State Politics & Policy Quarterly

VOLUME 23
NUMBER 2
JUNE 2023

CAMBRIDGE UNIVERSITY PRESS

The official journal of the State Politics and Policy section of the American Political Science Association

ISSN: 1532-4400

State Politics & Policy Quarterly

https://doi.org/10.1017/spq.2023.9 Published online by Cambridge University Press

State Politics & Policy Quarterly
Volume 23 Number 2 June 2023
The official journal of the State Politics and Policy section of the American Political Science Association

Contents

Articles

Short Article

https://doi.org/10.1017/spq.2023.9 Published online by Cambridge University Press

State Politics & Policy Quarterly (2023), **23**: 2, 117–139
doi:10.1017/spq.2022.16

ORIGINAL ARTICLE

Do Male and Female Legislators Have Different Twitter Communication Styles?

Daniel M. Butler[1], Thad Kousser[2] and Stan Oklobdzija[3]

[1]Department of Political Science, Washington University in St. Louis, St. Louis, MO
[2]Department of Political Science, University of California, San Diego, La Jolla, CA, USA
[3]School of Public Policy, University of California, Riverside, Riverside, CA, USA
Corresponding author: Thad Kousser, email: tkousser@ucsd.edu

(Received 12 February 2021; revised 31 August 2021; accepted 22 September 2021)

Abstract
Communication is a fundamental step in the process of political representation, and an influential stream of research hypothesizes that male and female politicians talk to their constituents in very different ways. To build the broad dataset necessary for this analysis, we harness the massive trove of communication by American politicians through Twitter. We adopt a supervised learning approach that begins with the hand coding of over 10,000 tweets and then use these to train machine learning algorithms to categorize the full corpus of over three million tweets sent by the lower house state legislators who were serving in the summer of 2017. Our results provide insights into politicians' behavior and the consequence of women's underrepresentation on what voters learn about legislative activity.

Keywords: state legislators; gender and Twitter; constituent communication

Elected officials' communication efforts are an important part of representation (Fenno 1977; Grimmer 2013). Historically, traditional media outlets have been a major way that politicians have reached their constituents. However, traditional media outlets have demonstrated bias against women, making it harder for them to reach voters (Baitinger 2015; Heldman, Carroll, and Olson 2005; Kahn 1992, 1994). This poses a challenge for women in politics; they need to reach voters but face obstacles in utilizing a primary avenue, traditional media outlets, for doing so. The rise of Twitter and other social media tools represents one way that women may be able to overcome bias in media coverage by allowing politicians to circumvent traditional media and directly reach voters. In this paper, we test whether women state legislators are more likely to use Twitter. We also explore whether gender predicts how politicians communicate with the public through Twitter.

We evaluate several hypotheses to learn about the differences between men and women. Some of these hypotheses tell us about what motivates politicians when they think about their election prospects. For example, previous work has argued that female politicians have strong incentives to portray themselves as conservative in order to counteract the stereotype that they are more liberal than their male, co-partisan colleagues (Koch 2000). Another stream of research shows that women work harder in political office (Kurtz et al. 2006), speak more in the legislature (Pearson and Dancey 2011), put more effort into their constituency service (Thomsen and Sanders 2020), and produce more legislation (Anzia and Berry 2011; Volden, Wiseman, and Wittmer 2013), suggesting that gender may affect how much effort state legislators will put into the time-consuming task of social media communication.

Other hypotheses have important implications for levels of descriptive representation. Previous work has found that politicians' communications can bias the information environment for voters. For example, Grimmer (2013) finds that politicians with more extreme preferences are more likely to communicate about policy issues and moderate politicians are more likely to discuss nonpolicy-related funding that they secure for their district. The differences in how these groups communicate allow extremist views to dominate the public policy debate. In a similar way, if men and women communicate differently, this has implications for what voters hear because women are underrepresented in office. Previous work has argued that women tend to work more on issues like education and health care (Foerstel and Foerstel 1996; Reingold 1992; Saint-Germain 1989; Swers 1998, 2002; Thomas 1994; Thomas and Welch 1991). If women also discuss these issues more, then electing more women will lead voters to hear and learn more about these issues. We test whether gender predicts how much legislators communicate on these issues.

To build the broad dataset necessary to undertake this analysis, we harness the massive trove of communication by American politicians through Twitter. Just as it has become a highly visible mode of political discourse in national politics (Garofoli 2018), social media is now one of the primary modes of political communication for state legislators. As we show, a majority of members of state lower houses have public Twitter handles, with the average lawmaker tweeting over 1,000 times. Together, the lower house state legislators we study produced over three million tweets in our period of study from October 2015 to July 2018.

This wealth of data presents both an opportunity and a challenge for state politics scholars. Lawmakers in statehouses all across the nation are speaking via the same medium, and doing so both during campaign seasons and while they are governing. Holding the medium constant, scholars can study what they have to say before, during, and after elections, whether the electoral rules under which they run affect their ideological positioning, whether citizen lawmakers speak differently from those in professional legislatures, or whether polarized statehouses produced more negative discourse. With user engagement data, scholars can determine what forms of political communication followers are most apt to like or retweet and whether this varies by state and party. But in order to answer such questions, researchers must make sense of a mountain of data (for review of prior work, see Jungherr 2016; Vergeer 2015).

The modern tools of machine learning can aid in the task of classifying the topics, tone, and content of the enormous amount of data that state legislators are producing every day on Twitter. Machine learning techniques for text analysis can be divided into two approaches. In the first, "unsupervised learning," researchers mine data for attributes, such as the topics that cluster together, and then attribute meaning to the output of these algorithms. We adopt the second approach, "supervised learning," a hybrid between qualitative and quantitative techniques that begins by applying human judgment to code texts and then uses these codings to train machine learning algorithms (see Grimmer and Stewart 2013; Peterson and Spirling 2018). Only after testing how precisely the algorithms can replicate human codings, and ensuring sufficient accuracy, do we move onto the stage of classifying the full set of state political tweets. This approach, which we detail below, has been used to study political tweets in gubernatorial elections by McGregor et al. (2016), in the Australian Parliament by Kousser (2019), and for US presidential candidates by Kousser and Oklobdzija (2018). Here, we apply it to state legislators, producing the largest set of classified tweets, including 3,580,727 spanning 49 distinct political systems, that we have seen in the literature.

Four main findings emerge from our analysis. First, women communicate more than men. They are more likely to have Twitter accounts and to use them. Second, in contrast to previous work, we find that female legislators' tweets have a more positive tone than male legislators' tweets. Third, women do discuss women's issues more than their male counterparts, tweeting about both education policy and about health care policy more often. Fourth, gender does not appear to predict the ideological content of tweets after we control for legislators' roll call records.

In what follows, we first draw hypotheses about gender and legislator communication from the previous literature. We then describe and validate our original dataset and use it to test the hypotheses we lay out. We summarize the findings and their implications in the conclusion.

Theories about Gender and Social Media Communication

Allocating Resources to Twitter Communication

Before politicians decide what to tweet, they must first decide whether they will tweet at all and how often they will do so. This choice is a strategic choice because committing to establishing a social media presence requires a significant investment of time. To study the "tweet styles" of Australian legislators, Kousser (2019) draws upon Fenno's (1977) classic work on the home styles adopted by members of Congress in their districts. Kousser makes an analogy between Fenno's concept of the allocation of resources that representatives devote to connecting with their districts and the allocation of effort that today's lawmakers devote to connecting through social media.

While tweeting does not require the pecuniary investments that are necessary to set up and staff a district office or to fly home to meet with constituents, social media communication taxes a lawmaker's most vital resource: time. According to Fenno (1977, 890), "Of all the resources available to the House member, the scarcest and most precious one, which dwarfs all others in posing critical allocative dilemmas, is his time." Tweeting consistently requires a significant investment of time and attention from lawmakers. The price of this investment is magnified because most state legislators typically author their own Twitter feeds. They must do so while still fulfilling a host of other job commitments. Kurtz et al.'s (2006) survey, conducted long before social media added yet another demand to the busy

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

lives of state legislators, demonstrates the immense time commitment required by serving in a statehouse, even one that pays a small salary and is considered a part-time body.

We argue that allocating time toward tweeting is a costly activity whether the lawmaker communicates directly or indirectly to constituents.[1] How should a lawmaker's gender impact this allocational decision? We expect that female legislators will be more likely to establish a social media presence—both by creating a public Twitter account and by tweeting more often —than male legislators. There are a few reasons this might occur.

First, electoral discrimination might lead to "sex-based selection."[2] Anzia and Berry (2011) argue that "if voters discriminate against female candidates, only the most talented, hardest working female candidates will win elections" (478; see also Fulton 2012, 2014; Pearson and McGhee 2013). Consistent with this argument, Anzia and Berry (2011) find that female members of Congress in fact outperform men when it comes to securing district funding and sponsoring and co-sponsoring legislation (see also Volden et al. 2013). If this same sex-based selection mechanism operates in state legislatures, we should expect female lawmakers to work harder when it comes to social media communication.

Second, women in state politics may be more motivated to devote time to tweeting because they are simply responding to the demands that constituents are making of them. In a field experiment conducted in collaboration with state legislators, Butler et al. (2020) find that when men and women legislators make the same outreach to constituents, constituents are more likely to ask women legislators to do more work. Legislators are motivated by a desire to win reelection and so craft their homestyles in order to please voters. If constituents are asking more of women, women may in turn do more in order to be responsive. Although many studies show that female candidates perform very well in general elections (Burrell 1994; Fox 2006; Newman 1994; Seltzer, Newman, and Leighton 1997), this may be because they are doing more to meet voter demands rather than because voters are not demanding more of them.

Third, traditional media outlets might be biased against women (Baitinger 2015; Heldman, Carroll, and Olson 2005; Kahn 1992, 1994). Women legislators might prefer to reach their constituents through traditional news outlets but prior studies reveal that they are simply not be covered at the same rates as men (Heldman, Carroll, and Olson 2005; Kahn 1992). If they face obstacles to reaching voters through traditional news outlets, women may get around this issue by using Twitter to directly reach out to voters. Twitter thus allows them to circumvent the agenda power of media and communicate to voters on their own terms. While this is an advantage for both men and women, the gender bias in the media should make this a relatively more attractive option for women, leading to greater uptake of Twitter among female legislators.

[1]When state politicians tweet, they are likely speaking directly to their immediate audience of followers but also indirectly others in their communities. Statehouse journalists increasingly cover tweets, and followers share tweets, either digitally or physically, through their social networks. Rosenstiel et al. (2015) show that a majority of non-Twitter users have seen tweets; they are exposed to them primarily on television, through friends, and in newspaper articles.

[2]This is a disputed position in the literature, with many studies finding that women candidates are not discriminated against in elections (see reviews in Brooks 2013; Hayes and Lawless 2016; Lawless 2015; Thomsen 2020).

Whatever mechanism is at work, we predict that female legislators will work harder than men to establish a social media presence. If this is true, it will be consistent with Kurtz et al.'s (2006, 332) finding that women in state legislators "devote an additional 7 percent of a full-time job to their legislative work compared with men." It will also fit with Evans and Clark's (2016) finding that female candidates tweeted more often than male candidates in the 2012 congressional elections and Thomsen and Sanders' (2020) study showing that women put more effort into their constituency service. In the social media realm, we set forth two empirical hypotheses to test the idea that female state lawmakers put forth more effort in this realm than their male counterparts.

Hypothesis 1: Compared to men, female state lawmakers will be more likely to establish a public Twitter account.

Hypothesis 2: Compared to men, female state lawmakers will tweet with greater frequency.

Sentiment

Prior research has tested whether gender predicts how negative politicians are in their public communications (e.g., Evans and Clark 2016). Gender stereotypes are a reason why gender might be correlated with the tone of communication. Society stereotypes women as being more helpful and kind and men as more aggressive and forceful (Fridkin and Kenny 2009; Huddy and Terkildsen 1993). If voters hold these stereotypes, this can shape what voters expect from them and how they respond to the tone of politicians' communications.

The effect of legislators' tone on voter evaluations is unclear. Some scholars conclude that voters punish women when they act in ways that are counter to existing stereotypes (e.g., Kahn 1996). Yet others conclude that taking a more negative tone helps women because it challenges those stereotypes (Lau and Pomper 2004).

In looking at social media, Evans et al. (2014) and Evans and Clark (2016) find that women are more likely to send more negative attack tweets (cf., Parmelee and Bichard 2012). Evans and Clark (2016) also find that the number of negative tweets (coming from both men and women) increases with the number of women in the race. One reason that women may be negative in their tone is that they are more likely to be attacked (Lazarus and Steigerwalt 2018). This may lead them to defend themselves with tweets that have a more negative tone because they are trying to deal with a more hostile political environment. On the other hand, women may feel more pressure regarding reelection (Krook 2020), leading them to try to win over constituents using a more positive tone in their tweets. We test whether this relationship identified at other levels of office holds among state legislators.

Hypothesis 3: Compared to men, female state lawmakers' tweets will be more likely to have a negative tone.

Issue Focus

Men and women may also differ in the policy content of their communication. At the most basic level, they might differ because they work on different issues. Previous

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

studies have identified health, welfare, and education as "women's issues" (Saint-Germain 1989; Swers 1998, 2002). Other studies have instead focused on specific issues: for example, focusing on funding for breast cancer as opposed to all health care funding generally (Osborn 2012; Reingold 2000). We focus on the general categories, in part, because of the data. There are few tweets on any given specific issue. Indeed, there are some major categories that are rarely tweeted about. Looking at the more general categories provides greater variation for analysis. However, using general categories is a noisier measure. This is why "[s]tudies that adopt a more specific definition of women's issues, or those issues that directly affect women, find a closer connection between women's presence and policy outputs benefitting women" (Osborn 2012, 27). For this reason, our test is a harder test of gender differences in issue coverage.

Theoretically, research suggests that women may be more likely to work on these issues because they have more knowledge about these issues or simply because they personally prioritize these issues (Foerstel and Foerstel 1996; Reingold 1992; Thomas 1994; Thomas and Welch 1991). Either way, previous studies have found that women are more active in policy making on women's issues.

During the committee stage, women are more likely to advocate for women's interests (Swers 2002). And committees with more women are more likely to produce legislation that incorporates women's interests (Berkman and O'Connor 1993; Norton 1999; Swers 2002). Further, female legislators in both the United States (Thomas 1994; Thomas and Welch 1991) and elsewhere (Considine and Deutchman 1994; Heath, Schwindt-Bayer, and Taylor-Robinson 2005) are more likely to serve on committees that deal with issues traditionally considered women's issues.

Even after legislation leaves the committee, women are more active on these issues with female legislators in Congress participating more in floor debates on women's issues (Swers 2000; Tamerius 1995). The focus is not simply because women have more opportunities to work on women's issues. When men and women are given the same requests for help, female legislators are more likely to work on women's issues than are their male counterparts (Butler 2014).

In identifying women's issues, we use general categories rather than specific issues. In order to classify tweets by issue areas, we follow the categorizations created by the US Policy Agendas Project Codebook (see Adler and Wilkerson 2014).[3] "Education" tweets are ones that fit the topics in 600: Education in that coding system, and include primary and higher education as well as tweets about universal pre-K. Our "Healthcare" category includes policies fitting into 300: Health, including references to Medicaid, the Affordable Care Act, Medicare for All, and prescription drug prices. We did not use a separate "Welfare" category, because tweets about this policy area were so rare. In our training dataset of 10,104 hand-coded tweets, only five used the word "welfare," with three of these being references to corporate welfare and thus included under macroeconomic policy. After classifying all state tweets, following the machine learning process that we detail below, we test whether women are more likely than men to discuss women's issues by comparing their rates of tweeting about education and health care.

[3]The codebook used by Adler and Wilkerson (2014), which we followed and adapted, is available at https://comparativeagendas.s3.amazonaws.com/codebookfiles/Topics_Codebook_2014.pdf, first accessed in June 2016.

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

Hypothesis 4: Compared to men, female state lawmakers' tweets will include more content on education and health care policy.

Ideology

In this study, we look at how legislators communicate with their constituents. In order to win elections, politicians want to publicly take positions that appeal to voters. This can lead politicians to try to shape their public record to appear to be more in line with their voters' preferences. Gender can affect this dynamic because voters think female, Republican legislators are more liberal than male, Republican legislators and they think female, Democratic legislators are more liberal than male, Democratic legislators (Koch 2000). If this leads voters to think that the voters are out-of-step with their constituents, then legislators have incentives to engage in more conservative position-taking in order to compensate for voters' stereotypes and present themselves as in-line with their constituents.

Hypothesis 5: Controlling for their positions on roll calls, female state lawmakers' tweets will be more conservative than male state lawmaker's tweets.

We might see a partisan difference in how legislators take positions because of their incentives to appeal to primary voters (Brady, Han, and Pope 2007). Democratic women may not try to appear more conservative than their voting record simply because being viewed as liberal can help them in the primary election (Sides et al. 2020). Thus, we may not see any relationship between gender and Twitter ideology among Democrats once we control for their roll call-based ideology. Republican women, by contrast, have incentives to appear more conservative in order to win their primary elections (Koch 2000). This suggests that we may only see Republicans engaging in position-taking to make themselves appear more conservative than they really are.

Hypothesis 6: Among Republicans, female state lawmakers' tweets will be more conservative than male state lawmakers' tweets when controlling for their roll call record.

Case Selection

We study legislators' tweets—and focus on the states—for four reasons. First, legislators control their tweets. In contrast to coverage in traditional media, the legislators are able to control what they write. This is important because it may be that the media systematically covers female politicians differently than male politicians. If we look at the media coverage, then it is unclear if we are measuring the actions taken by the legislators, the biases of the media, or a combination of both. Because we are interested in how politicians choose to portray themselves, looking at Twitter—a communication form they control—allows us to do that (see also Pearson and Dancey 2011; Pearson and McGhee 2013).

Second, at least in some legislatures, women face institutional constraints that affect their ability to influence legislation or other outcomes (Hawkesworth 2003). Twitter is a tool that is not controlled by legislative leaders or legislative institutions

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

and therefore allows us to measure the legislators' activity free from any biases against them or constraints placed on them.

Third, tweets are public information. We need access to what legislators say in order to measure how legislators portray themselves. Twitter has this information. Also, legislators cannot microtarget tweets. It is not the case that women can send messages only to female followers and men only to male followers. If they could, we might worry that the differences in content might reflect the specific group they were microtargeting. This is not the concern because we are getting the public tweets that they use to speak to all constituents, the media, and fellow legislators.

Fourth, social media and Twitter are an increasingly important form of communication. They are used extensively not only by American state legislators but also by politicians all around the world (Alles and Jones 2016; Jungherr 2016; Vergeer 2015). Understanding how politicians communicate through a medium that they use nearly every day is critical to understanding how they choose to portray themselves—and how the public perceives them—in the modern era.

Focusing on *state* legislators in particular provides a strong empirical ground in which to study the impact of gender on communication styles. In our dataset, there are 1,391 female lawmakers, making up 25.7% of state legislators overall. This includes 535 Republican women and 845 Democratic women. By contrast, during the 116th Congress, only 101 women served in the House, including 88 Democrats and just 13 Republicans.[4] Compared to Congress, studying the states provides more opportunity to identify systematic patterns and to differentiate the effects of gender from those of party. And studying Twitter in state legislatures can provide a particularly unfiltered view of political communication. The scarcity of staff resources makes it more likely that state legislators send tweets themselves rather than relying upon staff, relative to members of Congress. When studying the impact of an individual attribute such as gender, this ability to observe direct personal behavior is valuable.

Data Collection

In order to combine human coding with machine learning techniques to classify the tweets of all lower house state legislators, we proceeded in four steps:

- Classifying and validating a "training set" of 10,104 hand-coded tweets
- "Pre-processing" the tweets to focus on their essential linguistic characteristics
- Training machine learning algorithms to replicate the hand codes, and testing their accuracy
- Classifying the full corpus of 3,580,727 tweets.

We began by creating a training set of tweets by American politicians over the last several years, categorized by their ideology, sentiment, whether they contained explicitly political subjects or not, the policy area that they address, and whether they constituted an opinion or a factual claim. We did so by building on the work done by Kousser and Oklobdzija (2018) who had a team of multiple research

[4]See "Women Serving in the 116th Congress (2019–2021)," Center for American Women and Politics, Rutgers University, https://cawp.rutgers.edu/list-women-currently-serving-congress.

Table 1. Measures of intercoder reliability: humans agreeing with humans

	Agreement rate	Cohen's kappa
Is the Tweet a factual claim or an opinion?	0.78	0.42
Ideology (liberal, neutral, or conservative)	0.78	0.66
Sentiment (negative, neutral, or positive)	0.75	0.60
Is the Tweet political or personal?	0.91	0.55
Topic: Immigration	1.00	0.87
Topic: Macroeconomics	0.97	0.73
Topic: Defense	0.96	0.77
Topic: Law and Crime	0.99	0.86
Topic: Civil Rights	0.98	0.71
Topic: Environment	1.00	0.84
Topic: Education	1.00	0.83
Topic: Health	0.99	0.82
Topic: Government operations	0.98	0.23
Topic: No policy content	0.91	0.78
Asks for a donation?	0.99	0.66
Asks to watch, share, or follow?	0.95	0.65
Asks for miscellaneous action?	0.93	0.57

Note: Based on an analysis of 1,217 tweets coded by rotating pairs of research assistants.

assistants (RAs) hand code a random sample of 8,206 tweets by the 2016 presidential candidates and their SuperPACs. These tweets were downloaded from Twitter's public API every week from October 2015 to July 2018.[5]

We then supplemented the database from Kousser and Oklobdzjia (2018) with 1,898 tweets from upper house state legislators and statewide officeholders that we coded for this project. We coded the tweets from the upper house state legislators because we wanted to make sure we included communications from those serving in a legislative context in our training dataset, but to train our algorithms on a set of tweets that was distinct from the lower house tweets in the full corpus that we later analyze. We downloaded the tweets from state legislators and state officials beginning in June 2018. We also had a group of RAs hand code the tweets using the same procedure as Kousser and Oklobdzjia (2018). In particular, all RAs worked from the same codebook and met regularly but coded tweets independently and were given only the text of the tweet, with no information about who sent it.

Table 1 provides data demonstrating that these coders reliably agreed in their independent categorizations. Using a subset of 1,217 presidential tweets, which were assigned to overlapping pairs of coders, we report two measures of intercoder reliability: the rate of agreement between coders and the Cohen's Kappa, measuring how much more likely our coders were to agree than two coders would be by random chance alone. Our rates of agreement range from 75% on our three-category sentiment measure to perfect agreement on three of our subject areas, with the Cohen's Kappa measures ranging from "fair" to "almost perfect" agreement levels for all but one of our variables. For our measure of ideology, the coders agreed 78% of the time, with a Cohen's kappa of 0.66, demonstrating that they could make this subjective judgment in a reliable, replicable manner.

[5]Kousser and Oklobdzjia (2018) found that their coding led to Cohen's kappa, which measures how much more likely our coders were to agree than two coders would be by random chance alone, range from 75% on the three-category sentiment score to perfect agreement on three of the subject areas.

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

Table 2. Examples of variables and tweets in each category

What issue did the tweet address?		
Education issue (3.4%)	Health care issue (1.6%)	
Thank you for helping me reach 1,400 followers. I appreciate your support for my campaign to #StopCommonCore in #AZ	The #opioidcrisis is very much here. This is a great initiative by @GovernorTomWolf and steps to link people to…	
Thanks to @SenatorBaldwin for fighting to expand access to #HigherEd	I've personally witnessed how Medi-Cal has changed lives. 1/3 of CA's population (13 million children & adults) is on Medi-Cal	
RT @HouseDemsIL: Superintendents call on @GovRauner to do his job and #SignSB1 to ensure our children receive a quality education	Live call w/ @ChrisMurphyCT thanking him for leading opposition to #TrumpCare. Watch. https://t.co/WZH1Ot0LUa @ActTogetherCT @womensmarchct	
Ideology		
Liberal (8.3%)	Neutral (72.1%)	Conservative (19.6%)
Medicaid enrollment has slowed in recent years, but it still serves nearly a quarter of our population	Top 5 AZ consumer fraud complaints and 5 warning signs: https://t.co/fznWrvoBr6	I was proud to stand up for the #righttolife again this week! #prolife #alpolitics
Workers' Rights: Check this video out w/ local verizon workers on strike in EB and what their cause is about. http://t.co/0FPMMr9 via @youtube	Last night, we went to Seussical. My cousin Wes was one of the leads and all three of my neighbors were in it. They all did great!	New approach to business tax filings greatly reduces gov't red tape—was recommended by our streamlining initiative
Sentiment		
Negative (19.0%)	Neutral (22.0%)	Positive (59.1%)
Lots that could have been fixed if WI GOP had not quit early: student loan debt, transpo funding, voucher accountability… #WICanDoBetter	Discrimination Lawsuit Filed Against Used Car Dealership in Mesa	Pa. Medical Marijuana legislation back on track #SenAHW #CoSponsor #SB3 #MedicalMarijuana #PA via
In Senate Approps being asked to spend $1.9 million to cover @SchuetteOnDuty's fight against marriage equality. Waste of taxpayer dollars	Intel predicts a $7 trillion self-driving future	Governor Brown signs 5 #EquityAndJustice bills today! A BIG step to promote safety, rehabilitation & family cohesion

In Table 2, we provide examples of tweets by state politicians that fit into the key categories that we focus on in this analysis. We show what types of text would highlight to our coders that a tweet had liberal, neutral, or conservative ideological content, as well as whether it conveyed negative, neutral, or positive sentiment. We report tweets that were identified as falling into the education or health care policy realms, two types of "women's issues" highlighted by prior research (Saint-Germain 1989; Swers 1998, 2002) and following the policy categories used by the Policy Agendas Project to code federal bills by Adler and Wilkerson (2014). For each category, we report how prevalent it was in the full corpus of state lower house tweets, according to our classifications.

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

Table 3. Measures of classification accuracy: computers replicating humans

	Final testing accuracy	Cohen's kappa
Is the Tweet an opinion versus a factual claim?	0.65	0.29
Ideology (liberal, neutral, or conservative)	0.62	0.40
Sentiment (negative, neutral, or positive)	0.60	0.38
Is the Tweet political or personal?	0.83	0.38
Topic: Education	0.92	0.03
Topic: Health	0.97	0.20
Topic: Immigration	0.98	0.29
Topic: Macroeconomics	0.93	0.27
Topic: Defense	0.93	0.10
Topic: Law and crime	0.98	0.07
Topic: Civil rights	0.97	0.26
Topic: Environment	0.20	0.00
Topic: Government operations	0.98	0.07
Topic: No policy content	0.79	0.56
Asks for a donation?	0.99	0.00
Asks to watch, share, or follow?	0.95	0.50
Asks for miscellaneous action?	0.29	0.02

Note: Based on an analysis of a final testing set of 1,010 tweets after training on 8,084 tweets and testing on 1,010 tweets.

With this training set in hand, we then pre-processed the tweets through a series of steps that are commonly used in text analysis. We made every word lower case, removed URLs as well as additional links and emails, and deleted all alphanumeric text. Depending on whether it improved prediction accuracy for individual characteristics of tweets, we also removed unnecessary stop words such as "the," "a," or "an," and removed screen names.

We then used the remaining text of each tweet, along with the human codings of their characteristics, to train a set of algorithms that fit models connecting the text to the codings. The algorithms that we used in this stage of the analysis were all taken from the scikit-learn Python library.[6] To train the algorithms, we divided our training set of 10,104 tweets, using 80% of them to train, 10% to test the accuracy and select the most accurate algorithm, and 10% to use as a "final testing set," which avoids over-fitting a model. Table 3 reports the results of these final tests. The first column shows accuracy for each variable, which is rate at which the algorithm was able to correctly replicate the human coding. The second column reports the Cohen's kappa, which is the improvement in accuracy over what we would expect by random chance if the algorithm always placed a tweet in the most prevalent category. For the policy variables, which take on only two values, accuracy is consistently high, registering over 90% in final testing accuracy for every policy area other than environmental policy (which does not feature in our analysis). While accuracy is lower for sentiment and ideology, reflecting the increased difficulty of correctly coding a variable that takes on three values, Cohen's kappa values fall just over or just below the "moderate" threshold (Landis and Koch 1977, 165) for both variables.

[6]This library can be accessed at https://scikit-learn.org/stable/. We used the algorithm that produced the best accuracy for each tweet characteristic, including Multinomial Naïve Bayes (for sentiment, political, ideology, no policy content, factual claims or opinions, and whether a tweet made a miscellaneous ask), Bagging Classifier (for immigration, macroeconomic, health care, national security, crime, and whether a tweet asked for donations), and Linear SVC (for civil rights, governance, and whether a tweet asked a follower to watch, share, or follow). We also adjusted the tuning parameters to identify the best fit for each model.

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

Finally, we used the trained algorithms to classify an original dataset of the tweets of all lower house state legislators. To collect these tweets, we began by working with undergraduate RAs to search for the Twitter handles of all legislators serving in lower houses of 49 states—excluding Nebraska's unicameral, nonpartisan house—in the summer of 2017. The RAs generally started by first performing a search for a one of the legislators on their list. In some states, especially more professional states, the handles were sometimes publicly listed together. More often, the RAs first found the Twitter account for one legislator and then found that the other legislators in their party in the state often linked to that account.

The RAs were unable to find accounts for many legislators, even after using several variants of the legislator's names in the search. If the RAs were unable to find an account after searching for several minutes, they moved on to the next account. RAs also limited the sample to publicly listed accounts because we are interested in how legislators portray themselves to the public. Once the RAs identified a likely match, they looked at several tweets in the accounts to confirm that they had correctly identified the legislator's account. In a few cases, the legislator did not have an account, but accounts were set up to parody the legislator. In other cases, legislators had multiple accounts. RAs looked through these accounts and identified the account(s) that were used as the legislator's account during the legislative session. In some cases, legislators had multiple accounts that met these criteria; in those cases, all accounts were used in the study. In many of these cases, the dates the accounts were used did not overlap, suggesting that it may simply have been a case where the legislator forgot their password and decided to simply create a new account.

The RAs also recorded the genders,[7] party affiliations, and districts represented by these 5,413 state legislators.[8] Of these lawmakers, 2,014 (37%) did not have a public Twitter handle that we could identify. For the 3,399 (63%) of state legislators who did have a social media presence, we collected all available tweets from Twitter's public API that were available in April 2019. This produced a dataset of 3,580,727 tweets. We then classified the features of these tweets and then calculated the average rates of each type of tweet for each tweeting legislator, along with their total tweet count. We merge this dataset with our data on legislator characteristics, successfully matching 3,129 state legislators to their tweet records. Finally, we appended data ideology based on statehouse roll call voting and national survey responses for all state legislators elected before 2016, using Shor and McCarty's (2011), updated with data from all legislators elected before 2016 through their website.[9]

In order to explore the validity of the classifications the algorithms produced, we can compare our tweet-based ideology score with the roll call-based measures for state legislators collected by Shor and McCarty (2011). Although lawmakers may choose to vote and to communicate in slightly different ways (e.g., Hypotheses 5 and 6), there should be a strong correlation between the ideological positions that legislators take on the floor and the images that they convey on social media.

[7]In internet searches, we have not identified any state legislators who have made a public declaration of a non-binary gender identification as of May 2019.

[8]This is two more legislators that serve at any one time in state lower houses (5,411), likely because some legislators who were selected in special elections to replace others were also included.

[9]We collected roll call ideology scores in May 2019 from https://americanlegislatures.com/data/.

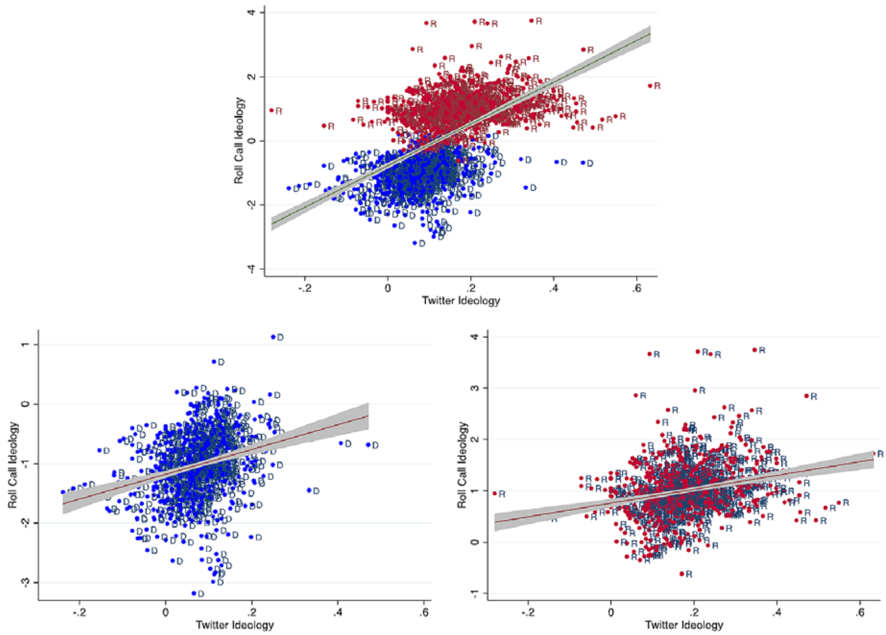


Figure 1. Testing the validity of tweet-based ideology measure.
Notes: All graphs compare our measure of the average ideology of each legislator's tweets with her roll call ideology, taken from updates to the dataset originally collected by Shor and McCarty (2011). Observations are all state lower house legislators elected before 2016 with more than 10 tweets.

As Figure 1 shows, a legislator's Twitter ideology score[10] is positively correlated with her roll call ideology score produced by Shor and McCarty (2011). The scatterplot in the top panel, which includes a fit line with a 95% confidence interval, combines the data for both Democrats and Republicans and shows there is a positive relationship between the two measures. A regression of roll call ideology on Twitter ideology, reported in Supplementary Appendix Table 1, shows that this relationship is statistically significant at the 99% confidence level. Importantly, our tweet-based scores also predict roll call ideology within parties. The two scatterplots in the lower panel of Figure 1 show that this link holds within the Democratic Party and Republican Party. Regressions in Supplementary Appendix Table 1 demonstrate that these relationships are statistically significant. In fact, this relationship holds even within party for a model with state fixed effects. That means that when two legislators are in the same party and members of the lower house in the same state, the

[10]We calculate this score by multiplying a legislator's liberal tweets by negative one, neutral tweets by zero, and conservative tweets by one, and taking the sum. This yields a score that can range from negative one to one, with larger values representing a higher rate of conservative versus liberal tweeting. In Figure 1, we display data only for legislators with more than 10 tweets, to guard against small-sample outliers, and also remove the −0.75 score of Mississippi legislators Earle Banks, the most liberal frequent tweeter in our dataset, whose Twitter record consists almost entirely of campaign messages during his 2012 run for state Supreme Court.

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

lawmaker who tweets more conservatively is also likely to vote more conservatively, an important validation of this measure.

Analysis

In using the information on the 5,413 state legislators and their tweets to test our hypotheses, we compare the raw, bivariate differences between female and male legislators and then present full multivariate tests (see Oklobdzija, Kousser, and Butler 2022 for replication code and data). In these tests, we control for legislators' party affiliation and also include state fixed effects. These fixed effects capture the impact of all measurable features of a state—its level of legislative professionalism, its political culture, the party balance in its statehouse—along with all idiosyncratic characteristics that are fixed. These multivariate models with controls for party and with state fixed effects are the focus of our main analysis. Later, we present extensions that add additional factors to probe the robustness of the impact of gender and to explore other social media dynamics, often with a subset of our cases. We look at the impact of a legislator's racial and ethnic identity, as well as its intersection with gender, in a section of the paper devoted to this question. We explore the effects of a state's legislative professionalism (Squire 2017) and of how recently a lawmaker was elected in analyses reported in our Supplementary Appendix. Each of these reveals important lessons but does not alter the clear relationship between gender and social media activity and messaging. That is the central focus of our main analysis presented below.

We first test relationship between gender and a lawmaker's allocation of time to establishing a social media presence. Hypotheses 1 and 2, respectively, predict that women will be more likely to have a Twitter handle and will tweet more frequently if they do. Regarding Hypothesis 1, 71.9% of the 1,391 female state legislators in our dataset had a public Twitter account we could identify. For the 4,022 male state legislators, this figure was only 60.0%. This significant gender gap in social media presence also holds when we estimate regressions that control for the legislator's partisanship and state fixed effects. According to this model, reported in Table 4, women establish handles at a 10.6% higher rate, all else equal, a result that is significant at the 99% confidence level.

Conditional on establishing an account, women also appear to tweet more often, tweeting an average of 1,200 times compared to 1,032 for men over the full history of their political Twitter account. That represents a 16% increase in how much more often women tweet compared to their male counterparts. In a multivariate model of tweet frequency, conditional on having an account, we also see that women send an estimated 121 more tweets, which again is strongly significant. The final model in Table 4 ties these two aspects of social media together into a single estimation. Our assumption here is that there is a latent variable measuring each lawmaker's "tweet effort." For those who take on high enough values of this variable to establish an account, we can directly observe their effort through their number of tweets. For those who have no account, our observation of their effort is censored at zero tweets. This sort of censoring can be corrected for by a "Tobit" maximum likelihood model, with left-censoring at zero. The estimated impact of gender from this model, which is determined both by women's higher rates of tweeting and the greater likelihood that

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

Table 4. Does gender affect twitter activity?

	Does the legislator have a handle?	Tweet count	Tobit model
Female legislator	0.106** (0.014)	118.771** (46.580)	327.635** (49.510)
Democratic legislator	0.053** (0.013)	314.705** (45.036)	354.528** (45.836)
State fixed effects	Included	Included	Included
Observations	5,422	3,134	5,422
Adjusted R-squared	0.159	0.081	

Notes: Observations are all state lower house legislators in the first and third models, and all state legislators with Twitter accounts in the second model. Standard errors in parentheses, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent variables are a dichotomous measure of whether a legislator had an official Twitter handle (in the first column) and the count of tweets from that handle (in the second).

Table 5. Does gender affect sentiment and attention to “women’s issues”?

	Sentiment score	Women’s issues	Education	Health care
Female legislator	3.774** (0.949)	0.716** (0.148)	0.308** (0.104)	0.408** (0.106)
Democratic legislator	−4.972** (0.978)	0.876** (0.135)	0.280** (0.093)	0.596** (0.097)
State fixed effects	Included	Included	Included	Included
Observations	3,134	3,134	3,134	3,134
Adjusted R-squared	0.107	0.080	0.086	0.052

Notes: Observations are all state lower house legislators with Twitter accounts. Standard errors in parentheses, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent variables are the percentage of tweets: (1) with a positive sentiment, (2) addressing women’s issues, (3) addressing education issues, and (4) addressing health care issues.

they will establish a social media presence in the first place, is that female state legislators score an estimated 330 tweets higher on this scale of tweet effort.

Our test of Hypothesis 3 explores the finding of Evans et al. (2014) and Evans and Clark (2016) that female candidates tweet with a more negative tone than their male counterparts. At first glance in our dataset, it appears that gender has little impact on sentiment, with women registering a 48.8% in our summary measure of sentiment[11] and men a 47.1%. Yet the multivariate analysis reported in Table 5 shows that gender does have an apparent effect that is hidden, in a bivariate comparison, by its correlation with party affiliation. Women tweet an estimated 3.8 percentage points more positively, while Democrats tweet 5 percentage points more negatively. The full extent of this gender gap is only revealed when we control for party because the majority of women serving in office belong to the Democratic Party and Democrats tweet more negatively. So when we compare female legislators to their partisan counterparts, we see that women exhibit a more positive sentiment in their tweets. Even with state fixed effects, these effects are significant at the 99% confidence level. Sentiment patterns in our sample of state legislator tweets run contrary to the patterns observed in congressional campaigns by Evans et al. (2014) and Evans and Clark (2016); when tweeting from statehouses, it appears, female legislators strike a more positive tone than their male counterparts.

[11]We calculate this score by multiplying a legislator’s negative tweets by negative one, neutral tweets by zero, and positive tweets by one, and taking the sum. A higher score denotes a higher frequency of positive tweets relative to negative tweets.

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

Table 5 also reports our tests of whether female legislators are more likely to tweet about education and health care (i.e., Hypothesis 4). These are policy realms in which they may have a great interest and expertise, and where their opportunities to claim credit for their work may be magnified if voters view them as more expert in these areas. To be sure, tweeting about these issues or indeed any specific policy realm is a rare occurrence for legislators of either gender. Over 66% of state legislative tweets have no clearly identifiable policy content, a trend that Kousser (2019) also identifies among members of the Australian Parliament. Still, the rate of tweeting about education or health care does significantly vary by gender in the states. Men address these issues in 3.5% of their tweets, while women do so in 4.4% of tweets. Controlling for party and for state fixed effects, our multivariate model estimates that female lawmakers tweet about women's issue 0.7 percentage points more often, a different that is significant at the 99% confidence level. Our models also show that this is because they tweet more about each women's issue significantly more often. Women tweet about education 0.3 percentage points more often and about health care 0.4 percentage points more often than men.

Our final tests look at whether women communicate differently than men about their ideology. Because they may seek to counter the stereotype that they are more liberal than their co-partisans (see Koch 2000), women of both parties may take more conservative positions on Twitter than men (Hypothesis 5). Because we are interested in how the present themselves *relative* to their roll call positions, we control for their roll call-based ideology in these tests.

In our test of Hypothesis 6, we explore the possibility that these incentives may operate differently for Republican and Democratic women. Republican women should have consistent incentives to take positions on social media that are more conservative, because this will position them well for Republican primary elections. Democrats face a countervailing incentive to appear more liberal to improve their chances in the primary election, which may push take them to tweet more liberally than expected given their roll call behavior.

The first model in Table 6 shows no apparent relationship between gender and Twitter ideology when we hold party affiliation and roll call-based ideology constant. As Figure 1 already showed, there is a positive correlation between legislators' twitter ideology and their roll call-based ideal points. Table 6 confirms this relationship with the positive coefficient on the roll call-based measure for legislators' ideal points. Significantly, when we control for their roll call record, party and gender do not predict the ideological content of their tweets. When we estimate separate models for each party (see columns 2 and 3), we also see no relationship between gender and the ideological content of their tweets. In sum, once we control for the legislators' ideal point, gender does not predict the ideological content of tweets. Women are not portraying themselves more conservatively in their tweets than they are in their roll call votes.

Race, Ethnicity, and Intersectionality

Recent studies suggest the importance of taking a broader view of identity. Looking at tweet activity by members of the Congressional Black Caucus during the 2013–2024 session, Tillery (2019) finds that gender was the single most powerful predictor of how often a caucus member tweeted about racial issues, with women tweeting

Table 6. Does gender predict Twitter ideology?

	All legislators	Only democrats	Only republicans
Roll call ideology	0.039** (0.007)	0.025 (0.013)	0.047** (0.010)
Female legislator	−0.009 (0.007)	−0.017 (0.010)	0.003 (0.009)
Democratic legislator	−0.012 (0.016)		
State fixed effects	Included	Included	Included
Observations	1722	782	937
Adjusted R-squared	0.216	0.063	0.149

Notes: Observations are all state lower house legislators elected before 2016 with Twitter accounts. Standard errors in parentheses, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent variable in all models is average ideology of each legislator's tweets.

significantly more frequently about race. Barrett's (1995) investigation demonstrates that the policy priorities of Black women in statehouses are shaped by both aspects of their identity, and Fraga et al. (2006) find significant differences between male and female Latinx state lawmakers in the coalitions that they form and how often members of other groups seek their expertise.

Using data collected by the Reflective Democracy Campaign (2017) on the race and ethnicity of state legislators in the 2016–2017 session, we were able to record the race and ethnicity of 3,355 members of our full dataset of 5,422 state lawmakers, including 2,077 of the 3,144 lawmakers with Twitter handles. For this exploratory analysis, we initially analyzed members of each racial and ethnic group individually and found that Black and Latinx legislators were distinct from other legislators. To streamline the analysis and preserve statistical power in the analysis below, we combine the members of the nation's two largest racial and ethnic minority groups.

An initial, descriptive analysis reveals that Black and Latinx women in state legislatures are by far the most active group, with 82.2% having a political Twitter handle, compared with 72.3% of women who are white or members of other groups. Among male state lawmakers, 71.6% of male Black or Latinx representatives tweet, compared with 63.6% of men who are white or members of other groups. Legislators from these racial and ethnic minority groups also tweet more often than their white or other counterparts of the same gender, with Black and Latinx female lawmakers sending 1,209 versus 1,149 tweets and Black and Latinx men sending 1,344 versus 1,040. The multivariate analysis of tweet activity reported in Table 7 shows that this is intertwined with partisan differences. Democratic state legislators are much more active on social media, and Black and Latinx lawmakers are much more likely to be members of the Democratic Party. Controlling for partisanship, we do not find a significant impact of race/ethnicity or its interaction with gender on tweet activity. Yet these patterns raise the question of whether the higher levels of social media engagement by Democrats in state legislatures is partly a function of the more diverse makeup of this party.

Our analyses of the content of tweets, reported in Table 8, again reveal mixed findings but do show a significant effect of race and ethnicity on attention to health care issues as well as the persistent influence of gender. Just as we found in our main analysis, women are more likely to tweet with a positive sentiment and focus more on education and on health care. For the latter issue, Black and Latinx lawmakers are significantly more attentive than white legislators or members of other groups. The

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

Table 7. How do race, ethnicity, and gender affect Twitter activity?

