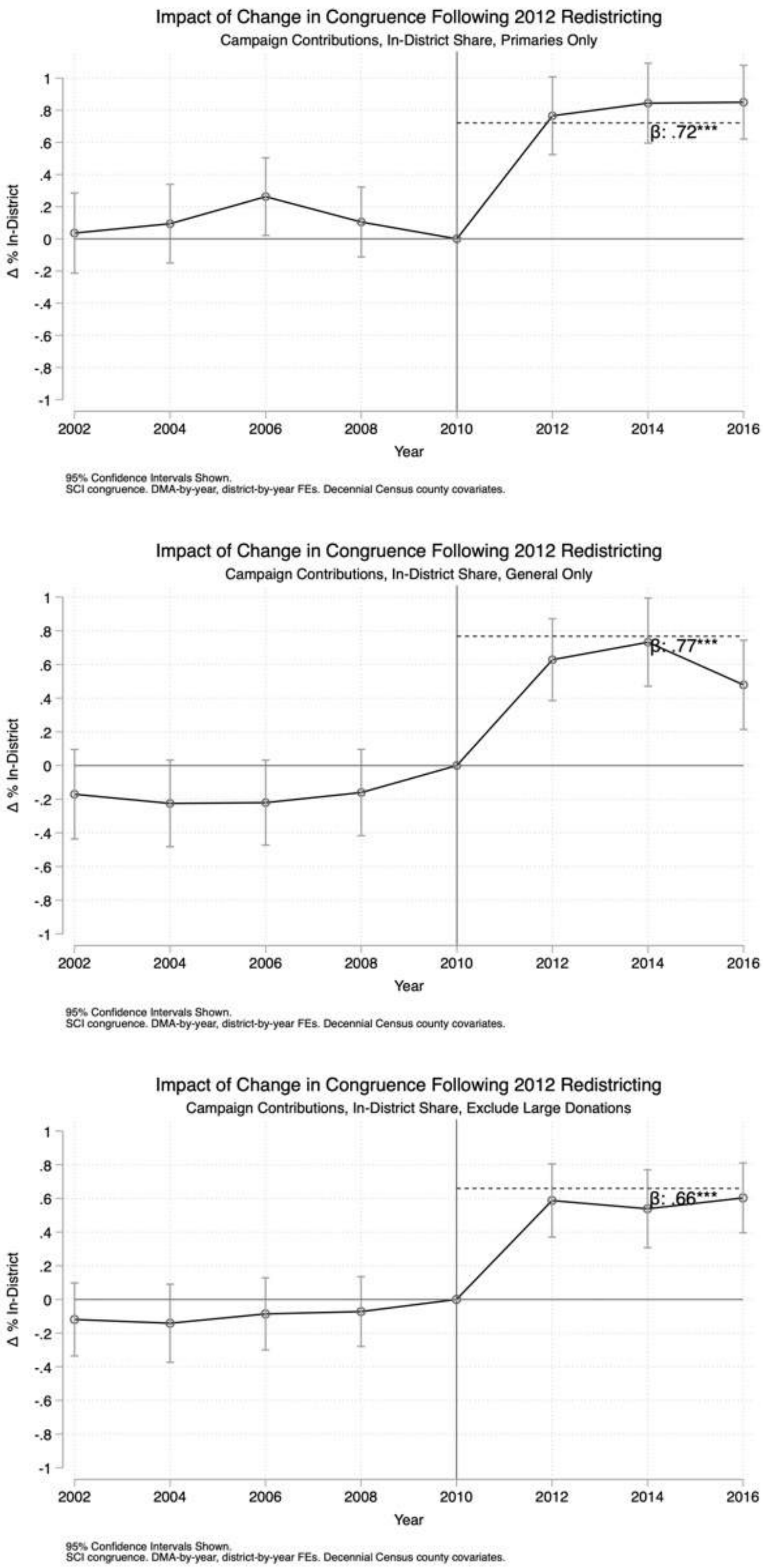


Figure 17: Effects of Increase in Congruence on Share of Campaign Contributions to In-District Candidates