	Does the legislator have a handle?	Tweet count	Tobit model
Female legislator	0.090**	91.402	248.057**
	(0.019)	(60.531)	(64.946)
Black or Latinx legislator	−0.025	39.148	−114.067
	(0.031)	(106.190)	(108.246)
Female × Black or Latinx legislator	−0.004	−243.474	−76.890
	(0.045)	(151.930)	(161.517)
Democratic legislator	0.063**	351.917**	422.495**
	(0.018)	(59.200)	(59.957)
State fixed effects	Included	Included	Included
Observations	3,355	2,077	3,355
Adjusted *R*-squared	0.149	0.099	

Notes: Observations are all state lower house legislators with race/ethnicity data in the first and third models, and all state legislators with Twitter accounts in the second model. Standard errors in parentheses, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent variables are a dichotomous measure of whether a legislator had an official Twitter handle (in the first column) and the count of tweets from that handle (in the second).

Table 8. How do race, ethnicity, and gender affect sentiment and attention to "women's issues"?

	Sentiment score	Women's issues	Education	Health care
Female legislator	4.676**	0.903**	0.371**	0.532**
	(1.317)	(0.190)	(0.141)	(0.126)
Black or Latinx legislator	0.787	0.480	0.100	0.379*
	(2.066)	(0.344)	(0.274)	(0.220)
Female × Black or Latinx legislator	0.905	−1.248**	−0.167	−1.081**
	(2.951)	(0.485)	(0.419)	(0.268)
Democratic legislator	−5.348**	0.791**	0.232	0.560**
	(1.264)	(0.166)	(0.117)	(0.112)
State fixed effects	Included	Included	Included	Included
Observations	2,077	2,077	2,077	2,077
Adjusted *R*-squared	0.100	0.061	0.044	0.064

Notes: Observations are all state lower house legislators with race/ethnicity data and Twitter accounts. Standard errors in parentheses, $^{**}p < 0.01$, $^{*}p < 0.05$. Dependent variables are the percentage of tweets: (1) with a positive sentiment, (2) addressing women's issues, (3) addressing education issues, and (4) addressing health care issues.

significant interaction between gender and race/ethnicity shows that this effect is strongest among male Black and Latinx lawmakers and demonstrates the value of taking an intersectional approach to studying Twitter behavior.

Discussion

We have looked at how state legislators use Twitter. We studied this increasingly important way to communicate because it allows us to directly learn about how legislators communicate without looking at how legislators' efforts are filtered through media. We collected data on state legislative Twitter communication to test hypotheses in four areas.

The largest difference we observed related to the level of effort legislators put into communicating on Twitter. The data show that female legislators are more likely to have Twitter accounts and use them. Previous researchers have argued that women have to work harder in order to get elected. Consistent with that argument, prior

https://doi.org/10.1017/spq.2022.16 Published online by Cambridge University Press

studies have found that women put more effort into their jobs along various dimensions (Anzia and Berry 2011; Kurtz et al. 2006; Pearson and Dancey 2011; Thomsen and Sanders 2020). We confirm that this pattern holds when looking at efforts to use Twitter to communicate with voters.

Our results also confirm previous work regarding how gender relates to the issues that politicians work on. Previous findings have shown that women work more than men on health, education, and other issues considered to be women's issues (Saint-Germain 1989; Swers 1998, 2002). Our data show that women also discuss health and education more on Twitter. Men discuss these issues in 3.5% of their tweets, while women do so in 4.4% of tweets. This means that women discuss these issues 26% more than men do (0.9/3.5 = 0.26).

One benefit of having more women in office is that voters learn about more issues. Politicians' communications are an important way for voters to learn about issues and to form evaluations (Arceneaux 2006). If politicians never focus on issues like health care and education, then voters are likely to pay less attention to those issues. If politicians descriptively represent the population, they are more likely to cover a wider range of issues that allows voters to learn about a wider range of issues.

Other results from our analysis contradict previous findings. Evans et al. (2014) and Evans and Clark (2016) found that among candidates for Congress, women were more likely to take a negative tone in their communication. When looking at the basic comparison between men and women we see the same pattern—i.e., women are more negative in tone. However, this is confounded with party. Women are more likely to be Democrats and Democrats are more negative in tone. When we control for partisanship, we find that, among state legislators, women are more positive than men in their Twitter communication.

Also, we find that gender no longer predicts the ideological positions that politicians take after accounting for their actual position. Previous work has found that politicians had incentives to take portray themselves as more conservative in order to counter the stereotype that women are more liberal. We find no evidence for this. Controlling for the legislators' position based on their roll call votes, gender does not predict how legislators portray themselves on Twitter. There is no evidence that women politicians are trying to counter gender stereotypes in their communication.

Future research could explore the determinants in legislators' tone in many ways. Among other things we can think more about how majority status could affect the level of negativity. Politicians in the majority might have more reasons to be more positive. We may have found that Democrats were more negative simply because of their status in the chamber. Alternatively, it might simply have been that our dataset covers the beginning of the Trump presidency (October 2015 through July 2018). Democratic politicians may have simply been responding to Trump and this could have led to a more negative tone.

More generally, research might test whether these patterns will hold in future time periods. Again, our data come from the period when Donald Trump transformed political communication by making Twitter his central means of reaching voters and attracting media attention (Kreis 2017; Ott 2017). It is also a time in which Trump was the center of attention, especially on Twitter and his sexist behavior may have influenced how women and men legislators used this communication tool (Scotto di Carlo 2020).

Social media is an increasingly important tool for legislators to use to communicate with voters. It also provides a fruitful opportunity for researchers to learn more about representation because the data are public and the legislators are in direct control of the content. As a result, we can directly observe what politicians want to communicate with voters. We have used this tool to study how gender relates to legislators' use of this communication form. Future work will expand this research in many and varied directions.

Supplementary Materials. To view supplementary material for this article, please visit http://doi.org/10.1017/spq.2022.16.

Data Availability Statement. Replication materials are available on SPPQ Dataverse at https://doi.org/10.15139/S3/MHAAZV (Oklobdzija, Kousser, and Butler 2022).

Acknowledgments. We thank Calvin Tam, Lauryn Johnson, Mohammed Al Elew, Brittany Van Ryder, Arsham Aliaskari, Eamon Aalipour, Chris Trinh, Treya Parikh, Adarsh Parthasarathy, Emily Yeh, Zaire Bailey, Kexin Chen, and Rachel Schoner for their research assistance.

Funding Statement. The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Adler, E. Scott, and John Wilkerson. 2014. Congressional Bills Project: 2014. NSF 00880066 and 00880061. The views expressed are those of the authors and not the National Science Foundation.

Alles, Santiago, and Mark P. Jones. 2016. "Tweeting in the Lone Star State: Locating State Politicians in a Common Space Using Twitter Followers." State Politics and Policy Conference.

Anzia, Sarah F., and Christopher R. Berry. 2011. "The Jackie (and Jill) Robinson Effect: Why Do Congresswomen Outperform Congressmen?" *American Journal of Political Science* 55 (3): 478–93.

Arceneaux, Kevin. 2006. "Do Campaigns Help Voters Learn? A Cross-National Analysis." *British Journal of Political Science*, 36 (1), 159–173.

Baitinger, Gail. 2015. "Meet the Press or Meet the Men? Examining Women's Presence in American News Media." *Political Research Quarterly* 68 (3): 579–92

Barrett, Edith J. 1995. "The Policy Priorities of African American Women in State Legislatures." *Legislative Studies Quarterly* 20: 223–47.

Berkman, Micahel B., and Robert E. O'Connor. 1993. "Do Women Legislators Matter? Female Legislators and State Abortion Policy." *American Politics Quarterly* 21 (1):102–124.

Brady, David W., Hahrie Han, and Jeremy C. Pope. 2007."Primary Elections and Candidate Ideology: Out of Step with the Primary. Electorate?" *Legislative Studies Quarterly* 32 (1): 79–105.

Brooks, Deborah Jordan. 2013. *He Runs, She Runs: Why Gender Stereotypes Do Not Harm Women Candidates*. Princeton, NJ: Princeton University Press.

Burrell, Barbara C. 1994. *A Woman's Place Is in the House: Campaigning for Congress in the Feminist Era*. Ann Arbor: University of Michigan Press.

Butler, Daniel M. 2014. *Representing the Advantaged: How Politicians Reinforce Inequality*. New York: Cambridge University Press.

Butler, Daniel M., Elin Naurin, and Patrik Öhberg. 2020. "Constituents Ask Female Legislators to Do More." Paper Presented at the 2020 APSA Annual Meeting.

Considine, Mark, and Iva Ellen Deutchman. 1994. "The Gendering of Political Institutions: A Comparison of American and Australian State Legislators." *Social Science Quarterly* 75 (4): 854–866.

Evans, Heather K., Victoria Cordova, and Savannah Sipole. 2014. "Twitter-Style: An Analysis of How House Candidates Used Twitter in their 2012 Campaigns" *PS: Political Science and Politics* 47 (2): 454–462.

Evans, Heather K., and Jennifer Hayes Clark. 2016. "'You Tweet Like a Girl!': How Female Candidates Campaign on Twitter." *American Politics Research* 44 (2): 326–52.

Fenno, Jr., Richard F. 1977. "U.S. House Members in Their Constituencies: An Exploration." *The American Political Science Review* 71 (3): 883–917.

Fox, Richard L. 2006. "Congressional Elections: Where are We on the Road to Gender Parity?" In *Gender and Elections*, eds. Susan J. Carroll and Richard L. Fox, 97–116. New York: Cambridge University Press.

Foerstel, Karen, and Herbert N. Foerstel. 1996. *Climbing the Hill: Gender Conflict in Congress*. Westport, CT: Praeger.

Fulton, Sarah A. 2012. "Running Backwards and in High Heels: The Gendered Quality Gap and Incumbent Electoral Success." *Political Research Quarterly* 65 (2): 303–14.

Fulton, Sarah A. 2014. "When Gender Matters: Macro-dynamics and Micro-mechanisms." *Political Behavior* 36 (3): 605–30.

Fraga, Luis Ricardo, Linda Lopez, Valerie Martinez-Ebers, and Ricardo Ramírez. 2006. "Gender and Ethnicity: Patterns of Electoral Success and Legislative Advocacy Among Latina and Latino State Officials in Four States."*Journal of Women, Politics & Policy* 28 (3-4): 121–145.

Fridkin, Kim L., and Patrick J. Kenney. 2009. "The Role of Gender Stereotypes in U.S. Senate Campaigns" *Politics and Gender* 5: 301–324.

Garofoli, Joe. 2018. "After Trump, Politics is All About Social Media," *San Francisco Chronicle*, July 31, 2018.

Grimmer, Justin. 2013. "Appropriators Not Position Takers: The Distorting Effects of Electoral Incentives on Congressional Representation." *American Journal of Political Science* 57 (3): 624–42.

Hawkesworth, Mary. 2003 "Congressional Enactments of Race–Gender: Toward a Theory of Raced–Gendered Institutions." *American Political Science Review* 97 (4): 529–50.

Hayes, Danny, and Jennifer L. Lawless. 2016. *Women on the Run: Gender, Media, and Political Campaigns in a Polarized Era*. New York: Cambridge University Press.

Heath, Roseanna Michelle, Leslie A. Schwindt-Bayer and Michelle M. Taylor-Robinson. 2005. "Women on the Sidelines: Women's Representation on Committees in Latin American Legislatures." *American Journal of Political Science* 49 (2): 420–436.

Heldman, Caroline, Susan J. Carroll, and Stephanie Olson. 2005. "'She Brought Only a Skirt': Print Media Coverage of Elizabeth Dole's Bid for the Republican Presidential Nomination." *Political Communication* 22 (3): 315–35.

Huddy, Leonie, and Nayda Terkildsen. 1993. "Gender Stereotypes and the Perception of Male and Female Candidates." *American Journal of Political Science* 37 (1): 119–147.

Jungherr, Andreas. 2016. "Twitter Use in Election Campaigns: A Systematic Literature Review." *Journal of Information Technology & Politics* 13 (1):72–91.

Kahn, Kim Fridkin. 1992. "Does Being Male Help? An Investigation of the Effects of Candidate Gender and Campaign Coverage on Evaluations of U.S. Senate Candidates." *The Journal of Politics* 54 (2): 497–517.

Kahn, Kim Fridkin. 1994. "Does Gender Make a Difference? An Experimental Examination of Sex Stereotypes and Press Patterns in Statewide Campaigns." *American Journal of Political Science* 38 (1): 162–95.

Kahn, Kim Fridkin. 1996. *The Political Consequences of Being a Woman: How Stereotypes Influence the Conduct and Consequences of Political Campaigns*. New York: Columbia University Press.

Koch, Jeffrey. 2000. "Do Citizens Apply Gender Stereotypes to Infer Candidates' Ideological Orientations?" *Journal of Politics* 62(2): 414–29.

Kousser, Thad. 2019. "Tweet Style: Campaigning, Governing, and Social Media in Australia." *Australian Journal of Political Science* 54: 183–201; doi:10.1080/10361146.2019.1568966.

Kousser, Thad, and Stan Oklobdzija. 2018. "The Supply and Demand of Fact v. Opinion in Presidential Tweets." Paper presented at the 2018 Meetings of the American Political Science Association.

Kreis, Ramona. 2017. "The 'tweet politics' of President Trump." *Journal of Language and Politics* 16 (4): 607–18.

Krook, Mona Lena. 2020. *Violence Against Women in Politics*. New York: Oxford University Press.

Kurtz, Karl T., Gary Moncrief, Richard G. Niemi, and Lynda W. Powell. 2006. "Full-Time, Part-Time, and Real Time: Explaining State Legislators' Perceptions of Time on the Job." *State Politics & Policy Quarterly* 6 (3): 322–38.

Landis, J. R, and Gary G. Koch. 1977. "The measurement of observer agreement for categorical data." *Biometrics* 33 (1): 159–174.

Lau, Richard R. and Gerry Pomper. 2004. Negative Campaigning: An Analysis of U.S. *Senate Elections*. Lanham, MD: Rowman Littlefield.

Lawless, Jennifer L. 2015. "Female Candidates and Legislators." *Annual Review of Political Science* 18: 349–66.

Lazarus, Jeffery, and Amy Steigerwalt. 2018. *Gendered Vulnerability: How Women Work Harder to Stay in Office*. Ann Arbor, MI: University of Michigan Press.

McGregor, Shannon C., Regina G. Lawrence, and Arielle Cardona. 2016 "Personalization, Gender, and Social Media: Gubernatorial Candidates' Social Media Strategies." *Information, Communication and Society* 20 (2): 264–83.

Newman, Jody. 1994. *Perception and Reality: A Study Comparing the Success of Men and Women Candidates*. Washington, DC: National Women's Political Caucus.

Norton, Noelle H. 1999. "Committee Influence Over Controversial Policy: The Reproductive Policy Case." *Policy Studies Journal* 27 (2): 203–216.

Oklobdzija, Stan, Thad Kousser, and Daniel Butler. 2022. "Replication Data for: Do Male and Female Legislators Have Different Twitter Communication Styles?" *UNC Dataverse*, V1, UNF:6:CZoFF/yyo4C7HlTRja/mmA== [fileUNF]. https://doi.org/10.15139/S3/MHAAZV.

Osborn, Tracy. 2012. *How Women Represent Women*. New York: Oxford University Press.

Ott, Brian L. 2017. "The Age of Twitter: Donald J. Trump and the Politics of Debasement." *Critical Studies in Media Communication* 34 (1): 59–68.

Parmelee, John H., and Shannon L. Bichard. 2012. *Politics and the Twitter Revolution: How Tweets Influence the Relationship Between Political Leaders and the Public*. Lanham, MD:Lexington Books.

Pearson, Kathryn, and Logan Dancey. 2011. "Elevating Women's Voices in Congress: Speech Participation in the House of Representatives." *Political Research Quarterly* 64 (4): 910–23.

Pearson, Kathryn, and Eric McGhee. 2013. "What it Takes to Win: Questioning 'Gender Neutral' Outcomes in US House Elections." *Politics & Gender* 9 (4): 439–62.

Peterson, Andrew, and Arthur J. Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis* 26 (1):120–128.

Reflective Democracy Campaign. 2017. *Reflective Democracy Research Findings 2016–2017*. Accessed at https://wholeads.us/research/reflective-democracy-research-findings-2016-2017/ in July 2021.

Reingold, Beth. 1992. "Concepts of Representation among Female and Male State Legislators." *Legislative Studies Quarterly* 17 (4): 509–537.

Reingold, Beth. 2000. *Representing Women: Sex, Gender, and Legislative Behavior in Arizona and California*. Chapel Hill, NC: University of North Carolina Press.

Rosenstiel, Tom, Jeff Sondermon, Kevin Locker, Maria Ivancin, and Nina Kjarval. 2015. *Twitter and the News: How People Use the Social Network to Learn About the World*. Arlington, VA: American Press Institute.

Saint-Germain, Michelle A. 1989. "Does Their Difference Make a Difference? The Impact of Women on Public Policy in the Arizona Legislature." *Social Science Quarterly* 70 (4): 956–970.

Scott di Carlo, Giuseppina. 2020. "Trumping twitter: Sexism in President Trump's Tweets." *Journal of Language and Politics* 19 (1): 48–70.

Seltzer, Richard A., Jody Newman, and M. Voorhees Leighton. 1997. *Sex as a Political Variable*. Boulder, CO: Lynne Reinner.

Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105 (3): 530–51.

Sides, John, Christopher Tausanovitch, Lynn Vavreck, and Christopher Warshaw. 2020. "On the Representativeness of Primary Electorates." *British Journal of Political Science* 50 (2): 677–685.

Squire, Peverill. 2017. "A Squire Index Update." *State Politics and Policy Quarterly* 17 (4): 361–71.

Swers, Michele L. 1998. "Are Congresswomen More Likely to Vote for Women's Issue Bills Than Their Male Colleagues?" *Legislative Studies Quarterly* 23 (3): 435–448.

Swers, Michele L. 2002. *The Difference Women Make: The Policy of Women in Congress*. Chicago: University of Chicago Press.

Tamerius, Karin L. 1995. "Sex, Gender and *Leadership* in the Representation of Women." In *Gender Power, Leadership*, and Governance, eds. Georgia Duerst-Lahti and Rita Mae Kelly. Ann *Arbor*, MI: University of Michigan Press.

Thomsen, Danielle M. 2020. Ideology and Gender in U.S. House Elections. *Political Behavior* 42: 415–42.

Thomsen, Danielle M., and Bailey K. Sanders. 2020. Gender Differences in Legislator Responsiveness. *Perspectives on Politics* 18: 1017–30.

Thomas, Sue. 1994. *How Women Legislate*. New York: Oxford University Press.

Thomas, Sue, and Susan Welch. 1991. "The Impact of Gender on Activities and Priorities of State Legislators." *Western Political Quarterly* 44: 445–56.

Tillery, Alvin B. 2019. "Tweeting Racial Representation: How the Congressional Black Caucus Used Twitter in the 113th Congress." *Politics, Groups, and Identities* 9 (2): 219–38.

Vergeer, Maurice. 2015. "Twitter and Political Campaigning." *Sociology Compass* 9 (9):745–60.

Volden, Craig, Alan E. Wiseman, and Dana E. Wittmer. 2013. "When are Women More Effective Lawmakers Than Men?" *American Journal of Political Science* 57: 326–41.

Author Biography. Dan Butler is a Professor of Political Science at the Washington University in St. Louis. His research focuses on understanding representation.

Thad Kousser is a Professor of Political Science at the University of California, San Diego and co-directs the Yankelovich Center for Social Science Research.

Stan Oklobdzija is a Visiting Assistant Professor of Public Policy at the University of California, Riverside where his research focuses on state/local politics, housing policy, and interest group behavior. He holds a PhD in Political Science from the University of California, San Diego and his prior research has appeared in Electoral Studies, Research & Politics, Socius, and State Politics & Policy Quarterly.

Cite this article: Butler, Daniel M., Thad Kousser, and Stan Oklobdzija. 2023. Do Male and Female Legislators Have Different Twitter Communication Styles? *State Politics & Policy Quarterly* **23** (2): 117–139, doi:10.1017/spq.2022.16

State Politics & Policy Quarterly (2023), **23**: 2, 140–165
doi:10.1017/spq.2022.17

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

The US Political Economy of Climate Change: Impacts of the "Fracking" Boom on State-Level Climate Policies

Sam Zacher

Department of Political Science, Yale University, New Haven, CT, USA
Email: sam.zacher@yale.edu

(Received 12 January 2022; revised 13 May 2022; accepted 05 July 2022)

Abstract

In the face of the intensifying global climate crisis, the US has failed to implement comprehensive policies to mitigate greenhouse gas emissions. During the 2000s, the shale oil and gas extraction (i.e., "fracking") revolution highlighted the American energy economy. Is the fracking boom partially to blame for US lagging on climate policy? Political economy theory suggests that economic resources are primary drivers of policy outcomes. In this paper, I originally evaluate that claim in the context of the American states, the governments most powerful to mitigate emissions while the federal government faces gridlock. I first introduce an original measure of one state-level climate policy: adoption of the low-emission vehicle (LEV) policy from 1991 to 2015. I then frame the US fracking boom of the mid-to-late 2000s as a natural experiment, employing a difference-in-difference design to compare the effects of fracking on two climate policies across the American states – LEV and renewable electricity policy. Results yield evidence of a causal impact of the fracking boom on state LEV adoption and more suggestive evidence of an impact on renewable electricity mandates. I conclude by arguing that efforts to evaluate the influence of business on policy should account for "structural power" mechanisms.

Keywords: Environmental Policy; Federal/State; Interest Groups

Introduction

Hydraulic fracturing technologically shocked the US energy economy in the 2000s, resulting in rapid oil and natural gas production increases – also known as the "fracking boom." US domestic natural gas production levels had been relatively stagnant from the 1980s through the 2000s, until 2005, when gas production began growing, nearly doubling from 2005 through 2020.[1] Because of this increase, natural

[1]"Natural gas explained." US Energy Information Administration. 2021. https://www.eia.gov/energyexplained/natural-gas/where-our-natural-gas-comes-from.php.

gas overtook coal in 2015 as the plurality energy source used to generate electricity in the US.[2] The fracking boom was such a big deal that former President Obama, who was supported by environmental groups, tried to take credit in 2018, saying, "suddenly America's like the biggest oil producer and the biggest gas – that was me, people."[3] Throughout the same period – the late 2000s through 2010s – public opinion and social movement activity supporting more aggressive climate policies intensified (e.g., Caniglia et al. 2015; Roser-Renouf et al. 2015). And yet, the US federal and (most) state governments did not enact significant policies to mitigate planet-warming emissions (e.g., Bang 2015), despite this increased political pressure and escalating certainty and alarm among climate scientists.[4]

The literature on the political economy of climate and environment emphasizes that the entrenchment of fossil fuel extraction and consumption interests is the main driver behind why governments across the world have not enacted more emissions-mitigating policies (e.g., Colgan, Green, and Hale 2021; Mildenberger 2020; Stokes 2020). However, quantitative, causally identified evidence falsifying these propositions is relatively thin (for one exception, see Cooper, Kim, and Urpelainen 2018). In this paper, I aim to fill that gap, leveraging the exogenous nature of the distribution of shale underneath US states as a treatment variable, to examine the possible impacts of fossil fuel endowments on climate policies at the state government level in the US. The fracking boom is a potentially fruitful setting to study causal effects, given the relatively short period of time when fracking technology rapidly spread across certain subsections of the country, opening up new fossil fuel reserves that had previously been untapped (Deutch 2011).

There is a rich literature in the study of state politics that makes a strong case for state governments as critical actors in the formation of American climate policy (for a review, see, e.g., Konisky and Woods 2018). While scholars have occasionally entertained "race to the bottom" theories that expect state governments to compete for minimal environmental regulation, states have been found to sometimes implement regulations more stringent than the federal government (Potoski 2001). During President Barack Obama's time in office, after climate legislation stalled in Congress, the primary attempt to use executive action to mitigate greenhouse gas emissions (via the Clean Power Plan) would have given state governments significant authority of how to specifically achieve regulatory targets (Konisky and Woods 2016). Independently, state governments have created their own unique climate or environmental regulations – often in part due to lack of significant federal policy changes – in areas such as fossil fuel drilling (Davis 2017), renewable electricity (Carley et al. 2018), carbon pricing (Rabe 2018), and clean water rules (Fowler and Birdsall 2021), among others. The institutionally federalist US political system makes state governments critical to the American ability to mitigate emissions.[5]

[2]Murphy, Tom. "Natural gas surpasses coal as biggest US electricity source." July 13, 2015. Associated Press. https://apnews.com/article/59a30fadd58e42f08280e4cdd198653c.

[3]Richardson, Valerie. "Obama takes credit for U.S. oil-and-gas boom: 'That was me, people'." November 28, 2018. Washington Times via Associated Press. https://apnews.com/article/business-5dfbc1aa17701ae219239caad0bfefb2.

[4]To be sure, "partisan polarization" is a relevant factor in explaining why US governments have not enacted more emissions-mitigating policy, but polarization is likely only a "mediating" variable in climate politics, exacerbated by fossil fuel politics and other forces.

[5]State governments also learn from each other when crafting renewable energy policies (Parinandi 2020).

In this paper, I study two climate policies (as outcome variables) that state governments have exercised control over for decades: the level of electricity generated by renewable sources (sometimes called "Renewable Portfolio Standards") and low-emission vehicle (LEV) policy adoption. While renewable electricity targets may be familiar to readers, the LEV policy may not be: When a state government adopts a LEV policy, the vehicles that are sold in the state must subsequently meet more stringent pollution regulations.[6] The increased stringency of both policies would mitigate greenhouse gas emissions (and associated pollutants).[7] Fossil fuel extraction companies are known to oppose more stringent climate policies across the board (e.g., Brulle 2018; Stokes 2020). Employing a difference-in-difference ("diff-in-diff") research design, I show that fracking negatively impacted LEV adoption and more suggestive evidence that fracking negatively impacted renewable electricity mandates. The latter evidence is more suggestive given its lack of extensive pretreatment outcome data to verify the common trends assumption (which is necessary for diff-in-diff research designs). To summarize the findings in other words, having the ability extract oil and/or natural gas via fracking within their geographic boundaries caused state governments to be less likely to adopt climate policies (LEV and renewable electricity policy).

My findings in this paper contribute to the climate political economy literature that shows that fossil fuel interests are one major factor that prohibits governments from mitigating societal emissions (Colgan, Green, and Hale 2021; Cooper, Kim, and Urpelainen 2018; Hughes and Urpelainen 2015; Mildenberger 2020; Stokes 2020). In this particular case, fracking for shale oil and gas has prohibited American state governments from passing more stringent climate policies. Very few existing studies have produced quantitative (causal inference) evidence that fossil fuel resources have exerted effects on policymaking institutions (Cooper, Kim, and Urpelainen 2018 is an exception), so this paper adds novel evidence on this front. Further, the findings in this paper join a small body of American politics research that has demonstrated political impacts of the fracking boom (Bishop and Dudley 2017; Cooper, Kim, and Urpelainen 2018; DiSalvo and Li 2020; Fedaseyeu, Gilje, and Strahan 2015; Mallinson 2014; Sances and You 2022). However, nearly all of those other studies focus on campaign contributions, election winners, or voter turnout; only one studies something close to policy outcomes – environmental ratings of federal legislators (Cooper, Kim, and Urpelainen 2018). This paper is the first to show that the fracking boom impacted *overall policy outcomes* (at the state government level) in particular. In addition, I offer the unique contribution of an original measurement of LEV policy adoption per state–year – a newly measured variable that may prove useful for other scholars of climate politics and policy.

In the following sections, I first review relevant literature, laying out the theoretical cases for economic interests influencing climate politics and the empirical gaps that I

[6]As explained further in the Research Design section, states have not had the legal freedom to create their own unique LEV policy. They have all faced a choice between just two options when deciding rules governing air pollution from vehicles: default to federal regulations (e.g., Corporate Average Fuel Economy, "CAFE") or adopt California's specific LEV policy.

[7]California's most recent version of the regulation (LEV III) is estimated to reduce greenhouse gas emissions from new vehicles by 40% by 2025 (from 2012 levels). ("Advanced Clean Cars Program." 2022. California Air Resources Board. https://ww2.arb.ca.gov/our-work/programs/advanced-clean-cars-program/about).

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

fill. Then, I clarify this paper's theory and describe my original empirical design, including the inferential steps necessary to interpret this observational data analysis as evidence of causal effects. Finally, I argue that research which aims to test the influence of "business" on policy should account for "structural power" mechanisms, as this paper's treatment (fossil fuel extraction) can be plausibly conceived to encompass the structural power of (fossil fuel) business activity.

Climate political economy and the American states

There is general agreement in the literature on the political economy of climate and environmental issues that economic interests drive policy outcomes. Specifically, these economic interests are energy endowments and the corporate, labor, and consumer interests that follow from the endowments. Much of this climate PE work is theoretical and qualitative; only a small handful of publications use quantitative causal inference research designs to show effects on politics of fossil fuel endowments and/or extraction. Some have specifically studied the American fracking boom, but none have examined state-level political outcomes nor actual policy enactment by governments.

Some recent prominent publications exhibit the importance of fossil fuel interests as a primary driver of climate politics across the world. Colgan, Green, and Hale (2021) argue that domestic climate policy outcomes can be best characterized by battles between those who own assets that drive climate change (e.g., fossil fuels) and assets vulnerable to climate impacts. Mildenberger (2020) argues that carbon-intensive businesses and even fossil fuel labor unions drive policy outcomes in many countries, including the USA, Australia, and some European countries. Stokes (2020) traces how organized fossil fuel interests used various tactics to roll back renewable energy policies at the state level in the US (an effect that this analysis will show using diff-in-diff analysis) – in crucial states where those industries perceive future profits to be made from more fossil fuel extraction. Hughes and Urpelainen (2015) argue that domestic climate politics is generally shaped by energy-intensive sectors (and constrained to some degree by public sentiment). Cory, Lerner, and Osgood (2021) analyze lobbying data to show that companies in all sectors of the US economy lobbied against federal climate legislation in the US because of their interdependence on carbon-intensive supply chains. All of these books and papers theoretically argue and show some qualitative or descriptive evidence that fossil fuel interests play some role in preventing more aggressive climate policy enactment.

Other more recent work applies quantitative methods to show plausible causal effects of economic interests driving environmental policy outcomes. Cooper, Kim, and Urpelainen (2018), also studying some effects of the US fracking boom, employ a local regression discontinuity design to show that a congressional district's exposure to shale (at least in the Pennsylvania area) seemed to cause its federal representative to be more likely to vote against all environmental policy after the fracking boom's height. Another recent prominent (but non-climate) paper shows a similar political-economic finding: Dasgupta (2020) shows that agricultural interests in certain irrigation technology seemed to cause those jurisdictions to vote for conservative economic policies in 20th century American politics. This emerging quantitative causal inference work shows how economic interests can affect elections and policymaking regarding environmental politics at the federal level, but this line of research is still rather thin, only federally focused, and has not studied *specific* policy outcomes.[8]

To be sure, there are forces other than fossil fuel interests that scholars argue matter in determining climate policies. Various kinds of institutional arrangements are likely to structure political outcomes in different ways at all levels of government (Hughes and Urpelainen 2015). Specific to American climate politics at the state level, the degree of unified Democratic state government control or advantage in Democratic partisan identification at the voter level may increase a state's likelihood of an increasingly stringent climate policy (Trachtman 2020). Public opinion on environmental policy – and particularly government environmental *spending* – seems to be represented somewhat by state elected leaders (e.g., Fowler 2016; Johnson, Brace, and Arceneaux 2005). Organized interest group representation on the pro-climate side (i.e., in opposition to fossil fuel interests) in the form of social movement protest activity may also lead a state government to implement stricter climate policies and decrease overall emissions (Muñoz, Olzak, and Soule 2018).

Particular to the substantive study of the effects of the fracking boom on climate politics, there are a few other papers in addition to Cooper, Kim, and Urpelainen (2018). One study showed that shale gas extraction in congressional districts seems to cause Republicans to win national electoral races at a higher rate (Fedaseyeu, Gilje, and Strahan 2015). Similarly, fracking has appeared to increase campaign donations to Republican candidates (DiSalvo and Li 2020; Sances and You 2022) but decrease overall voter turnout (Sances and You 2022). Using basic correlational methods, papers have (unsurprisingly) found evidence that state legislators who receive more campaign contributions from oil and gas companies are more likely to support pro-fracking policies (Bishop and Dudley 2017; Mallinson 2014). Overall, the fracking boom's political effects have seemed to advantage US Republicans, electorally. But this niche research area can be pushed further in terms of its methodological rigor and its focus on *policies* as outcomes.

In summary, the comparative and US politics literatures agree that energy interests are one significant driver of climate policy outcomes. Regarding the US fracking boom in particular, there is some evidence that it helped Republicans win more elections, and it seemed to cause voting by members of Congress in the Pennsylvania–Ohio–West Virginia–New York shale region to cast more anti-environmental votes (Cooper, Kim, and Urpelainen 2018). However, there is not any existing evidence that the fracking boom affected (A) specific policy outcomes, (B) outcomes systematically across the entire geographic US, or (C) state-level government policy. In this paper, I aim to fill all those gaps.

In this paper, I set out to test the theory that the fracking boom – the rapid increase in shale oil and/or gas extraction from a subset of US states – decreased the stringency of climate policy in the states that were "treated" by fracking. As I will further describe in the Research Design section, I measure the treatment as shale coverage underneath a state's boundaries, which is more properly understood as the *potential to frack*. This measurement is advantageous because it is more exogenous to political outcomes than the *actual* extraction of gas or oil, which is endogenous to the preexisting level of

[8]In another political-economic analysis of policymaking, Trachtman (2021) employed a causal inference method to show that marijuana sector growth at the state level affected federal lawmaker voting on marijuana policy.

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

regulation, itself shaped by a state's prior political battles. Further, the potential to frack better captures the future profit incentives that may drive corporate political behavior.

I test the impact of fracking on two state climate policy outcomes. Therefore, the two distinct hypotheses are as follows. First, an increased fracking potential in a state caused a decrease in the chances of that state government to adopt the LEV policy (i.e., a regulation that mitigates the global warming impacts from passenger vehicles). Second, an increased fracking potential in a state caused that state government to decrease its mandated share of electricity generation to come from renewable (i.e., climate-friendly) energy sources. Both of these hypotheses follow from the climate political economy literature, which agrees that the economic incentives associated with incumbent energy industries cause carbon-intensive sectors to oppose stringent climate policies and take political actions to prevent climate policy enactment (e.g., Cory, Lerner, and Osgood 2021; Stokes 2020).

One economic aspect inherent to the commodity of shale fossil fuels is worth mentioning. Oil or natural gas produced by fracking can be transported across state lines. Therefore, if a fracking company extracts natural gas from a given state, it can transport and sell it elsewhere. This means there is less of an incentive to exert political pressure over climate policy in the same state where a fuel is extracted – so there may be an attenuated (i.e., less substantively large) effect of fracking in policy on the states where extracted, if companies are able to sell their product across all states. To be sure, it is still less expensive for a company to produce and sell gas or oil in the same state – all else equal – so there is still some incentive for the entire industry or an individual company to influence policy in that state where the fuel is extracted. This theoretical understanding does not bias the research design. It should only make us expect that fracking's effects on policy may not be as large as they might be, if the commodity could only be sold in the state where it was extracted.

My analysis in this paper does not directly test any potential causal mechanisms. I simply test *whether* fracking potential seemed to cause policy outcomes to change, not *how* it might do so. However, to improve the intuitive plausibility of the paper, it can help to imagine mechanisms through which this causal effect may be operating.

Overall, the fracking boom increased the current profits of fossil fuel companies and their incentive to benefit from more future extraction. One large category of mechanisms includes classic theories of direct, organized business influence on policy. With increased current and future profits, it is possible that these companies explicitly intended to influence policy by spending more money on lobbying (e.g., Hacker and Pierson 2010), direct campaign contributions (e.g., Hall and Wayman 1990), or outside "dark" money (e.g., Gilens, Patterson, and Haines 2021). These ways of using "instrumental power" may alter the political calculations of elected legislators, replace existing legislators with those friendlier to business interests, persuade the public on an issue relevant to business, provide more policy expertise that affects the eventual makeup of implemented policy, or have other effects that increase the chances that policy aligns with the interests or preferences of organized business.

Another category of mechanisms captures more indirect effects of such an economic interest. It is possible that fossil fuel companies influenced public opinion on climate policy in various states (as they seemed to have done nationally; e.g., Oreskes and Conway 2011). It may be that fossil fuel companies used their power to more subtly change what issues matter to their workers and voters in any state (e.g., Gaventa 1982). It is possible that unorganized "structural power" may be at work:

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

policymakers' own political interests have become dependent on fossil fuel companies' provision of jobs and tax-generating economic production, and that is why politicians may do what fracking companies want (e.g., Culpepper 2015; Lindblom 1977; Przeworski and Wallerstein 1988). Citizen interests are in part driven by their source of jobs and income, which, in this case, become somewhat in tension with more stringent climate policy. Overall, there are many possible mechanisms through which an increase in fossil fuel extraction may cause a decrease in the stringency of climate policy. My analysis in this paper does not distinguish between possible influence mechanisms.

Research design

The hypothesis I test in this paper is that fracking caused state climate policies to be less stringent than they otherwise would have been in the counterfactual absence of fracking. I employ a difference-in-difference (diff-in-diff, DiD, or D-i-D) research design, which relies on particular assumptions. In this section, I describe the measurements of treatment, outcome, and control variables, and I justify the relevant assumptions. Crucially, I also describe the original dataset that I introduce and employ in this paper: LEV policy adoption by state year.

Treatment: Distribution of shale deposits

The concept underlying the treatment variable is the fracking of shale oil or gas. However, the empirical analogue (i.e., the measured quantity) I employ in this paper is simply the distribution of a shale deposit, which is more accurately understood as the *potential to frack.* This measurement is more advantageous than actual gas or oil extraction data, since the shale distribution is plausibly exogenous to state political variables that are likely independent drivers of the relevant policy outcomes (to be discussed further in the Additional Assumption subsection). More specifically, the measurement of the treatment here is the share of any given state's geography covered (underneath) by shale.[9] It may be more accurate to measure the treatment variable as the actual *volume* of shale oil or gas reserves (by geographic area), but these data do not appear to exist for shale reserves, specifically; fortunately, the same measurement I employ (i.e., two-dimensional area) was used by Cooper, Kim, and Urpelainen 2018) in their study of causal impacts of fracking on congressional voting.[10] Here, I code treatment as shale distribution in multiple ways: greater than zero proportion of

[9]My decision to measure different levels of fracking based on percent land coverage follows Cooper, Kim, and Urpelainen (2018). These data, just as in Cooper, Kim, and Urpelainen (2018), come from the International Energy Agency (see "Shapefile" options): https://www.eia.gov/maps/maps.htm.

[10]One reason why volume of shale oil/gas reserves (i.e., three-dimensional) may be a more difficult quantity to measure is that the volume of extractable fuel is often unknown. Therefore, I rely on the more basic measure of whether or not (dichotomously) there is shale oil or gas underneath any given land area. One prominent example of unknown estimates of the volume of shale gas reserves is a 2008 Pennsylvania State University scientist broadcasting that the Marcellus shale area (PA) may have far more natural gas than previously estimated. ("Unconventional natural gas reservoir could boost U.S. supply." January 17, 2008. Penn State press release. https://www.psu.edu/news/research/story/unconventional-natural-gas-reservoir-could-boost-us-supply/). The US EIA publishes a range of data on "proved" shale gas reserves, but much of the data only exists since 2017: https://www.eia.gov/naturalgas/crudeoilreserves/.

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

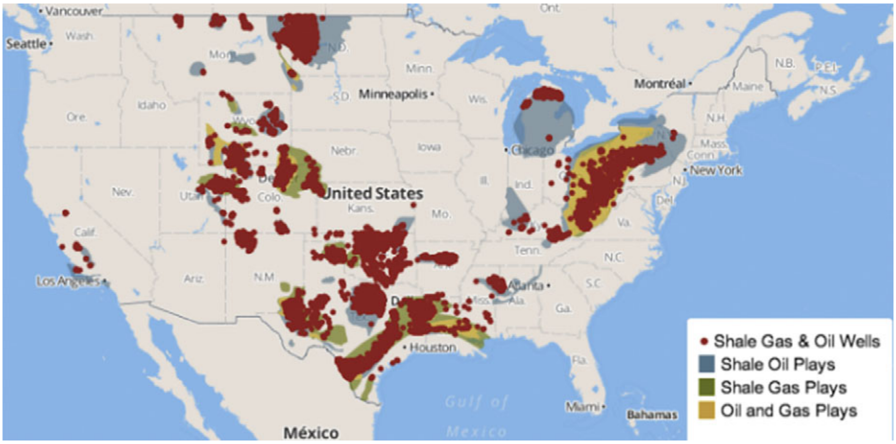


Figure 1. Geographic distribution of shale deposits in the US.
Source: Post Carbon Institute. https://shalebubble.org/dbd-map/.

a state covered by shale (which includes 29 states), greater than 5% of a state (21 states), greater than 10% of a state (16 states), and greater than 20% of a state (10 states). I study 49 out of 50 states – Texas is dropped for inferential reasons (see footnote).[11] Since the estimation strategy is difference-in-difference, a unit's status of being in the treated or untreated group takes on a binary value (0/1) for each of these possible treatment statuses.

The map in Figure 1 shows the distribution of shale reserves that are known to have natural gas or oil.

The potential to frack (and actual fracking wells, also shown on the map) is distributed around multiple parts of the country – concentrated in a few regions but still affecting many states. Table 1 lists which groups of states fall into which treatment status categories.

I conservatively take the beginning of the treatment period to be 2004.[12] The actual extraction of shale oil and gas slowly increased through the mid-2000s, really picking up closer to 2010. Therefore, I employ multiple treatment period options: 2004–2006, 2004–2008, and 2004–2010.[13] 2004–2006, the shortest treatment period, is most likely to avoid capturing other sorts of (time-variant) relevant political dynamics changing in states. Alternatively, 2004–2010 is most likely to pick up effects of a "stronger" treatment, since more shale oil and gas would have been extracted over that period, becoming more integral to a state's political economy. Thus, I use multiple treatment time periods to estimate the relevant regression quantities.

[11]This group of treated states does not include Texas, which is dropped from all analyses, per Cooper, Kim, and Urpelainen (2018)'s explanation that the threat of fracking into shale oil and gas existed earlier in Texas than in all other states: "shale gas extraction began much earlier in…Barnett, Texas, and cannot be considered exogenous for the purposes of identification" (635).

[12]Cooper, Kim, and Urpelainen (2018) consider 2004 to still be *prior* to the fracking boom, but their paper only studied the shale deposit surrounding Pennsylvania. Given that some shale extraction data (from Fedaseyeu, Gilje, and Strahan 2015) show some states with shale gas wells in 2004, I claim that the "pre-treatment" period only lasts through 2003.

[13]Cooper, Kim, and Urpelainen (2018) use 2005–2010 as their treatment period.

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

Table 1. List of state groups by treatment status, by different level of treatment

Treatment status: >0% shale coverage (excludes Texas)	
Treated states (29)	Untreated states (20)
Alabama, Arkansas, California, Colorado, Georgia, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maryland, Michigan, Mississippi, Missouri, Montana, Nebraska, New Jersey, New Mexico, New York, North Dakota, Ohio, Oklahoma, Pennsylvania, South Dakota, Tennessee, Utah, Virginia, West Virginia, Wyoming	Alaska, Arizona, Connecticut, Delaware, Florida, Hawaii, Idaho, Iowa, Maine, Massachusetts, Minnesota, Nevada, New Hampshire, North Carolina, Oregon, Rhode Island, South Carolina, Vermont, Washington, Wisconsin
Treatment status: >5% shale coverage (excludes Texas)	
Treated states (21)	*Untreated states (28)*
Alabama, Arkansas, Colorado, Indiana, Kansas, Kentucky, Louisiana, Maryland, Michigan, Mississippi, Montana, New Mexico, New York, North Dakota, Ohio, Oklahoma, Pennsylvania, Utah, Virginia, West Virginia, Wyoming	Alaska, Arizona, California, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Iowa, Maine, Massachusetts, Minnesota, Missouri, Nebraska, Nevada, New Hampshire, New Jersey, North Carolina, Oregon, Rhode Island, South Carolina, South Dakota, Tennessee, Vermont, Washington, Wisconsin
Treatment status: >10% shale coverage (excludes Texas)	
Treated states (16)	*Untreated states (33)*
Alabama, Arkansas, Colorado, Indiana, Kentucky, Louisiana, Michigan, Mississippi, Montana, New York, North Dakota, Ohio, Oklahoma, Pennsylvania, West Virginia, Wyoming	Alaska, Arizona, California, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Iowa, Kansas, Maine, Maryland, Massachusetts, Minnesota, Missouri, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, North Carolina, Oregon, Rhode Island, South Carolina, South Dakota, Tennessee, Utah, Vermont, Virginia, Washington, Wisconsin
Treatment status: >20% shale coverage (excludes Texas)	
Treated states (10)	*Untreated states (39)*
Indiana, Kentucky, Louisiana, Michigan, New York, North Dakota, Ohio, Pennsylvania, West Virginia, Wyoming	Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Iowa, Kansas, Maine, Maryland, Massachusetts, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, North Carolina, Oklahoma, Oregon, Rhode Island, South Carolina, South Dakota, Tennessee, Utah, Vermont, Virginia, Washington, Wisconsin

Figure 2 shows the timing of shale gas extraction increases. The notable increase in shale gas extraction began around 2005–2006 and steadily increased. Extraction of *oil* from shale reserves similarly began in the mid-2000s and then steeply increased around 2010.[14] It would be ideal to be able to separate – by geography and time – the

[14]US Energy Information Administration, Annual Energy Outlook 2013 Early Release. https://www.eia.gov/outlooks/aeo/er/pdf/0383er(2013).pdf.

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

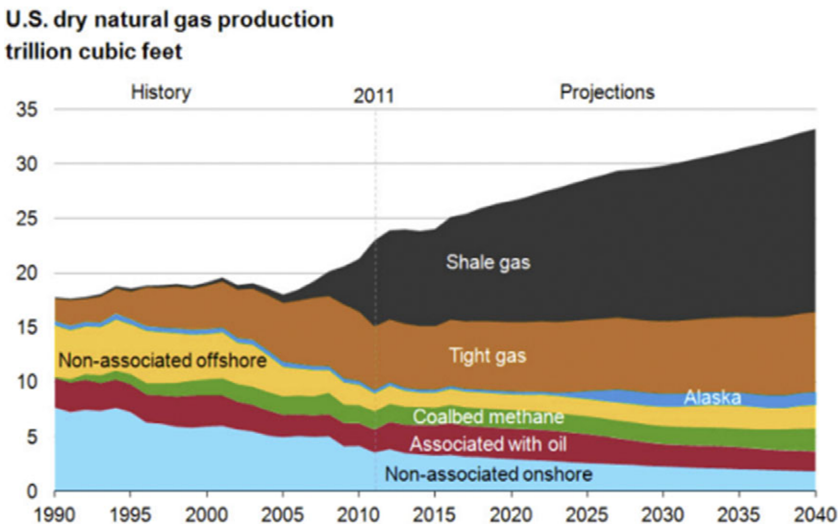


Figure 2. Over-time increase in shale gas production in the USA.
Source: US Energy Information Administration, Annual Energy Outlook 2013 Early Release. https://www.eia.gov/outlooks/aeo/er/pdf/0383er(2013).pdf.

extraction of shale oil from shale gas in the data, as they are energy sources used in distinct sectors of the economy and are therefore of difference relevance to unique policies. In this paper's case, LEV policy governs cars, which use *mostly* gasoline (from oil), and RPS policies govern electricity, which is powered by an array of sources (today by a far higher share of natural gas than oil; but in the early- and mid-2000s, a non-negligible share of US electricity came from petroleum/oil).[15] However, data limitations appear to make this impossible. I am forced to keep shale oil and gas bundled together as one treatment. Although some nuance is lost by keeping them bundled, it is nonetheless theoretically possible that any fossil fuel company (whether fracking for oil or gas) will oppose and fight any climate policy, whether it aims to mitigate emissions from automobiles (LEV) or electricity generation (RPS). Further, companies that extract oil also extract gas, so the political goals of oil and gas are often intertwined.[16] We therefore may not gain much unique treatment measurement by quantitatively separating oil from gas. Unfortunately, the data at hand cannot distinguish between these various theoretical possibilities.

Outcome 1: Original data on LEV policy adoption

The first contribution I make in this paper is to introduce an originally collected dataset that reports which states adopted the LEV policy in which year – which is the

[15]"Electricity explained." April 19, 2022. US Energy Information Administration. https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php.

[16]A 2009 EIA report shows that some of the biggest oil producers were also the biggest gas producers in the USA. "Performance Profiles of Major Energy Producers 2009." February 2011. US Energy Information Administration. https://www.eia.gov/finance/performanceprofiles/pdf/020609.pdf.

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

first of two policy outcome variables I employ.[17] These LEV policy data take the form of state–year and spans 1991–2015, which encapsulates the beginning of the policy's history in the states.[18]

LEV is a policy that states have been able to adopt in order to decrease the air pollution and greenhouse gas impacts of vehicles sold in their state. However, most states do not have the legal ability to craft a LEV to their liking. California created LEV in 1990 – updating the policy over time to include different pollutants and increased levels – and all other states can only choose to adopt portions of California's LEV; alternatively, they can default to federal vehicle emission standards. This legal regime exists because the Clean Air Act does not allow states to preempt federal regulations, except that California is specifically exempted from that rule (per section 209 of the Clean Air Act) because it took very early state action (in 1966) to mitigate pollution from vehicles.[19] Therefore, all states (other than California) face the policy choice between just two options: *default to federal vehicle emission regulations or adopt California's LEV*. Therefore, the measurement of LEV adoption in this paper dichotomously only takes the value of 0 (non-adoption, defaulting to federal rules) or 1 (adopting California's LEV).

The US federal government has had vehicle fuel standards in place since 1975, first for passenger cars and a few years later for "light trucks" (e.g., sport utility vehicles, pickup trucks). These regulations did not change much over a few decades (until the 2007 Energy Independence and Security Act). California, seeking to do more to mitigate the impact of vehicles on its air quality (and in later LEV iterations, the impact on the global climate), created the first LEV regulation in 1990, submitting a formal waiver to the federal government to allow the state to go above and beyond the federal rules.[20]

This first LEV regulation affected vehicles in model years 1994–2003. It created maximum levels for many pollutants – including carbon monoxide, nitrogen oxide, and others – emitted from passenger cars, light-duty vehicles (e.g., minivans, SUVs), and medium-duty vehicles (e.g., box trucks, school buses). A full list of regulated pollutants is available in the source in the footnote. California then updated its LEV regulations in 1998 and 2004, increasingly the stringency of preexisting pollutant maxima and regulating new pollutants (e.g., carbon dioxide in its 2004 iteration). During this period of decades, the federal government did have some regulations governing similar pollutants, but California's LEV policy was always more stringent.[21]

[17] I owe thanks to Srinivas Parinandi, assistant professor at University of Colorado Boulder, for the idea to collect this original data when we began a project together in 2018. These data have not been used for any published work before this paper.

[18] These data end in 2015 only because the original collection happened in 2018.

[19] Sources include: (1) "U.S. State Clean Vehicle Policies and Incentives." 2019. Center for Climate and Energy Solutions. https://www.c2es.org/document/us-state-clean-vehicle-policies-and-incentives/. (2) "42 U.S. Code 7543 – State standards." Legal Information Institute, Cornell Law School. https://www.law.cornell.edu/uscode/text/42/7543. (3) "History." California Air Resources Board. https://ww2.arb.ca.gov/about/history.

[20] Sources include: (1) "Federal Vehicle Standards." 2021. Center for Climate and Energy Solutions. https://www.c2es.org/content/regulating-transportation-sector-carbon-emissions/. (2) "A Brief History of US Fuel Efficiency Standards." 2017. Union of Concerned Scientists." https://www.ucsusa.org/resources/brief-history-us-fuel-efficiency#.WxWy2dPwbq0.

[21] The Zero-Emission Vehicle (ZEV) policy is a portion that California added onto LEV to encourage electric car sales. Other states have also adopted ZEV over time, but ZEV is not a focus of this dataset or analysis. Sources include: (1) "The California Low-Emission Vehicle Regulations. (With Amendments

Immediately after California adopted the first LEV iteration in 1990, other states began adopting LEV, as well (e.g., New York and Massachusetts in 1991 and Maine in 1993, even before the EPA formally approved California's waiver in 1993), which was allowed under Section 177 of the Clean Air Act. Some state legislatures passed bills to adopt LEV (e.g., Connecticut and New Jersey in 2004) while some governors perceived they had the legal authority to adopt LEV via executive action, without legislation (e.g., Arizona in 2006, Delaware in 2010).[22] In total, original data collection yielded 15 total states (including California; see footnote) who have adopted LEV by 2015.[23] LEV adoption is coded as a binary outcome per state–year – 0 for non-adoption, 1 for adoption.[24]

Outcome 2: Mandated renewable share of state electricity generation

The second outcome variable employed in this analysis is the share of a state's electricity generation mix mandated to come from renewable sources. A higher share mandated to come from renewables is a more stringent climate policy, thus mitigating more greenhouse gas emissions. This is measured using data from the Lawrence Berkeley National Lab.[25] More precisely, it is total megawatt-hours (MWh) of electricity mandated to come from renewable sources, per state (and year), divided by total MWh generation per state (of all electricity sources),[26] to produce the percentage of renewable electricity mandated, per state and year. These renewable energy data include various definitions – decided by each individual state – of what counts as a "renewable" source (e.g., some states classify biomass as renewable). This variation in energy creates a bit of measurement error for this outcome variable. However, this should not create inferential problems: Even though the definition of "renewable" can vary, it does not appear to ever include natural gas, however classified, mitigating concerns that this outcome variable could also measure some of the treatment variable (i.e., fracking). Some combination of state legislatures and executive agencies has primary decision-making power over the level of renewable energy mandated. These outcome variable data are

---

Effective October 1, 2019) "2019. California Air Resources Board. (2) "CALIFORNIA: LIGHT-DUTY: LOW EMISSION VEHICLES." 2018. TransportationPolicy.net http://www.transportpolicy.net/standard/california-light-duty-low-emission-vehicles/.

[22]A dichotomous control variable for Democratic governor is included in the diff-in-diff regressions.

[23]The LEV-adopting states (through 2015) include: Arizona, California, Connecticut, Delaware, Maine, Maryland, Massachusetts, New Jersey, New Mexico, New York, Oregon, Pennsylvania, Rhode Island, Vermont, and Washington. A couple other states such as Texas and North Carolina adopted small portions of LEV, such as only heavy-duty vehicle diesel pollutant requirements. This original dataset only counts a state as adopting LEV if the passenger car and light-duty vehicle portions of LEV were adopted. Colorado is one state who adopted LEV after 2015 (in 2018). Other states may have as well, although the original dataset only exists until 2015.

[24]Sources include: (1) "California's Light-Duty Vehicle Emissions Standards: The Clean Air Act Waiver, Standards History, and Current Status." 2017. Issue Brief, MJ Bradley & Associates, LLC. (2) "Sales of California-certified 2008–2010 Model Year Vehicles (CrossBorder Sales Policy)." 2007. Letter from US EPA to Manufacturers. https://iaspub.epa.gov/otaqpub/display_file.jsp?docid=16888&flag=1.

[25]The source for this data is a 2018 report from Lawrence Berkeley National Lab: https://emp.lbl.gov/publications/us-renewables-portfolio-standards-1.

[26]Data from EIA again: https://www.eia.gov/electricity/data.php.

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

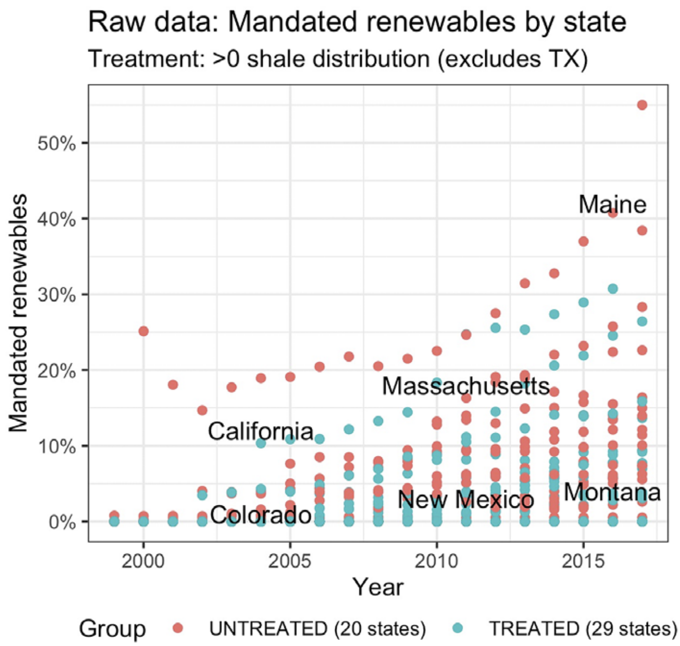


Figure 3. For >0% shale coverage treatment status: raw data (entire *y*-axis shown).

available for 1999–2017. Figure 3 shows the entire raw data to visually depict the variation over time for this variable (and labels for a few high- and low-percentage states).

Difference-in-difference assumption: Common trends

The primary assumption employed in diff-in-diff estimation is common trends – that the trend of the outcome variable for the treatment group would have continued, posttreatment, at the same trend that the control group's observed outcomes did, in the absence of treatment. While this is an assumption and so cannot be seamlessly empirically verified, common trends charts (displaying the over-time trend in the outcome variable *prior to treatment*) can visually provide some increased level of confidence in this assumption. Figures 4 and 5 show common trends for the "strongest" two treatments (>20% and >10% shale coverage) for the first outcome variable, LEV adoption from 1991 to 2015 (the entire history of the data); common trends for other treatment types are shown in the Supplementary Material. (Note that the *y*-axis is cut down, as distinct from Figure 4, in these and some following charts, in order to visually center the group means.)

Figures 4 and 5 show that prior to the treatment period beginning (which is 2004; the vertical line sits at 2003.5), the levels of both treated and untreated groups moved at similar trends until the treatment period began, at which point, the untreated states (as a group) increased their LEV adoption rate. This common trend chart shows that the causal impact of fracking that plausibly happened was in fact one that caused

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

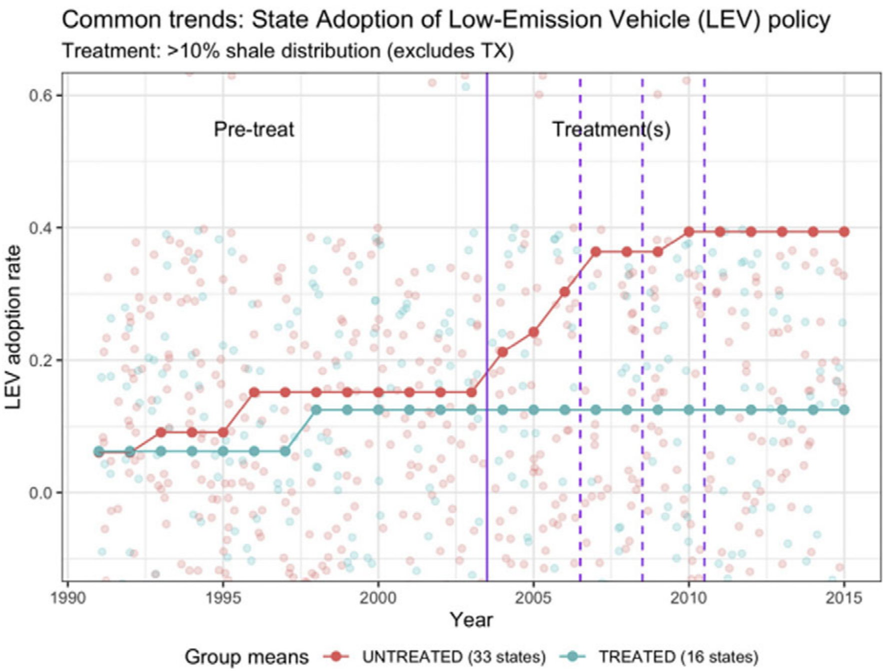


Figure 4. For >10% shale coverage treatment status: Common trends before, during, and after the height of the fracking boom.
Vertical lines (2003, 2013) indicate breaks between pretreatment (1999–2003), treatment (2004–2006, 2004–2008, or 2004–2010), and posttreatment (2007–2011, 2009–2013, or 2011–2015) periods. Raw data are plotted in the background of group means and trends. Multiple vertical dotted lines indicate the multiple treatment timing periods tested.

treated states (those with shale) to not follow untreated states in increasing the adoption rate in the mid-late 2000s. I do not claim that the *lack of fracking potential* in untreated states is the proximate cause of their LEV adoption rate increase, but rather that fracking had a causal impact on holding back the group of treated states from an increased likelihood of adopting LEV.

Figure 6 plots the common trend for the second policy outcome variable, renewable electricity percentage mandates, for just one treatment status: >10% shale coverage (common trends for other treatment statuses are shown in the Supplementary Material). Since these policy data exist only starting in 1999, there are not much historical data to be able to see pretreatment common trends. Pretreatment, the treated and untreated states clearly diverge by small outcome variable percentages (<2% vs. 0%) in the trends of the outcome variable averages. Unfortunately, the lack of pretreatment data for this policy outcome variable (due to its recent development) leaves us with little ability to visually study these pretreatment common trends.

Additional assumption: Treatment exogeneity

The difference-in-difference research design does not rely on an exogenous treatment; it only relies on the assumption of common trends in the outcome variable. Overall, we want to be confident that the changed outcome of the treated group was

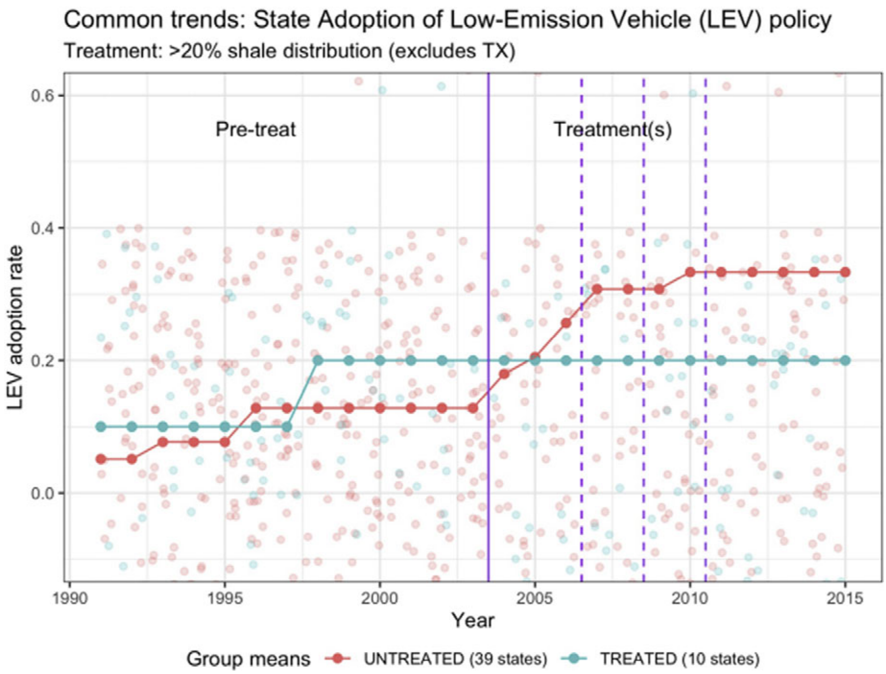


Figure 5. For >20% shale coverage treatment status: Vertical lines (2003, 2013) indicate breaks between pretreatment (1999–2003), treatment (2004–2006, 2004–2008, or 2004–2010), and posttreatment (2007–2010, 2009–2012, or 2011–2014) periods.

truly caused by the treatment. However, employing the additional identifying assumption of exogeneity – that *the geographic distribution of shale* is orthogonal to political development prior to the fracking boom – strengthens our confidence that the diff-in-diff design is producing a causal effect. Cooper, Kim, and Urpelainen (2018), in employing shale geography as an exogenous treatment, write: "the definition of a shale play is ideally suited for an identification strategy based on the exogenous distribution…it does not require the onset of extraction activity or consider possible regulatory issues" (637).

If it is true that the assignment of shale to a state was exogenous, that means that no other variable relevant to this political context caused its observed variation. It is clear that shale was distributed by the geologic processes of the earth, which happened prior to the political development of each state. However, it is possible that the shale distribution could be correlated with similar resource development, like oil, which could affect a state's political development. For that reason, regressions include one control variable that measures the prior decade of oil production (as a share of gross state product, or GSP).

A second way that the distribution of shale per state may be endogenous to renewable energy policies is if state policymakers – legislative or executive – knew about and anticipated the effect of fracking on a state's politics. As previously discussed, fracking technology became widely available during the mid-2000s; before that, there was virtually zero shale oil or gas extraction. However, it also appears that there was little knowledge or political attention to fracking. Helpful qualitative

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

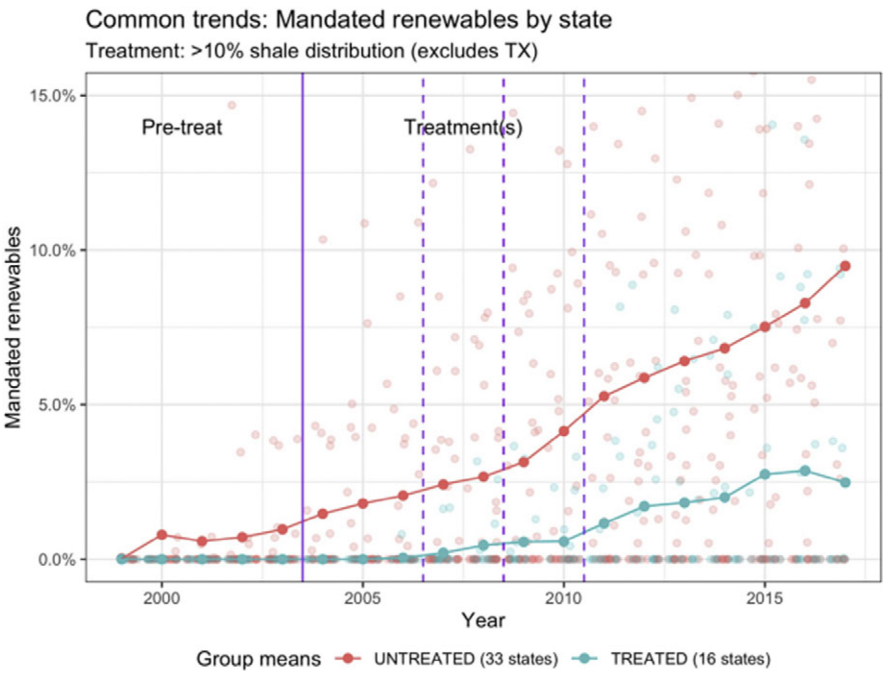


Figure 6. For >10% shale coverage treatment status: Vertical lines (2003, 2013) indicate breaks between pretreatment (1999–2003), treatment (2004–2006, 2004–2008, or 2004–2010), and posttreatment (2007–2010, 2009–2012, or 2011–2014) periods.

research on the brief history of the fracking boom from Cooper, Kim, and Urpelainen (2018) supports this assertion. Cooper and coauthors read journalistic accounts and searched media headlines, finding little evidence of attention to potential shale gas drilling prior to the late 2000s outside of Texas: "[we] found no evidence of widespread interest in shale gas outside Texas by the end of 2004" (Cooper, Kim, and Urpelainen (2018, 638). Even though some states clearly knew they had shale basins underneath their geographic boundaries, estimates of how much gas could be extracted were quite low. Texas is dropped from this analysis.

Pretreatment control variables

Given the orthogonal causes of the geographic distribution of shale, we can have some confidence that the assignment of the treatment of shale for fracking is plausibly exogenous for this research design's purposes. However, it is still not the case that shale distribution was assigned to states such that the treated and untreated groups are balanced by all imaginable relevant traits. The common trends assumption may be reasonable, but I still choose to employ control variables in the diff-in-diff regressions, to make the estimate of the effects of fracking more precise. This is similar logic to using pretreatment covariates in an experimental setting – to improve the precision of the treatment effect estimates, especially when the treated and untreated groups are not necessarily balanced by relevant covariates. It is also similar to using regression weights based on pretreatment covariates. In other words, we

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

need to verify that it was not the case that states already (predisposed) likely to adopt less stringent climate policies were simply given shale. By different logic, the use of pretreatment covariates is also a hedge against the possibility that the common trends assumption seems less reasonable to some readers. Given the pretreatment common trends Figures 4 and 5, it appears far more valid for the LEV policy than for mandated renewables, given that pretreatment outcome data exist further back (historically) for LEV. However, in the Supplementary Material I display two-way fixed effects (i.e., state-time) regressions that do not use these pretreatment covariates; those regressions produce largely similar results.

In this paper's primary regressions, I include a handful of covariates (i.e., control variables) that could plausibly drive differences in outcomes. One pair of control variables is unified Democratic state government control and gap in partisan affiliation of a state's electorate (i.e., Democrats minus Republicans),[27] because a prior study (Trachtman 2020) found that those both correlate quite well with a state adopting stringent renewable energy policy. This is intuitive, as it is generally understood in American politics that the Democratic Party coalition contains climate and environmental groups, while the Republican Party virtually does not. Crude oil extracted per state[28] – divided by state GSP[29] – is also included as a covariate, as a measure of a state's preexisting economic reliance on oil. State GSP (per capita) on its own is also included, in the chance that richer states have more leeway to adopt potentially costly climate policies. State government ideology is also included,[30] since more liberal state governments are more likely to adopt more stringent climate policy. Democratic governor is added as an additional control, since some state executive branches perceived they had the legal ability to adopt these policies without consent of the legislature.

A final key pretreatment covariate is a state population's environmental policy opinion,[31] since the public's opinion may certainly affect this policy outcome. The exact question wording is, "I support pollution standards even if it means shutting down some factories," and answers range from 1 to 6 in levels of support. This is not a survey precisely about clean car or renewable electricity policy, but it may be a good enough proxy.[32] These rare state-level data are only available until 1998, so this variable is measured as the average of a state's opinion from 1990 to 1998. Crucially, this control variable allows more weight to the claim that the evidence in this paper of the influence of fracking on policy change is *contrary* to this particular sense of the public interest – stated environmental policy preferences measured *before* the possibility of fracking.[33]

[27]Data from Caughey and Warshaw (2018).

[28]Data from EIA: https://www.eia.gov/dnav/pet/pet_crd_crpdn_adc_mbbl_m.htm.

[29]Data from US Bureau of Economic Analysis: https://www.bea.gov/data/gdp/gdp-state.

[30]Data from Berry et al. (1998), updated.

[31]Environmental policy opinion is measured by DDB Life Style survey data, which ends in 1998: http://bowlingalone.com/?page's_id=7.

[32]This seems to be the best available measure of public opinion on environmental issues measured before 2003 that exists for nearly all states. Hawaii and Alaska are not included, so those two states are also dropped from regressions (along with Texas, mentioned previously).

[33]See the Supplementary Material for a discussion on the possibility of an instrumented difference-in-difference method.

Results

The following equation is the general form that all regressions employed for this research design use.

$$\text{Policy} = \beta_0 + (\beta_1 \text{Treated}) + (\beta_2 \text{Time}) + (\beta_{3-i} \text{Pre Treatment Controls})$$
$$+ (\beta_k \text{Treated} \times \text{Time}) + e$$

Therefore, β_k estimates the causal effect of interest. The following table shows main diff-in-diff regression results for LEV policy, outcome 1, showing different treatment statuses (>0%, >5%, >10%, and > 20% shale coverage). Only the 2004–2006 treatment timing period is shown; the Supplementary Material shows similar tables for 2004–2008 and 2004–2010 treatment periods (results are very similar). In the Supplementary Material, I also display multiple two-way fixed effects models – that yield similarly significant results – to corroborate the main diff-in-diff results.

Results in Table 2 show evidence of statistically significant correlations for the interaction term (i.e., treated × time), the relevant coefficient for diff-in-diff designs, for treatment statuses of >10% and >20% (but not >0% nor >5%). Results are displayed without pretreatment control variables results (full regression results tables shown in the Supplementary Material). This suggests that the possible causal effect of fracking potential on LEV policies in states operated at more significant levels of

Table 2. Effect of fracking on LEV policy, 2004–2006 as treatment period.

	Shale > 0%	Shale > 5%	Shale > 10%	Shale > 20%
	(29 states)	(21 states)	(16 states)	(10 states)
Time (0/1)	0.278*	0.231*	0.258**	0.216**
	(0.109)	(0.084)	(0.080)	(0.069)
Treated_0 (0/1)	0.006			
	(0.111)			
Treated_0 × Time	−0.174			
	(0.123)			
Treated_05 (0/1)		0.033		
	(0.114)			
Treated_05 × Time		−0.136		
		(0.107)		
Treated_10 (0/1)			0.093	
			(0.131)	
Treated_10 × Time			−0.258**	
			(0.080)	
Treated_20 (0/1)				0.237
				(0.147)
Treated_20 × Time				−0.216**
				(0.069)
R^2	0.406	0.398	0.412	0.410
Adj. R^2	0.334	0.325	0.341	0.339
Num. obs.	94	94	94	94
Root mean square error (RMSE)	0.347	0.350	0.345	0.346
N Clusters	47	47	47	47

Abbreviation: LEV, low-emission vehicle.
**$p < 0.01$;
*$p < 0.05$.

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

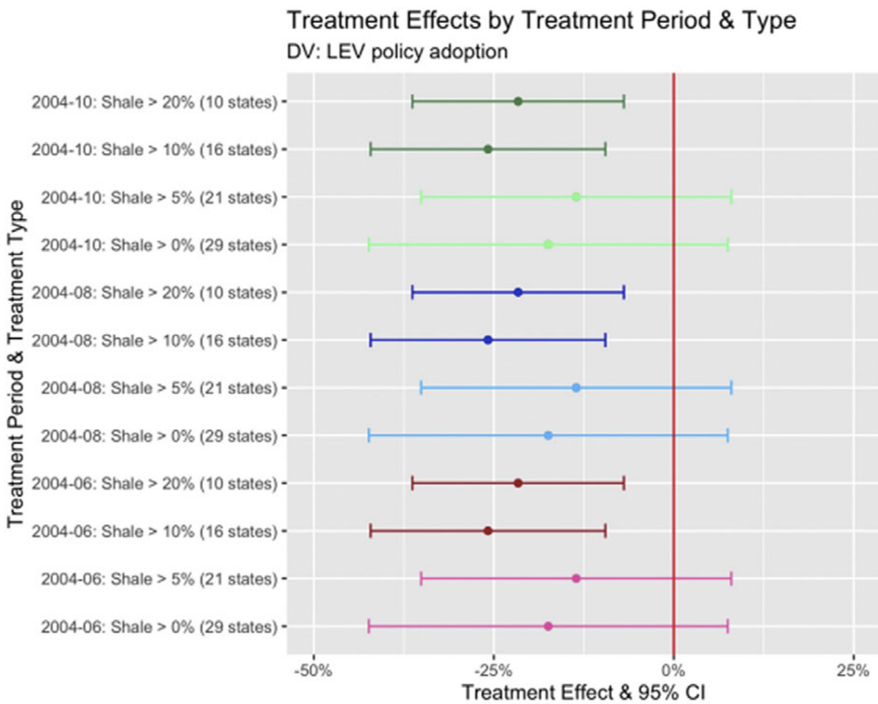


Figure 7. Error bars show 95% confidence intervals.
State control variables are included in the regressions that produced these estimates.

treatment (i.e., higher fracking share of a state). The statistically significant coefficients of −0.258 (>10% shale) and −0.216 (>20% shale) mean that being treated with fracking caused a 26 percentage point and 22 percentage point decrease, respectively, in the chances of adopting a LEV policy for the group of states. This is substantively large.

The following coefficient plot in Figure 7 shows the regression results (including all the treatment timing periods) in more visual terms.

Table 3 shows main diff-in-diff regression results for mandated renewable electricity, outcome 2, showing different treatment statuses (>0%, >5%, >10%, and > 20% shale coverage) for the 2004–2006 treatment timing period.

Results in Table 3 show evidence of statistically significant correlations for the interaction term (i.e., treated × time), the relevant coefficient for diff-in-diff designs, for all treatment statuses of >5%, >10%, and >20% (but not >0%). This suggests that the possible causal effect of fracking potential on renewable electricity mandates in states operated at more significant levels of treatment, similar to the effects on LEV policy adoption. The statistically significant coefficients of −0.022, −0.024, and −0.022, respectively, mean that being treated with fracking caused a roughly 2 percentage points drop in mandated renewable portion of state electricity generation. This is a substantively meaningful coefficient size, as the average level of renewables mandated for the treated group (when treatment is measured as >10%

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

Table 3. Effect of fracking on renewable electricity policy, 2004–2006 as treatment period

	Shale > 0% (29 states)	Shale > 5% (21 states)	Shale > 10% (16 states)	Shale > 20% (10 states)
Time (0/1)	0.032**	0.031**	0.030**	0.026**
	(0.008)	(0.008)	(0.007)	(0.006)
Treated_0 (0/1)	−0.006			
	(0.008)			
Treated_0 × Time	−0.016			
	(0.011)			
Treated_05 (0/1)		−0.004		
		(0.007)		
Treated_05 × Time		−0.022*		
		(0.009)		
Treated_10 (0/1)			−0.002	
			(0.007)	
Treated_10 × Time			−0.024**	
			(0.008)	
Treated_20 (0/1)				0.000
				(0.006)
Treated_20 × Time				−0.022**
				(0.007)
(Intercept)	0.016	0.044	0.053	0.030
	(0.111)	(0.115)	(0.118)	(0.112)
R^2	0.253	0.263	0.259	0.243
Adj. R^2	0.162	0.174	0.170	0.152
Num. obs.	94	94	94	94
Root mean square error (RMSE)	0.035	0.034	0.035	0.035
N Clusters	47	47	47	47

$^{**}p < 0.01$;
$^{*}p < 0.05$.

shale, as in Figure 6) in 2015 is only 5% total. However, given the lack of pretreatment data to show some validation of the common trends assumption, we are left with only assuming no unobserved confounding for testing the effect of fracking on renewable electricity policy. Therefore, this evidence of fracking's impact on state renewable electricity proportions is more suggestive than the LEV adoption evidence.

Unlike the LEV adoption (outcome 1) regression results, the analyses for fracking's potential impact on mandated renewables (outcome 2) seem to increase as the treatment period lengthens. Therefore, Table 4 shows analyses for the 2004–2010 treatment period. Results for 2004–2008 treatment timing are shown in the Supplementary Material.

The following coefficient plot in Figure 8 shows the regression results (including all the treatment timing periods) in more visual terms.

Placebo test results shown in the Supplementary Material – to test whether treatment may have affected the outcomes *before* 2004 – show null results.

While I have aimed to evaluate the influence of fracking on state climate policy, the shale oil and gas extraction happened at a particular moment in US political history. Therefore, it is worth noting some theoretical conditions that can help with hypothesizing about the generalizability of these findings. First, fracking happened decades and centuries after states had politically developed, so this influence happened at a particular point in the development of policy – there may have already been some

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

Table 4. Effect of fracking on renewable electricity policy, 2004–2010 as treatment period

	Shale > 0% (29 states)	Shale > 5% (21 states)	Shale > 10% (16 states)	Shale > 20% (10 states)
Time (0/1)	0.061**	0.059**	0.058**	0.052**
	(0.015)	(0.014)	(0.013)	(0.011)
Treated_0 (0/1)	−0.005			
	(0.010)			
Treated_0 × Time	−0.026			
	(0.019)			
Treated_05 (0/1)		0.001		
		(0.009)		
Treated_05 × Time		−0.032		
		(0.017)		
Treated_10 (0/1)			−0.000	
			(0.009)	
Treated_10 × Time			−0.039*	
			(0.015)	
Treated_20 (0/1)				0.001
				(0.010)
Treated_20 × Time				−0.034*
				(0.014)
(Intercept)	0.018	0.035	0.077	0.045
	(0.152)	(0.158)	(0.159)	(0.152)
R^2	0.322	0.324	0.335	0.316
Adj. R^2	0.240	0.243	0.255	0.234
Num. obs.	94	94	94	94
Root mean square error (RMSE)	0.050	0.050	0.050	0.051
N Clusters	47	47	47	47

$^{**}p < 0.01$;
$^{*}p < 0.05$.

level of entrenched influence in the policy prior the initial time period. Further, this influence under study is the potential effect of an economic interest in just one sector, and the independent variable is not often (for most states) the growth from zero fossil fuel corporation presence; instead, it is usually from some nonzero level of economic activity to a higher level. In these ways, these results estimate a "local" treatment effect in this paper – local in time, and local in the stage of fossil fuel development.

Is this evidence of business influence?

In this paper, I show some evidence that the fracking boom seemed to diminish climate policy stringency at the state level. Traditionally, "business influence" is usually thought to include lobbying, campaign spending, and occasionally attempts to sway public opinion. However, if we take seriously Lindblom's (1977) argument that corporations may bias government policy in their favored direction via "structural power," then we can conceive of this evidence of fracking's influence on policy as the influence of business.

It is true that I do not test any hypotheses about which mechanisms may be at play, exerting the influence of fracking on policy. It could be traditional methods of organized business influence attempts, such as lobbying or campaign spending; it could be that more voters in the state saw their economic well-being as connected

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

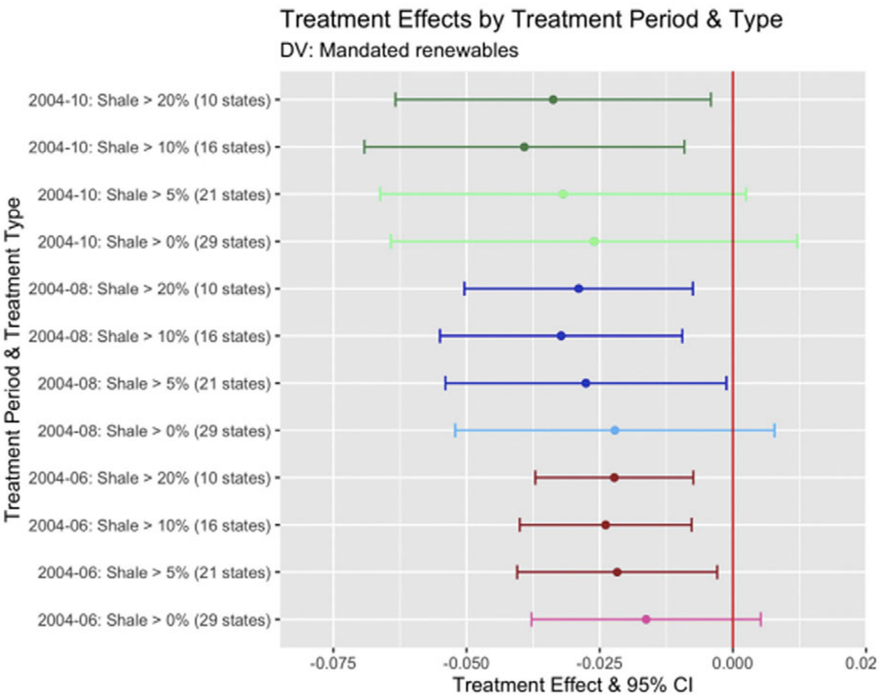


Figure 8. Error bars show 95% confidence intervals.
State control variables are included in the regressions that produced these estimates.

with the interests of fossil fuels and therefore opposed climate policy; it could be that fracking companies directly persuaded the public against climate policy through advertising campaigns; it may be that Lindblom's (1977) "structural power" was at work, causing state legislators to see their political fortunes as interdependent with the jobs and tax revenue that the state fossil fuel industry was providing.

The point is that the production of a commodity seemed to bias government policy in a direction that favored that commodity's future production. Under a more expansive definition of "business influence" – that Lindblom would argue for – I have uncovered evidence of the influence of business on policy. Therefore, future empirical attempts that aim to test hypotheses about business' impact on politics and policy should ensure to account for multiple possible influence mechanisms (Hacker and Pierson 2002). A subliterature in American politics sometimes argues that business may not actually influence policy all that much (e.g., Ansolabehere, De Figueiredo, and Snyder 2003; Fowler, Garro, and Spenkuch 2020; Smith 2000), but those studies do not often account for some mechanisms that may, in fact, be at work.

Conclusion

The fracking boom generated enormous economic activity in some American states in the 2000s and 2010s. Contemporaneously, the climate crisis intensified – as did

https://doi.org/10.1017/spq.2022.17 Published online by Cambridge University Press

public opinion and social movement activity in support of more aggressive emissions-mitigating policies. While US emissions have begun falling, experts agree that far more stringent policies to further mitigate emissions will be necessary to stave off the worst effects of global warming.

In this paper, I have advanced our knowledge about the particular impact of fracking on politics and policy, showing that it harmed the ability for American state governments to enact emissions-mitigating policies. Specifically, the ability to extract shale oil or gas caused state governments to be less likely to enact the LEV policy and less likely to mandate a higher share of electricity generation to come from renewable energy. This research provides novel quantitative (causal inference) evidence that the fracking boom caused state-level *climate policy outcomes* to be less stringent than they otherwise would have been (in the absence of fracking). Existing studies have provided suggestive evidence that fracking has seemingly strengthened Republican political forces (DiSalvo and Li 2020; Fedaseyeu, Gilje, and Strahan 2015; Sances and You 2022) and has caused members of Congress to vote in more anti-environmental ways, across a host of issues (Cooper, Kim, and Urpelainen 2018). Those papers provide initial looks at some effects of fracking, and I do not disagree with them. My argument in this paper goes deeper: I point to the influence of fossil fuel interests on the eventual prize of climate politics battles at the American subnational level, *policies.* To my knowledge, I add the (heretofore) only quantitative causal inference evidence of this. And after all, subnational governments (primarily states) are key to the US's ability to mitigate the climate crisis overall, particularly in the face of federal government gridlock.

More broadly, my findings in this paper highlight the core assertion from political economy research: that economic resources are prominent drivers of policy outcomes. Other scholars have argued that fossil fuel interests – and their organized political strategies – have been primary drivers of climate politics (Colgan, Green, and Hale 2021; Cooper, Kim, and Urpelainen 2018; Hughes and Urpelainen 2015; Mildenberger 2020; Stokes 2020). Together, my paper and this body of work imply that incumbent economic industries will be resistant to change, will likely fight in the organized political realm (as Oreskes and Conway 2011; Stokes 2020 and others have shown in the organized battle of American climate politics), and may win policy battles more often than normative democratic theory portends. It is not public opinion or protest activity that primarily drives climate politics – concentrated economic interests dominate. This phenomenon may generalize from climate politics – not just federally, but also at the American state level – to other policy arenas for which large, profitable industries have a stake in the status quo economic regime: the regulation of finance, technology, healthcare, and many more. Policymakers and other powerful decision-makers should take note when an industry begins to carve out its share of the American economy – that industry may have also begin carving out its influence over public policy.

Lastly, I have conceptually argued that this paper's evidence – of a commodity shock biasing policy in the direction of that commodity's future economic well-being – should be considered evidence of the influence of business in politics. This conception of business influence takes the notion of "structural power" (Culpepper 2015; Lindblom 1977; Przeworski and Wallerstein 1988) seriously, that certain industrial sectors or individual corporations may have a deep hold over politicians by the nature that they can control jobs, consumer prices, and the raising of some tax revenue. Prior studies that have claimed to find little to no influence of money in politics. And while understanding specific influence mechanisms is outside the scope of this paper, my research design does account for all forms of instrumental and structural power influence. Therefore, future research would do well to theoretically imagine these more expansive pathways of influence and empirically test for mechanisms that include forms of structural power.

Supplementary Materials. To view supplementary material for this article, please visit http://doi.org/10.1017/spq.2022.17.

Data availability statement. Replication materials are available on SPPQ Dataverse at https://doi.org/10.15139/S3/UGCS7Y (Zacher 2022).

Acknowledgments. The author thanks Ian Turner for formally advising this paper in its initial stages, Jacob Hacker and Bryan Garsten for overseeing the course for which this paper was developed, and Jacob for comments throughout its various stages. Further, I appreciate other feedback on this project by participants at the American politics graduate student workshop and Research & Writing final presentation.

Funding statement. The author received general financial support for the research publication of this article from Yale University's graduate program.

Conflict of interest. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Ansolabehere, Stephen, John M. De Figueiredo, and James M. Snyder Jr. 2003. "Why is There So Little Money in US Politics?" Journal of Economic Perspectives 17 (1): 105–30.

Bang, Guri. 2015. "The United States: Obama's Push for Climate Policy Change". In *The Domestic Politics of Global Climate Change*. Northampton, MA: Edward Elgar.

Berry, William D., Evan J. Ringquist, Richard C. Fording, and Russell L. Hanson. 1998. "Measuring Citizen and Government Ideology in the American States, 1960-93." *American Journal of Political Science* 42: 327–48.

Bishop, Bradford H., and Mark R. Dudley, 2017. "The Role of Constituency, Party, and Industry in Pennsylvania's Act 13". *State Politics & Policy Quarterly* 17 (2): 154–79.

Brulle, Robert J., 2018. "The Climate Lobby: A Sectoral Analysis of Lobbying Spending on Climate Change in the USA, 2000 to 2016." *Climatic Change* 149 (3): 289–303.

Caniglia, Beth Schaefer, Robert J. Brulle, and Andrew Szasz. 2015. "Civil society, social movements, and climate change." *Climate change and society: Sociological perspectives* 1: 235–268.

Carley, Sanya, Lincoln L. Davies, David B. Spence, and Nikolaos Zirogiannis. 2018. "Empirical Evaluation of the Stringency and Design of Renewable Portfolio Standards." *Nature Energy* 3 (9): 754–63.

Caughey, Devin, and Christopher Warshaw. 2018. "Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 1936–2014." *American Political Science Review* 112 (2): 249–66.

Colgan, Jeff D., Jessica F. Green, and Thomas N. Hale. 2021. "Asset Revaluation and the Existential Politics of Climate Change." *International Organization* 75 (2): 586–610.