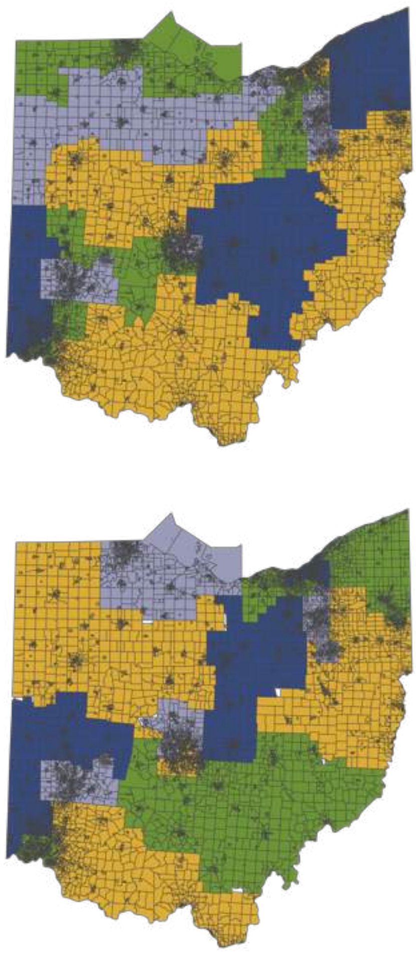


Figure 18: Left: Current Enacted Map in Ohio; Right: OCRC Map

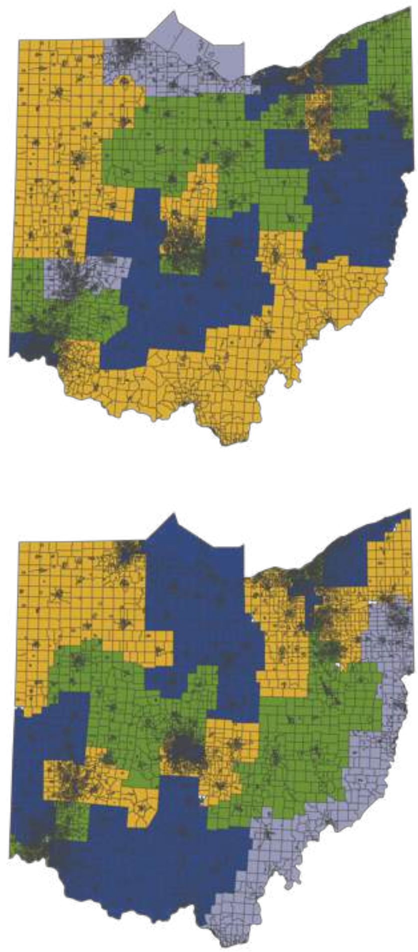


Figure 19: Partisan Maps Proposed in State Legislature Commissions. Left: Democrats' Map; Right: Republicans' Map

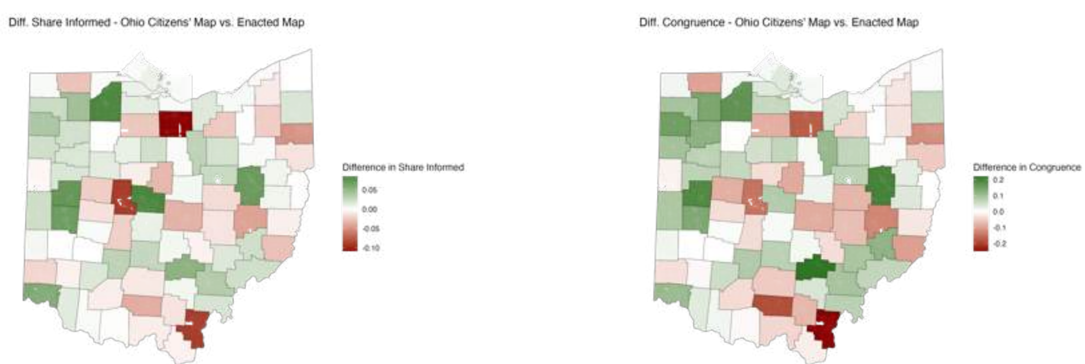


Figure 20: OCRC map vs. Enacted Map. Left: Difference in Congruence; Right: Difference in Share Informed.

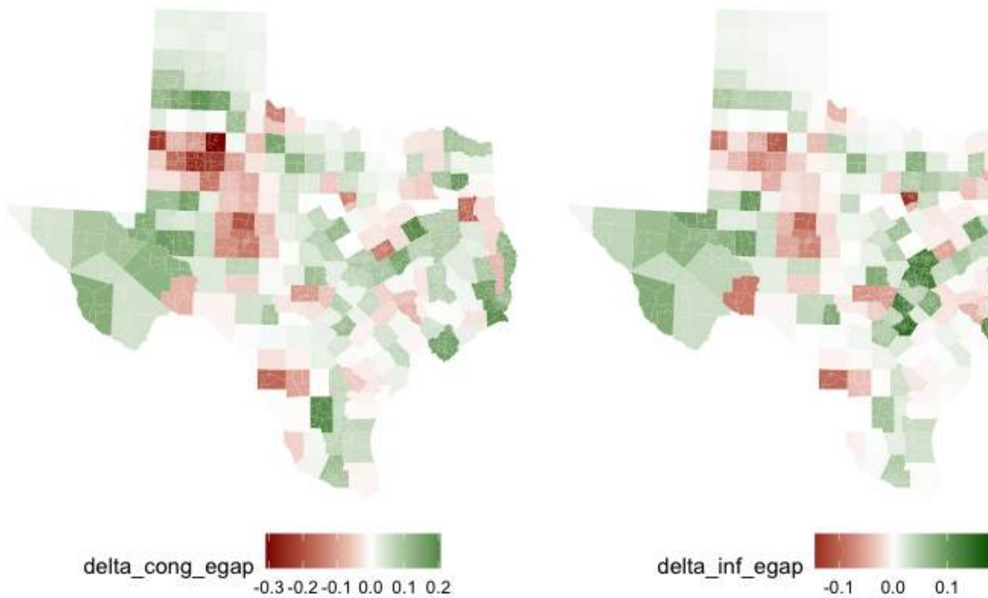


Figure 21: Change in (left) congruence, (right) share informed, from current map to efficiency gap minimizing map

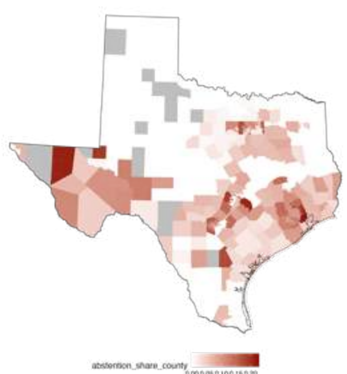


Figure 22: Simulated County Abstention Rates in Texas Under Current Map