Cooper, Jasper, Sung Eun Kim, and Johannes Urpelainen. 2018. "The Broad Impact of a Narrow Conflict: How Natural Resource Windfalls Shape Policy and Politics." *The Journal of Politics* 80 (2): 630–46.

Cory, Jared, Michael Lerner, and Iain Osgood. 2021. "Supply Chain Linkages and the Extended Carbon Coalition." *American Journal of Political Science* 65 (1): 69–87.

Culpepper, Pepper D. 2015. "Structural Power and Political Science in the Post-Crisis Era." *Business and Politics* 17 (3): 391–409.

Dasgupta, Aditya. 2020. "Explaining Rural Conservatism: Technological and Political Change in the Great Plains." Working paper.

Davis, Charles. 2017. "Fracking and Environmental Protection: An Analysis of US State Policies." *The Extractive Industries and Society* 4 (1): 63–8.

Deutch, John. 2011. "The Good News about Gas: The Natural Gas Revolution and Its Consequences." *Foreign Affairs* 90:82–93.

DiSalvo, Richard, and Zhao Li. 2020. "Economic Geography and Special Interest Entrenchment: Evidence from the Fracking Boom and State Campaign Finance." Working paper.

Fedaseyeu, Viktar, Erik Gilje, and Philip E. Strahan. 2015. "Voter Preferences and Political Change: Evidence from Shale Booms (No. w21789)." National Bureau of Economic Research.

Fowler, Luke. 2016. "The States of Public Opinion on the Environment." *Environmental Politics* 25 (2): 315–37.

Fowler, Luke, and Chris Birdsall. 2021. "Does the Primacy System Work? State versus Federal Implementation of the Clean Water Act." *Publius: The Journal of Federalism* 51 (1): 131–60.

Fowler, Anthony, Haritz Garro, and Jörg L. Spenkuch. 2020. "Quid Pro Quo? Corporate Returns to Campaign Contributions." *The Journal of Politics* 82 (3): 844–58.

Gaventa, John. 1982. *Power and Powerlessness: Quiescence and Rebellion in an Appalachian Valley*. Urbana: University of Illinois Press.

Gilens, Martin, Shawn Patterson, and Pavielle Haines. 2021. "Campaign Finance Regulations and Public Policy." *American Political Science Review* 115: 1074–81.

Hacker, Jacob S., and Paul Pierson. 2002. "Business Power and Social Policy: Employers and the Formation of the American Welfare State." *Politics & Society* 30 (2): 277–325.

Hacker, Jacob S., and Paul Pierson. 2010. *Winner-Take-All Politics: How Washington Made the Rich Richer--And Turned its Back on the Middle Class*. New York: Simon and Schuster.

Hall, Richard L., and Frank W. Wayman. 1990. "Buying Time: Moneyed Interests and the Mobilization of Bias in Congressional Committees." *American Political Science Review* 84 (3): 797–820.

Hughes, Llewelyn, and Johannes Urpelainen. 2015. "Interests, Institutions, and Climate Policy: Explaining the Choice of Policy Instruments for the Energy Sector." *Environmental Science & Policy* 54: 52–63.

Johnson, Martin, Paul Brace, and Kevin Arceneaux. 2005. "Public Opinion and Dynamic Representation in the American States: The Case of Environmental Attitudes." *Social Science Quarterly* 86 (1): 87–108.

Konisky, David M., and Neal D. Woods. 2016. "Environmental Policy, Federalism, and the Obama Presidency." *Publius: The Journal of Federalism* 46 (3): 366–91.

Konisky, David M., and Neal D. Woods. 2018. "Environmental Policy." In *Politics in the American States*, eds. Virginia Gray, Russell L. Hanson and Thad Kousser, 451–77. Washington, DC: CQ Press.

Lindblom, Charles. 1977. *Politics and Markets: The World's Political-Economic Systems*. New York: Basic Books.

Mallinson, Daniel J. 2014. "Upstream Influence: The Positive Impact of PAC Contributions on Marcellus Shale Roll Call Votes in Pennsylvania." *Interest Groups & Advocacy* 3 (3), 293–314.

Mildenberger, Matto. 2020. *Carbon Captured: How Business and Labor Control Climate Politics*. Cambridge, MA: MIT Press.

Muñoz, John, Susan Olzak, and Sarah A. Soule. 2018. "Going Green: Environmental Protest, Policy, and CO2 Emissions in US States, 1990–2007." *Sociological Forum* 33 (2): 403–21.

Oreskes, Naomi, and Erik M. Conway. 2011. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues From Tobacco Smoke to Global Warming*. New York: Bloomsbury.

Parinandi, Srinivas C. 2020. "Policy Inventing and Borrowing among State Legislatures." *American Journal of Political Science* 64 (4), 852–68.

Potoski, Matthew. 2001. "Clean Air Federalism: Do States Race to the Bottom?" *Public Administration Review* 61 (3): 335–43.

Przeworski, Adam, and Michael Wallerstein. 1988. "Structural Dependence of the State on Capital." *American Political Science Review* 82 (1): 11–29.

Rabe, Barry G. 2018. *Can We Price Carbon?* Cambridge, MA: MIT Press.

Roser-Renouf, Connie, Neil Stenhouse, Justin Rolfe-Redding, Edward Maibach, and Anthony Leiserowitz. 2015. "Engaging Diverse Audiences with Climate Change: Message Strategies for Global Warming's Six Americas." In *The Routledge Handbook of Environment and Communication*, 388–406. London: Routledge.

Sances, Michael W., and Hye Young You. 2022. "Voters and Donors: The Unequal Political Consequences of Fracking." *The Journal of Politics* 84 (3). doi: 10.1086/717084.

Smith, Mark. 2000. *American Business and Political Power: Public Opinion, Elections, and Democracy*. Chicago, IL: University of Chicago Press.

Stokes, Leah Cardamore. 2020. *Short Circuiting Policy: Interest Groups and the Battle over Clean Energy and Climate Policy in the American States*. New York, NY: Oxford University Press.

Trachtman, Samuel. 2020. "What Drives Climate Policy Adoption in the US States?" *Energy Policy* 138: 111214.

Trachtman, Samuel. 2021. "State Policy and National Representation: Marijuana Politics in American Federalism." *Legislative Studies Quarterly* 0: 1–33.

Zacher, Sam. 2022, "Replication Data for: The U.S. Political Economy of Climate Change: Impacts of the "Fracking" Boom on State-Level Climate Policies." UNC Dataverse. V1. https://doi.org/10.15139/S3/UGCS7Y.

Author Biography. Sam Zacher is a PhD candidate at Yale University studying the politics of inequality, redistribution, and the climate crisis in contemporary America. His current research is aimed at understanding the nature of economic interests in the Democratic Party.

Cite this article: Zacher, Sam. 2023. The US Political Economy of Climate Change: Impacts of the "Fracking" Boom on State-Level Climate Policies. *State Politics & Policy Quarterly* **23** (2): 140–165, doi:10.1017/spq.2022.17

State Politics & Policy Quarterly (2023), **23**: 2, 166–186
doi:10.1017/spq.2023.1

ORIGINAL ARTICLE

Preaching to the Choir or Proselytizing to the Opposition: Examining the Use of Campaign Websites in State Legislative Elections

Joshua Meyer-Gutbrod

University of South Carolina, Department of Political Science, Columbia, SC 29208, USA
Email: meyerguj@mailbox.sc.edu

(Received 08 March 2022; revised 26 May 2022; accepted 22 July 2022)

Abstract

The Internet has spawned a renewed hope for facilitating increased access to candidate information for voters. However, the nationalization and polarization of constituents have left many candidates averse to the risks of personalized campaigns, especially in subnational elections. Under what conditions are state candidates willing to establish a personalized web presence as opposed to relying on partisanship? This study introduces a novel dataset of campaign website presence for the 2018 and 2020 state legislative elections. During this time, approximately one-third of state legislative candidates opted to forgo a personalized campaign website. District-level constituent ideology was significantly correlated with the website use, even when controlling for district education, income, age, and race, and the candidate's competitive position. District ideological homogeneity encouraged website use across both parties, while adversarial district ideology corresponded to low website use among Republicans. The results indicate that state legislative candidates, especially Republican candidates, are far more likely to preach to their partisan choir rather than incur the risks of proselytizing among their partisan opposition. The results reiterate the divergent responses of the political parties regarding partisan polarization and shed light on the impact of nationalization within state legislative campaigns.

Keywords: campaigns; political communication; ideology; parties and elections; mass media

Introduction

The Internet has altered the lives of individuals across the globe and political campaigns in the US have not been spared from this revolutionizing force. According to Pew, by 2008, 55% of the US adult population already turned to the Internet for political information (Smith 2009). A year later, Druckman, Kifer, and Parkin (2009, 343) declared campaign websites essential arguing that they "provide an unmediated, holistic, and representative portrait of campaigns". Modern campaign websites have expanded to increase donor access, facilitate interaction with candidates, and foster personalized constituent experiences (Bimber 2014; Gibson, Ward, and Lusoli 2002;

Smith and Duggan 2012). While this revolution spawned significant early research, recent efforts have shifted to concerns over the role of social media platforms, taking websites as perfunctory requirement for a campaign.

While the early literature on digital campaigning saw the use of websites as a forgone conclusion, nationalized, partisan polarization has thrown this assumption into question. The polarization of the national parties has resulted in more centralized control of party messaging alongside a nationalizing citizenry, and an increasing divide between opposing partisans (Abramowitz and Webster 2016; Cox and McCubbins 2005; Hopkins 2018). This increase in partisan polarization has left some candidates facing more ideologically homogenized constituencies. Candidates running in ideologically adversarial or moderate districts need to simultaneously appeal to both highly partisan primary constituents and moderate or adversarial general election constituencies. Under these polarized conditions, are candidates willing to employ substantial personalized campaign websites at the risk of alienating polarized voters, or are websites merely another tool to preach to the partisan choir in safe districts?

To better understand the interaction between polarization and state legislative campaign websites, this study introduces a novel data set that identifies the presence of campaign websites for all general election, and state legislative campaigns during the 2018 and 2020 campaigns. I argue that the growth of national partisan polarization has encouraged state legislative candidates to forgo websites in favor of the anonymity of partisan heuristics, especially in districts that leave them divided between polarized constituencies and general election appeals. Across the nation, approximately one-third of state legislative candidates do not establish a clearly identifiable website with significant variation across the candidates. To assess the drivers of individual variation, I model website presence against constituent demographics, resources, partisan contestation, and district-level citizen ideology. Competition and constituent preferences, rooted in age, income, race, and education, all contribute to higher likelihoods of website use. However, the more striking effect was the impact of constituent polarization on website presence. While candidates from both parties are more likely to employ websites in more homogenous districts, only Republicans are particularly averse to website use in ideologically opposing districts.

The effects present a pessimistic picture regarding the influence of polarization on campaign website use in the states. First, the low use of campaign websites provides a strong indication that candidates are willing to forgo individualized campaign content. In addition, the data reveals that candidates are averse to the risks associated with proselytizing within diverse districts and for Republicans in opposing districts. Candidates instead favor the safety of preaching to the partisan choir in safe districts. When candidates opt for anonymity or micro-targeting in situations of potential conflict within their districts, it reduces citizen exposure to potential policy distinctions between state-level candidates and the national political parties. Combined with the decline in local news coverage, this can present a serious challenge for local information to shift perceptions of nationalized partisan agendas, furthering citizen polarization (Moskowitz 2021). A careful analysis of these competing effects is necessary in order to reinvigorate the potential for digital campaign content to advance norms of democratic citizenship by diminishing polarization and nationalization of state political agendas.

Websites as reward and risk

The Internet revolution renewed hope in democratic norms by facilitating increased information and communication. In terms of engagement, the Internet has a mixed record for facilitating the mobilization of non-traditional groups or encouraging participation beyond donations (Bimber 1998; 2001; Bimber and Copeland 2013). However, scholars have viewed websites as a key tool for campaigns, fostering the growth of consultants dedicated to facilitating digital content (Benoit and Benoit 2005; Johnson 2002; McKelvey and Piebiak 2018). Candidates and parties establish clear web presences to communicate campaign information, target critical donor communities, and enhance participation for core constituent groups (Bimber and Davis 2003; Druckman *et al.* 2009; Druckman, Kifer, and Parkin 2010; 2018; Gaynor and Gimpel 2021; Gibson *et al.* 2002). Studies have shown that access to these online campaign portals corresponds to the increased likelihood of voting (Tolbert and McNeal 2003). As early as 2002, scholars highlighted how parties around the world had established a web presence and by 2008 scholars examining the US declared that "most if not all political campaigns will develop and maintain a presence on the Internet" (Gibson *et al.* 2002; Latimer 2009, 1036). Subsequent studies of congressional websites have found them to be consistent in spite of significant technological shifts (Druckman *et al.* 2018; Druckman, Kifer, and Parkin 2014). The growth of digital resources effectively facilitated the expansion of low-cost, substantial campaign content, with most scholars taking website use as a perfunctory component of a viable campaign.

However, little work has been done on evaluating the role of polarization and nationalization in shaping personalized campaign content through websites. Candidates, especially incumbents, have historically been averse to taking strong policy stances during campaigns that could prove unpopular in future elections (Arnold 1990; Druckman *et al.* 2009). While this aversion to concrete statements is a unique feature in individualized American politics, the centralization of party power under national polarization heightens this risk. The ideological polarization of America's national parties has corresponded to a more direct and centralized control of partisan agendas and messaging from congressional leadership (Cox and McCubbins 2005; Lee 2016; Sinclair 2011; 2014). This has resulted in increased ideological distance between the two major parties and increased internal homogeneity within each party (Hare and Poole 2014; Poole and Rosenthal 2011). The nationalization of partisan agendas has trickled down into constituencies across the nation, with an increased focus on national issues at the expense of local issues and a growing animosity for the opposition party (Abramowitz and Webster 2016; Hopkins 2018). The resulting heightened risk of significant opposition might discourage candidates from running individualized digital campaigns due to the risk of alienating such strongly opposed partisan camps.

This risk is especially pronounced for candidates in state legislative races due to the lack of local control of partisan agendas. The twenty-first century has seen an increase in the influence of national partisan trends in state elections, with presidential elections dominating state races (Rogers 2016). At the same time, a significant decline in local news and the effective growth of intermittent "flashlight" coverage for local politics has reduced the potential for accountability in state politics (Abernathy 2020; Conerly 2013; Graber 1989; Moskowitz 2021). The nationalization of media sources within American politics has shifted the concerns and opinions of citizens, resulting in an almost exclusive focus on national partisan agendas (Hopkins 2018). Modern voters reflect these shifts and are less interested in robust local campaign information but instead prefer some middle ground between concrete policy and simple partisan heuristics (Lipsitz *et al.* 2005). Nationalizing citizen preferences corresponding to national control of partisan agendas, combined with the lack of local news, should discourage website use within state elections. In addition, recent studies have found that state legislative candidates are more likely to engage in digital advertising on Facebook, allowing them to engage in more partisan and directly targeted advertising and avoid concrete policy stances (Fowler *et al.* 2021). The result is that state-level candidates should neglect personal campaign websites to avoid individual campaign content, relying instead on partisan cues or targeted advertising.

However, candidates may feel compelled to adopt websites in spite of this competing interest. Early website adoption by congressional candidates in the late 1990s was frequently correlated with incumbency status, resources, and the presence of competition (Klotz 1997; Ward and Gibson 2003). Further, the availability of this new resource created additional social pressure forcing candidates into the digital arena (Gibson *et al.* 2002). In a single study targeting state legislative campaigns, Herrnson, Stokes-Brown, and Hindman (2007) used survey evidence from 1759 campaigns in 1998 and 2000 to conclude that website adoption in the states was rooted in competition, professionalization, and district demographics including race, age, and education. Finally, similar factors have contributed to the adoption of Facebook as an alternative digital medium, with early adoption favoring democrats, and more affluent and educated districts (Williams and "Jeff" Gulati 2013).

These studies have provided a suite of potential explanations for the expansion of website use, but most focus on the role of early adoption and none engage with the growth of polarization and nationalization in the modern political environment. While website use has been accepted as a forgone conclusion for successful campaigns, the nationalization and centralization of partisan agendas creates a credible incentive for candidates to avoid individualization through websites. This is especially problematic in low salience elections, like state legislative races. This can undermine accountability and enhance citizen reliance on partisan cues as opposed to substantial policy claims, contributing to nationalization and polarization. Fully understanding the competing influences of partisan polarization and the expectations of digital campaigning at the state level requires a thorough assessment of these competing factors.

Constituent pressure, competition, and polarization

The polarization and nationalization of voters may leave state legislative candidates with a strong incentive to avoid individualization through campaign websites. This incentive against campaigning may not be felt equally across the states or districts. Significant variation in levels of partisan polarization within state legislatures provides an indicator that the nationalization of agendas, while common, is not ubiquitous (Shor and McCarty 2011). Further, using survey data and post-stratification techniques, scholars have identified significant variation in constituent ideology both across states and across state legislative districts (Tausanovitch and Warshaw 2013;

https://doi.org/10.1017/spq.2023.1 Published online by Cambridge University Press

Warshaw and Rodden 2012). Candidates should react differently to districts that share their ideology as opposed to districts in which they are running opposed to the dominant ideology. However, understanding the impact of district ideology on websites requires separating two distinct components of polarization: the cohesion within each district and the district's level of ideological extremism.

From a national standpoint, higher partisan cohesion tends to indicate more centralized control of partisan priorities and clearer partisan agendas. This is no different at the state level, where centralized cohesive parties can discourage candidate deviation. However, candidates that confront highly cohesive constituents run few risks in taking strong stances. With little variation, campaign agendas should be clear and candidates should feel free to preach to the proverbial choir, leading to the *Ideological Homogeneity Hypothesis*:

> *Ideological Homogeneity Hypothesis*: Higher levels of homogeneity in constituent ideology within a district will correspond to higher levels of website use among candidates.

Alternatively, more extreme partisan ideologies can produce competing effects. Generally, higher levels of ideological extremism indicate significant buy-in into nationalized partisan polarization. Under this scenario, higher levels of ideological extremism within a district could encourage candidates to conceal themselves in the anonymity of a partisan label. However, this effect will likely vary. A district with significant ideological support for a candidate would likely encourage a more active position taking due to the low level of risk associated with the constituent opposition. Alternatively, a candidate confronting an oppositional ideological district is forced to balance the competing demands of partisan and general election constituencies. Risk aversion, in this situation, would lead to the avoidance of strong policy stances on either side, resulting in a decline in website use, leading to the *Ideological Extremism Hypothesis*:

> *Ideological Extremism Hypothesis*: A greater distance between the district's and candidate's ideology, will correspond to lower levels of website use among candidates.

Measuring website use

Identifying the influence of constituent demands, competition, and polarization on website use in state legislative elections requires a systematic assessment of the presence of campaign websites across state legislative districts. Websites were identified for all major party candidates for upper and lower state legislative races during the 2018 and 2020 general elections. Websites were identified using the Google search engine. Search terms included the candidates' full name, chamber, state, and campaign year. All searches were conducted between six and two weeks prior to the general election date. The first twenty Google results were included in each search and all searches were conducted with history-driven search improvements disabled. In addition, official Facebook pages returned in the Google search were subsequently searched for links to external campaign websites. This search method focused on websites most likely to be identified by constituents, assuming that a well-concealed website is functionally no website.

A number of specific types of websites were excluded from the results. First, official legislative office websites, denoted by a .gov address or with a consistent attribution statement to the legislature, were not included as personalized campaign content. Second, official websites produced explicitly by the state party, identified by a common attribution statement and consistent design, were excluded. Third, partisan donation-specific websites, including ActBlue and WinRed, that include donor portals for multiple candidates were excluded. In all of these instances, candidates lacked exclusive control over website content and are seen as relying on partisan signals as opposed to personalized agendas. Finally, Facebook, Twitter, and other social media sites were excluded. While the current trend in campaigns has been to increase the use of social media to both advertise and engage with constituents, the types of engagement are distinct from a major website. While both accomplish the goal of outreach, including announcing campaign calendars, and fundraising, substantial policy engagement requires the unlimited space associated with a campaign website. The fluidity and space limitations of the Facebook and Twitter platforms lend themselves more to platitudes than policy proposals, though they remain a potentially fruitful avenue of future research (Fowler *et al.* 2021).

In total, 7,074 candidate websites were identified amongst the 10,483 candidates in 2018 and 6,721 websites were identified amongst the 9,874 candidates in 2020. Across both years, approximately one-third of state legislative candidates opted not to set up an accessible campaign website. These low numbers provide significant evidence on the desire of state legislative candidates to avoid engaging in personalized campaigning that could put them at risk with polarized constituencies. The following analysis will restrict the sample to single-member legislative districts to avoid more complicated interactions with partisanship, web presence, and even shared web space between multiple candidates.[1] Among single-member districts, websites were not evenly distributed across the nation. Tables 1 and 2 provide the percentage of candidates with websites for 2018 and 2020, respectively, for single-member districts by the state for uncontested incumbents, contested incumbents, uncontested challengers, and contested challengers, and totals. As the table indicates, as competition levels increase both in terms of contested elections and for challengers facing incumbent opposition, the likelihood of website use increases. However, this increase is not consistent across all states, with some states exhibiting significantly higher rates of website use than others.

Figures 1 and 2 map the presence of websites across lower and upper chamber state legislative districts by the party for the 2018 and 2020 campaign years. You can find higher-resolution regional breakdowns of these figures in the Supplementary Material. The maps illustrate a few potential trends. First and foremost, more rural populations, especially in the South and Midwest, appear less likely to have a strong website presence from either party, as indicated by the blocks of gray. This could be a combination of uncontested elections, a decline in constituent expectations regarding web presence, or simply reduced availability of Internet access among constituents. Secondly, there is a complex interaction between competition and professionalization, with states like Florida, California, Wisconsin, Minnesota, and even Colorado

[1]A number of states were not included in the analysis because of either lack of elections or lack of clear district dynamics, as is the case with multi-member districts. NH and VT were excluded for district dynamics in both election years. NJ, MS, and VA did not have elections in 2018 and NJ, NE, AL, LA, MD, MS, VA, and WA did not have elections in 2020

https://doi.org/10.1017/spq.2023.1 Published online by Cambridge University Press

Table 1. Percentage of candidates with websites by incumbency and competition, 2018 election

State	Total	Incumbent-uncontested	Incumbent-contested	Challenger-uncontested	Challenger-contested
Alabama	55% (205)	24% (59)	48% (42)	56% (16)	80% (88)
Alaska	80% (80)	55% (11)	77% (22)	86% (7)	88% (40)
Arizona	94% (52)	67% (3)	94% (16)	100% (1)	97% (32)
Arkansas	51% (166)	32% (41)	51% (51)	33% (9)	66% (64)
California	94% (186)	100% (13)	92% (75)	100% (13)	94% (85)
Colorado	90% (155)	100% (3)	96% (50)	(0)	87% (102)
Connecticut	48% (350)	15% (13)	29% (147)	0% (1)	66% (189)
Delaware	80% (82)	80% (10)	68% (28)	100% (4)	88% (40)
Florida	86% (231)	41% (22)	86% (77)	83% (6)	94% (126)
Georgia	75% (326)	60% (87)	72% (120)	100% (8)	89% (111)
Hawaii	57% (86)	45% (22)	59% (29)	100% (5)	57% (30)
Idaho	71% (55)	64% (11)	84% (19)	0% (1)	67% (24)
Illinois	85% (239)	81% (54)	88% (74)	100% (7)	83% (102)
Indiana	54% (212)	29% (21)	40% (87)	50% (2)	70% (101)
Iowa	64% (214)	30% (20)	46% (80)	75% (4)	82% (110)
Kansas	69% (185)	58% (31)	73% (75)	50% (4)	71% (75)
Kentucky	63% (221)	33% (12)	55% (82)	0% (1)	71% (126)
Maine	29% (345)	29% (7)	28% (115)	50% (6)	30% (212)
Maryland	82% (124)	73% (11)	84% (45)	89% (9)	81% (59)
Massachusetts	79% (259)	69% (109)	79% (61)	88% (16)	92% (73)
Michigan	76% (297)	(0)	84% (75)	(0)	74% (222)
Minnesota	86% (264)	0% (1)	86% (109)	100% (1)	87% (153)
Missouri	61% (311)	55% (22)	61% (92)	71% (14)	61% (183)
Montana	39% (209)	21% (14)	34% (73)	40% (10)	45% (109)
Nebraska	85% (27)	85% (13)	100% (2)	75% (8)	100% (4)
Nevada	92% (88)	100% (8)	100% (28)	80% (5)	87% (47)
New Mexico	64% (102)	28% (18)	70% (40)	50% (2)	74% (42)
New York	54% (353)	44% (50)	44% (142)	83% (6)	65% (155)
North Carolina	76% (336)	(0)	74% (146)	(0)	77% (190)
North Dakota	20% (44)	0% (3)	12% (16)	0% (1)	29% (24)
Ohio	75% (217)	80% (5)	74% (69)	50% (2)	76% (141)
Oklahoma	72% (220)	42% (19)	62% (53)	60% (5)	80% (143)
Oregon	91% (129)	100% (8)	95% (59)	100% (1)	87% (61)
Pennsylvania	72% (373)	51% (47)	65% (142)	67% (6)	83% (178)
Rhode Island	44% (154)	31% (51)	48% (44)	80% (10)	45% (49)
South Carolina	52% (169)	37% (49)	53% (62)	83% (6)	63% (51)
South Dakota	34% (71)	0% (4)	41% (27)	0% (1)	33% (39)
Tennessee	67% (200)	50% (22)	59% (64)	100% (6)	73% (108)
Texas	89% (271)	84% (56)	91% (87)	100% (3)	90% (125)
Utah	75% (164)	57% (7)	74% (61)	100% (4)	76% (92)
Washington	91% (45)	71% (7)	93% (15)	100% (4)	95% (19)
West Virginia	44% (117)	0% (2)	33% (48)	100% (1)	53% (66)
Wisconsin	80% (187)	73% (11)	72% (86)	100% (4)	88% (86)
Wyoming	40% (101)	30% (40)	27% (22)	40% (5)	59% (34)

Note. Percentages are calculated based on all single-member districts across both chambers in the state. Category totals are shown in parenthesis. Categories without candidates are denoted by a (0).

showing a strong web presence over the last few years. While the maps provide an interesting cursory glance, they underscore the complexity of determining which factors motivate campaign websites use.

To fully assess competing explanations, I model website presence as a logistic regression where the dependent variable is coded as 1 for the use of a website and 0 for not. Independent variables are designed to test whether polarization has an impact on

Table 2. Percentage of candidates with websites by incumbency and competition, 2020 election

State	Total	Incumbent-uncontested	Incumbent-contested	Challenger-uncontested	Challenger-contested
Alaska	73% (71)	38% (13)	75% (24)	100% (5)	83% (29)
Arizona	75% (55)	50% (4)	84% (19)	(0)	72% (32)
Arkansas	46% (165)	20% (45)	43% (60)	50% (2)	69% (58)
California	90% (189)	93% (14)	88% (69)	100% (10)	90% (96)
Colorado	89% (157)	100% (3)	95% (60)	100% (1)	84% (93)
Connecticut	54% (347)	21% (14)	38% (156)	(0)	70% (177)
Delaware	74% (78)	75% (16)	67% (30)	(0)	81% (32)
Florida	86% (242)	14% (7)	80% (83)	60% (5)	94% (147)
Georgia	74% (352)	56% (86)	75% (120)	100% (11)	84% (135)
Hawaii	64% (74)	40% (5)	62% (26)	100% (2)	66% (41)
Idaho	70% (50)	64% (11)	83% (18)	100% (1)	60% (20)
Illinois	73% (204)	62% (47)	67% (75)	100% (4)	83% (78)
Indiana	56% (209)	17% (23)	42% (93)	67% (3)	79% (90)
Iowa	56% (216)	58% (19)	38% (88)	67% (3)	72% (106)
Kansas	71% (275)	47% (36)	73% (94)	78% (9)	74% (136)
Kentucky	66% (187)	56% (16)	60% (83)	100% (4)	71% (84)
Maine	45% (359)	67% (6)	37% (136)	43% (7)	50% (210)
Massachusetts	83% (245)	78% (113)	88% (68)	100% (11)	83% (53)
Michigan	66% (220)	(0)	70% (84)	(0)	63% (136)
Minnesota	82% (395)	50% (4)	79% (173)	(0)	84% (218)
Missouri	65% (271)	48% (25)	54% (97)	88% (17)	73% (132)
Montana	46% (206)	14% (14)	31% (64)	62% (13)	57% (115)
Nevada	86% (86)	100% (11)	96% (27)	100% (2)	76% (46)
New Mexico	74% (194)	26% (27)	78% (69)	67% (3)	84% (95)
New York	60% (353)	46% (46)	46% (127)	100% (8)	72% (172)
North Carolina	77% (333)	(0)	78% (149)	(0)	77% (184)
North Dakota	33% (39)	0% (7)	36% (14)	(0)	44% (18)
Ohio	65% (216)	50% (4)	60% (88)	(0)	69% (124)
Oklahoma	75% (174)	46% (26)	74% (81)	100% (4)	86% (63)
Oregon	82% (141)	86% (7)	77% (53)	100% (1)	84% (80)
Pennsylvania	68% (371)	40% (47)	60% (156)	100% (7)	84% (161)
Rhode Island	46% (149)	42% (48)	47% (45)	67% (12)	43% (44)
South Carolina	64% (250)	52% (67)	64% (88)	100% (4)	70% (91)
South Dakota	46% (61)	0% (5)	39% (23)	100% (1)	56% (32)
Tennessee	68% (162)	59% (27)	60% (77)	100% (3)	82% (55)
Texas	93% (29)	100% (3)	92% (12)	(0)	93% (14)
Utah	80% (149)	54% (13)	80% (61)	67% (3)	85% (72)
West Virginia	48% (112)	33% (3)	41% (44)	(0)	54% (65)
Wisconsin	79% (178)	50% (10)	71% (75)	(0)	89% (93)
Wyoming	48% (96)	30% (33)	52% (21)	46% (13)	66% (29)

Note. Percentages are calculated based on all single-member districts across both chambers in the state. Category totals are shown in parenthesis. Categories without candidates are denoted by a (0).

website use, while controlling for constituent demographics and competitive position. Polarization within the district is measured using Tausanovitch and Warshaw's 2013 state legislative district ideology estimates, as the most recent estimates using the 2018 and 2020 districts (Tausanovitch and Warshaw 2013). District extremism is measured using their MRP mean estimated ideology within the district. However, to account for the ideological position of candidates, it is calibrated against the candidate's party. The new measure positions Democratic candidates at the most liberal point and Republican candidates at the most conservative point and then measures the district's ideological distance from that point. Therefore, a Democratic candidate

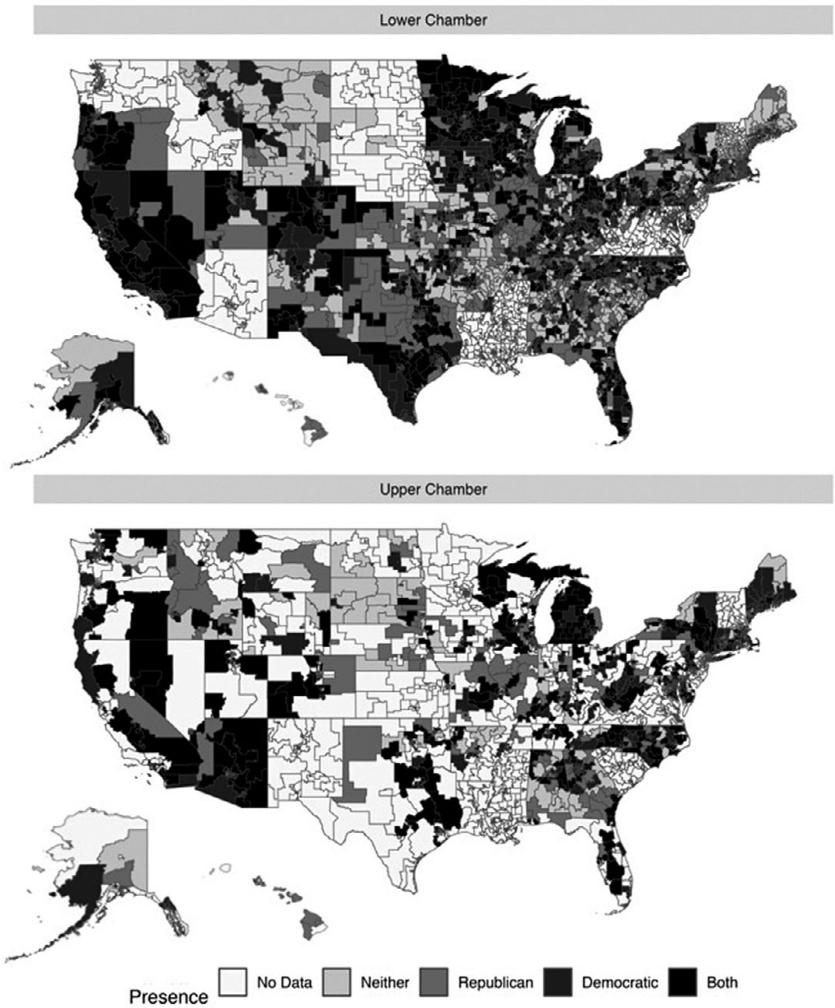


Figure 1. Distribution of Website Use by Party for 2018.
Figure 1 plots the map of state legislative, single-member districts, for 2018 races. Shading indicate the presence of websites by, partisanship. Missing data, including districts with no race that year are shown in white. Alaska and Hawaii are shown not to scale to visualize the districts. Regional maps with clearer district lines can be found in the Supplementary Material.

facing an extremely liberal district will have a score of 0 and a Republican facing an extremely conservative district will have that same score. As the ideology score increases, the *Ideological Extremism Hypothesis* would dictate that we would be less likely to see a website due to the risks of alienating partisan constituents. District homogeneity in ideology is measured using the standard deviation for their MRP ideology estimates. According to the *Ideological Homogeneity Hypothesis*, increases in the standard deviation of the district ideology would result in a lower likelihood to have a website due to the risks associated with alienating more diverse constituents.

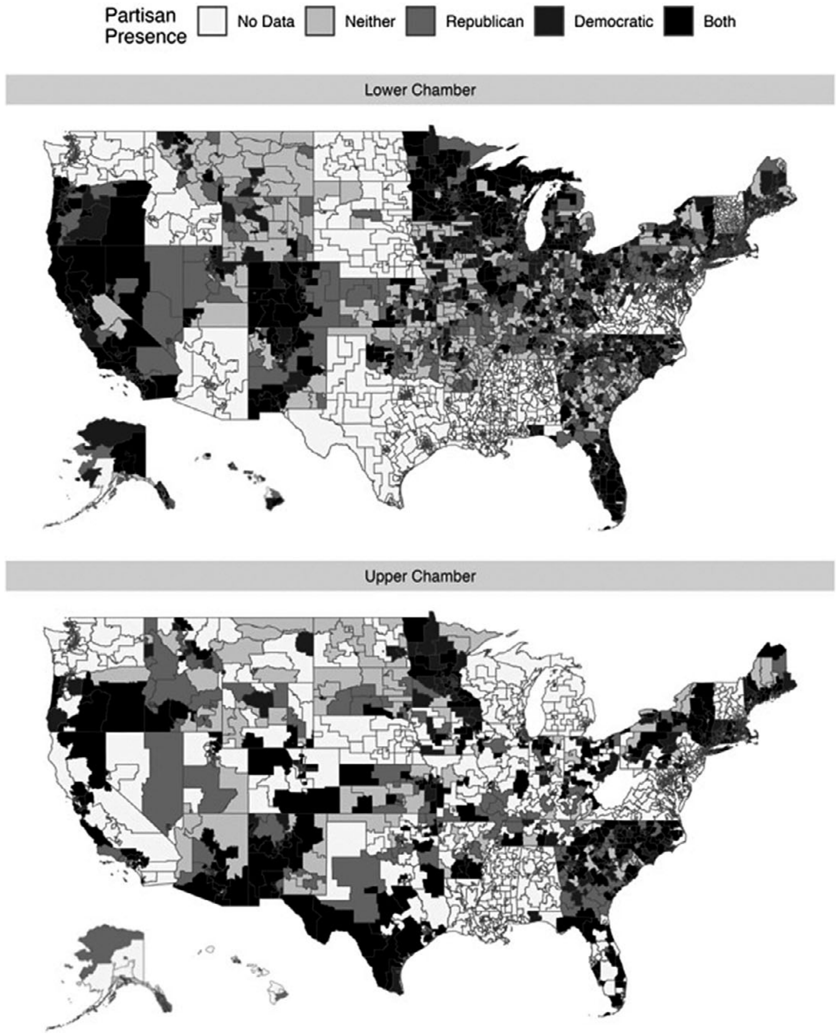


Figure 2. Distribution of Website Use across Party for 2020.
Figure 2 plots the map of state legislative, single-member districts, for 2020 races. Shading indicates the presence of websites by, partisanship. Missing data, including districts with no race that year are shown in white. Alaska and Hawaii are shown not to scale to visualize the districts. Regional maps with clearer district lines can be found in the Supplementary Material.

In addition to district ideology, two binary variables are included to assess the impact of the competitive environment on a candidate's decision, including an indicator if the candidate is a challenger and an indicator if the candidate faced a contested general election through major party opposition.[2] In addition, an

[2]Contestation in general elections is used as a proxy for competition both due to low levels of contestation in state legislative elections and the general difficulty of accurately measuring actual competitive elections.

https://doi.org/10.1017/spq.2023.1 Published online by Cambridge University Press

interaction effect between these two variables is included under the assumption that challengers might approach competitive and non-competitive elections differently.

Finally, controls for district demographic composition include district education level, median income level, median age, and percentage of the population who identify as "white" are all calculated using the Census Bureau's American Community Survey 2016 five-year estimates (United States Census Bureau 2016). In addition, controls are included for candidate party and small district size, measured as districts with a population smaller than 33,300. While district population has been shown to impact early website use (Herrnson *et al.* 2007), it actually holds a non-linear relationship with website use in 2018 and 2020. Instead, the data reveals a stepwise shift at around the first quartile of population sizes, capping at 33,300. This is unsurprising given that the marginal impact of websites given larger populations significantly diminishes at a certain point. The model controls for professionalization, operationalized as the state's most recent Squire Professionalization Index, a measure of the resources available to state legislators relative to the resources available to Congress, and a binary variable for upper chamber races (Squire 2017). Missing from these controls is a measure of Internet availability among constituents. The US Census Bureau's American Community Survey measures Internet access by the household at the county level but does not provide a similar measure at the state legislative district level. Approximately 80% of households nationally have some form of within-home Internet (United States Census Bureau 2016). However, the disparity in mapping between county lines and state legislative districts across states renders inferences from this measure problematic. However, there is a positive and significant correlation between county population and county Internet access rates ($r[3218] = 0.43$, $p < 0.01$) rendering population size a significant proxy for Internet access.

Models are run including and excluding relevant interaction terms and include robust standard errors clustered at the state level and separate models are run for the 2018 and 2020 campaigns to allow for variation between presidential and midterm elections across all variables. The model formulation is robust to alternative controls for the effect of states. Alternative models, including a model using state fixed effects and state random effects, are included in the Supplementary Material.

The competing effects of polarization

Results for the 2018 and 2020 models are presented in Table 3. The first column for each model shows the coefficient and robust, clustered standard error and the second column lists the odds ratio for each coefficient. Model years are separated and separate models are run within each year, excluding and including key interaction terms.

The non-interaction baseline models support many of the existing theories of website use with most of the control variables being statistically significant and substantial. In addition, Democratic candidates, challengers, and candidates with major party competition are all more likely to use websites. Finally, the statistically significant coefficients for the measures of district heterogeneity and candidate relative distance lend support for both the *Ideological Homogeneity Hypothesis* and *Ideological Extremism Hypothesis*. More ideologically diverse districts and candidates confronting more adversarial districts both corresponded to a reduced likelihood of website use by candidates in those races. However, the statistically significant

Table 3. Logistic regression results for website use

	Model 1 – non-interaction		Model 2 – interactions	
2018 Model	Coefficient (SE)	Odds ratio	Coefficient (SE)	Odds ratio
(Intercept)	0.47 (0.62)	1.592	0.8 (0.77)	2.229
Citizen ideological heterogeneity	−6.47 (1.78)*	0.002*	−5.95 (2.26)*	0.003*
Citizen ideological relative distance	−0.42 (0.13)*	0.659*	−0.94 (0.27)*	0.392*
Democrat (binary)	0.6 (0.13)*	1.83*	−0.43 (0.52)	0.651
Democrat × ideological hetero.	—	—	−0.88 (1.61)	0.416
Democrat × ideological distance	—	—	1.15 (0.48)*	3.155*
Challenger (binary)	0.52 (0.11)*	1.685*	1.27 (0.15)*	3.545*
Competitive (binary)	0.27 (0.12)*	1.304*	0.4 (0.13)*	1.499*
Professionalization (squire)	0 (0.01)	1.003	0.01 (0.01)	1.009
Upper chamber race (binary)	0.05 (0.1)	1.055	0.06 (0.1)	1.058
Median age (district)	−0.03 (0.01)*	0.971*	−0.03 (0.01)*	0.971*
% Bachelors or more (district)	0.014 (0.006)*	1.014*	0.021 (0.005)*	1.021*
Median income (district)	0.017 (0.005)*	1.017*	0.014 (0.005)*	1.015*
Percentage white (district)	0.011 (0.003)*	1.011*	0.007 (0.003)*	1.007*
Small district (binary)	−0.91 (0.21)*	0.404*	−0.82 (0.22)*	0.439*
Competitive × challenger	—	—	−0.84 (0.16)*	0.431*
N	8,883		8,883	
AIC	9,903.4		9,852.9	
	Model 1 – non-interaction		**Model 2 – interactions**	
2020 Model	Coefficient (SE)	Odds ratio	Coefficient (SE)	Odds ratio
(Intercept)	0.34 (0.43)	1.408	0.84 (0.52)	2.307
Citizen ideological heterogeneity	−4.58 (1.47)*	0.01*	−5.33 (1.96)*	0.005*
Citizen ideological relative distance	−0.74 (0.17)*	0.479*	−1.17 (0.26)*	0.31*
Democrat (binary)	0.81 (0.12)*	2.24*	−0.57 (0.53)	0.568
Democrat × ideological hetero.	—	—	1.91 (1.89)	6.784
Democrat × ideological distance	—	—	1.03 (0.38)*	2.806*
Challenger (binary)	0.75 (0.09)*	2.114*	1.68 (0.24)*	5.369*
Competitive (binary)	0.39 (0.13)*	1.473*	0.5 (0.14)*	1.649*
Professionalization (squire)	−0.01 (0.01)	0.992	0 (0.01)	0.996
Upper chamber race (binary)	0.01 (0.08)	1.013	0.02 (0.09)	1.017
Median age (district)	−0.02 (0.01)*	0.979*	−0.02 (0.01)*	0.979*
% Bachelors or more (district)	0.02 (0.005)*	1.02*	0.027 (0.006)*	1.027*
Median income (district)	0.012 (0.004)*	1.013*	0.01 (0.004)*	1.01*
Percentage white (district)	0.01 (0.004)*	1.01*	0.007 (0.003)*	1.007*
Small district (binary)	−0.88 (0.18)*	0.414*	−0.83 (0.18)*	0.436*
Competitive × challenger	—	—	−1.02 (0.23)*	0.362*
N	8,318		8,318	
AIC	9,323.7		9,277.5	

Note. The first set of models for each year excludes interaction terms, while the second column incorporates a series of theoretically important interaction terms. The first column for each model lists logistic regression coefficients with robust standard errors clustered by state in parenthesis. The second column lists converted odds-ratios.
*$p < 0.05$ two tailed.

interaction effects reveal a more complex story across parties and across incumbent-challenger dynamics.

Both interaction models show support for the argument that constituent demographics can have a substantial impact on the likelihood of website use. District median age, education level, median income, and racial composition are significant for both models. A single standard-deviation increase in the percentage of the district population with a bachelor's degree (14% increase) corresponds to an increase of approximately 6% in the likelihood of having a campaign website in 2018 and 2020, controlling for other variables. With regard to income, a one standard-deviation shift or an increase of $20,000 in median income also corresponded to an increase in the likelihood by approximately 6% in 2018 and 3% in 2020. A one standard-deviation increase in the percentage of the population who identify as "white" (a 20% increase) corresponds to a 3% increase in the likelihood of website use for both models, holding other variables constant. Finally, a 10-year increase in median age corresponded to a decline in the likelihood of having a website by 6% and 5% in 2018 and 2020, respectively.

Lurking behind all of these shifts is the important role of small population sizes as a control variable. Districts under 30,000 citizens are significantly less likely to have campaign websites, with a drop-in the likelihood of having a website by 16% in the midterm election and 14% in presidential election years, even controlling for demographic variables. This is unsurprising as smaller districts may facilitate more personal campaign styles while larger sizes do not provide a significant marginal advantage and are correlated with other more predictive variables. However, there is no strong evidence that significant expectations are produced by the professionalization of the state legislature or the office. Both upper chamber races and measures of state legislative professionalization failed to achieve statistical significance. These results indicate that citizen expectations are a localized phenomenon, rooted in district demographics, and not broader professional standards across the state.

While constituent demographics clearly have a significant impact on the presence or absence of websites, the competitive environment in which a candidate is running has an equally robust impact. The coefficients for competition and incumbency status are both statistically significant and in the expected direction. Figure 3 plots the predicted probabilities of having a website based on both models for different levels of competition, incumbency, and party.

The likelihood that an incumbent facing a contested election employs a campaign website is 8% higher in 2018 and 11% higher in 2020, relative to unchallenged incumbents. Further, within contested elections, challengers see an 8% increase in the likelihood of website use in 2018 and a 10% increase in 2020 compared to incumbents. This results in an increase in the likelihood of having a website by 16% in 2018 and 21% in 2020 for challengers in contested elections compared to incumbents in non-contested elections. The statistically significant interaction term between contestation and incumbency serves as an indicator that the bulk of this shift occurs within the realm of incumbency. The difference between competitive and non-competitive challengers is much smaller compared to these other shifts. Finally, as Figure 3 illustrates, in both 2018 and 2020 Democrats were more likely to employ websites compared to Republicans, with an 11% increase in the likelihood of having a website in 2018 and a 12% increase in 2020.

While constituent demographics and competitive environment both play a strong role in forcing candidates to take visible campaign positions through a personalized campaign website, the impact of polarization is more nuanced and dynamic. As Figure 4 illustrates, there is a strong correlation between relative ideological opposition within the district and the likelihood of website use. However, this distinction is asymmetric, in a manner typical of conversations concerning polarization. For safe districts, with relative distance scores less than 0.6, both parties seem more than willing to engage in digital campaigning. These candidates confront friendly districts and are comfortable preaching to the partisan choir. However, in adversarial districts

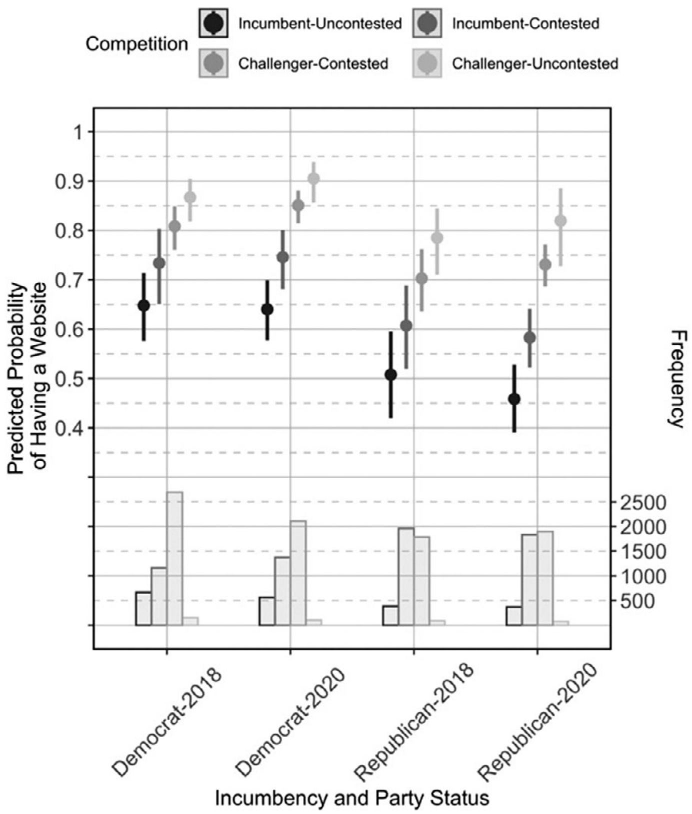


Figure 3. Predicted Percentage Chance of Website Use by Party, Incumbency, and Competition. Figure 3 plots the predicted percentage chance of having a website against the candidate's partisanship, incumbency status, and presence of major party competition. All other variables are held at their mean or mode. The left axis corresponds to the predicted percent probability dot plot. The right axis corresponds to the histogram. Predictions are obtained using Model 2-Interactions in Table 3 and state-clustered standard errors are shown using bars around the dot plot.

where the relative ideological position of the district moves further from the candidate's party, there is a divergent response between the two parties. Democrats are just as likely, if not more likely to employ a campaign website in these adversarial districts. This is an indication that Democrats are willing to reach out and proselytize amongst this strong opposition. On the contrary, among Republicans, the likelihood of a campaign website drops precipitously as the ideological composition of the district moves from safe to adversarial. A 1 standard-deviation increase in adversarial ideology (+0.36 shift) corresponds to a 9% drop in website use in 2018 and 2020. Republican candidates confront the risks of engagement with oppositional districts differently and appear unwilling to engage in personalized digital campaigning in this environment (Figure 4).

In addition to the relative ideological separation between a candidate and their district, the model also illustrates the importance of district homogeneity for both parties. Figure 5 plots the model-predicted probability of having a website against the

https://doi.org/10.1017/spq.2023.1 Published online by Cambridge University Press

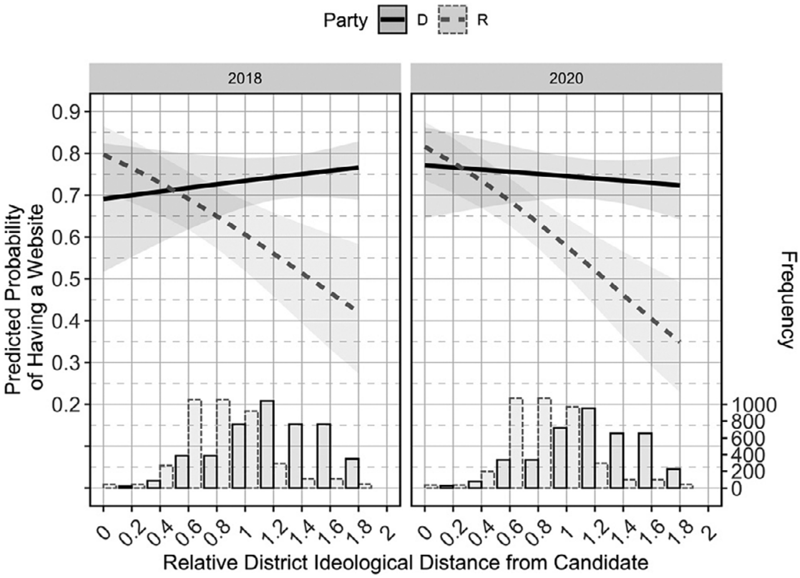


Figure 4. Predicted Percentage Chance of Website Use by District Ideological Adversity.
Figure 4 plots the predicted probability of having a website against relative district ideological distance from a candidate, by party. Ideological distance scores of 0 indicate a strongly supportive district while higher scores indicate an ideologically opposed district. Candidate variables are set as an incumbent in the chamber majority party in a competitive election. All other variables are held at their mean or mode. The left axis corresponds to the predicted percent probability line plot. The right axis corresponds to the histogram. Predictions are obtained using Table 3 Model 2-Interactions for 2018 (Panel 1) and 2020 (Panel 2). State-clustered standard errors are shown using shading around the linear plot.

range of district ideological heterogeneity measured as the standard deviation for the MRP estimated district ideology. Again, the bar plot along the bottom shows a histogram with the frequency of ideology scores.

Unlike the effects of ideological distance, the effects of ideological heterogeneity are similar across both parties. A 1 standard-deviation increase in heterogeneity (a 0.05 shift) corresponds to approximately a 7% decline in the probability of having a website for both parties and in both years.

The results from both Models 1 and 2 and across both years show that a number of demographic variables and increased competition can increase the likelihood of campaign website use in these subnational elections. However, even controlling for district demographics and competition, district ideological polarization has a strong potential effect on website use. When considering both district extremism and district ideological homogeneity, the safer a candidate feels in a district, the more likely they are to take up a personalized campaign web presence. More homogenous districts and districts with the mean ideology that is closer to the candidate's party both correspond to an increased likelihood of website use. Candidates in these districts feel safe preaching to the choir, with little risk of angering large groups of constituents by playing to a clear party line. However, in more heterogenous districts, candidates confront increased risks due to the diversity of opinion. This is especially pronounced for Republican candidates in

https://doi.org/10.1017/spq.2023.1 Published online by Cambridge University Press

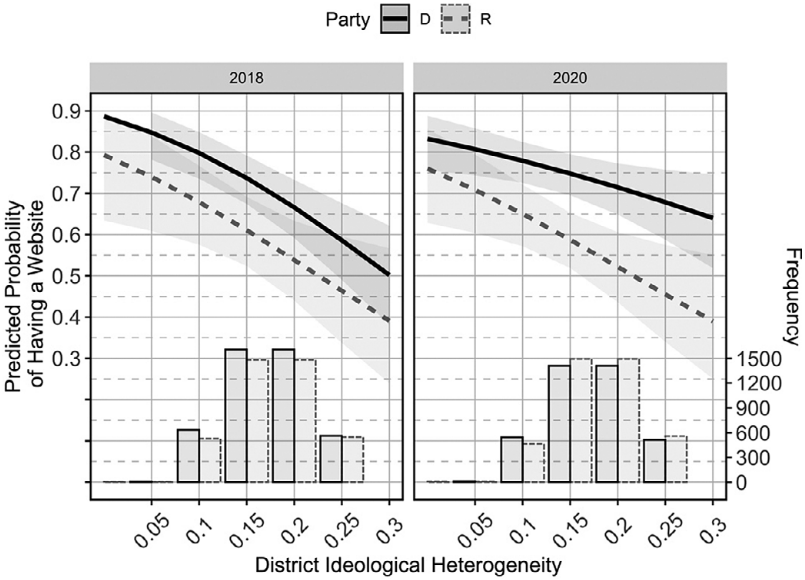


Figure 5. Predicted Percentage Chance of Website Use by District Ideological Heterogeneity. Figure 5 plots the predicted probability of having a website against district ideological heterogeneity by party. Ideological distance scores of 0 indicate a homogenous district while higher scores indicate an ideologically diverse district. Candidate variables are set as an incumbent in the chamber majority party in a competitive election. All other variables are held at their mean or mode. The left axis corresponds to the predicted percent probability line plot. The right axis corresponds to the histogram. Predictions are obtained using Table 3 Model 2-Interactions for 2018 (Panel 1) and 2020 (Panel 2). State-clustered standard errors are shown using shading around the linear plot.

adversarial districts, where proselytizing among the competition risks angering both the primary and general election constituencies. The same effect does not apply to their Democratic counterparts who seem more willing to engage in this activity. This one-sided aversion to engagement undermines critical opportunities for state-level policy development and understanding among citizens and instead implicitly reinforces national partisan platforms.

Conclusion

The expansion of the Internet at the turn of the twentieth century provided renewed hope among scholars and journalists for a more active and engaged citizenry with abundant access to information. Scholars have continued to examine the role of the Internet in shaping political activity and engagement, and have found that the promise of the Internet largely fell short (Bimber and Copeland 2013). Even in terms of information flows, recent scholarship has focused on the ability of the Internet to foster the propagation of misinformation (Anderson and Rainie 2017; Mitchell *et al.* 2021). However, the general consensus at the turn of the twentieth century was that this new medium would, at a minimum, provided campaigns with an easy, cost-effective mechanism to convey information to citizens. As a result, websites should

https://doi.org/10.1017/spq.2023.1 Published online by Cambridge University Press

have become the new standard for campaigning in a digital world, and for national campaigns, this has largely been true (Druckman *et al.* 2018).

However, far from becoming a standard of professionalization, state legislative candidates frequently forgo campaign websites entirely with one-third of state legislative campaigns in 2018 and 2020 failing to establish an easily identifiable website. Within low-information settings, like state legislative campaigns, candidates can avoid websites to enjoy some degree of anonymity, relying instead on either partisan heuristics or microtargeting messages at particular communities. This strategy has been fostered by the nationalization and polarization of the electorate and the low rates of competition in state legislative campaigns, relative to their national counterparts (Abramowitz and Webster 2016; Hopkins 2018). By avoiding public, personalized, campaign websites, candidates can avoid concrete policy statements, which may alienate more diverse electorates.

With only two-thirds of state legislative campaigns opting for personal campaign websites, there is strong evidence that the trend toward anonymity in down-ballot elections is strongly at work. However, there is substantial evidence for local factors impacting the use of campaign websites, most notably, constituent expectations rooted in demographic variation and the competitive environment a candidate faces. The models show strong evidence that higher education, younger, wealthier, and more Caucasian districts were more likely to encourage campaigns to use websites. This is largely driven by the expectations of citizens, with white, young, affluent, and educated populations being more likely to obtain news and campaign information on the Internet (Smith 2009). However, the impact of more professionalized state legislatures is not statistically significant, indicating that norms of professionalization largely vary across district, and not necessarily state.

Finally, two additional variables loomed large in understanding district-level variation in website use. First, district population sizes within the first quartile of the population distribution (less than 33,000 residents) are less likely to correspond to website use. These districts may favor more personalized campaigns due to the lower populations. Low populations are also associated with low rates of in-home Internet access. According to data from the 2018 American Community Survey, approximately 85% of counties that fall into the first quartile of the population distribution have in-home Internet access rates less than 60%, a stark contrast compared to the 80% national average (United States Census Bureau 2016). Low Internet access within a district may correspond to reduced incentives to communicate with constituents through digital means. In addition, the competitive environment a candidate confronts may compel them to engage in more visible campaign strategies, including the use of a personal campaign website. In 2018 and 2020, a challenger was 16%–21% more likely to use a website than an uncontested incumbent. In addition, incumbents in contested races were approximately 9% more likely to use a campaign website than an uncontested incumbent.

While constituent expectations and competition can push candidates out of their preferred state of anonymity, the ideological environment within a candidate's district can also have a significant impact on state legislative campaign website use. Ideologically diverse districts have a negative correlation with website use across both parties. The risks associated with alienating diverse populations through concrete policy statements deter this type of engagement from candidates. Instead, candidates can rely on simple partisan heuristics, name recognition, and in more professionalized campaigns, microtargeting of messages to constituent groups. Further, Republicans are especially unwilling to engage in website use in adversarial districts, where they risk alienating either primary or general election constituencies and imperiling their electability. In both instances, candidates are willing to preach to the partisan choir, but averse to engaging in proselytizing among diverse and occasionally adversarial constituencies.

The failure of the digital revolution of the twentieth century to drag state legislative campaigns into public scrutiny is deeply problematic for a number of reasons. Most importantly, a lack of distinguishing information about candidates and a reliance on nationalized partisan heuristics reinforces citizen polarization. State candidates confront a difficult task of navigating local concerns with the tangible influence of nationalized partisan agendas within the constituency. When state candidates are not compelled to endorse local concerns over national interests in a public capacity, citizens are not confronted with variation in partisan agendas. Instead, they are encouraged to view things through the nationalized red-and-blue dichotomy associated with polarization. The result enhances the potential one-sided affective citizen polarization and discourages citizen engagement with state and local politics, permitting state governments to operate their tremendous policy influence with reduced accountability.

In addition, while the Internet has facilitated the potential for additional avenues of news and information, it has also undermined specifically local coverage of state politics (Abernathy 2020; Graber 1989; Hopkins 2018). With the decline in local newspapers, citizens must increasingly rely on campaign information from candidates themselves, which they are loath to provide barring the scrutiny of competition, constituent demand, or the impact of a state polarization. This combination is especially problematic when low access to information via news media combines with low website use.

While this study highlights a potential failure of the Internet to expand information access to campaigns, the changing digital media environment further complicates this position. Missing from this study is the expanded use of digital campaign content beyond a candidate's personalized website, including social media platforms and official websites. While individual candidates are averse to consistently employing and updating campaign websites, state governments experience significant pressure to establish official legislator pages. These pages provide a key opportunity for incumbents to credit claim on significant legislation for campaign purposes while avoiding making assertive statements on undecided and contested agenda items. In addition, the growth of social media has also provided an opportunity for candidates to reach out to citizens without providing a substantial campaign agenda. Through Facebook and Twitter, candidates can respond to constituent expectations regarding web presence. However, the truncated nature of both mediums lends itself more to platitudes of support and campaign announcements than substantial policy claims. Recent research has highlighted this important distinction, especially within state legislative campaigns, showing that campaign advertisements on Facebook are more targeted to particular constituents, more identifiably partisan, and also less likely to engage in concrete policy discussions (Fowler *et al.* 2021). These alternative mediums have provided candidates the opportunity to gain some of the clear benefits of web presence without incurring the risks of a full, individual, campaign website. This is especially beneficial to candidates who wish to avoid controversial statements in diverse or ideologically opposing districts.

While this study effectively dispels the myth that campaign websites are standard practice for down-ballot elections, it raises important questions in terms of campaign content at this level. Political science has focused extensively on the role of polarization in shaping national campaigns and partisan agendas. Campaign websites at the state level reveal a potentially alterative avenue to better explore the impact of polarization for candidates in low-information settings. First and foremost, subsequent studies should target the use of campaign websites in primary elections as well. Primary elections often include fewer professional candidates, but may spur increased use of websites when they present significant competition in spite of potentially nonexistent competition in the general election. In addition, access to a consistent collection of campaign websites across the years creates opportunities to examine campaign content. While this study highlights the significant variation in website use, it does not engage in evaluating the content of campaign websites, which could provide clues as to the degree that websites individualize state candidates in the face of strong national parties. Measures of ideology and space dedicated to particular issues should be examined across not only geographic boundaries but district demographics including urban-rural distinctions. Finally, what influence do candidate qualities have on not only web presence but content as well? While the Internet has facilitated a renewed hope in the potential for information to enhance citizen engagement and democratic accountability, this study underscores the prevailing interests of politicians to avoid facilitating improved access. Future studies should continue to explore this tendency as it relates to campaigns, with an eye to better understanding the potential for state variation to undermine national partisan polarization and the polarization and nationalization of the electorate.

Supplementary Materials. To view supplementary material for this article, please visit http://doi.org/10.1017/spq.2023.1.

Data Availability Statement. Replication materials are available on SPPQ Dataverse at https://dataverse.unc.edu/dataset.xhtml?persistentId=doi:10.15139/S3/DNNRIZ (Meyer-Gutbrod 2022).

Funding Statement. This material is based upon work supported by the National Science Foundation SBE Postdoctoral Research Fellowship under Grant No. 1808962. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Additional funding was provided by the University of South Carolina, College of Arts and Sciences Faculty Research Initiative.

Conflict of Interest. The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Abernathy, Penelope Muse. 2020. "News Deserts and Ghost Newspapers: Will Local News Survive." *Center for Innovation and Sustainability in Local Media, School of Media and Journalism, University of North Carolina at Chapel Hill.* https://www. usnewsdeserts. com/wp-content/uploads/2020/06/2020_News_Deserts_and_Ghost_Newspapers. pdf.

Abramowitz, Alan I., and Steven Webster. 2016. "The Rise of Negative Partisanship and the Nationalization of US Elections in the 21st Century." *Electoral Studies* 41: 12–22.

Anderson, Janna, and Lee Rainie. 2017. "The Future of Truth and Misinformation Online." *Pew Research Center.* https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online/ (May 21, 2021).

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

Benoit, Pamela J., and William L. Benoit. 2005. "Criteria for Evaluating Political Campaign Webpages." *Southern Journal of Communication* 70 (3): 230–47.

Bimber, Bruce. 1998. "The Internet and Political Transformation: Populism, Community, and Accelerated Pluralism." *Polity* 31 (1): 133–60.

Bimber, Bruce. 2001. "Information and Political Engagement in America: The Search for Effects of Information Technology at the Individual Level." *Political Research Quarterly* 54 (1): 53–67.

Bimber, Bruce. 2014. "Digital Media in the Obama Campaigns of 2008 and 2012: Adaptation to the Personalized Political Communication Environment." *Journal of Information Technology & Politics* 11 (2): 130–50.

Bimber, Bruce, and Lauren Copeland. 2013. "Digital Media and Traditional Political Participation over Time in the US." *Journal of Information Technology & Politics* 10 (2): 125–37.

Bimber, Bruce, and Richard Davis. 2003. *Campaigning Online: The Internet in US Elections*. New York: Oxford University Press.

Conerly, Bill. 2013. "*The Death of Newspapers: A Third Nail in the Coffin." Forbes.* https://www.forbes.com/sites/billconerly/2013/06/21/the-death-of-newspapers-a-third-nail-in-the-coffin/ (October 9, 2017).

Cox, Gary W., and Mathew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives*. Cambridge: Cambridge University Press.

Druckman, James N, Martin Kifer, and Michael Parkin. 2009. "Campaign Communications in U.S. Congressional Elections." *American Political Science Review* 103 (3): 343–66.

Druckman, James N., Martin J. Kifer, and Michael Parkin. 2010. "Timeless Strategy Meets New Medium: Going Negative on Congressional Campaign Web Sites, 2002–2006." *Political Communication* 27 (1): 88–103.

Druckman, James N., Martin J. Kifer, and Michael Parkin. 2014. "US Congressional Campaign Communications in an Internet Age." *Journal of Elections, Public Opinion & Parties* 24 (1): 20–44.

Druckman, James N., Martin J. Kifer, and Michael Parkin. 2018. "Resisting the Opportunity for Change: How Congressional Campaign Insiders Viewed and Used the Web in 2016." *Social Science Computer Review* 36 (4): 392–405.

Fowler, Erika Franklin, Michael M. Franz, Gregory J. Martin, Zachary Peskowitz, and Travis N. Ridout. 2021. "Political Advertising Online and Offline." *American Political Science Review* 115 (1): 130–49.

Gaynor, SoRelle Wyckoff, and James G. Gimpel. 2021. "Small Donor Contributions in Response to Email Outreach by a Political Campaign." *Journal of Political Marketing* 1–25.

Gibson, Rachel, Stephen Ward, and Wainer Lusoli. 2002. "The Internet and Political Campaigning: The New Medium Comes of Age?" *Representation* 39 (3): 166–80.

Graber, Doris A. 1989. "Flashlight Coverage: State News on National Broadcasts." *American Politics Quarterly* 17 (3): 277–90.

Hare, Christopher, and Keith T. Poole. 2014. "The Polarization of Contemporary American Politics." *Polity* 46 (3): 411–29.

Herrnson, Paul S., Atiya Kai Stokes-Brown, and Matthew Hindman. 2007. "Campaign Politics and the Digital Divide: Constituency Characteristics, Strategic Considerations, and Candidate Internet Use in State Legislative Elections." *Political Research Quarterly* 60 (1): 31–42.

Hopkins, Daniel J. 2018. *The Increasingly United States: How and Why American Political Behavior Nationalized*. Chicago, IL: University of Chicago Press.

Johnson, Dennis W. 2002. "Campaign Website: Another Tool, but No Killer App." *Journal of Political Marketing* 1 (1): 213–15.

Klotz, Robert. 1997. "Positive Spin: Senate Campaigning on the Web." *PS: Political Science and Politics* 30 (3): 482–86.

Latimer, Christopher. 2009. "Understanding the Complexity of the Digital Divide in Relation to the Quality of House Campaign Websites in the United States." *New Media & Society* 11 (6): 1023–40.

Lee, Frances E. 2016. *Insecure Majorities: Congress and the Perpetual Campaign*. Chicago, IL: University of Chicago Press.

Lipsitz, Keena, Christine Trost, Matthew Grossmann, and John Sides. 2005. "What Voters Want from Political Campaign Communication." *Political Communication* 22 (3): 337–54.

McKelvey, Fenwick, and Jill Piebiak. 2018. "Porting the Political Campaign: The NationBuilder Platform and the Global Flows of Political Technology." *New Media & Society* 20 (3): 901–18.

Meyer-Gutbrod, Joshua. 2022. "Replication Data for: Preaching to the Choir or Proselytizing to the Opposition: Examining the Use of Campaign Websites in State Legislative Elections." UNC Dataverse, V1, UNF:6:FpFWRyR+41AJt42JrpgynA== [fileUNF]. doi: 10.15139/S3/DNNRIZ.

Mitchell, Amy, Mark Jurkowitz, J. Baxter Oliphant and Elisa Shearer. 2021. "Misinformation and Competing Views of Reality Abounded Throughout 2020." *Pew Research Center*. https://www.journalism.org/2021/02/22/misinformation-and-competing-views-of-reality-abounded-throughout-2020/ (May 21, 2021).

Moskowitz, Daniel J. 2021. "Local News, Information, and the Nationalization of US Elections." *American Political Science Review* 115 (1): 114–29.

Poole, Keith T., and Howard L. Rosenthal. 2011. *Ideology and Congress*. New Brunswick, NJ: Transaction Publishers.

Rogers, Steven. 2016. "National Forces in State Legislative Elections." *The ANNALS of the American Academy of Political and Social Science* 667 (1): 207–25.

Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105 (03): 530–51.

Sinclair, Barbara. 2011. *Unorthodox Lawmaking: New Legislative Processes in the U.S. Congress*. CQ Press. Washington D.C.

Sinclair, Barbara. 2014. 10 *Party Wars: Polarization and the Politics of National Policy Making*. University of Oklahoma Press.

Smith, Aaron. 2009. "The Internet's Role in Campaign 2008." *Pew Research Center*. https://www.pewresearch.org/internet/2009/04/15/the-internets-role-in-campaign-2008/.

Smith, Aaron, and Maeve Duggan. 2012. "Presidential Campaign Donations in the Digital Age." *Pew Research Center*. https://www.pewresearch.org/internet/2012/10/25/presidential-campaign-donations-in-the-digital-age/ (May 13, 2021).

Squire, Peverill. 2017. "A Squire Index Update." *State Politics & Policy Quarterly* 17 (4): 361–71.

Tausanovitch, Chris, and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *The Journal of Politics* 75 (2): 330–42.

Tolbert, Caroline J., and Ramona S. McNeal. 2003. "Unraveling the Effects of the Internet on Political Participation?" *Political Research Quarterly* 56 (2): 175–85.

United States Census Bureau. 2016. "*American Community Survey, 5-Year Estimates*." https://www.census.gov/programs-surveys/acs/data.html.

Ward, Stephen, and Rachel Gibson. 2003. "On-Line and on Message? Candidate Websites in the 2001 General Election." *The British Journal of Politics and International Relations* 5 (2): 188–205.

Warshaw, Christopher, and Jonathan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" *The Journal of Politics* 74 (1): 203–19.

Williams, Christine B., and Girish J. "Jeff" Gulati. 2013. "Social Networks in Political Campaigns: Facebook and the Congressional Elections of 2006 and 2008." *New Media & Society* 15 (1): 52–71.

Author Biography. Joshua Meyer-Gutbrod is teaching faculty in the Department of Political Science at the University of South Carolina, Columbia campus. His research focuses on the impact of national polarization in shaping state-level politics and he maintains the Digital Campaigns Project, tracking and archiving state legislative campaign websites.

Cite this article: Meyer-Gutbrod, Joshua. 2023. Preaching to the Choir or Proselytizing to the Opposition: Examining the Use of Campaign Websites in State Legislative Elections. *State Politics & Policy Quarterly* **23** (2): 166–186, doi:10.1017/spq.2023.1

State Politics & Policy Quarterly (2023), **23**: 2, 187–217
doi:10.1017/spq.2022.26

ORIGINAL ARTICLE

The Dynamics of Hidden Partisanship and Crossover Voting in Semi-Closed Primaries

Matthew P. Thornburg

Department of History, Political Science, and Philosophy, University of South Carolina Aiken, Aiken, SC, USA
Email: matthewth@usca.edu

(Received 27 July 2021; revised 17 June 2022; accepted 11 July 2022)

Abstract

Among US states with party registration, many allow the unaffiliated to choose either the Democratic or Republican primary. States with these semi-closed rules thus provide an option to voters with greater choice than registering with a single political party. Using the synthetic control method, I find that the introduction of semi-closed primaries is associated with growth in unaffiliated registration. However, the likelihood of unaffiliated registration is not even across the electorate in semi-closed states. I show that it is most common where a voter's party is not competitive and the access unaffiliated registration provides to the strong party's primary is valuable. Consistent with this instrumental motive, unaffiliated voters in semi-closed states use their freedom of choice to vote in the primary of the stronger party in the electorate. This leads to significant crossover voting among unaffiliated voters who do not identify with that party such as Democrats in red states or Republicans in blue states. These findings show the unintended consequences of electoral institutions and find primary crossover voting is more common under some circumstances than others.

Keywords: primary elections; election rules; voting behavior; elections; party identification; party registration

Introduction

Three-fifths of US states regulate primary election participation through party registration. Citizens register with the Democratic, Republican, or minor parties or remain unaffiliated. This choice determines which party primary they may vote in with "closed" primaries requiring voters to register with a party to participate in its primary. However, 12 "semi-closed" states also allow unaffiliated voters to vote in the primaries of either major party.

States with party registration instituted this system to prevent primary crossover voting (Ware 2002). Yet closed and semi-closed states rely on a critical assumption to deter it: a voter's party registration matches their party identification. This may not be true and, under some circumstances, a mismatch is widespread in the electorate (Arrington and Grofman 1999; Key 1949; Thornburg 2018; 2019).

Some instances of this mismatch (termed "hidden partisanship") are instrumental and driven by a lack of interparty competition. There is evidence that voters who live in states with closed primaries and identify with the minority party strategically register with the strong party in order to gain access to its primary elections and select the nominees of the party favored to win the general election (Arrington and Grofman 1999; Key 1949).

If this is the case, then semi-closed primaries, where unaffiliated voters may choose to vote in either the Democratic or Republican primary, represent a more attractive prospect still for instrumental hidden partisanship. In states with semi-closed primaries, voters who identify with the electorally weaker of the major parties in a state but remain unaffiliated gain access to the primary of the stronger party and preserve access to their own party's primary. Voters identifying with the weak party in semi-closed states should remain unaffiliated.

The greater instrumental utility of unaffiliated registration in semi-closed primaries should lead states that implement semi-closed laws to increase the share of registrants that is unaffiliated. To establish causality, I use the opening of Arizona and North Carolina's primary elections to unaffiliated voters as comparative case studies. Using the synthetic control method (Abadie, Diamond, and Hainmueller 2010), I generate a counterfactual version of the treated states that did not open its primary elections. Compared to this synthetic control, the proportion of both states' electorate that is unaffiliated significantly increased in the decade after the implementation of semi-closed primaries there. The ratios of partisan registrants to identifiers also decreased, supporting the notion of hidden partisanship, especially Democrats in Republican Arizona.

I model the decision to register with a political party or remain unaffiliated. The model predicts that we observe hidden partisanship in uncompetitive states among voters identifying with the minority party in the electorate and that unaffiliated hidden partisanship by partisans in semi-closed states will be the most common pattern observed. Using the 2018 Cooperative Congressional Election Study (CCES), I examine Democrats, Republicans, and independents in semi-closed and closed primary states. I confirm that (1) hidden partisanship is most commonly observed in semi-closed states and (2) the probability of hidden partisanship grows as a voter's party becomes weaker in the state. There is weak evidence for such hidden partisanship in states with closed primaries with the exception of independents, who register with the stronger party in the state.

Unaffiliated voters in semi-closed states also know that they may vote in either the Democratic or Republican primaries and take advantage of this choice. Voters respond to the partisanship of their state in deciding which primary to participate in. Unaffiliated primary voters in blue states are most likely to vote in the Democratic primary, and likewise with the Republican primary in red states. This includes a significant minority of unaffiliated partisans who engage in crossover voting in states where they do not identify with the stronger party. These individuals participate in the primary most likely to yield the eventual officeholders, consistent with an instrumental motive.

While primary elections may have been opened to unaffiliated voters in hopes of bringing independents into the partisan ranks, they lead to strategic behavior by partisans. Most of the literature on primary crossover voting finds it to be rare and inconsistent. Among unaffiliated partisans in politically unfriendly semi-closed primary states, crossover voting is more common. The findings highlight the neglected role that institutions and political context play in motivating strategic behavior in party registration and primary voting.

Background and theory

Hidden partisanship is the phenomenon where a voter's party registration fails to match their party self-identification (Arrington and Grofman 1999). This can include partisans registered with a different party than the one with which they identify or who remain unaffiliated. It can also include independents registered with a political party. The literature identifies several mechanisms that may result in this hidden partisanship.

Some hidden partisanship is "unintentional." Thornburg (2018) uses local changes of address and the subsequent required re-registration to show that many voters in Oklahoma would switch party registration from the Democratic Party if they could. The author theorizes that the realignment of the state from Democratic to Republican has "stranded" many individuals registered with the Democratic Party who now identify as Republican. Similarly, Thornburg (2019) finds that in counties that have realigned and are located in states where it is difficult to change party registration, there is greater difference between presidential vote share and aggregate party registration, indicating voters located in these counties may be registered with one party but identify with (or at least vote for) a different one. These studies suggest that this unintentional hidden partisanship is most prevalent where a partisan realignment has occurred. In such places, more people are registered with the weakening party than now identify with it.

Hidden partisanship may also be due to "social pressure." Voters living in areas dominated by one party while they identify with the minority party might conceivably register with the dominant party out of this pressure. Ansolabehere and Hersh (2012) and Bell and Buchanan (1966) show that voters may misreport their party registration compared to validated measures. Bell and Buchanan (1966) theorize this misreport is due to the greater prestige of some party registration statuses and the role social pressure plays. Social pressure hidden partisanship should exist where one party is in the clear majority and a voter self-identifies with the weaker party. It should take the form of a voter registering with the dominant party as opposed to the weak party.

Hidden partisanship may also be due to *instrumental* factors. In such a case, primary crossover voting drives hidden partisanship. The general consensus in the literature is that crossover voting – choosing to vote in a primary where the voter does not identify with the party – is rare in the aggregate nationwide (Norrander 2018). However, evidence also shows that the rate of crossing over is not consistent across elections and responsive to context. Reported rates of primary crossover voting vary widely in ways corresponding to election-specific factors (Alvarez and Nagler 1997; 2002; Burden and Jones 2009). For example, Burden and Jones (2009) find the percentage of the primary electorate composed of crossover voters ranges from 18% to 49% (including independents) among a number of studies and contests.

One type of instrumental hidden partisanship may be related to "maximizing options." The literature on primary turnout shows that voters are more likely to vote in a primary that is competitive compared to one that is uncompetitive or uncontested (Ezra 2001; Jewell 1977; Kenney 1983; Kenney and Rice 1986). With this in mind, voters might register in ways that afford the freedom to switch party primaries to the contest that is most competitive, such as remaining unaffiliated in semi-closed states.

Another form of instrumental hidden partisanship – and the focus of this paper – is driven by "impact voting." In examining the California blanket primary in 1998, Alvarez and Nagler (2002) find drastically different rates of primary crossover voting across state legislative districts. In this study, the highest rates of crossover voting among partisans were in districts where one party held a clear advantage and the general election race was perceived to be safe. The crossover behavior here by identifiers with the electorally weak party accords with strategic attempts to maximize the impact of one's vote. Confirming this, Weaver (2015) finds rates of crossover voting among Democrats were highest in Republican areas of the state during the North Carolina 2010 primary elections. Among related research on primary turnout, the partisan balance of a state (Ezra 2001; Jewell 1977; Kenney 1983, 1986; Kenney and Rice 1986) drives turnout. Hanks and Grofman (1998) examine primary turnout in the one-party South, where primary election turnout relative to general election turnout increased with competitive primary races and low levels of interparty competition. Taken together with the research on primary crossover voting, these studies show that voters gravitate toward primaries where the winner of the primary will be likely to win the general election.

Impact voting hidden partisanship follows naturally from this. Key (1949) observed an extreme case in the Solid South in an era where the Democrat was usually the foregone winner of the general election. In North Carolina, Republicans registered as Democrats because the Republican nominees were sure to lose the general election, making participation in the Republican primary of little instrumental value. Arrington and Grofman (1999) confirm this by examining party registration totals in North Carolina localities versus the actual support that parties receive at the ballot box. Fewer individuals register with the electorally weak party relative to its actual electoral support. The authors conclude that voters identifying with the less competitive party choose not to register with it. This suggests that the local political context in an electorate – specifically the level of interparty competition there – drives the party a voter registers with. Voters should, all else equal, assume registration statuses that grant access to the other party's primary election when the voter's own party is not competitive in the general election and participation in its primary is of little instrumental value.

Because these four forms of hidden partisanship are driven by different behavioral mechanisms, we should observe different patterns with each. Unintentional hidden partisanship is due to the barriers in place to changing party registration. It should therefore be the least responsive to changes in electoral institutions, competitiveness of primaries, or levels of interparty competition and should lag changes in aggregate party identification, such as electoral realignments. Maximizing options hidden partisanship is instrumental and responds to changes in electoral institutions. Voters engaging in this form of hidden partisanship desire the freedom to choose between the parties and will gravitate toward the party primary that is most competitive. Voters maximizing options will register as unaffiliated in semi-closed primary states and do so regardless of the level of interparty competition (i.e., whether their party is strong or weak). They are not necessarily motivated by the futility of voting in their own party primary if they are in the minority (except insofar as the dominant party's primaries are usually more competitive) so much as which intraparty contest is most competitive. Impact voting hidden partisanship, in contrast, is affected by which party is dominant in the electorate. These voters register as unaffiliated or with the other major party where their own party is weak to gain access to the dominant party's primary. Among these voters, we should observe significant amounts of crossover voting where their party is weak. Finally, social pressure hidden partisanship is prevalent where a voter believes there are social or professional consequences from registering with their own political party. This should also occur where a voter's party is weak in the electorate. Social pressure hidden partisanship is distinguished from impact voting in that the latter is characterized by crossover voting, while the former is not.

Finally, among the purposeful forms of hidden partisanship (those which are not unintentional), it is certainly possible more than one mechanism affects a particular individual. Voters maximizing options may also engage in impact voting or be affected by social pressure. Especially among the instrumental motives, it is conceivable that voters who gravitate toward the most competitive primary to best "spend" their vote will also gravitate toward the primary of the party most likely to yield the eventual general election winner.

At the same time, other factors drive the decision to register (or not) with a party. Large numbers of voters in semi-closed states register with the parties, even though it is not necessary to do so and registering with a party actually restricts the voter to just one party primary rather than granting a choice of primary. Therefore, a psychic benefit to party registration also exists. Thornburg (2014) calls the act of registering with a party a "constitutive norm," validating an individual's party identification through official recognition. Constitutive norms serve as "the very actions that lead others to recognize an actor as having a particular identity" (Abdelal *et al.* 2006, 697). Thus, these norms signal to the individual and others engaging in them that they are members of a group. In examining the concept of what it means to be "American," Schildkraut (2007) finds that official status as an American citizen is among the most important signifiers of identity as an American. Similarly, research has found that individuals who are "formal" members of a group relate to the group differently, holding a weaker sense of autonomy but a notably stronger sense of differentiation from others (Sheldon and Bettencourt 2002). Other research also indicates that the act of registering with a political party reinforces an individual's party identification (Burden and Greene 2000; Finkel and Scarrow 1985; Gerber, Huber, and Washington 2009).

Other plausible non-instrumental factors influencing an individual's choice of party registration exist. Gerber *et al.* (2017) suggest that many voters hold exclusionary beliefs about who should vote in primary elections which may discourage hidden partisanship. They find 44% of individuals they surveyed believed that partisans should not engage in crossover voting and 23% believed independents should not participate in primary elections. Thus, a strong social norm exists for individuals to register with their own party, discouraging partisans from registering with the other party or remaining unaffiliated. These exclusionary beliefs increase with the strength of partisanship, perhaps leading to greater resistance to hidden partisanship from strong partisans compared to weak partisans or leaners.

The results from Gerber *et al.* (2017) as well as research on constitutive norms suggest that we should observe different psychic benefits among different party registration states. For a partisan, especially a strong one, registration with one's own party provides the greatest psychic benefit, first because this serves to validate a voter's existing party identification as well as because it does not violate an individuals's exclusionary beliefs about participation in another party's primary election. Registration with the other major party provides the smallest psychic benefit as this directly contradicts an individual's partisan identity and may violate social norms.

Model of party registration

Based on this prior research, I present a simple model of party registration here.

A voter, i, identifies with a political party in a state with party registration. They must decide which party registration state to select from $j \in \{D, R, I\}$ (i.e., Democratic, Republican, or unaffiliated). Without loss of generality, i supports the Democratic Party. Their decision is based primarily on two forms of utility: *psychic utility* and *instrumental utility* – the latter driven here by impact voting. Psychic utility represents the perceived psychological and expressive benefit from their party registration state, j. Instrumental utility describes the utility derived by i from the access to the primaries in choosing the eventual office holders. The overall utility has the following function:

$$U_{ij}\left(s_i, p_j, \mathbf{d}_i\right) = s_i r_j + (1 - s_i) v_j + p_j + \mathbf{b}_j \cdot \mathbf{d}_i + \varepsilon_{ij}. \tag{1}$$

Here, s_i is an indicator, equaling 1 if an individual is a strong partisan and 0 otherwise. r_j and $v_{j,}$ respectively, are values giving the psychic benefit from state j, which is assumed to differ between those with strong and weak partisanship. Without loss of generality, for Democratic i, we assume that $r_D > r_I > r_R$ and $v_D > v_I > v_R$ and set $r_R = 0$ and $v_R = 0$. In addition, we assume that $r_D > v_D$ and $r_D - r_I > v_D - v_I$. p_j is an expression of the value of primary access provided by j. For $j \in \{D, R\}$, p_j equals the proportion of partisans in the electorate identifying with the party. It is assumed that $p_D = 1 - p_R$. In a closed primary state, $p_I = 0$. In a semi-closed primary state, $p_I = 1$. In addition, $\mathbf{d}_i$ is a vector of individual characteristics multiplied by vector of coefficients $\mathbf{b}_j$. ε_{ij} is a random disturbance term taking on a type-I extreme value distribution.

Due to the distribution of the random disturbance term, the probability that i chooses j, ρ_{ij} can be expressed as

$$\rho_{ij}\left(s_i, p_j, \mathbf{d}_i\right) = \frac{e^{s_i r_j + (1 - s_i) v_j + p_j + \mathbf{b}_j \cdot \mathbf{d}_i}}{\sum_j e^{s_i r_j + (1 - s_i) v_j + p_j + \mathbf{b}_j \cdot \mathbf{d}_i}}. \tag{2}$$

Previous literature suggests that as the Democratic Party becomes less competitive in i's electorate, registering as a Republican or remaining unaffiliated will become a more attractive prospect. Taking the derivative of ρ_{iD} with respect to p_R, we obtain

$$\frac{\partial \rho_{iD}}{\partial p_R} = \rho_{iD}^2 - \rho_{iD} - \rho_{iD}\rho_{iR}. \tag{3}$$

This derivative will always be negative, indicating that the probability of registering as a Democrat decreases as the Republican Party grows stronger in their electorate. It also follows:

$$\frac{\partial \rho_{iR}}{\partial p_R} = \rho_{iR} - \rho_{iR}^2 + \rho_{iD}\rho_{iR} \tag{4}$$

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

and

$$\frac{\partial \rho_{iI}}{\partial p_R} = \rho_{iD}\rho_{iI} - \rho_{iR}\rho_{iI}. \tag{5}$$

It is clear that ρ_{iR} is strictly increasing with p_R as expected. For the probability of remaining unaffiliated (in either closed or semi-closed states), as long as $\rho_{iD} > \rho_{iR}$, the probability is increasing with respect to p_R. This is likely given the important psychic role that party registration plays (especially registering with one's own party) as well as the social norms in place discouraging strategic registration.

Finally, because $p_I > p_R$ in all cases in semi-closed primary states as well as the fact that $r_I > r_R$ and $v_I > v_R$, $\rho_{iI} > \rho_{iR}$ in semi-closed states in all semi-closed situations.

The effect of semi-closed primaries on hidden partisanship

Based on the foregoing model, I distinguish semi-closed from closed primaries using three criteria: (1) semi-closed primary states allow voters who are unaffiliated on primary election day to participate in party primaries while closed primary states do not; (2) semi-closed primary states give unaffiliated voters access to the primaries of both major parties (rather than just one); and (3) individuals registered with a political party on primary election day may only vote in that party's contest in both closed and semi-closed states[1]. In essence, semi-closed states provide greater instrumental utility to unaffiliated voters on primary election day compared to unaffiliated voters in closed primary states or voters registered with a political party in either closed or semi-closed states. In the former case, an unaffiliated voter in a semi-closed primary state accesses both major party primaries compared to an unaffiliated voter in a closed primary state who may not participate in any party primary. An unaffiliated voter in a semi-closed state also has greater choice compared to an individual registered with a political party; the option to choose either party primary exists compared to just one party's contest for those registered with a party.

If registration with a political party is not simply a declaration of one's party identification or independence but instead a decision informed by the instrumental utility this choice provides in selecting one's representatives, then we should see differences in aggregate registration counts among states with different primary election laws. A larger portion of the electorate in semi-closed primary states will be unaffiliated compared to states with closed primaries.

Table 1 lists the mean state percentage of registered voters in party registration states who were unaffiliated at the time of the 2018 general election. Closed primary states are distinguished from states with semi-closed primaries. Averages are *not* weighted by state population and include all registered voters in the state. States used in this paper's analysis and the coding for them are in Table 2. The table includes the year a state became fully semi-closed where applicable.

The large difference between the means in Table 1 suggests a greater tendency to register as unaffiliated in semi-closed states as opposed to closed primary states. However, this does not by itself show a causal relationship, nor does it indicate hidden partisanship. Observational studies purporting to demonstrate a causal effect of

[1]Please see Section A of the Supplementary Material for a greater discussion of definitions and justification for individual classification of states.

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

Table 1. Unaffiliated registration in closed and semi-closed primary states

	Closed states	Semi-closed states
Mean % unaffiliated	21.8	37.1
Number of states	13	12

Table 2. Party registration states with closed or semi-closed primaries

	Primary format	Year semi-closed	Synthetic control?	Competitiveness
Arizona	Semi-closed	2000	✓	C
Colorado	Semi-closed	1982	×	C
Connecticut	Closed		✓	D
Delaware	Closed		✓	D
Florida	Closed		✓	C
Idaho	Semi-closed	2011	×	R
Kansas	Semi-closed	1980	×	R
Kentucky	Closed		✓	R
Massachusetts	Semi-closed	1903	×	D
Maryland	Closed		✓	D
Maine	Semi-closed	1985	×	C
North Carolina	Semi-closed	1995	✓	C
Nebraska	Closed		×	R
New Hampshire	Semi-closed	1987	×	C
New Jersey	Semi-closed	1975	×	D
New Mexico	Closed		✓	D
Nevada	Closed		✓	D
New York	Closed		✓	D
Oklahoma	Closed		✓	R
Oregon	Closed		✓	D
Pennsylvania	Closed		✓	C
Rhode Island	Semi-closed	1974	×	D
South Dakota	Closed		✓	R
Utah	Semi-closed	2000	×	R
West Virginia	Semi-closed	2007	×	C

Abbreviations: D, Democratic; C, competitive; R, Republican.

electoral institutions on political behavior may suffer from endogeneity as the behavior of the electorate drives implementation of election laws (Hanmer 2009). Semi-closed primaries may not lead to hidden partisanship and registration as an unaffiliated voter, especially given that some elected policymakers implement semi-closed primaries in hopes of increasing independent support for their party (Madden 1986; Sinclair 2013). Norrander (1989) also shows that wide variation exists in independent *identification* among the states. We should not assume every state has the same proportion of its electorate identifying as independents (and therefore the same proportion remaining unaffiliated by default). With the possibility that unaffiliated voters drive semi-closed primary laws rather than the other way around, we need more sophisticated methods of establishing causality.

I use the introduction of semi-closed primary elections in Arizona and North Carolina as quasi-experiments. Starting in 2000, unaffiliated registrants residing in Arizona on primary election day could vote in either the Democratic or Republican non-presidential primary elections. And in North Carolina, following a change to state law, the Republican Party opened its primary to unaffiliated voters in 1988 with the Democratic Party following suit in 1995.

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

This test of the causal effect of election laws on hidden partisanship is intended to separate unintentional from other forms of hidden partisanship and show that one of the others is at work with semi-closed primaries. Unintentional hidden partisanship results from the stickiness and slow change of party registration in the aggregate. If it drives hidden partisanship, a change in electoral rules should not greatly affect the aggregate party registration of a state. In contrast, social pressure, maximizing options, and impact voting hidden partisanship emphasize responsiveness to the partisan context and/or electoral rules in place. With these three types of hidden partisanship, voters purposely avoid registering with their own party, either out of social pressure or to gain access to other primary options. If the introduction of semi-closed primaries in these two states led to the share of the electorate that is unaffiliated to increase, then the evidence will support the presence of one of these three types of hidden partisanship.

Furthermore, I theorize much of the effect of semi-closed primary laws on hidden partisanship is due to a desire to engage in impact voting. However, this form of hidden partisanship is observationally equivalent to social pressure hidden partisanship in terms of aggregate party registration. Both forms of hidden partisanship are most prevalent among supporters of the less competitive party in a state. If either of these forms of hidden partisanship is present, I expect that the ratio of registered Democrats to Democratic identifiers to exhibit a greater decrease compared to the ratio of registered Republicans to Republican identifiers with the introduction of semi-closed primaries in Arizona. While Arizona has become more competitive politically in recent years, at the time of the introduction of semi-closed primaries, it was considered strongly Republican and has remained Republican-leaning well into the twenty-first century. Thus, more Democrats in Arizona should engage in hidden partisanship. I also evaluate the Democratic and Republican registrant/identifier ratios in North Carolina, though predictions are less straightforward with that state.

Methods exist for causal inference of a policy change that is not randomly assigned. In Arizona and North Carolina, the decision was a result of conscious changing of the laws. Because of the fundamental problem of causal inference (Imbens and Rubin 2015), we are unable to simultaneously observe these two states post-treatment both with and without the ability of unaffiliated voters to vote in primary elections. Thus, we are not able to conclusively determine whether a difference between closed states and the two treated states post-introduction of semi-closed primaries is due to the causal effect of election laws on the latter.

I employ the synthetic control method. Synthetic control methods work well for comparing the effect of a policy treatment or other intervention in a single aggregate unit to other units that did not receive the treatment (Abadie, Diamond, and Hainmueller 2010). In this case, the synthetic control method generates composite "counterfactual states" against which to compare the treated states before and after the implementation of semi-closed primaries observed in Arizona and North Carolina (Abadie, Diamond, and Hainmueller 2010; Abadie and Gardeazabal 2003). The weighted average of the pool of closed primary states forms the synthetic control, with weights assigned to each member of the pool ranging between zero and one and summing to one. The weights are chosen such that relevant characteristics of Arizona and North Carolina pre-treatment are most closely approximated by the synthetic control.

If $\mathbf{X_1}$ comprises a $(k \times 1)$ vector of relevant pre-intervention characteristics for Arizona (North Carolina), and $\mathbf{X_0}$ is a $(k \times J)$ matrix containing the values of these characteristics for the pool of J closed primary states, then the vector of weights chosen, $\mathbf{W}^*$, minimizes

$$\sum_{m=1}^{k} v_m(\mathbf{X_{1m}} - \mathbf{X_{0m}}\mathbf{W})^2.$$

Here, v_m is a weight assigned to the mth variable (characteristic of the states). A number of methods exist for determining the variable weights (v_m); in this case, I choose weights based on their ability to predict the dependent variable during the pretreatment period. The choice of v_m minimizes the mean-squared prediction error (MSPE) for the dependent variable of the treatment and control states for the time period of 1980 to the election year before semi-closed primaries were fully implemented in each states' analysis.

The choice of predictor variables is particularly important for creating an accurate counterfactual state. Norrander (1989) examines the wide variation in independent/unaffiliated registration and identification among states and identifies characteristics affecting independence among voters. In particular, the analysis cites state location in the South, state political competitiveness, turnout of the electorate, and the strength of the party system as predicting independent/unaffiliated registration. I use seven predictor variables, measuring these characteristics of the state and the lagged dependent variable. For presence in the South, I include a dummy variable (including Kentucky and Oklahoma, which exhibit similar party registration patterns to other southern states). To measure the strength of the state's party system, I use the two variables Morehouse and Jewell (2005) created, measuring the weakness of the state's parties by the divisiveness of gubernatorial nominations and the ability of parties to formally endorse candidates. Both of these variables are measured on a three-point scale with larger numbers indicating a weaker state party system or less ability to endorse. To measure the degree of political competition, I use the 10-year average folded Ranney Index for each state and year and the 10-year average Ranney Index for each state and year (Klarner 2013). Larger values of the folded Ranney Index indicate more competition and larger values of the Ranney Index indicate a more Democratic state. For each year, I also include the two-year average of both the percentage of the citizen voting age population that is registered and that voted in that election. In addition, I follow the recommendation of Abadie, Diamond, and Hainmueller (2010) and include the lagged value of the dependent variable as a predictor.

I initially examine the proportion of the registered voters in each state not affiliated with a major political party as my dependent variable, covering the time period from 1980 to 2010 on even (election) years. I separately compare Arizona and North Carolina to all states maintaining closed primaries over this period with the exception of Nebraska for which suitable data are not available. There is no reason to expect spillover effects among voters when registering as unaffiliated. I match from 1980 to 1998 in the case of Arizona and from 1980 to 1994 in the case of North Carolina. Table 3 displays descriptive statistics for Arizona and North Carolina, the mean of the pool of control states and their synthetic controls as well as values of v_m in each case.

A comparison of the synthetic control to the treated states in question shows a good match on covariates compared to the unweighted pool of controls, especially for variables weighted heavily in determining unaffiliated registration (v_m). The composition of the composite control states is given in Table 4.

Table 3. Predictor variable values and weights

	Arizona	Synthetic Arizona	Pool of controls	v_m
Party system strength	2.00	1.88	1.92	0.01
Party endorsement	3.00	2.21	2.33	0.00
Southern state	0.00	0.01	0.25	0.01
Folded Ranney Index	0.93	0.92	0.86	0.01
Regular Ranney Index	0.43	0.56	0.61	0.00
CVAP turnout	52.77	52.82	56.38	0.05
CVAP registration	63.83	63.85	68.98	0.05
Lagged unaffiliated	0.11	0.11	0.12	0.87
	North Carolina	Synthetic North Carolina	Pool of controls	v_m
Party system strength	1.00	2.68	1.92	0.00
Party endorsement	3.00	2.99	2.33	0.21
Southern state	1.00	0.68	0.25	0.00
Folded Ranney Index	0.78	0.78	0.85	0.20
Regular Ranney Index	0.72	0.72	0.62	0.10
CVAP turnout	50.11	50.27	57.27	0.00
CVAP registration	64.32	64.49	69.28	0.14
Lagged unaffiliated	0.05	0.05	0.12	0.35

Table 4. State weights for synthetic control

State	Arizona Synthetic weight	North Carolina Synthetic weight
Connecticut	0.12	0.00
Delaware	0.01	0.00
Florida	0.00	0.00
Kentucky	0.01	0.45
Maryland	0.00	0.00
Nevada	0.31	0.31
New Mexico	0.00	0.00
New York	0.00	0.00
Oklahoma	0.00	0.23
Oregon	0.00	0.00
Pennsylvania	0.55	0.01
South Dakota	0.00	0.00

With these weights composing the synthetic Arizona and North Carolina, the top two plots of Figure 1 show the proportion of the electorate in the treated states and their synthetic controls that is composed of unaffiliated voters over the time period measured. The unaffiliated proportion in the synthetic controls approximate the actual Arizona and North Carolina from 1980 through the implementation of semi-closed primaries, indicating their suitability as a counterfactual. However, after the introduction of semi-closed primaries in the two states, treatment and control diverge. The proportion of voters who are unaffiliated increases significantly in Arizona and North Carolina over the next decade. The bottom two plots of Figure 1 show the gap between treatment and control for Arizona and North Carolina and further demonstrate divergence. It is important to recall that in the case of North Carolina, the Republican Party actually opened its primary to unaffiliated voters in 1988, seven years prior to the Democratic Party and thus prior to the shift of the state

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

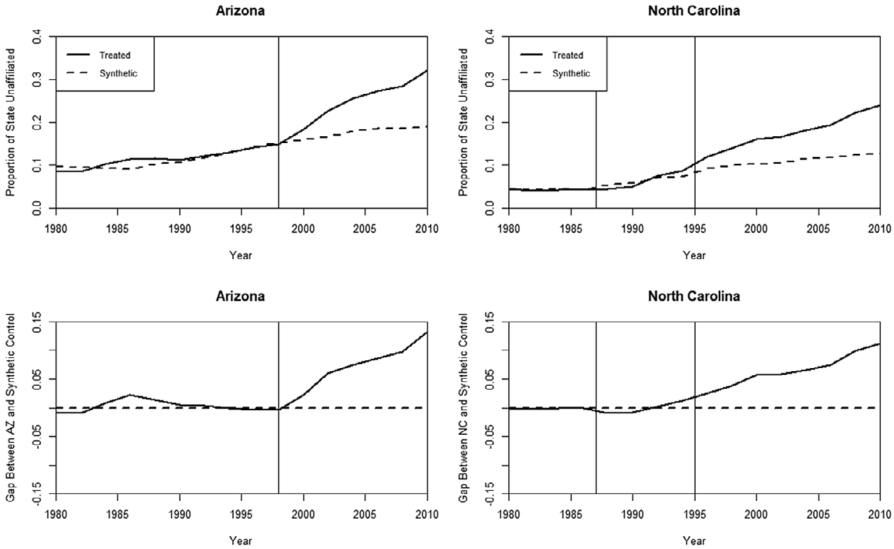


Figure 1. Comparison of treated states and synthetic controls.

to semi-closed under the definition of this paper. The growth of North Carolina's unaffiliated population relative to the state's synthetic control in the period 1988 to 1994 may thus be due to the Republicans' earlier shift.

Besides a purely visual comparison of the treated states and their synthetic controls, Abadie, Diamond, and Hainmueller (2015) recommend an analysis of the post-/pre-treatment MSPE ratio for the treated units. If the introduction of semi-closed primaries does indeed increase the proportion of the electorate that is unaffiliated post-treatment, then we should witness a divergence between the treated states and their synthetic controls *after* the implementation of the rules change. On the other hand, prior to the implementation of the semi-closed rules, the synthetic control and treated state should closely match each other. The gap between treated state and synthetic control is measured as the MSPE. Therefore, the ratio of the MSPE post-treatment to the MSPE pre-treatment gives an intuitive measurement of both synthetic control fit prior to treatment and effect of the treatment. I also follow Abadie, Diamond, and Hainmueller (2015)'s recommendation to conduct a placebo test on donor states. I generate a synthetic control for each closed donor state and compute the post-/pre-treatment MSPE ratio for each one, comparing the closed states to Arizona and North Carolina.

Figure 2 shows that the post/pre MSPE ratio is large for both Arizona and North Carolina compared to the closed "control" states in the placebo tests. The only placebo states that exceed Arizona or North Carolina is Florida in the case of North Carolina. While I do not have a conclusive reason why the proportion of unaffiliated registrants increased significantly in Florida around 1995, Norrander (1989) in the analysis of unaffiliated/independents in states finds uncompetitive southern states in 1989 to differ significantly from the rest of the country in unaffiliated registration. Florida, as a "rim South" state growing more competitive may have been witnessing a surge in unaffiliated registration as they transitioned from one-party governance.

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

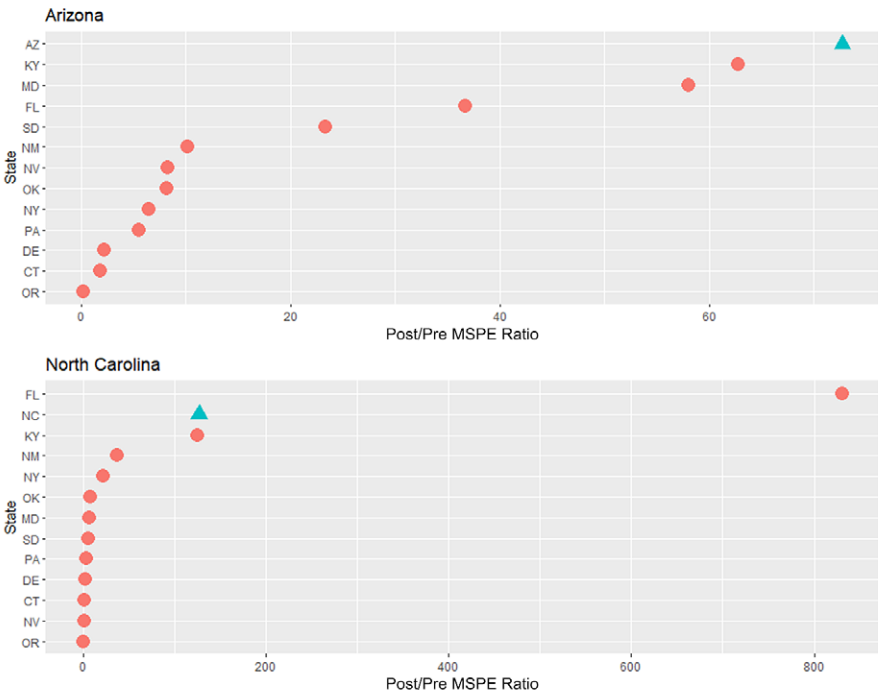


Figure 2. Post-/pre-treatment MSPE for treated states.

I have shown that two states significantly increased their proportion of unaffiliated voters compared to synthetic controls after they implemented semi-closed primary elections. This is consistent with social pressure, maximizing options, and impact voting hidden partisanship. However, my theory that hidden partisanship is driven by impact voting also predicts that in politically uncompetitive states, a greater proportion of identifiers with the weaker party will register as unaffiliated because of the instrumental utility this status provides in accessing the primary of the stronger party[2]. I test this theory on both Arizona and North Carolina with my dependent variables being the ratio of registered Democrats (Republicans) to Democratic (Republican) identifiers in the state. I utilize the measures of aggregate state party identification from Enns and Koch (2013) which are available from 1980 to 2010 and average these measures of party ID from the previous three election years. These state party identification measures utilize multilevel regression with poststratification (MRP) and survey aggregation to create estimates of party identification for each state in every year during this time span.

I predict that in Arizona, the ratio of registered Democrats to Democratic identifiers should decrease after the implementation of semi-closed primary elections there compared to the Republican ratio, as a greater proportion of Democrats chose to register unaffiliated in what was a strongly Republican state. My predictions for

[2]Once again, this same pattern of hidden partisanship is observed for social pressure, though for a different reason.

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

North Carolina are less clear and complicated by three factors: First, the North Carolina Republican Party opened its primary to unaffiliated voters in 1988, seven years before the Democrats. This gradual roll-out of semi-closed primaries in the state may complicate an easy analysis of hidden partisanship. Second, North Carolina, like many southern states, displayed a significant degree of segmented partisanship with its voters supporting Republicans at the federal level and Democrats at the state and local level (Wekkin 1991). Finally, the period from 1980 to 2010 was one of significant realignment while the state moved from fully Democratic to competitive. Table 5 displays descriptive statistics for the Democratic and Republican ratios of registrants to identifiers in Arizona and North Carolina, including the treatment states, pools of controls, and synthetic controls. Table 6 shows the weights assigned to each of the donor states forming the synthetic controls in the comparisons of ratios.

Figure 3 shows the path and gap plots for the Democratic and Republican ratios in Arizona and Figure 4 shows these plots for North Carolina. Examining Figure 3 shows support for the hypothesis that the Democratic ratio of registrants to identifiers showed a greater decrease relative to the Republican ratio after the implementation of semi-closed primaries in Arizona. The Democratic ratio decreases consistently from 2000 to 2010, indicating that the ratio of registered Democrats to self-identified Democrats went down over this time period compared to the synthetic control. Fewer Democrats were registered with their party relative to those identifying as Democrats in Arizona after the implementation of semi-closed primaries in the state – consistent with social pressure or impact voting hidden partisanship. While the ratio decreases for Republicans as well, the gap between Arizona and the synthetic control is smaller in that case.

Figure 5 quantifies the difference between Democratic and Republican ratios in Arizona. The post-/pre-treatment MSPE plots show that the MSPE for the Democratic ratio post-implementation of semi-closed primaries is over 50 times the MSPE

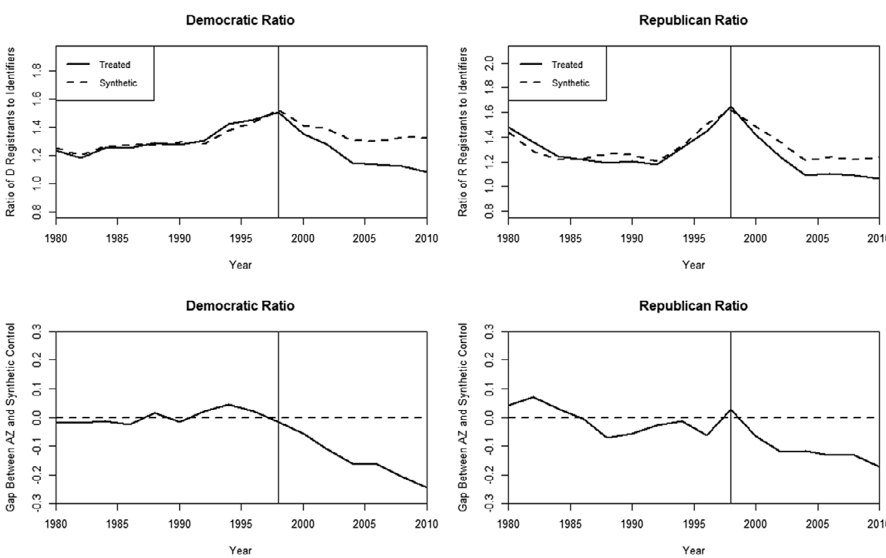


Figure 3. Comparison of Arizona and synthetic controls.

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

Table 5. Predictor variable values and weights

	Arizona	Synthetic Arizona	Pool of controls	v_m
Democratic ratio				
Party system strength	2.00	2.00	1.92	0.00
Party endorsement	3.00	2.09	2.33	0.02
Southern state	0.00	0.00	0.25	0.06
Folded Ranney Index	0.93	0.96	0.86	0.00
Regular Ranney Index	0.43	0.49	0.61	0.15
CVAP turnout	52.77	53.83	56.38	0.22
CVAP registration	63.83	65.71	68.98	0.01
Lagged D ratio	1.30	1.29	1.28	0.54
Republican ratio				
Party system strength	2.00	2.12	1.92	0.00
Party endorsement	3.00	2.97	2.33	0.08
Southern state	0.00	0.16	0.25	0.00
Folded Ranney Index	0.93	0.86	0.86	0.00
Regular Ranney Index	0.43	0.64	0.61	0.00
CVAP turnout	52.77	52.82	56.38	0.04
CVAP registration	63.83	63.80	68.98	0.05
Lagged R ratio	1.31	1.31	1.40	0.83
	North Carolina	**Synthetic North Carolina**	**Pool of controls**	v_m
Democratic ratio				
Party system strength	1.00	2.44	1.92	0.01
Party endorsement	3.00	2.99	2.33	0.03
Southern state	1.00	0.98	0.25	0.29
Folded Ranney Index	0.78	0.78	0.85	0.06
Regular Ranney Index	0.72	0.72	0.62	0.03
CVAP turnout	50.11	53.88	57.27	0.03
CVAP registration	64.32	67.97	69.28	0.00
Lagged D ratio	1.41	1.40	1.25	0.55
Republican ratio				
Party system strength	1.00	2.68	1.92	0.00
Party endorsement	3.00	2.99	2.33	0.16
Southern state	1.00	0.99	0.25	0.10
Folded Ranney Index	0.78	0.77	0.85	0.04
Regular Ranney Index	0.72	0.73	0.62	0.03
CVAP turnout	50.11	54.16	57.27	0.01
CVAP registration	64.32	68.50	69.28	0.00
Lagged R ratio	1.00	1.02	1.37	0.66

Table 6. State weights for synthetic control.

State	Arizona D ratio weight	Arizona R ratio weight	North Carolina D ratio weight	North Carolina R ratio weight
Connecticut	0.00	0.00	0.00	0.00
Delaware	0.00	0.03	0.00	0.00
Florida	0.00	0.00	0.54	0.31
Kentucky	0.00	0.16	0.19	0.21
Maryland	0.00	0.00	0.00	0.00
Nevada	0.00	0.59	0.01	0.00
New Mexico	0.00	0.00	0.00	0.00
New York	0.00	0.00	0.00	0.00
Oklahoma	0.00	0.00	0.25	0.48
Oregon	0.00	0.22	0.00	0.00
Pennsylvania	0.91	0.00	0.00	0.00
South Dakota	0.09	0.00	0.00	0.00

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

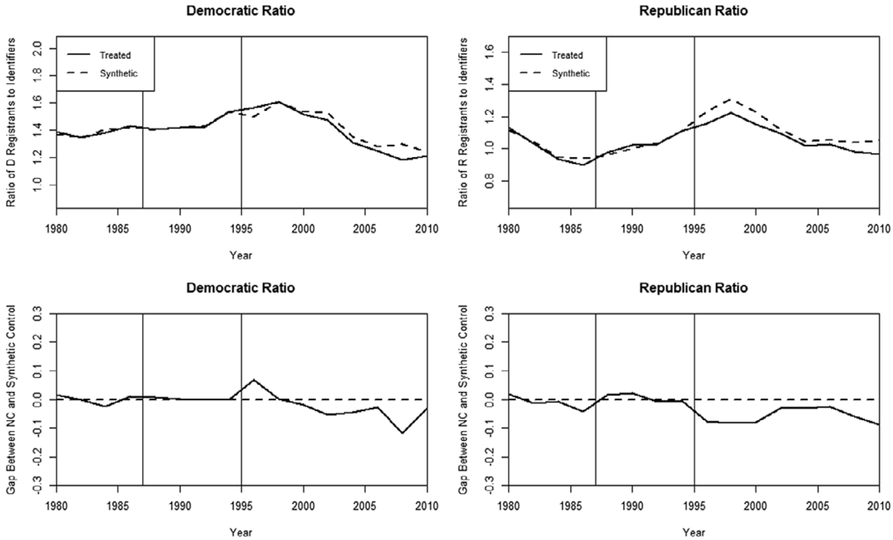


Figure 4. Comparison of North Carolina and synthetic controls.

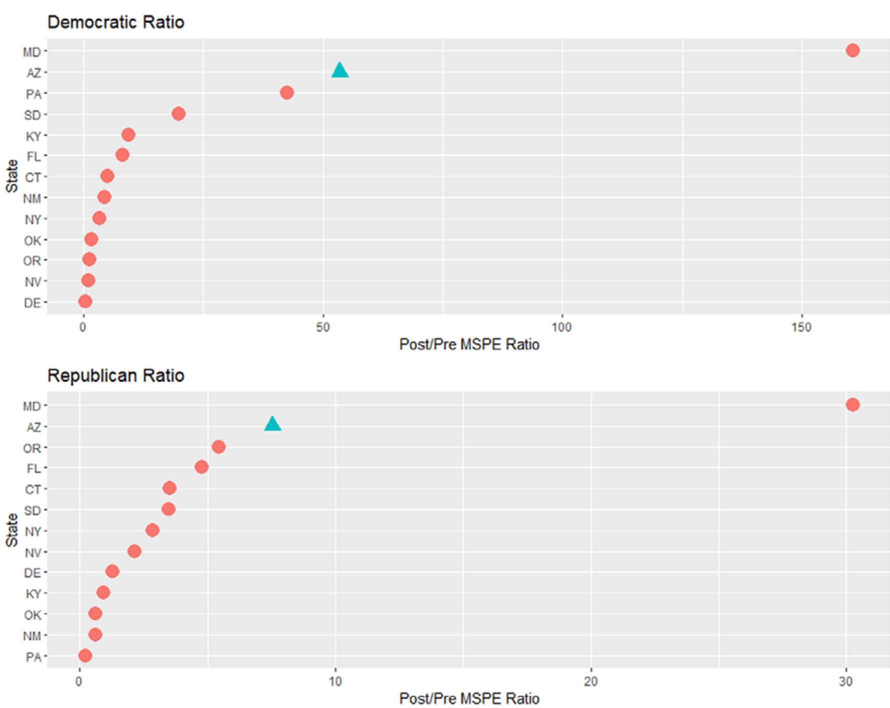


Figure 5. Post-/pre-treatment MSPE for Arizona.

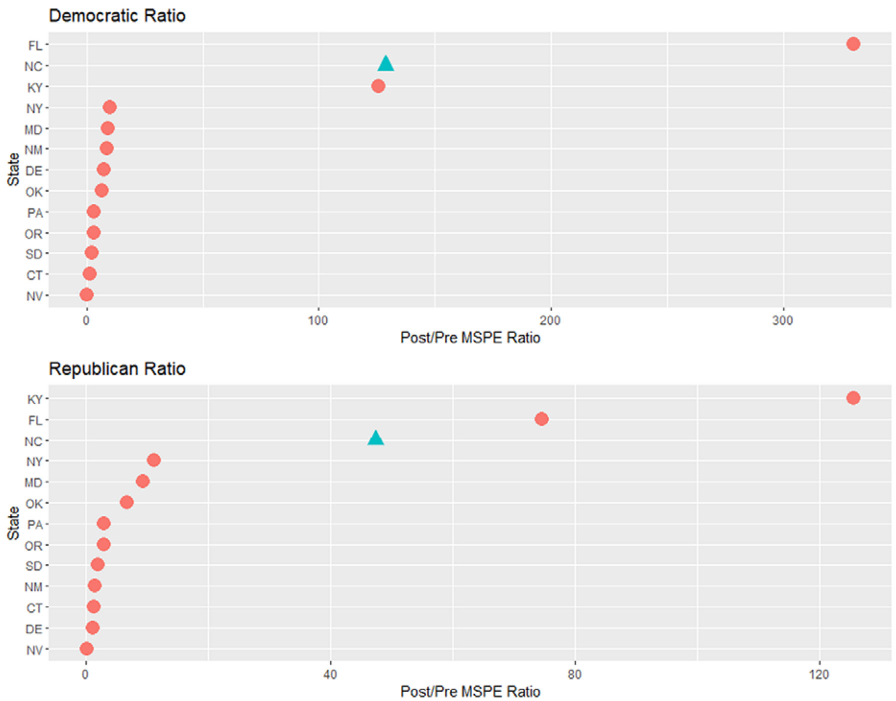


Figure 6. Post-/pre-treatment MSPE for North Carolina.

for the Democratic ratio pre-implementation of semi-closed primaries. This indicates strong matching to the synthetic control prior to 2000 and a major divergence from the control afterward. In contrast, the Republican ratio is smaller, indicating a much less clear treatment effect and divergence from the synthetic control after unaffiliated voters were able to vote in the semi-closed primaries of the state.

As expected, the picture is less clear for North Carolina in Figure 4. While the Democratic and Republican registrant-to-identifier ratios decreased after the opening of primaries to unaffiliated voters, the magnitude of the change is much smaller than in Arizona. In other words, after both Democrats and Republicans opened their primaries to unaffiliated voters, fewer individuals registered as Democrats (Republicans) relative to Democratic (Republican) identification in the state. Figure 6 shows that North Carolina Democratic and Republican post-/pre-treatment MSPE ratios are both high but less so than Florida and Kentucky.

Overall, the analysis of aggregate party registration data from Arizona and North Carolina supports the social pressure, maximizing options, and impact voting theories of hidden partisanship. Implementation of semi-closed primaries in both states clearly leads to a large increase in the proportion of the electorate that chooses to register as unaffiliated, as shown in Figures 1 and 2. The large post-/pre-treatment MSPE ratio for both states indicates a good fit between the synthetic control and treated states prior to the treatment followed by a major divergence (in the expected direction) after the unaffiliated are able to vote in both Democratic and Republican primaries. In addition, in the clearly Republican state of Arizona, there is strong

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

evidence that the ratio of Democratic registrants to identifiers decreases more relative to the ratio of Republican registrants to identifiers after the implementation of semi-closed primaries in 2000. The latter finding is in keeping with an impact voting motive for registering as unaffiliated in semi-closed primary states. At this point in my analysis, I also cannot rule out social pressure as a motive either.

Patterns of hidden partisanship in semi-closed primary states

In states with semi-closed primary elections, voters are more likely to be unaffiliated with a political party compared to closed primary states. A comparison of Arizona and North Carolina's implementation of semi-closed primaries to synthetic counterfactuals shows this causal relationship. In comparison with the synthetic control, the treated states' introduction of semi-closed primaries significantly increased the proportion of unaffiliated voters in the state. To many voters, unaffiliated registration, which promises the ability to access the other party's primary as well as one's own, is a more attractive prospect than registration with a political party. Evidence also points to this sort of hidden partisanship being most prevalent among identifiers with the weaker party in the electorate where one party dominates, consistent with impact voting or social pressure theories of hidden partisanship.

The theory of this paper postulates that voters choose to remain unaffiliated in semi-closed primaries to gain access to the strong party's primary elections (i.e., impact voting hidden partisanship). By itself, the fact that semi-closed primaries lead to a greater number of unaffiliated voters does not necessarily show this. For example, semi-closed rules might instead allow the growing number of independents to express their true identity as unaffiliated rather than be required to register with a party in order to vote in primary elections. Evaluating the impact of voting explanation for semi-closed hidden partisanship requires analysis of individual-level information, such as a voter's party identification. At this point in the analysis, I cannot yet rule out observationally equivalent social pressure explanations.

Aside from Key (1949), previous studies of hidden partisanship in the literature (Arrington and Grofman 1999) utilize ecological inference of aggregate registration and vote shares. However, large-N datasets exist that measure the relevant variables among individual voters and contain sufficient statistical power to examine patterns of party registration at sub-national levels. Given the well-documented issues with the ecological fallacy, I directly test the formal model of party registration at the individual level.

I hypothesize that in keeping with the instrumental motivation for engaging in hidden partisanship, clear patterns will be evident in its occurrence. First, building on the results from the previous section, I predict that the probability of hidden partisanship increases as a voter's party grows less competitive within their state. Registering with the opposite party (in closed primary states) or remaining unaffiliated (in semi-closed primary states) which grants access to the majority party's primary elections become increasingly attractive options in such states. I also predict that hidden partisanship is more common in semi-closed primary states compared to closed primary states and remaining unaffiliated in semi-closed states is the most common form of hidden partisanship. The greater instrumental *and* psychic benefit from remaining unaffiliated compared to registering with the other major party make this form of hidden partisanship the most attractive.

To evaluate these hypotheses it is necessary to use an individual-level dataset measuring both a voter's party identification and party registration. My theory predicts that hidden partisanship will be most common among the rarest voters: its prevalence increases as a voter's party shrinks in the electorate. Thus, an adequate test of the theory requires a sufficient sample of identifiers with the electorally weak party in a state. I utilize the 2018 CCES. The CCES survey uses a matched random sample of members from an opt-in panel managed by YouGov Polimetrix. It is administered in two waves; the first takes place in September of the election year with a post-election wave occurring in November (Schaffner and Ansolabehere 2018). The study measures a variety of political attitudes and demographic characteristics in a sample that regularly exceeds 50,000 respondents. It also measures party registration. The 2018 survey validates party registration using voter files from the Catalist data service.

In this analysis, party registration is the dependent variable. I measure it as a nominal variable taking on three values: unaffiliated/independent, Democratic, and Republican. The option of voters to affiliate with third-parties and which parties receive recognition is idiosyncratic to individual states and only 1.4% of the 2018 sample in party registration states registered with a third-party. Thus, I exclude these individuals from the analysis. I exclude individuals registered as "independent" if they were registered with the Independent Party but not if this was the state's signifier of unaffiliated status.

Table 7 shows weighted crosstabs for the 2018 CCES among Democrats, Republicans and independents (including leaners with partisans). I include both percentages and raw numbers in parentheses. I distinguish between closed and semi-closed states and the partisanship of the state they reside in. To estimate the latter, I coded states as safe Republican if greater than 55% of the partisan identifiers in the state were Republican and safe Democratic if greater than 55% of partisan identifiers were Democrats with other states labeled "competitive." These categorizations are found in the rightmost column of Table 2. I determined the partisan composition of these states via MRP on the 2018 CCES sample (Gelman and Little 1997). The hierarchical models in the MRP procedure estimate the probability of respondents identifying or leaning Democratic or Republican based on individual characteristics (race, gender, age, political interest, and education). The intercepts of these models vary by state through random effects. Following the convention of Hill (2015), I use the CCES poststratification weights in the MRP procedure. Details of the hierarchical models and estimates of state partisanship are available in the Supplementary Material.

I include leaners with partisans because of the documented effect of party registration on an individual's party identification (Burden and Greene 2000; Finkel and Scarrow 1985; Gerber, Huber, and Washington 2009; Thornburg 2014). Individuals registered as unaffiliated may identify as independent because of their party registration. Thus, excluding independent leaners from partisans provides an inaccurate estimate of hidden partisanship because some individuals may only identify as independent because of their choice to remain unaffiliated.

Table 7 confirms that the most common form of hidden partisanship among partisans is among Republicans in semi-closed safe Democratic states and among Democrats in semi-closed safe Republican states. Among self-identified Democrats living in semi-closed, safe Republican states, only about 60% of active registrants are actually registered with the Democratic Party. Among self-identified Republicans in semi-closed safe Democratic states, 55% of active registrants are registered with the

Table 7. Hidden partisanship among active registrants, 2018 CCES

	Safe R states		Competitive		Safe D states	
Semi-closed						
Self-identified Democrats						
Registered unaffiliated	28.66%	(80)	23.77%	(383)	24.36%	(253)
Registered Democratic	60.78%	(170)	73.51%	(1,184)	74.26%	(770)
Registered Republican	10.56%	(30)	2.72%	(44)	1.38%	(14)
Self-identified Republicans						
Registered unaffiliated	13.78%	(62)	23.88%	(344)	38.22%	(199)
Registered Democratic	2.42%	(11)	3.48%	(50)	6.71%	(35)
Registered Republican	83.80%	(378)	72.64%	(1,047)	55.07%	(287)
Self-identified independents						
Registered unaffiliated	54.36%	(63)	67.62%	(309)	72.10%	(195)
Registered Democratic	23.14%	(27)	10.78%	(49)	17.54%	(47)
Registered Republican	22.50%	(26)	21.60%	(99)	10.36%	(28)
Closed						
Self-identified Democrats						
Registered unaffiliated	8.48%	(43)	13.64%	(323)	13.81%	(356)
Registered Democratic	85.57%	(434)	82.46%	(1,950)	83.90%	(2,162)
Registered Republican	5.95%	(30)	3.90%	(92)	2.29%	(59)
Self-identified Republicans						
Registered unaffiliated	6.54%	(48)	10.31%	(213)	17.16%	(235)
Registered Democratic	11.21%	(82)	5.10%	(106)	7.70%	(106)
Registered Republican	82.25%	(599)	84.59%	(1,750)	75.14%	(1,030)
Self-identified independents						
Registered unaffiliated	45.18%	(59)	54.19%	(274)	56.09%	(307)
Registered Democratic	22.53%	(29)	21.38%	(108)	27.08%	(148)
Registered Republican	32.28%	(42)	24.43%	(123)	16.83%	(92)

Percentages are % of party identification group with indicated party registration.

Republican Party. Hidden partisanship is less prevalent in semi-closed states where a voter's party is strong. The effect of semi-closed primaries on unaffiliated registration is conditional on partisanship and political competition.

While the percentage of unaffiliated self-identified Democrats does not increase much moving from strongly Democratic semi-closed states to strongly Republican ones, the percentage of these voters registered with the Republican Party increases significantly. It is important to note that all of the semi-closed strongly Republican states (Idaho, Kansas, and Utah) change party registration of unaffiliated primary voters to registration with the party whose primary they voted in (voters are free to switch back later). It is thus possible that many of the Democrats registered with the Republican Party in these states recently participated in the GOP primary and have not yet switched back to unaffiliated.

It is hard to discern patterns of hidden partisanship in closed primary states. Even in the most Democratic closed primary states, at most 10% of Republicans register with the Democratic Party; likewise Democrats in Republican states.

I model the party registration of Democrats, Republicans, and independents in closed and semi-closed primary states as well. I perform this analysis separately on Democrats and Republicans (leaners included) as well as "pure" independents and separately for all three groups in semi-closed and closed primary states.

I control for whether the respondent self-identified as Black or Hispanic, the respondent's college education, their gender, strong partisanship (where applicable), high interest in news and politics, and their age divided by 100. Because some states transitioned to semi-closed primaries relatively recently, I include in the semi-closed

Table 8. Party registration for active registrants in semi-closed states, 2018 CCES

	Democrats		Republicans		Independents	
Variable	Unaffiliated	Republican	Unaffiliated	Democratic	Democratic	Republican
Instrumental utility	4.266*** (0.504)		3.185*** (0.659)		−0.899 (1.685)	
Years with semi-closed primary	1.399***	−0.131	1.458**	−0.240	−0.286	−1.849***
	(0.228)	(0.967)	(0.482)	(0.692)	(0.576)	(0.514)
Strong partisan	−1.617***	−1.199***	−1.974***	−1.415***	–	–
	(0.177)	(0.225)	(0.169)	(0.129)	–	–
Strong interest in news and politics	0.045	−0.795*	−0.274*	−0.507**	−0.201	0.240
	(0.162)	(0.318)	(0.123)	(0.180)	(0.385)	(0.302)
Black	−0.156	−0.749	−0.616	1.787*	1.282**	0.023
	(0.279)	(1.111)	(1.058)	(0.836)	(0.399)	(0.900)
Hispanic	0.045	−0.088	−0.396	−0.261	1.538**	0.173
	(0.201)	(0.361)	(0.546)	(0.751)	(0.579)	(0.635)
Age/100	−1.237**	0.845	−1.237***	0.651	0.382	0.687
	(0.472)	(0.702)	(0.354)	(0.486)	(1.121)	(0.811)
Female	0.038	−0.370	−0.199	0.087	−0.203	−0.142
	(0.070)	(0.266)	(0.127)	(0.378)	(0.316)	(0.293)
College graduate	0.140	0.737**	−0.580***	−1.049***	0.034	0.329
	(0.117)	(0.274)	(0.059)	(0.170)	(0.247)	(0.314)
(Constant)	−2.409***	−2.492***	−1.319***	−2.150***	−2.239*	−1.617*
	(0.402)	(0.501)	(0.319)	(0.538)	(1.024)	(0.641)
Log likelihood	−1,706.10		−1,459.72		−643.18	
Number of observations	2,927		2,414		843	

***$p < 0.001$.
**$p < 0.01$.
*$p < 0.05$.
†$p < 0.1$.

models the number of years the state has been semi-closed divided by 100. Finally, my primary independent variable of interest is p_j, instrumental utility. This variable is alternative specific, taking on a value of 0 or 1, respectively, for the unaffiliated alternative in closed and semi-closed primary states. For the Democratic and Republican alternatives, the variable equals the proportion of the state's partisans that identified or leaned with the respective party. I once again estimate the proportion of Democratic and Republican supporters in each state using MRP. Because the p_D gives the proportion of Democrats and Republicans identifying with the Democratic Party, $p_R = 1 - p_D$.

The multinomial logit model includes robust standard errors clustered on state. The reference category is registration with a voter's own party and the comparison groups are unaffiliated and registration with the opposite party. The estimates for Democrats, Republicans, and independents in semi-closed states are shown in Table 8 and for closed states in Table 9. Figure 7 plots the predicted probabilities of party registration for all six models as p_D changes.

Our chief concern is the variables coding for instrumental and psychic utility. The instrumental utility variable is statistically and substantively significant in the Democratic and Republican semi-closed models. As Figure 7 makes clear, the probability of hidden partisanship is highest among Republicans and Democrats in semi-closed

Table 9. Party registration for active registrants in closed states, 2018 CCES

	Democrats		Republicans		Independents	
Variable	Unaffiliated	Republican	Unaffiliated	Democratic	Democratic	Republican
Instrumental utility	0.686		0.581		1.593***	
	(0.765)		(0.953)		(0.421)	
Strong partisan	−1.701***	−1.101***	−1.864***	−0.981***	–	–
	(0.150)	(0.075)	(0.161)	(0.127)	–	–
Strong interest in news and politics	−0.156	−0.167	−0.345y	−0.488*	0.197	0.372
	(0.114)	(0.102)	(0.176)	(0.231)	(0.276)	(0.268)
Black	−0.762***	−1.453**	0.850**	1.886**	1.017***	−1.085y
	(0.147)	(0.525)	(0.275)	(0.616)	(0.272)	(0.581)
Hispanic	0.104	−0.202	0.839**	0.698	−0.128	−0.211
	(0.272)	(0.172)	(0.268)	(0.462)	(0.146)	(0.571)
Age/100	−2.029***	−0.390	−1.667***	−0.817†	1.243†	2.227*
	(0.241)	(0.843)	(0.315)	(0.456)	(0.721)	(0.969)
Female	−0.198*	−0.114	−0.308***	−0.050	0.472*	0.090
	(0.082)	(0.152)	(0.060)	(0.172)	(0.226)	(0.215)
College graduate	−0.048	−0.157	−0.363**	−0.375	0.074	0.115
	(0.128)	(0.143)	(0.115)	(0.252)	(0.131)	(0.193)
(Constant)	0.552	−1.960***	0.317	−1.253**	−2.740***	−2.896***
	(0.530)	(0.482)	(0.641)	(0.394)	(0.330)	(0.604)
Log likelihood	−2,546.71		−2,329.43		−988.05	
Number of observations	5,449		4,168		1,183	

$^{***}p < 0.001$.
$^{**}p < 0.01$.
$^{*}p < 0.05$.
$^{\dagger}p < 0.1$.

primary states, when a voter's party is electorally weak in a state. This supports the theory of impact voting hidden partisanship. However, it does not yet discount social pressure as an explanation either.

The plots also confirm that hidden partisanship is most frequently observed in semi-closed primary states with individuals remaining unaffiliated. Strong partisans are also less likely to engage in hidden partisanship compared to weak partisans or independent leaners.

Semi-closed primaries and crossover voting

Unaffiliated registration in semi-closed states is a popular choice among partisans who live where their own party is uncompetitive. This behavior accords with impact voting hidden partisanship: a desire to engage in impact crossover voting by participating in the primary of the party whose candidates are most likely to win the general election (Alvarez and Nagler 2002). Do these individuals use semi-closed primary laws to engage in crossover voting? While unaffiliated registration in semi-closed states provides greater instrumental utility than any other option where party registration exists, there are other reasons why a voter might remain unaffiliated in a semi-closed state. Voters might wish to avoid campaign contact from political parties but continue voting in primaries. Or they might be concerned that registration with the electorally weak party in a state will carry social or professional consequences

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

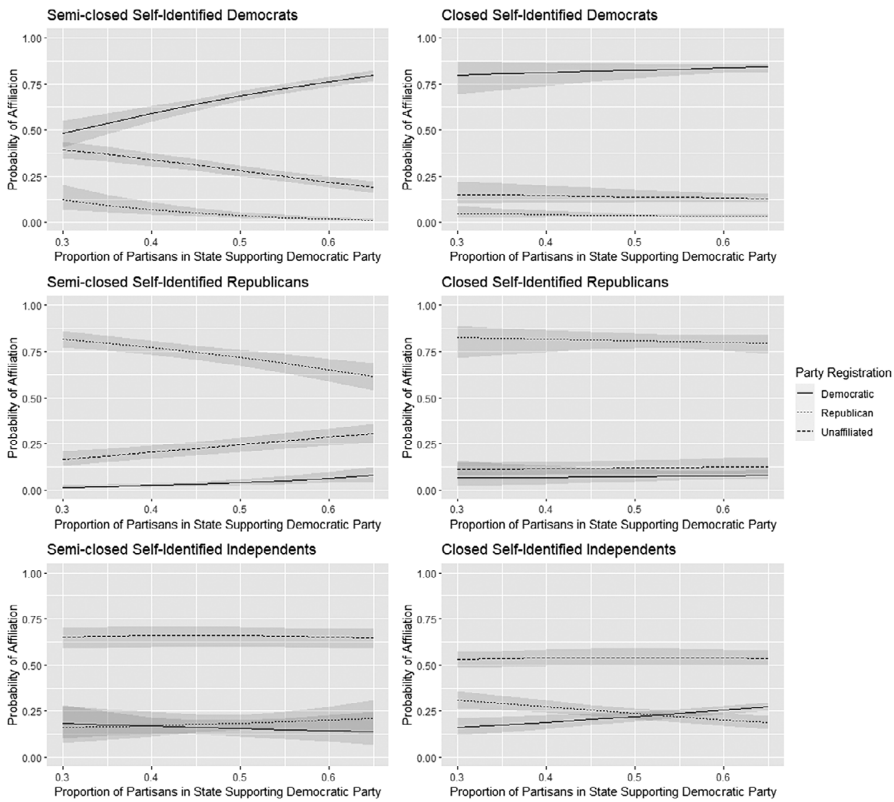


Figure 7. Party registration among actively registered voters in semi-closed and closed states.

(i.e., social pressure hidden partisanship) but wish to still preserve access to primary elections.

The only way to determine whether unaffiliated hidden partisanship in semi-closed states is impact voting rather than due to social pressure is to measure primary participation. I examine the party of the primary that voters choose and evaluate whether it is consistent with impact voting. In-depth analysis of crossover voting, such as comparison to rates among closed primary states and changes in the composition of primary electorates, is beyond the scope of this paper. I simply test whether unaffiliated voters in semi-closed states choose the party primary offering greater instrumental utility and whether unaffiliated partisans are willing to engage in crossover voting to do so. I predict that in blue states, a greater proportion of unaffiliated voters will vote in the Democratic primary, irrespective of their own party, compared with red states where the Republican primary will be most attractive.

I again use the 2018 CCES, this time to measure primary turnout and party of the primary. The 2018 CCES measures primary party turnout in two ways. The survey asks for self-reported party of primary voted. The CCES also includes voter file data (in the states where it is available) validating the party of the primary the voter participated in. An additional voter file validation is conducted for all semi-closed states of whether the voter participated in the primary (but does not report the party).

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

I construct two measures of primary turnout from these data. My self-reported measure examines individuals in semi-closed primary states who self-reported voting in the Democratic or Republican primary in response to the survey and have validated turnout. The validated measure uses the recorded party voted from the voter file. Two of the semi-closed primary states (North Carolina and Utah) did not have statewide Democratic and Republican primaries in 2018, meaning not all voters in these states had a choice between the parties. These states are excluded from analysis.

Both measures of party primary turnout have strengths and weaknesses. The self-reported measure includes respondents in all states and was asked of all participants. However, given the norms that exist against crossover voting (Gerber *et al.* 2017), it may have reliability problems and understate crossover voting. The validated measure avoids issues with self-reported voting but four semi-closed states do not record primary party in the voter file, leaving analysis of just six semi-closed states. For the sake of thoroughness, both measures are reported here.

Table 10 reports weighted self-reported and validated party of primary among those unaffiliated voters in semi-closed states who participated in the 2018 primary elections. I include percentages as well as raw numbers in parentheses. Voters are divided into "Safe Republican," "Competitive," and "Safe Democratic" states as in the previous section. We observe patterns of partisan and independent crossover voting consistent with impact voting: the share of voters voting in the Democratic primary is higher among all groups in Democratic states compared to Republican states. This pattern holds for validated party as well, although no safe Republican states are available with the validated party primary turnout measure.

I model the decision to vote in the Democratic or Republican primary among semi-closed unaffiliated primary participants using logistic regression models including robust standard errors clustered at the state level. The models include

Table 10. Party of primary voted among unaffiliated voters, 2018 CCES

	Safe R states		Competitive		Safe D states	
Self-reported primary turnout						
Self-identified Democrats						
Democratic primary	90.31%	(10)	87.87%	(85)	97.56%	(90)
Republican primary	9.69%	(1)	12.13%	(12)	2.44%	(2)
Self-identified Republicans						
Democratic primary	0.00%	(0)	2.19%	(2)	19.29%	(12)
Republican primary	100.00%	(2)	97.81%	(67)	80.71%	(49)
Self-identified independents						
Democratic primary	No Obs.		50.26%	(34)	55.26%	(32)
Republican primary	No Obs.		49.74%	(33)	44.74%	(26)
Validated primary turnout						
Self-identified Democrats						
Democratic primary	No Obs.		77.50%	(38)	97.75%	(90)
Republican primary	No Obs.		22.50%	(11)	2.25%	(2)
Self-identified Republicans						
Democratic primary	No Obs.		6.51%	(3)	16.53%	(10)
Republican primary	No Obs.		93.49%	(44)	83.47%	(50)
Self-identified independents						
Democratic primary	No Obs.		53.13%	(14)	50.07%	(28)
Republican primary	No Obs.		46.87%	(12)	49.93%	(28)

Percentages are % of party identification group voting in indicated primary.

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

Table 11. Democratic primary voting among unaffiliated semi-closed voters, 2018 CCES

Variable	Self-reported	Validated
Democratic partisanship of state	0.5541	−0.852
	(2.400)	(3.338)
Democrat	−1.503	−10.298***
	(2.456)	(0.980)
Republican	−14.364**	−7.304***
	(5.410)	(1.823)
Democrat × state partisanship	4.724	19.446**
	(4.495)	(2.146)
Republican × state partisanship	20.437*	9.626**
	(8.202)	(3.050)
Strong interest in news and politics	0.033	0.151
	(0.399)	(0.546)
Ideological distance from Democratic Party	−0.333***	−0.371***
	(0.075)	(0.079)
Ideological distance from Republican Party	0.365***	0.296*
	(0.082)	(0.118)
Black	3.203***	1.507***
	(0.743)	(0.330)
Hispanic	0.402	1.060
	(1.315)	(1.734)
Age/100	−1.272	−0.857
	(0.821)	(1.015)
Female	−0.104	0.165
	(0.392)	(0.353)
College graduate	0.794†	0.284
	(0.447)	(0.288)
(Constant)	0.222	1.008
	(1.324)	(2.543)
Log likelihood	−135.57	−111.80
Number of observations	457	332

***$p < 0.001$.
**$p < 0.01$.
*$p < 0.05$.
†$p < 0.1$.

unaffiliated voters in semi-closed primary states who voted in the Democratic or Republican primaries in 2018 according to the self-reported and validated party of primary measures. My primary variable of interest is the proportion of the two-party identifier share in the state that is Democratic. I predict that an increase in this share and a corresponding increase in the Democratic partisanship of the state will be associated with greater levels of voting in the Democratic primary among all unaffiliated voters in semi-closed states. I include dummy variables for Democratic or Republican identification (or leaning) and interaction terms for the Democratic partisanship share with Democratic and Republican identification. As before, I also control for whether the voter is Black or Hispanic, their age, whether they are female, college education, self-reported ideological distance from the Democratic and Republican parties, and high interest in news and politics. Table 11 shows the model estimates.

Figure 8 plots the probability of voting in the 2018 Democratic primary among unaffiliated members of all three groups of voters located in semi-closed primary states. For both self-reported and validated measures of Democrats and Republicans, increasing Democratic partisanship of the state leads voters to vote in the Democratic

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

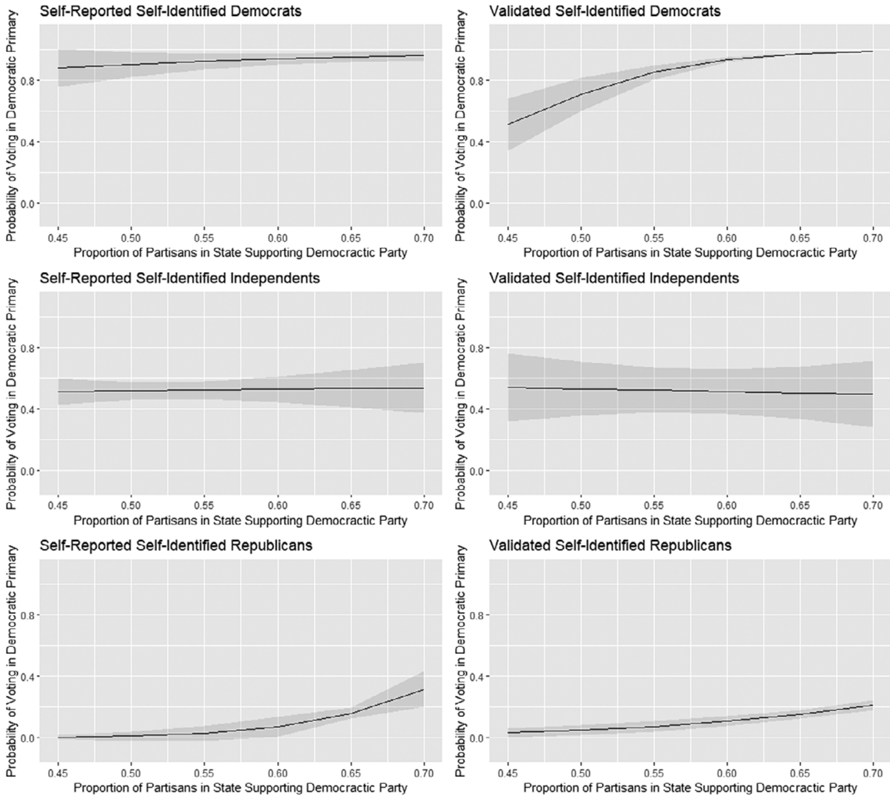


Figure 8. Participation in Democratic primary among validated unaffiliated primary voters in semi-closed states.

primary. Among unaffiliated partisans in semi-closed primary states, this means the probability of engaging in primary crossover voting is high when residing in a state where the voter identifies with the weak party (a blue state for Republicans and a red state for Democrats).

Unaffiliated voters in semi-closed states gravitate toward the party primary that, all things equal, is most likely to yield the general election winners. This is consistent with the impact voting observed by Alvarez and Nagler (2002) as well as the evidence from Key (1949) and Arrington and Grofman (1999). This supports the impact voting theory of hidden partisanship. While low levels of primary crossover voting are generally reported nationwide, in this particular circumstance, significant numbers of unaffiliated partisans in the most politically unfriendly states cross over into the other party's primary.

Discussion

The hidden partisanship Key (1949) and Arrington and Grofman (1999) observed reflects a desire for voters to maximize the instrumental utility of their primary vote. In states where a voter's own party is uncompetitive, the general election result may

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

be a foregone loss for that party. There is little instrumental value from nominating candidates who are certain to lose the general election. Therefore in states with party registration, registering with a voter's own party where it is uncompetitive provides little utility in affecting who eventually comes to represent a voter. In a closed primary state, affiliating with the stronger party provides the greatest amount of instrumental utility to the voter.

However, semi-closed primaries offer a better option for the voter whose party is not competitive. Voters registering as unaffiliated access both party primaries, enabling them to select the dominant party's nominees and presumptive representatives. The option remains to vote in one's own party primaries as well. In essence, unaffiliated registration in a semi-closed primary state transforms the election into an open primary.

The attractiveness of the unaffiliated option leads to larger numbers of registrants remaining unaffiliated for this primary access. The increase in unaffiliated registration after Arizona and North Carolina implemented semi-closed primaries has been large in those states, even during a time period when unaffiliated registration has been increasing among all states (McGhee and Krimm 2009). While large numbers of unaffiliated voters drive the decision to institute semi-closed primaries in some cases, it is also clear that the rules change affects the behavior of voters as well.

The attractiveness of unaffiliated registration varies across the electorate in semi-closed states. While all voters are more likely to register as unaffiliated in semi-closed states than in closed states, partisans' willingness to do so depends on the political conditions where they live. Democrats and Republicans (and independent leaners) register as unaffiliated in semi-closed states where their own party is electorally weak. This difference from closed primary states shows that the option to vote in either party primary afforded to unaffiliated voters is more attractive to individuals who expect their own party's nominees to lose in the general election. Semi-closed primaries, therefore, facilitate impact crossover voting. These patterns are confirmed using CCES survey data and partially confirmed with aggregate registration totals after the opening of Republican Arizona's primaries to unaffiliated voters.

Weaver (2015), examining North Carolina after the institution of semi-closed primaries, shows that crossover voting takes place where a voter's own party is electorally weak. I also show that among unaffiliated voters in semi-closed states, state partisanship affects the decision of whether or not to cross over. In Democratic states, unaffiliated voters participate in the Democratic primary; likewise with Republican states. This includes a minority of unaffiliated Democrats in red states and Republicans in blue states who engage in crossover voting.

These findings inform a longstanding debate on how primary election rules affect the composition and representativeness of primary electorates (Gerber and Morton 1998; Kanthak and Morton 2001; McGhee *et al.* 2014; Norrander and Wendland 2016). However, the question of whether hidden partisanship in semi-closed primaries leads to changes in the composition of primary electorates is beyond the scope of this paper. It is possible that the effect of this phenomenon will be inherently limited. Because impact voting hidden partisanship and primary crossover voting are only attractive where one party is weak and the other is strong, the number of partisans identifying with the weak party and crossing over will by definition be limited in number. Future research should explore the effect of primary crossover voting on the composition of electorates in semi-closed and other forms of primary.

https://doi.org/10.1017/spq.2022.26 Published online by Cambridge University Press

These findings are of interest to any scholar who uses party registration as a proxy for party identification. Because voter files provide large, easily obtainable datasets including demographic information and geographic location of individuals, they are increasingly utilized in political behavior research. In particular, much of the research dealing with geographic sorting has used party registration data to demonstrate geographic clustering of like-minded partisans (Carlson and Gimpel 2019; Martin and Webster 2020; Sussell 2013). The fact voters strategically engage in hidden partisanship where their own party is uncompetitive may lead analysis of party registration data to overstate geographic sorting.

The findings in this paper strongly suggest instrumental and strategic behavior regarding primary elections. It is nonetheless easy to overstate the case. The synthetic control analysis of North Carolina shows that unaffiliated registration increased after the state's parties opened their primaries to unaffiliated voters. However, it does not clearly show the ratio of Democratic nor Republican registrants to identifiers declined after this relative to the synthetic control. The North Carolina case is complicated by the fact the state gradually opened its primaries to the unaffiliated, the segmented partisanship of the state, as well as the realignment the state underwent. Nonetheless, these results should be interpreted with caution.

The patterns I observe suggest impact voting but do not necessarily rule out maximizing options either. Key (1949) famously noted that in the one-party South, where the general election was perfunctory, competition had shifted to the Democratic primary and the latter was "in reality the election" (p. 407). Key's observations and common sense suggest that where interparty competition is low, intraparty competition may be high. Therefore, voters seeking competitive primaries to cross over into may also find them where their own party is weak. For the present study, I simply do not have enough information about primary competitiveness up and down the ballot in 2018 to control for this factor. Even if it were so, instrumental hidden partisanship based on maximizing options is also an interesting finding and warrants further study in the future.

Also, while clear patterns are evident in crossover voting and hidden partisanship, the majority of partisans in semi-closed states register with their own political party, even where it makes instrumental sense to remain unaffiliated (i.e., where one's party is electorally weak). And even among unaffiliated partisans in semi-closed states, crossover voting is still not the norm. Instrumental party registration in closed primary states is also much less common than Key (1949) or Arrington and Grofman (1999) found. It is possible that the lower levels of closed primary hidden partisanship observed here compared to previous studies may be due to the greater political polarization that now exists between the parties. Research on the rise of affective partisanship finds that party identification now has an emotional component to it rather than just a policy one (Iyengar *et al.* 2019).

Instrumental hidden partisanship is consistent with impact crossover voting–casting a ballot for the most preferred candidate of the other party and thus having an effect on who comes to hold office. Rather than creating mischief or attempting to sabotage the other party, this form of crossover voting involves a serious consideration of which candidate seeking the other party's nomination is most attractive to the voter. As American voters increasingly exhibit negative feelings for candidates of the other party, they may eventually stop crossing over during the primary. The present study and others (Gerber *et al.* 2017) show that strong partisans are less likely to engage in hidden partisanship and/or crossover voting and thus the strengthening of party identification in America may dampen hidden partisanship. On the other hand, waning levels of local interparty competition across the country (Drutman 2020) place a growing number of Americans in a position where impact crossover voting may be their best chance to determine their elected representatives.

Overall, the results presented in this paper show the responsiveness of voters to electoral institutions in an instrumental manner. Where a registration option provides greater instrumental utility in selecting the next officeholder, many voters respond by selecting this option. Party registration is an unusual electoral feature as it is an official government record of an informal attitude. Registrants must state their political preferences honestly for the restriction to work as intended. If voters do not state such preferences, as shown here, then semi-closed election laws are limited in effectiveness.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/spq.2022.26.

Data availability statement. Replication materials are available on SPPQ Dataverse at https://doi.org/10.15139/S3/M3NXWH (Thornburg 2022).

Acknowledgments. I thank the editors and anonymous reviewers for their helpful comments. I also thank Mike P. McDonald for helping to develop this idea, Bernie Grofman for helpful comments, Eric McGhee for aggregate party registration data, and Brandy Horne for library assistance. All errors are my own.

Funding statement. The author received no financial support for the research, authorship, and/or publication of this article.

Conflict of interest. The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2015. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59 (2): 495–510.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–32.

Abdelal, Rawi, Yoshiko M. Herrera, Alastair Iain Johnston, and Rose McDermott. 2006. "Identity as a variable." *Perspectives on Politics* 4 (4): 695–711.

Alvarez, R. Michael, and Jonathan Nagler. 1997. "*Analysis of Crossover and Strategic Voting*." Technical report. Pasadena CA: California Institute of Technology.

Alvarez, R. Michael, and Jonathan Nagler. 2002. "Should I Stay or Should I Go?" In *Voting at the Political Fault Line: California's Experiment with the Blanket Primary*, eds. Bruce Cain and Elisabeth Gerber 107–123. Oakland CA: University of California Press.

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate." *Political Analysis* 20 (4): 437–59.

Arrington, Theodore S., and Bernard Grofman. 1999. "Party Registration Choices as a Function of the Geographic Distribution of Partisanship: A Model of 'Hidden Partisanship' and an Illustrative Test." *Political Geography* 18 (2): 173–85.

Bell, Charles G., and William Buchanan. 1966. "Reliable and Unreliable Respondents: Party Registration and Prestige Pressure." *Western Political Quarterly* 19 (1): 37–43.

Burden, Barry C., and Steven Greene. 2000. "Party Attachments and State Election Laws." *Political Research Quarterly* 53 (1): 63–76.

Burden, Barry C., and Philip Edward Jones. 2009. "Strategic Voting in the US." In *Duverger's Law of Plurality Voting*, eds. Shaun Bowler, André Blais, and Bernie Grofman 47–64. New York, NY: Springer.

Carlson, Caroline, and James G. Gimpel. 2019. "Political Implications of Residential Mobility and Stasis on the Partisan Balance of Locales." *Political Geography* 71: 103–14.

Drutman, Lee. 2020. *Breaking the Two-Party Doom Loop: The Case for Multiparty Democracy in America*. Oxford, UK: Oxford University Press.

Enns, Peter K., and Julianna Koch. 2013. "Public Opinion in the US States: 1956 to 2010." *State Politics & Policy Quarterly* 13 (3): 349–72.

Ezra, Marni. 2001. "A Reexamination of Congressional Primary Turnout." *American Politics Research* 29 (1): 47–64.

Finkel, Steven E., and Howard A. Scarrow. 1985. "Party Identification and Party Enrollment: The Difference and the Consequences." *The Journal of Politics* 47 (2): 620–42.

Gelman, Andrew, and Thomas C. Little. 1997. "Poststratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23 (2): 127–36.

Gerber, Alan S., Gregory A. Huber, Daniel R. Biggers, and David J. Hendry. 2017. "Why Don't People Vote in US Primary Elections? Assessing Theoretical Explanations for Reduced Participation." *Electoral Studies* 45: 119–29.

Gerber, Alan S., Gregory A. Huber, and Ebonya Washington. 2009. "Party Affiliation, Partisanship, and Political Beliefs: A Field Experiment." Technical report, National Bureau of Economic Research.

Gerber, Elisabeth R., and Rebecca B. Morton. 1998. "Primary Election Systems and Representation." *Journal of Law, Economics, & Organization* 14 (2): 304–24.

Hanks, Christopher, and Bernhard Grofman. 1998. "Turnout in Gubernatorial and Senatorial Primary and General Elections in the South, 1922–90: A Rational Choice Model of the Effects of Short-Run and Long-Run Electoral Competition on Relative Turnout." *Public Choice* 94 (3–4): 407–21.

Hanmer, Michael J. 2009. *Discount Voting: Voter Registration Reforms and Their Effects*. Cambridge: Cambridge University.

Hill, Seth J. 2015. "Institution of Nomination and the Policy Ideology of Primary Electorates." *Quarterly Journal of Political Science* 10 (4): 461–87.

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22: 129–46.

Jewell, Malcolm. 1977. "Voting Turnout in State Gubernatorial Primaries." *Western Political Quarterly* 30 (2): 236–54.

Kanthak, Kristin, and Rebecca Morton. 2001. "Congressional Primaries." In *Congressional Primaries and the Politics of Representation*, eds. Peter F. Galderisi, Marni Ezra, and Michael Lyons. 116–31. Lanham, MD: Rowman and Littlefield.

Kenney, Patrick. 1983. "Explaining Turnout in Gubernatorial Primaries." *American Politics Research* 11 (3): 315–26.

Kenney, Patrick. 1986. "Explaining Primary Turnout: The Senatorial Case." *Legislative Studies Quarterly* 11 (1): 65–73.

Kenney, Patrick, and Tom W. Rice. 1986. "The Effect of Contextual Forces on Turnout in Congressional Primaries." *Social Science Quarterly* 67 (2): 329.

Key, Valdimer O. 1949. *Southern Politics*. New York, NY: Vintage Books.

Klarner, Carl. 2013. "Other Scholars' Competitiveness Measures." https://doi.org/10.7910/DVN/QSDYLH.

Madden, Richard L. 1986. "*Parties Study Primary Ruling." New York Times.*

Martin, Gregory J., and Steven W. Webster. 2020. "Does Residential Sorting Explain Geographic Polarization?" *Political Science Research and Methods* 8 (2): 215–31.

McGhee, Eric, and Daniel Krimm. 2009. "Party Registration and the Geography of Party Polarization." *Polity* 41 (3): 345–67.

McGhee, Eric, Seth Masket, Boris Shor, Steven Rogers, and Nolan McCarty. 2014. "A Primary Cause of Partisanship? Nomination Systems and Legislator Ideology." *American Journal of Political Science* 58 (2): 337–51.

Morehouse, Sarah M., and Malcolm E. Jewell. 2005. "The Future of Political Parties in the States." Technical report, The Counsel of State Governments.

Norrander, Barbara. 1989. "Explaining Cross-State Variation in Independent Identification." *American Journal of Political Science* 33 (2): 516–36.

Norrander, Barbara. 2018. "The Nature of Crossover Voters." In *Routledge Handbook of Primary Elections*, eds. Robert G. Boatright 105. London, UK: Routledge.

Norrander, Barbara, and Jay Wendland. 2016. "Open Versus Closed Primaries and the Ideological Composition of Presidential Primary Electorates." *Electoral Studies* 42: 229–36.

Schaffner, Brian, and Stephen Ansolabehere. 2018. "CCES Common Content, 2018." https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910/DVN/ZSBZ7K

Schildkraut, Deborah J. 2007. "Defining American Identity in the Twenty-First Century: How Much "There" is There?" *The Journal of Politics* 69 (3): 597–615.

Sheldon, Kennon M., and B. Ann Bettencourt. 2002. "Psychological Need-Satisfaction and Subjective Well-Being Within Social Groups." *British Journal of Social Psychology* 41 (1): 25–38.

Sinclair, Andrew J. 2013. "Of Primary Importance: American Primary Elections 1945–2012." PhD thesis. California Institute of Technology.

Sussell, Jesse. 2013. "New Support for the Big Sort Hypothesis: An Assessment of Partisan Geographic Sorting in California, 1992–2010." *PS: Political Science & Politics* 46 (4): 768–73.

Thornburg, Matthew P. 2014. "Party Registration and Party Self-Identification: Exploring the Role of Electoral Institutions in Attitudes and Behaviors." *Electoral Studies* 36: 137–48.

Thornburg, Matthew P. 2018. "Sooner or Later: Oklahoma Party Registration and Delayed Realignment." *Oklahoma Politics* 28: 1–26.

Thornburg, Matthew P. 2019. "Party Registration Closing Date and Primary Turnout Among Democrats and Republicans." *New England Journal of Political Science* 11 (2): 160–96.

Thornburg, Matthew P. 2022. "Replication Data for: The Dynamics of Hidden Partisanship and Crossover Voting in Semi-Closed Primaries." https://doi.org/10.15139/S3/M3NXWH, UNC *Dataverse, V1, UNF:6: FGkQODMZ4UvBLW84hQhiUw== [fileUNF].*

Ware, Alan. 2002. *The American Direct Primary: Party Institutionalization and Transformation in the North*. Cambridge: Cambridge University.

Weaver, Russell C. 2015. "The Partisan Geographies of Sincere Crossover Voting Behavior: Evidence from North Carolina." *The Professional Geographer* 67 (2): 145–53.

Wekkin, Gary D. 1991. "Why Crossover Voters are not "Mischievous Voters" the Segmented Partisanship Hypothesis." *American Politics Quarterly* 19 (2): 229–47.

Author Biography. Matthew P. Thornburg is an Associate Professor of Political Science in the Department of History, Political Science, and Philosophy at University of South Carolina Aiken.

Cite this article: Thornburg, Matthew P. 2023. The Dynamics of Hidden Partisanship and Crossover Voting in Semi-Closed Primaries. *State Politics & Policy Quarterly* **23** (2): 187–217, doi:10.1017/spq.2022.26

State Politics & Policy Quarterly (2023), **23**: 2, 218–232
doi:10.1017/spq.2022.27

ORIGINAL ARTICLE

The Accuracy of Identifying Constituencies with Geographic Assignment Within State Legislative Districts

Tyler Steelman[1] and John A. Curiel[2]

[1]Office of Institutional Research and Assessment, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[2]Political Science, Ohio Northern University, Ada, OH, USA
Corresponding author: Tyler Steelman, email: tsteelman@unc.edu

(Received 11 December 2020; revised 08 March 2022; accepted 18 May 2022)

Abstract
Identifying the geographic constituencies of representatives is among the most crucial, yet challenging, aspects of state and local politics research. Regularly changing district lines, incomplete data, and computational obstacles can present barriers to matching individuals to their respective districts. Geocoding residential addresses is the ideal method for matching purposes. However, cost constraints can limit its applicability for many researchers, leading to geographic assignment methods that use polygonal units, such as ZIP codes, to estimate constituency membership. In this study, we quantify the trade-offs between three geographic assignment matching methods – centroid, geographic overlap, and population overlap matching – on the assignment of individual voters to state legislative districts. We confirm that population overlap matching produces the highest accuracy in assigning voters to their state legislative districts when polygonal location data are all that is available. We validate this finding by improving model estimates of lobbying influence through a replication analysis of Bishop and Dudley (2017), "The Role of Constituency, Party, and Industry in Pennsylvania's Act 13," *State Politics and Policy Quarterly* 17 (2): 154–79. Our replication suggests that distinguishing between out-of-district and in-district donations reveals a greater impact for in-district lobbying efforts. We make evident that population overlap assignment can confidently be used to identify constituencies when precise location data is not available.

Keywords: GIS/Spatial Analysis; Lobbying; Redistricting; Roll Call Voting

Introduction
The institutional design of representation in the US necessitates matching individuals to their respective legislative districts amidst an array of geographic boundaries that vary in both size and shape. Understanding a representative's geographic constituency is central to American state and local politics research (Fenno 1978), from progressive ambition (Rohde 1979) to policy responsiveness and lobbying influence

(Caughey and Warshaw 2018; Lewis 2013). However, the process of matching individuals to their representatives is a major constraint within the study of state and local politics; one that is further complicated by regularly changing district boundaries and the lack of perfectly nested sub-geographies inside politically relevant boundaries like legislative districts. The challenges arising from the need to match individual data points (like voters) to geographic units (like legislative districts) are experienced by numerous groups including both election administrators[1] and researchers.[2]

Although there have been significant advances in geocoding and database management to improve data sources like voter registration files (Amos and McDonald 2020; Ghitza and Gelman 2020), modern techniques can still be out of reach. If a researcher has access to precise location data like the full address of individuals – which is often not the case – the costs of locating those addresses inside a geographic unit can be prohibitively expensive in regard to time and money. Even with appropriately powerful software and hardware, large-scale geocoding of voter files, as employed by Amos and McDonald (2020), can take dozens of hours to complete.[3] Pay-as-you-go geocoding tools like Google API can be financially demanding when the $5.00/1,000 addresses rate is applied to state voter files with tens of millions of registrants.[4] Although universities can help bridge the resource gap needed to conduct large-scale spatial audits, less-resourced individuals may find the steps outlined by researchers like Amos and McDonald (2020) inaccessible. Fortunately, these costs can be overcome through geographic assignment matching – the process of assigning individuals to a higher level geography based on their inclusion in a nested lower level geography. Crucially for state politics scholars making use of legislative districts, there is no single geography that can be used to match individuals to districts. During the 2011 redistricting cycle, state lower and upper chambers split approximately 48% and 32% of the smallest unit of publicly known geographic units within the US – ZIP codes – respectively.[5] Failing to account for these geographic nuances can lead to error, yet the costs involved in addressing these sources of error can be intimidating. To date, the trade-offs to geographic assignment matching over more computationally intensive methods are presently unclear.

In this study, we first quantify these trade-offs by testing three geographic assignment matching methods – centroid, geographic overlap, and population overlap matching – on the assignment of individual voters to their state legislative districts. In doing so, we confirm that population overlap matching produces the most accuracy in assigning voters to their legal state legislative districts when

[1]In November 2017, some residents of Virginia's 94th state house district were inadvertently assigned to a neighboring district, and subsequently given the wrong ballot which were subsequently thrown out. The number of misassigned voters exceeded the margin of their legal state house race and could have changed the partisan control of the chamber.

[2]Applicable research includes attempts to impute individual level race data (Imai and Khanna 2016), estimate exposure on a geographic unit of interest (Marigalt 2011; Naman and Gibson 2015), or study the responsiveness of a politician to their donors (Gimpel, Lee, and Pearson-Merkowitz 2008).

[3]Amos and McDonald (2020) cataloged a duration of 5.5 hours to geocode the Florida.

[4]See Geocoding API Usage and Billing. Google Maps Platform. https://developers.google.com/maps/documentation/geocoding/usage-and-billing (accessed June 1, 2020).

[5]Estimated from Missouri Census Data Center (2018).

https://doi.org/10.1017/spq.2022.27 Published online by Cambridge University Press

polygonal location data (e.g., ZIP codes) are all that is available.[6] Additionally, so long as the effective number of districts within a lower level polygonal unit is under 1.3, population overlap matching can be used to locate individuals with confidence. We illustrate the applicability of geographic assignment in improving research by replicating and extending the effect of lobbying on legislator behavior by Bishop and Dudley (2017). We show that the burden associated with geocoding individual cases can be significantly reduced by first identifying which lower level geographies are not split between multiple higher levels (like state legislative districts). By better distinguishing between these areas, we discover that an average of 3% of the data is in question and requires use of a geographic assignment method for allocation. Due to the nature of the data, all three methods of controlling for district residency improve model estimates. Following these results, we conclude this study with a set of suggestions by which users can determine whether and how to implement geographic assignment methods within the US context.

The problem

Identifying the geographic constituency of an individual is a multi-stage problem. First, is there an existing data source that matches individuals to their legal representatives? Second, when such data is not present, do the necessary district boundaries in the form of geographic shapefiles or individual coordinates for constituents exist? Third, if coordinate data for individuals is not available, are the addresses in a format conducive to geocoding? And, finally, does the researcher have the means to pay the cost – in both time and money – to perform these computations?

Regarding these problems, the first issue tends to afflict any data that is not a state voter file or a proprietary equivalent. While there is much that can be done with voter files, as demonstrated by Ghitza and Gelman (2020) in improving upon multilevel regression with post-stratification, most research cannot use voter files – even when they are publicly available.[7] Furthermore, the second issue in identifying the geographic constituency of an individual lies in the quality (and existence) of residential address data or legislative district shapefiles for the chamber(s) being studied.[8]

Scholars addressing issue three – matching residential addresses to coordinates – have made many advances. Amos and McDonald (2020) demonstrate the process by which to employ ESRI and Google geocoders, which engage in fuzzy string matching, to geocode millions of addresses from state voter files. The process of hierarchical geocoding devised by Amos and McDonald (2020) improves upon the strengths of each geocoder while mitigating their shortcomings (Swift, Goldberg, and Wilson 2008) to the point of even identifying thousands of errors in misassigned voters

[6]In this study, we use the term ZIP code to describe the geographic unit of a ZIP Code Tabulation Area (ZCTA). ZIP codes are mail routes created by the U.S. Postal Service for efficient mail delivery. ZCTAs are a geographic approximation of those routes maintained by the U.S. Census Bureau. Our calculations use ZCTAs.

[7]States vary in regard to the accessibility of voter files. Although supposed to be free, Wisconsin, for example, charges $12,000, even when for research. Maine only makes voter files available to Maine residents or political action committees.

[8]State legislative districts date back to the 1990s from the US Census. The 2000s see better coverage, though see some gaps whenever a state redistricts mid-decade. Post-2010 data see legislative district boundaries as not as much of an issue. Local boundaries, such as electoral wards, varies in availability.

https://doi.org/10.1017/spq.2022.27 Published online by Cambridge University Press

within official Colorado and Florida state voter files.[9] Therefore, their process demonstrates that it is possible to locate constituents when address and boundary data are present.

The final issue is related to cost. Even when precise coordinate data is available, the issue of cost – in both time and money – often prohibits research from approaching the standards set by researchers like Amos and McDonald (2020). As Goplerud (2015) notes, addressing these issues related to geographic assignment and matching individuals to different levels of geography is consistently expensive in regard to programming skills or the purchase of proprietary software. For example, the hierarchical geocoding process employed by Amos and McDonald (2020) requires access to ESRI proprietary software, ArcGIS, and required 5.5 hours for the Florida voter file alone. If accuracy and non-missingness are a concern, the set of backup checks necessary with a suite of several geolocators can significantly increase time costs.[10] Furthermore, those interested in relying upon proprietary software, such as the Google API, will spend upward of tens of thousands of dollars to geocode a state as populous as Florida. Shepherd et al. (2021), in their recent work analyzing polling place access in North Carolina, relied on a service that provides unlimited geocoding, but at the cost of a $1,000 per month subscription.[11] Absent the resources of a larger research university – or one of the 175 universities with the infrastructure necessary to run graduate programs in geographic information systems (GIS)[12] – these costs can be prohibitively expensive. Therefore, the question naturally arises, are there any methodological shortcuts that might decrease the burden of relying upon hierarchical geocoding that does not sacrifice the quality of the research?

Within the US context, it is possible to avoid both geocoding and more advanced geographic assignment methods in cases where a small enough unit of geography can be identified that is fully nested within a larger geography. ZIP codes provide the smallest unit of publicly known geography and, in many cases, can fulfill this purpose. Figure 1 graphically shows the percentage of a state's population that lives in effectively wholly nested ZIP codes within in relation to state legislative districts. There are apparent differences between chambers, with the results ranging from a low of 5.6% in Rhode Island's state house to a high of 92% in Vermont's state senate. We additionally see that heavily populated states, such as California, Florida, and Michigan, have populations where well over 50% of the state's residents live in ZIP codes nested fully within both the lower and upper state legislative districts. Nationwide, 43% of the population lives in ZIP codes fully nested within state house districts and 61% in state senate districts. It is therefore possible to reduce the need and burden associated with geocoding and more complex methods of geographic assignment, though some of either method will still be necessary when locating individuals that live in non-nested ZIP codes.

[9]Their identification strategy returns to the issue of error in "correctly" geocoded voters.

[10]With a computer with 32 GB of RAM, it took approximately 137 hours to code several snapshots of the North Carolina voter file and its approximately 4.3 million unique addresses. The geocoding made use of ESRI's USA point address locator, street address locator, street name centroid locator, and five-digit ZIP code locator. Of these, 8.1% relied upon ZIP code centroids, which we will go into later in this study.

[11]"Straightforward, Affordable Pricing." Geocodio. https://www.geocod.io/pricing/ (accessed September 1, 2020).

[12]See AAG Guide to Geography Programs in the Americas. https://www.arcgis.com/apps/webappviewer/index.html?id=2f115c9f7ff74723a07aacb6e266b2af (accessed September 25, 2020).

https://doi.org/10.1017/spq.2022.27 Published online by Cambridge University Press

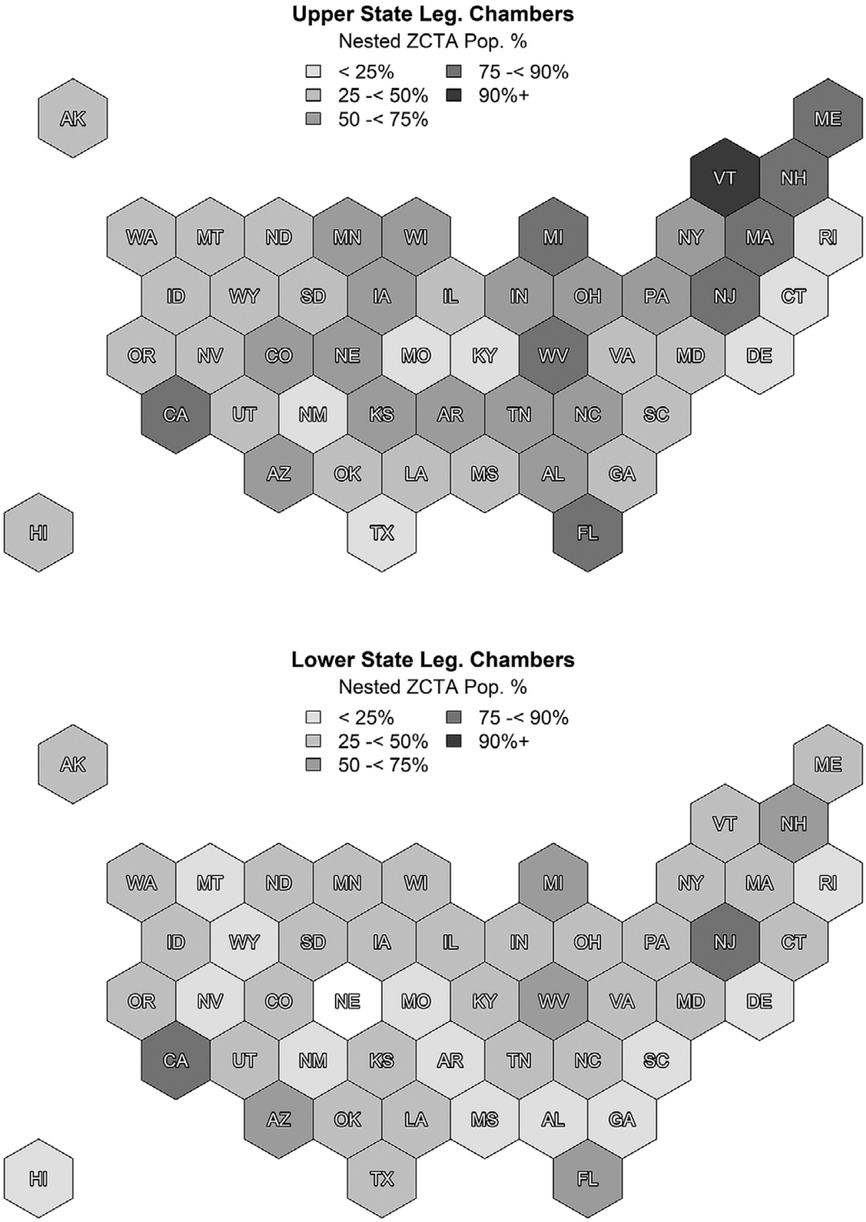


Figure 1. Percentage of state's population living within fully nested ZIP codes, by state legislative chamber.

Assignment of lower levels of geography to higher levels when the two are not perfectly nested has historically used one of three geographic assignment methods: centroid matching, geographic overlap, and population overlap. The centroid matching technique assigns an observation from a lower-level geography when its geographic center, or centroid, falls within the boundaries of a higher-level geographic

https://doi.org/10.1017/spq.2022.27 Published online by Cambridge University Press

unit. Geographic overlap matching assigns or weights a lower-level geography according to the shared area between it and its higher-level overlapping units. Population overlap matching assigns or weights a lower-level geography by using a third level of atomic tabulation units to estimate the population distribution for the overlap between the lower-level and higher-level geographic units.[13] While each of these methods is prone to some assignment error, Amos, McDonald, and Watkins (2017) determine the assignment accuracy of these three techniques is highest for population overlap and lowest for centroid matching. However, their analysis was conducted in aggregate. It did not distinguish between where these methods were most useful and when researchers could expect their results to be biased depending on the geographies employed. Some lower-level geographic units are heavily split between several higher-level geographic units, other lower-level units are wholly nested inside a higher-level unit. Knowing where and when these matching methods can accurately assign individuals without geocoding is necessary for its application to be used with confidence.

Figure 2 illustrates how these geographic matching methods are computed and the challenges that arise from their use. Pictured is the 27713 ZIP code in Durham, North Carolina. For individual data points originating from this ZIP code, there are three overlapping legislative districts.[14] For a ZIP code like 27713, assigning a voter to a legislative district using only this identifier is challenging. Using the centroid method, a researcher would place all individual data points from this ZIP code in the fourth legislative district (as denoted by the star in the center of the figure). Researchers using geographic overlap would also assign this ZIP code's data points to the fourth legislative district given the approximately 40% of geographic space that is shared between the ZIP code and the legislative district. Researchers using population overlap, though, would assign this ZIP code to the first legislative district because the majority of the ZIP code's population resides to the north. This divergence plagues research attempting to allocate individual data points to one geography based on the point's membership in a smaller level of geography. This study quantifies the trade-offs when using centroid matching, geographic overlap matching, or population overlap matching in these situations and confirms that population overlap matching is consistently more accurate than alternative geographic matching methods.

Validation

We probe the accuracy of geographic matching techniques against the validated and audited geocoded voter file used by Amos (2019). Their data provide the correct and known district residency for each voter within their voter file data (Amos and McDonald 2020). Using their data on correct legislative districts, we predict the correct assignment of voters to their (upper and lower) state legislative districts using only their ZIP codes in Colorado, Florida, Louisiana, New York, North Carolina, and Ohio. ZIP codes are the smallest publicly known geographies within state voter files.

[13]For more, see Amos, McDonald, and Watkins (2017), Duque, Laniado, and Polo (2018), Eicher and Brewer (2001), and Rao (2003).

[14]The main focus of this paper is assignment of individual data points to state legislative districts using ZIP codes. For illustrative purposes, this figure uses congressional districts.

https://doi.org/10.1017/spq.2022.27 Published online by Cambridge University Press

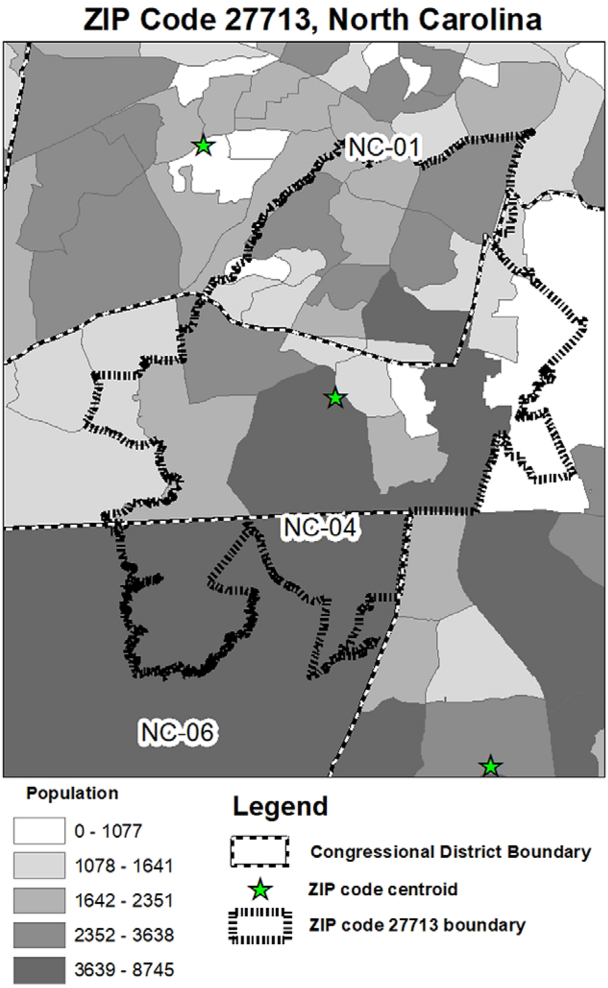


Figure 2. Example of difficulties matching ZIP codes to legislative districts, NC 27713.

Therefore, their use does not require geocoding for matching purposes and can significantly reduce the costs associated with geographic matching.[15]

The dichotomous dependent variable for this validation captures whether an individual within a voter file is assigned to the correct district (1) or not (0) for each matching method. To match each voter to their legal state legislative districts using centroid matching, we employed the ArcGIS feature-to-point tool to calculate the geographic center of each ZIP code (constrained to fit within the boundary of the ZIP code) and then overlaid these onto state legislative maps.[16] Geographic overlap

[15]Curiel and Steelman (2018) note that the population distribution of ZIP codes is on par with Census tracts, with a median population of approximately 3,000 people.

[16]This took under 1 minute to complete for all states using 16 GB of memory.

matching was accomplished using Missouri Census Data Center (2018), which produced dyads of each ZIP code/legislative district pairing as well as the degree of geographic overlap ranging from 0 to 1.[17] For the dichotomous assignment, we assign each ZIP code to the legislative district that it shares the most geographic overlap with.[18] Likewise, we make use of Missouri Census Data Center (2018) to produce dyads of each ZIP code/legislative district pairing as well as the degree of population overlap ranging from 0 to 1, with Census blocks as the atomic tabulation unit. We then follow the same dichotomous assignment of ZIP codes to state legislative districts as used in geographic overlap matching. This resulted in three models, one for each matching method.

We identify the context of where a matching method is most appropriate through a continuous measure of the degree of nestedness between ZIP codes and legislative districts. To do so, we employ the recommended measure as posited by Curiel and Steelman (2020) – the Herfindahl index. It is calculated on a 0 to 1 scale by taking the sum of squared proportions for all of the ZIP code-district dyadic population overlap scores to the ZIP code level. When a ZIP code is fully nested inside a legislative district, its Herfindahl index score is 1. As the effective number of districts inside a single ZIP code reaches infinity, the score approaches 0.[19] The scores are calculated from the GeoCorr output and represent every ZIP code's overlap for the state house and state senate district maps.[20]

Of the three methods, population overlap performs best in aggregate, ranging from 80% to 90% accuracy in predicted legislative district membership for the six states – in line with expectations from Amos, McDonald, and Watkins (2017). As evident in Table 1, geographic overlap performs on par or slightly better than centroid matching – both of which are less accurate than population overlap in assigning voters based solely on a ZIP code. Increased accuracy when using population overlap matching varies across states from a minimum of a single percentage point in Ohio to eight percentage points in Colorado. These results fall short of what is necessary for a full spatial audit consistent with the recommendations from Amos and McDonald (2020) for election administration. However, these findings support the notion that geographic assignment methods are generally helpful for research applications.

Table 1. Accuracy of geographic matching methods in six US states by matching method

State	Population overlap	Centroid	Geographic overlap
CO	0.83	0.75	0.75
FL	0.90	0.88	0.88
LA	0.80	0.74	0.75
NC	0.81	0.77	0.77
NY	0.84	0.81	0.82
OH	0.88	0.87	0.87

[17]This method, which uses a web service, took approximately 10 minutes for all states.

[18]It is possible to weight a given observation instead. However, for comparison to prior work (i.e., Winburn and Wagner 2010), we are employing simple dichotomous assignment.

[19]The inverse of the Herfindahl index provides the effective number of districts with a ZIP code.

[20]Data can be found on the SPPQ dataverse repository (Steelman and Curiel 2022).

In order to determine *where* these methods differ in accuracy, we predict the probability of correct assignment given the degree of nestedness between a ZIP code and its overlapping legislative districts. We conducted the analysis stratified by state, with the results not substantively different by state. As an example, we present the predicted probability plot for North Carolina in Figure 2. The x-axis presents the Herfindahl index to measure the degree of nestedness, with (1) equating to a ZIP code wholly within a legislative district and (0) representing a ZIP code that is infinitely split.

Looking at the left panel of Figure 3, we see that with a Herfindahl score of 0.90, the probability of correct matching exceeds 95% for all three methods. A score of 0.95 on the Herfindahl index corresponds to effectively 100% accuracy in assignment. Insofar as the accuracy starts to dip below 90% accuracy, it will occur for Herfindahl scores around the 0.75 to 0.79 range. Such a score is equivalent to an effective number of districts within a ZIP code being approximately 1.3. It is around this range that we also start to see the differences in the accuracy of each matching method diverge.

The right panel of Figure 3 plots the difference between the accuracy of population overlap matching compared to the centroid and geographic overlap methods. We see that population overlap reaches its maximum advantage over geographic overlap at a Herfindahl score of 0.41 by approximately 10% points. However, at such a score, the population overlap method is only accurate in 46% of cases. Compared to the centroid approach, population overlap performs its best at a Herfindahl score of 0.47, increasing 8.5% points in accuracy. At such a score, population overlap is estimated to have approximately 56% accuracy in assignment. It is also important to note that Herfindahl index values in this range represent a large portion of all ZIP code-legislative district pairs – as made evident from the density plot at the bottom of the panel. Therefore, while substantive disparities in matching methods arise,

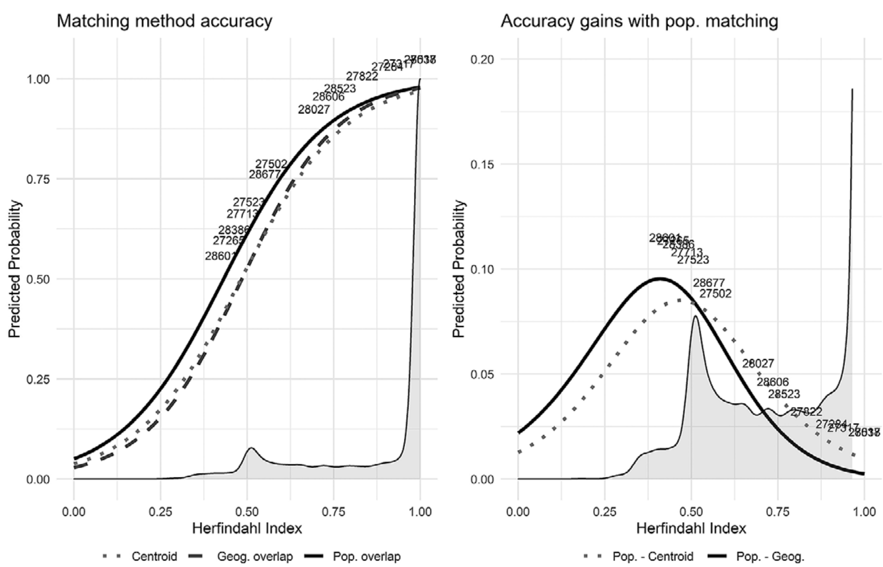


Figure 3. Predicted probability of correct matching.

https://doi.org/10.1017/spq.2022.27 Published online by Cambridge University Press

researchers should hierarchically geocode their data if possible and utilize population overlap matching when hierarchical geocoding is not possible.

The light gray shaded area reflects the distribution of ZIP code nestedness within the North Carolina data. The right panel limits the analysis to ZIP codes with a Herfindahl index under 0.95 for the purpose of focusing on the changes in accuracy across methods. ZIP codes hovering above the solid lines represent a sampling of ZIP codes representative of ZIP code-district nestedness.

Application

By utilizing geographic matching methods to distinguish between constituent and non-constituent influence in lobbying, we can apply the various techniques using a real-world situation. Bishop and Dudley (2017) research the influence of lobbying relative to constituency interests among Pennsylvania state legislators voting for pro-fracking legislation. Their case study selection allows for a critical test of matching techniques that minimizes the impact of endogeneity that typically accompanies research on lobbying and policy outcomes. Bishop and Dudley (2017) tackle the challenges posed by Ansolabehere, de Figueiredo, and Snyder Jr. (2003) in identifying a causal relationship between lobbying and policy outcomes head on, choosing a case where a lobby formed within a few years, effectively precluding a mistake in the causal direction. Additionally, the authors employ geocoded gas-well data to measure constituency reliance upon the natural gas industry.

In a roll call model of voting, negative values of the dependent variable reflect a more conservative voting record in favor of the natural gas industry. The explanatory variables of interest are, first, natural gas production within a legislator's district, and, second, donations from the natural gas industry's political action committee and associated individuals. In their original analysis, the authors find lobbying, in general, exerts a significant, albeit modest, impact on legislator voting.

The only measurement shortcoming of Bishop and Dudley (2017) can be captured in not having distinguished between donations that arise from in- or out-of-district sources. As theorized by Kingdon (1977), representatives attempt to minimize the tension between their stakeholders, ideally never choosing between influential lobbyists and their constituents, hence their preference for committees relevant to their district. For example, Kalla and Broockman (2016) find that the combination of being both a constituent and donor leads one to be more likely to secure meetings with representatives in their randomized field experiments. If one could separately estimate lobbying by in-district versus out-of-district sources, it would be possible to ascertain how much power and influence lobbyists had relative to a legislator's own voters. If all lobbying arose from in-district, it would suggest that Pennsylvania representatives were acting within their constituency's interests. If out-of-district funds still retain an effect, that would suggest a degree of power more associated with fears of corruption and responsiveness to corporations raised by critics of the expansion of natural gas industry goals.

The benefit of population overlap to this analysis is the ability to better distinguish constituency versus non-constituency interests measured using the ZIP codes reported within the contribution data. Pennsylvania is not the ideal state for wholly assigning individuals dichotomously to a district using only a ZIP code. Fortunately, population overlap analysis allows us to weight donations as in-district based upon the proportion of a ZIP code's population that is located within a legislative district to calculate the respective logged donations. Such coding allows us to estimate the impact of out-of-district lobbying. With these new data, we re-estimate the first model presented by Bishop and Dudley (2017) in their Table 3 (169), predicting legislator voting scores as measured by the Pennsylvania League of Conservation Voters (PLCV).

When analyzing the donation data, we find that approximately 97.2% of the donors to members of the state house lived in ZIP codes completely outside the state house member's district. The figure is 90.6% when looking at donors to members of the state senate. Approximately 0.4% of donors to members of the state house were completely nested within one district, and 7.5% of the state senate. This results in 97.6% of the donations made to members of the state house and 98.1% of the donations made to members of the state senate bypass the need for geocoding. In fact, only 4.7% of the donations made to members of the state house and 2.0% of the donations made to members of the state senate require the use of a geographic matching method to allocate them as in- or out-of-district. As a result, this suggests that model differences comparing geographic matching methods will be minimal.

Table 2 estimates separate models for the effect of in-district and out-of-district donations using each matching method for comparison. Furthermore, for population overlap and geographic overlap matching, we include separate models where donations are assigned wholesale to the legislative district with the greatest overlap with the donation's ZIP code and where assignment is weighted on the shared proportion of overlap between a ZIP code and its greatest overlapping legislative district.

Perhaps most telling about the usefulness of matching methods is the difference between coefficients for the effect of donations in the original model (column 1) to

Table 2. Comparisons in predicting PLCV scores

	Dependent variable					
	Original	Pop. weighted	Pop. plural	Geog. weighted	Geog. plural	Centroid
	(1)	(2)	(3)	(4)	(5)	(6)
Democrat	63.479***	63.367***	63.302***	63.416**	63.270***	63.497***
	(3.753)	(3.747)	(3.738)	(3.742)	(3.726)	(3.743)
NPAT score (ideology)	−4.801	−4.818	−4.947	−4.742	−4.895	−4.776
	(3.071)	(3.066)	(3.059)	(3.062)	(3.048)	(3.063)
Senate	−0.406***	−0.341**	−0.337**	−0.326**	−0.310*	−0.338**
	(0.153)	(0.16)	(0.158)	(0.161)	(0.158)	(0.160)
Log dist. gas prod.	−9.062***	−8.826	−8.820***	−8.863***	−8.684***	−8.818***
	(2.187)	(2.190)	(2.182)	(2.184)	(2.178)	(2.187)
Log industry donations	−0.613**					
	(0.269)					
Logged out of district industry donations		−0.551**	−0.540**	−0.544**	−0.533**	−0.553**
		(0.273)	(0.271)	(0.272)	(0.270)	(0.272)
Logged in district industry donations		−1.188**	−1.417***	−1.268**	−1.599***	−1.273**
		(0.504)	(0.530)	(0.502)	(0.530)	(0.519)
Constant	31.057***	30.929***	30.921***	30.894***	30.918***	30.857***
	(2.389)	(2.387)	(2.380)	(2.384)	(2.372)	(2.387)
Observations	248	248	248	248	248	248
R^2	0.877	0.878	0.878	0.878	0.879	0.878
Adjusted R^2	0.874	0.875	0.875	0.875	0.876	0.875

Note: *p < 0.1; **p < 0.05; ***p < 0.1.

https://doi.org/10.1017/spq.2022.27 Published online by Cambridge University Press

Table 3. Cost/benefit analysis of geographic matching methods

Method	Accuracy	Allocation method	Replication impact	Costs	Program
Centroid	Least	Dichotomous	Moderate	Moderate	ArcGIS
Geographic overlap	Moderate	Dichotomous or weighted	Moderate	Low	GeoCorr
Population overlap	High	Dichotomous or weighted	Moderate	Low	GeoCorr

the coefficients of in-district and out-of-district contributions in each of the subsequent models. It is clear that the original impact of industry donations are somewhat muted relative to estimates obtained when the source of donations are distinguished from one another. For example, when examining the impact of in- and out-of-district donations using population overlap matching, it is clear that in-district donations are having an outsized impact relative to donations coming from outside the legislator's district. The coefficients for in-district donations tend to be around double that of out-of-district donations. This pattern is found regardless of matching method and without respect to how donations are aggregated. The lack of significant differences between estimations is not surprising, given the aforementioned lack of donor ZIP codes split between districts. Therefore, these results suggest that while the nature of the data makes it less meaningful *how* the user assigns data, any type of control for donation source can add more nuance as to the differential impact of lobbying by source. Regardless, their original conclusion not only holds up, but is strengthened by utilizing geographic matching to distinguish the source of contributions.

Discussion

As demonstrated, there are situations where geocoding might be necessary to ascertain who represents an individual. However, in the context of American state and local politics, the use of geographic assignment matching methods can reduce the geocoding burden by at least a quarter in every state. Even when an area is split between multiple districts it is possible to confidently assign individuals to legislative districts by gauging the degree of lower-level geography nestedness to make appropriate decisions about the use of geographic assignment methods. These tools can be of use to scholars – and reviewers – in making the call of whether more rigorous methodologies must be employed when assigning voters to legislative districts. Furthermore, it is possible to review previous research, such as Gimpel, Lee, and Pearson-Merkowitz (2008), and determine methodological soundness given matching method and geographic context. In the case of Gimpel, Lee, and Pearson-Merkowitz (2008), we estimate approximately 75% of the nation's population to reside within ZIP codes fully nested within congressional districts during the 2000s, with the greatest inaccuracies arising within Maryland, Nevada, Florida, New York, and North Carolina.[21]

Following our results, we present a cost/benefit analysis of each geographic matching method in Table 3. We organize the results by method, accuracy, allocation method, impact on replication, and the program used to conduct each.

[21]The data acquired to estimate these are from Curiel and Steelman (2018).

https://doi.org/10.1017/spq.2022.27 Published online by Cambridge University Press

We ultimately find that in regard to accuracy, population overlap ranks highest, followed by geographic overlap and finally centroid. The impact of accuracy becomes most clear when a lower level geographic unit is split between two effective higher level units. Allocating individual data points using geographic matching proves to be most flexible when using population or geographic overlap, since each provides a continuous 0–1 score to allocate fractions of a value or assign an entire unit of geography based on the greatest degree of overlap. Costs were highest in terms of compromised accuracy and financial burden for the centroid method as it required access to the centroid geographic bounding tool through ArcGIS. This is compared to a free download of the program GeoCorr which can be used to facilitate geographic and population overlap.[22]

From Table 3, it is apparent that population overlap weakly dominates the other two methods, and centroid assignment is weakly dominated by both population and geographic overlap. In order to aid in future research, we suggest the following rules by which to implement geographic matching assignment.

First, are both levels of geography available from the US Census? If so, then it is possible to employ GeoCorr. If either level of data is not from the US Census, employ a package that can read in raw shapefiles and use geographic operations to find the population or geographic overlap. These might consist of an R CRAN available package by Goplerud (2015) or the recently developed arealOverlapr package (Curiel 2022).

Second, to what extent does the higher level of geography split the lower level? This should be determined by finding the Herfindahl index/effective number of higher geographic units nested within the lower level. To avoid the need of discarding data with accuracy under 90% as practiced by Enos (2015), it is recommended to weight data by dyadic overlap should the Herfindahl index fall below 0.75.

Finally, should the researcher feel uncomfortable with partial weighted geographic assignment and they prefer to geocode individual observations, it is recommended that the researcher lessen the burden of geocoding. By identifying those lower level geographic units completely nested within the higher level, researchers can subset the data they must geocode to only those observations that are not fully nested within the higher level geography being employed. This procedure can save researchers hours of time and potentially thousands of dollars. As illustrated in Figure 1 and the replication of Bishop and Dudley (2017), it might be the case that only few observations even need to be partially weighted or geocoded after properly identifying the data to be geocoded.

Although there will always be uncertainty in geographic assignment where different geographies do not nest within each other, we have improved the confidence that one can have when researching such matters and utilizing such matching techniques. The improvements in population overlap analysis highlight the usefulness of more recent advances in GIS capability and ease of access to individuals. We assert that population overlap analysis offers a valuable tool to anyone pursuing research questions involving the geographic assignment of inconsistently nested geographies.

[22]While free programs like R have the ability to find centroids, this approach might not be geographically bounded as they are in ArcGIS.

https://doi.org/10.1017/spq.2022.27 Published online by Cambridge University Press

We conclude by noting that even though the field of state and local politics is highly variable in regard to the quality of data available, it is possible to overcome the challenges of identifying geographic constituencies scientifically. Moreover, more nuanced identification of these constituencies via geographic assignment and weighting can in turn improve and expand our understanding of state and local politics.

Data Availability Statement. Replication materials are available on SPPQ Dataverse at https://doi.org/10.15139/S3/WIN7SM (Curiel and Steelman 2022).

Funding Statement. The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Amos, Brian. 2019. "Replication Data for: A Method to Audit the Assignment of Registered Voters to Districts and Precincts." Harvard Dataverse, V1. https://doi.org/10.7910/DVN/Y18MK5.

Amos, Brian, and Michael P. McDonald. 2020. "A Method to Audit the Assignment of Registered Voters to Districts and Precincts." *Political Analysis* 28 (3): 356–71.

Amos, Brian, Michael P. McDonald, and Russell Watkins. 2017. "When Boundaries Collide: Constructing a National Database of Demographic and Voting Statistics." *Public Opinion Quarterly* 81: 385–400.

Ansolabehere, Stephen, John M. de Figueiredo, and James M. Snyder Jr. 2003. "Why Is There so Little Money in Congress?" *Journal of Economic Perspectives* 17 (1): 105–30.

Bishop, Bradford H., and Mark R. Dudley. 2017. "The Role of Constituency, Party, and Industry in Pennsylvania's Act 13." *State Politics and Policy Quarterly* 17 (2): 154–79.

Caughey, Devin, and Christopher Warshaw. 2018. "Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 1936–2014." *American Political Science Review* 112 (2): 249–66.

Curiel, John A. 2022. "arealOverlapR." https://github.com/jcuriel-unc/arealOverlapr2.

Curiel, John A., and Tyler Steelman. 2018. "Redistricting Out Representation: Democratic Harms in Splitting Zip Codes." *Election Law Journal* 17 (4): 328–53.

Curiel, John A., and Tyler Steelman. 2020. "A Response to "Tests for Unconstitutional Partisan Gerrymandering in a Post-Gill World" in a Post-Rucho World." *Election Law Journal* 19 (1): 101–9.

Duque, Juan C., Henry Laniado, and Adriano Polo. 2018. "S-maup: Statistical Test to Measure the Sensitivity to the Modifiable Areal Unit Problem." *PLoS One* 13 (11): 1–25.

Eicher, Cory L., and Cynthia A. Brewer. 2001. "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation." *Cartography and Geographic Information Science* 28: 125–38.

Enos, Ryan. 2015. "What the Demolition of Public Housing Teaches Us about the Impact of Racial Threat on Political Behavior." *American Journal of Political Science* 60 (1): 123–42.

Fenno, Richard. 1978. *Home Style: House Members in Their District*. Boston, MA: Little, Brown and Company.

Ghitza, Yair, and Andrew Gelman. 2020. "Voter Registration Databases and MRP: Toward the Use of Large-Scale Databases in Public Opinion Research." *Political Analysis* 28 (4): 507–31.

Gimpel, James G., Frances E. Lee, and Shanna Pearson-Merkowitz. 2008. "The Check is in the Mail: Interdistrict Funding Flows in Congressional Elections." *American Journal of Political Science* 52 (2): 373–94.

Goplerud, Max. 2015. "Crossing the Boundaries: An Implementation of Two Methods for Projecting Data Across Boundary Changes." *Political Analysis* 24: 121–9.

Imai, Kosuke, and Kabir Khanna. 2016. "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Record." *Political Analysis* 24 (2): 263–72.

Kalla, Joshua L., and David E. Broockman. 2016. "Campaign Contributions Facilitate Access to Congressional Officials: A Randomized Field Experiment." *American Journal of Political Science* 60 (3): 545–58.

Kingdon, John. 1977. "Models of Legislative Voting." *Journal of Politics* 39: 563–95.

Lewis, Daniel C. 2013. "Advocacy and Influence: Lobbying and Legislative Outcomes in Wisconsin." *Interest Groups and Advocacy* 2 (2): 206–26.

Marigalt, Yotam. 2011. "Costly Jobs: Trade-Related Layoffs, Government Compensation, and Voting in U.S. Elections." *American Political Science Review* 105 (1): 166–88.

Missouri Census Data Center. 2018. "Geocorr 2018: Geographic Correspondence Engine." http://mcdc.missouri.edu/applications/geocorr2018.html.

Naman, Julia Marie, and Jacqueline MacDonald Gibson. 2015. "Disparities in Water and Sewer Services in North Carolina: An Analysis of the Decision-Making Process." *American Journal of Public Health* 105 (10): 20–6.

Rao, J. N. K. 2003. *Small Area Estimation*. Hoboken, NJ: John Wiley and Sons.

Rohde, David W. 1979. "Risk-Bearing and Progressive Ambition: The Case of Members of the United States House of Representatives." *American Journal of Political Science* 23: 1–26.

Shepherd, Michael E., Adriane Fresh, Nick Eubank, and Joshua D. Clinton. 2021. "The Politics of Locating Polling Places: Race and Partisanship in North Carolina Election Administration, 2008–2016." *Election Law Journal* 20 (2): 155–177.

Steelman, Tyler S., and John A. Curiel. 2022. "Replication Data for: The Accuracy of Identifying Constituencies with Geographic Assignment Within State Legislative Districts." UNF:6:hf3thMP7B4gHG7GM91WVwA== [fileUNF]. https://doi.org/10.15139/S3/WIN7SM.

Swift, Jennifer N., Daniel W. Goldberg, and John P. Wilson. 2008. "Geocoding Best Practices: Review of Eigth Commonly Used Geocoding Systems." https://spatial.usc.edu/wp-content/uploads/2014/03/gislabtr10.pdf.

Winburn, Jonathan, and Michael W. Wagner. 2010. "Carving Voters Out: Redistricting's Influence on Political Information, Turnout, and Voting Behavior." *Political Research Quarterly* 63 (2): 373–86.

Author biographies. Tyler Steelman is the survey research analyst for the Office of Institutional Research and Assessment at UNC Chapel Hill. He completed his PhD in political science and political psychology at the University of North Carolina at Chapel Hill and focuses his research on marginalized identities in contemporary American politics and surrogate representation.

John A. Curiel is an assistant professor of Political Science at Ohio Northern University and earned his PhD in political science from the University of North Carolina at Chapel Hill in August of 2019. He primarily researches how American institutions at the national and state levels mediate representation, in addition to Bayesian and spatial methods.

Cite this article: Steelman, Tyler, and John A. Curiel. 2023. The Accuracy of Identifying Constituencies with Geographic Assignment Within State Legislative Districts. *State Politics & Policy Quarterly* **23** (2): 218–232, doi:10.1017/spq.2022.27

State Politics & Policy Quarterly (2023), **23**: 2, 233–243
doi:10.1017/spq.2023.5

CAMBRIDGE UNIVERSITY PRESS

SHORT ARTICLE

Local Candidate Roots and Electoral Advantages in US State Legislatures

Charles Russell Hunt[1,*] and Stella Rouse[2,3]

[1]Political Science, School of Public Service, Boise State University, Boise, ID, USA
[2]Political Science, Department of Government and Politics, University of Maryland, College Park, MD, USA
[3]Center for Democracy and Civic Engagement, University of Maryland, College Park, MD, USA
*Corresponding author: Charles Russell Hunt, email: charleshunt@boisestate.edu

(Received 10 August 2022; revised 07 October 2022; accepted 11 November 2022)

Abstract

A growing literature has revealed a notable electoral advantage for congressional and gubernatorial candidates with deep local roots in their home districts or states. However, there is a dearth of research on the presence and impact of local roots in state legislative races. In this paper, we close that gap by demonstrating the consistent and significant electoral impacts that state legislators' local roots have on their reelection efforts. We use data capturing a representative cross-section of state legislative incumbents ($N = \sim 5{,}000$) and calculate a novel index measuring the depth of their local roots modeled after Hunt's (2022, *Home Field Advantage: Roots, Reelection, and Representation in the Modern Congress*) measure for the US House. We present evidence that state legislators with deep local roots in the districts they represent run unopposed in their general elections nearly twice as often as incumbents with no such roots. Of those who do attract challengers in their reelection efforts, deeply rooted incumbents enjoy an average of three extra percentage points of vote share. Our results have important implications for candidate emergence in state legislative elections during a time when so many are uncontested. They also demonstrate the limits of electoral nationalization for understanding state politics.

Keywords: political geography; state legislatures; representation; elections; candidate-centered elections

Introduction

A growing literature is uncovering substantively meaningful effects of the place-based connections that elected officials have with the geographic areas they represent. The consequence most scholars have concerned themselves with is the electoral advantage that candidates enjoy when they possess deep local ties to their home districts or states. These ties encompass various forms of personal biographical roots, such as being born and raised, attending school, or having worked or raised in a family within the geographic boundaries of the jurisdiction a candidate is running to represent.

These effects have been demonstrated at the experimental (Munis 2021; Schulte-Cloos and Bauer 2021) and observational levels (Evans *et al.* 2017; Hunt 2021a, 2021b; 2022; Stevens *et al.* 2018) at multiple levels of government, and in both American and international settings. However, to the best of our knowledge, no scholarship has spoken to the effects of local candidate roots at the state legislative level.

Alongside traditional theories as to why local roots should positively impact legislators' electoral fortunes generally, we propose here a novel framework for understanding why local ties should be just as, if not more, impactful in state legislative races. We pair this framework with an original dataset of nearly 5,000 sitting state legislators from all 50 states, for all of whom we have collected unique measures of their local roots, such as where they were born, and where they went to college and postgraduate school. We use these measures to create a Local Roots Index (LRI) modeled after Hunt's (2022) measure, used originally for the US House of Representatives, with the expectation that incumbents with higher scores on this index will be more electorally successful than their more "carpetbagging" counterparts.

Our results are in line with previous findings for other offices and levels of government. Given the notoriously high proportion[1] of uncontested state legislative seats (Squire 2000), we first model the likelihood of incumbents running unopposed in their general elections, conditional on the depth of their local roots and other key variables like partisan balance of the district, chamber seniority, prior political experience, and race and gender. We find that although state legislative incumbents with few to no local roots in their districts run unopposed about a quarter of the time, their counterparts with deep local roots do so more than 40% of the time. This is an effect size that rivals that of district partisanship, and speaks to deep local ties as a major factor in discouraging potential challengers to state legislative incumbents. We also find that even among those incumbents who *do* attract a challenger from the opposing party, local roots offer statistically significant advantages of as many as three percentage points of vote share, a finding consistent with federal offices like the US House. These combined effects of local roots at both the candidate emergence and general election stages demonstrate not only that local roots are highly impactful in state legislative races; but that they continue to influence them even amidst increasing nationalization and polarization of state legislatures that has characterized the modern era of American politics.

Local roots and electoral advantages

It is well established by the literature that local ties are desirable attributes for candidates that voters appreciate and reward on Election Day. Early work demonstrated what V.O. Key dubbed the "friends-and-neighbors" effect, in which candidates for office pull higher-than-expected voter support in their area of residence compared with other areas in their jurisdictions (Aspin and Hall 1987; Key 1949; Parker 1982; Tatalovich 1975). Other work has captured these effects via home state advantages in presidential elections (Garand 1988; Lewis-Beck and Rice 1983). More recent work has confirmed these findings in a variety of ways (Campbell *et al.* 2019; Panagopoulos, Leighley, and Hamel 2017; Put, von Schoultz, and Isotalo 2020),

[1]A total of 32% of incumbents in our sample ran uncontested; many states have rates of state legislative non-contestation higher than 50%.

https://doi.org/10.1017/spq.2023.5 Published online by Cambridge University Press

indicating that the electoral power of geographic closeness remains a factor in modern American elections.

Other literature has demonstrated jurisdiction-wide advantages associated with local ties and, conversely, the pitfalls of candidacy in an area with which one has no background, commonly referred to as "carpetbagging" (Galdieri 2019). Previous work has shown that voters' deep geographic ties to one's jurisdiction is a desirable candidate trait (Munis 2021), resulting in consistent electoral advantages for local candidates in both congressional primaries (Hunt 2021b) and general elections (Evans *et al.* 2017; Hunt 2021a; Stevens *et al.* 2018).

In *Home Field Advantage* (2022), Hunt offers a framework for understanding how these electoral benefits emerge from deep local ties. One set of mechanisms is practical in nature, focusing on heightened local name recognition; more extensive social, economic, and political networks in the jurisdiction; and a homegrown knowledge of the community, including the issues its voters prioritize the most. A second set of mechanisms is more symbolic, drawing on classic findings on phenomena like home styles, representational trust, and the "personal vote" (Cain, Ferejohn, and Fiorina 1987; Fenno 1978; Fiorina and Rohde 1991). This work posits that voters are largely in search of representatives who are "like them," and thus can be trusted to have their best interests at heart once in office. Local roots in a particular home area, when shared between candidate and voter, can create such a connection and imbue the relationship with trust. More recent work has drawn out these symbolic mechanisms via what has come to be called "place identity" – the representational connection individuals feel with particular places or types of places – as the bedrock of why voters consistently choose homegrown candidates at higher rates (Jacobs and Munis 2018; Munis 2021; Schulte-Cloos and Bauer 2021).

Local roots in state legislatures

Somewhat surprisingly, the literature on local roots or place identity, to date, has focused largely on federal or statewide offices. As noted above, this work has found that local roots are an important component of candidate assessment in these contests. In this paper, we argue that state legislative elections can also be fruitful venues for observing the influence of local roots.

First, state legislative elections encompass smaller constituencies (districts), which are more parochial and provide a fertile environment to sow the seeds of stronger place-based attachments. Second, the platform of state legislative candidates is much more focused on state and local issues that often uniquely or disproportionately affect that particular constituency. The ability to address these issues may be more contingent on the place-based attachment a candidate has to that community. Third, it is much easier for state-level candidates to personalize their campaigns and emphasize their local credentials, particularly when they can directly reach many or most constituents through door-to-door campaigning or townhall meetings. Furthermore, voters are more likely to personally know a state-level candidate and be familiar with them outside of politics; most candidates elected to state legislatures continue to hold nonpolitical jobs, since serving as a state legislator is mostly a part-time commitment. Finally, while politics at all levels has become increasingly nationalized (Hopkins 2018), this effect is still less prevalent at the state level. Candidates for state legislatures

https://doi.org/10.1017/spq.2023.5 Published online by Cambridge University Press

can deliver a more localized message, rather than exclusively having to tow the party line.

The effects of local roots do not operate in an electoral vacuum; rather, they also work to translate electoral success into representational benefits (Hunt 2022). This may be even more conspicuous at the state level because legislative candidates can benefit from more familiar personal connections to garner a coalition of loyal voters. Constituents are more receptive when they are actual participants in a mutual relationship with their representative who they feel has specific qualities that will make them more responsive to their needs (Germany 2008). Therefore, deep local roots act as a form of descriptive representation wherein a candidate makes their place identity as "one of them," a group connection that could have a similar influence to that of shared gender, race, and ethnicity (Bratton and Haynie 1999; Rouse 2013). State legislators are also more responsive in this relationship and work to communicate this responsiveness (Jewell 1982) because they see it as valuable, not only for present but future electoral and representative benefits; state legislative service provides ample opportunities for progressive ambition (Maestas 2003).

The factors above indicate that the attachment to place and homegrown candidates should have a significant effect in the electoral success of state legislative candidates. Therefore, based on previous work on congressional elections that has uncovered the importance of local roots and the work on state-level elections that may predict a similar if not stronger effect on state legislative contests, we consider the following hypotheses:

H1: Locally rooted state legislators will be more likely to run uncontested in their general elections.

H2: If they do face opposition, more locally rooted state legislators will receive higher two-party vote share than their less-rooted counterparts.

Data and methods

In this study, we utilize a new dataset compiled by the State Legislators Data Service from *KnowWho*, a commercial data analytics firm that collects and sells background information on state lawmakers.[2] Our dataset uses *KnowWho*'s available data on serving state legislators as of 2018, when the data were obtained by the authors. Although *KnowWho* provided substantial baseline of data for most sitting state legislators, additional coding and data collection efforts on the part of the authors were required for several independent variables, most notably the measurements for legislators' local roots. The result was a cross-section of nearly 5,000 state legislator observations, which represents just under 70% of all sitting legislators at the time.[3]

[2]More information about KnowWho can be found at: https://kw1.knowwho.com/.

[3]This sample was highly representative of the total population of state legislators in terms of race, gender, and party affiliation. There was also very little difference between the groups in terms of the dependent variables: The sample and full population ran unopposed 32% and 33% of the time, respectively; and those in contested races received an average of 59% and 58% of the general-election vote, respectively. As a result, we have little reason to believe that the sample is biased in any systematic way. See Table A3 in the Supplementary Material for the full comparison on key variables between the sample and the full cross-section of legislators provided by *KnowWho*.

To capture the depth of legislators' local roots, we employ a modified version of Hunt's (2022) LRI, which is a summed index of several legislator-specific local roots indicators. We use four indicators in this analysis: whether the legislator was born in their home state; whether more specifically they were born within their district boundaries; went to college in their home state; or obtained postgraduate education in their home state.[4] Although most prior work has used single indicators like birthplace or current residency, the use of an index more comprehensively captures local roots at many points across a legislator's life prior to their service.[5]

KnowWho's data contained the requisite information for some of the local roots indicators (about 50% of sitting legislators), but extensive candidate-level research was necessary to gather more complete information on these indicators, and to capture the broadest possible cross-section of sitting legislators. We were able to increase this sample to just under 70% of sitting legislators using their campaign websites, social media pages, news articles, and official biographies on their official state legislature websites. The combination of *KnowWho*'s data and our own coding efforts yielded city/state locations for birthplace, undergraduate, and postgraduate education for each of these legislators. [6] To determine whether a legislator was actually born in their district, we used GIS tools to intersect this city/state location with state legislative district shapefiles; if the city intersected with the district the legislator represented, they were coded as having been born in their district (see Hunt 2022 for further details).

Our legislator-level data also include each legislator's most recent election results,[7] which allowed us to parse two separate dependent variables corresponding with Hypotheses 1 and 2 respectively: first, whether or not the incumbent legislator ran uncontested in their last general election (H1); and second, the legislator's eventual share of the vote in that election.[8] We argue that the higher an incumbent's score on the LRI, the more likely they will be to run unopposed; and the higher the general election vote share they will receive if they do face a challenger. We also include a

[4]Our multivariate models also include controls for whether the state legislator obtained undergraduate or postgraduate education anywhere, to ensure that we are not simply picking up any electoral advantages associated with higher levels of education.

[5]Table A2 in the Supplementary Material demonstrates that the four component indicators when modeled individually and separately have effects consistent with the combined index. See Hunt (2022), Chapter 3 for more detailed arguments as to the advantages of an additive index.

[6]Although Hunt's version of the LRI for members of Congress included other indicators such as high school attendance or whether the legislator owned a local business in their district, these measures were neither available nor feasible to collect for a large enough sample of state legislators for the reasons discussed earlier. However, as Figure A1 in the Supplementary Material indicates, the four indicators provide substantial variation between legislators; and as Figure A2 in the Supplementary Material indicates, the LRI we use here is quite normally distributed on a scale from 0 (coded as nonlocal on all four indicators) to 4 (coded as local on all four). We believe this LRI represents the fullest possible extent of observational data on local roots that could be obtained for such a large sample of state legislators.

[7]Because term lengths vary for state legislators (either two or four years), these results were either from 2014 or 2016. However, models run separately (or with interactions) based on the legislator's chamber, as well as whether they were elected in 2014 or 2016, yielded robust results across the board.

[8]Incumbents who did not attract challengers in their races were not included in the models testing H2 because their dependent variable values would all be at or just below 100%, thus biasing the sample and the distribution of the dependent variable. However, doing so raises the possibility of selection issues, since the sets of incumbents who do and do not attract challengers is far from random. As a result, selection models run using a Heckman correction (see Table A1 in the Supplementary Material) indicate robust results.

covariate for district magnitude (operationalized using the logged total district population) on the suspicion that local roots may be more or less meaningful depending on the size of the district in question.

We also must control for factors that independently condition the likelihood of challengers emerging to face incumbents, and the eventual outcomes of the general election race. Chief among these is the partisan balance of the state legislative district, captured here using DailyKos's measures of the vote share for the incumbent party's most recent presidential nominee. Higher figures for this variable, therefore, indicate a friendlier partisan environment for the incumbent. We also include control variables for state legislator seniority in their chamber, scaled across states from 1 to 100; whether they are a Democrat, female, or nonwhite; whether they had elected experience prior to their state legislative service; whether they previously held nonelective roles in the party organization or other campaigns; whether they represent a multimember district; and the level of professionalization in their state legislature (Squire 2007). We also include Hinchliffe and Lee's (2016) statewide measure for whether the state has a traditional party organization system. For our modeling techniques, we utilize standard logistic regression (for H1) and ordinary least squares regression (for H2), with standard errors clustered by state in order to account for any nonrandom uncaptured likeness between legislators from the same state.[9]

Results

We first investigate whether state legislators with deep local roots are more likely to run uncontested in their general elections. The logistic regression results in Table 1 strongly suggest that this is the case. At high levels of statistical significance, legislators with higher LRI's are far more likely to run unopposed in their districts.

Figure 1, which generates predicted probabilities based on the model in Table 1, tells us that these effects are substantively as well as statistically significant. State legislators with the deepest local roots are predicted to run uncontested in the general election a little over 40% of the time. They are nearly twice as likely to do so than their unrooted counterparts, who run uncontested less than 26% of the time. Figure 1 also offers important context for the size of this effect. Although local roots and carpetbagging are by no means as impactful as district partisanship in predicting electoral fortunes or the emergence of potential challengers, the effects are in the same ballpark: the safest state legislative incumbents, based on presidential performance in the district, run unopposed about 56% of the time, compared to 11% for those running in districts that heavily favor the opposing party.

Although incumbents can put themselves in strong positions to deter potential challengers from running, the latter's decisions to do so are ultimately out of the incumbent's hands. And so, are incumbents' local roots still impactful even when they do attract a general election challenger? The results in Table 2 again suggest that they are. The finding is more substantively modest, but still statistically significant: deeply rooted incumbents accrue on average about three additional percentage

[9]Although they were not included in the final models, we also ran versions that included a control for Shor–McCarty ideological extremism (not included due to collinearity with presidential vote share), and a control for which state legislative chamber the legislator served in. Neither had any conditioning effects on the LRI's impact on the dependent variables, and so were excluded for simplicity, but are available upon request.

Table 1. Likelihood of running uncontested in general election

Dependent variable	Ran uncontested
Local Roots Index	0.19***
	(0.04)
Any undergrad education	−0.17**
	(0.09)
Any postgrad education	−0.13*
	(0.09)
District partisan safety	4.31***
	(0.56)
Multimember district	−2.93***
	(0.93)
District magnitude	0.06
	(0.13)
SL professionalization	−1.77
	(2.05)
Traditional party org.	0.05
	(0.09)
Democrat	0.22*
	(0.16)
Prev. elected experience	−0.02
	(0.10)
Prev. party/campaign experience	0.00
	(0.12)
Chamber seniority	1.44***
	(0.14)
Female	−0.39***
	(0.08)
Nonwhite	0.15
	(0.19)
Constant	−1.76
	(1.66)
Pseudo R-squared	0.15
N	4,945

Results found using standard logistic regression; SEs clustered by state.
*p < 0.1
**p < .05
***p < 0.01

points of vote share in their general elections compared to their "carpetbagging" counterparts.

This finding is consistent with the effect sizes found by Hunt (2022) in his investigations of the US House. A vote share effect of three percentage points represents a six-point spread in terms of vote margin. For many incumbents, this will not likely change the actual outcome of the election. However, nearly 600 incumbents in our sample of state legislators fell within this six-point margin of victory or defeat (that is, they garnered between 47% and 53% of the vote). In addition, a several-point improvement in electoral fortunes has positive effects for incumbents on the amount of campaigning, fundraising, and spending required of them to feel electorally comfortable. Even if the electoral boost gained by local ties is not decisive in the race, it can grow the margin of victory such that the outcome scares away future potential challengers. In this sense, this result can partially help explain the results on uncontested races found in Table 1 and Figure 2. Beyond candidacy and campaign effects, parties and outside groups are also invested in these margins. With more

https://doi.org/10.1017/spq.2023.5 Published online by Cambridge University Press

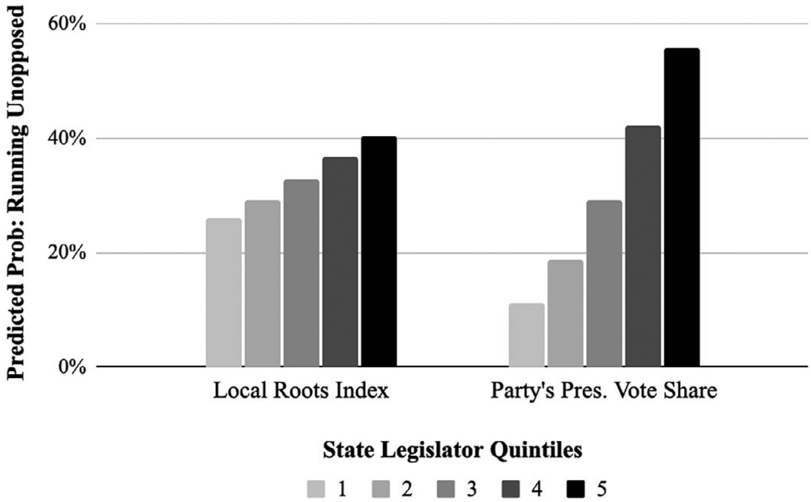


Figure 1. Predicted probability of state legislator running unopposed in their general election based on five-point Local Roots Index (left) and partisan safety of the district (right).

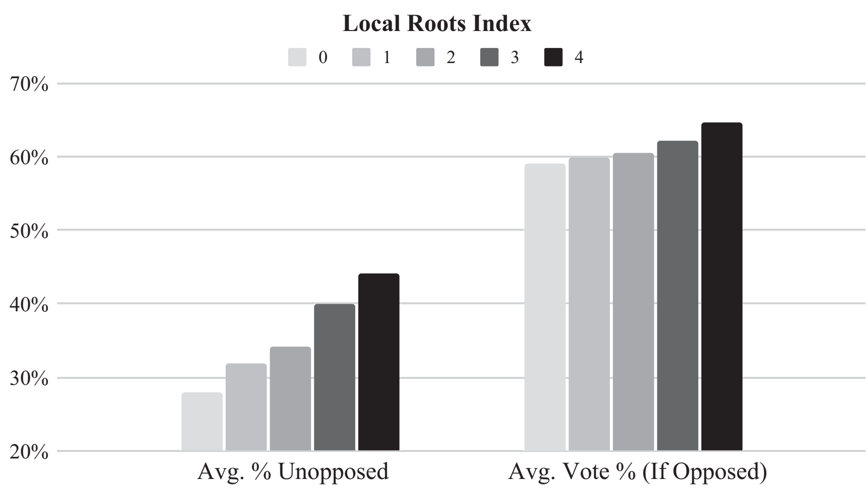


Figure 2. Average percentage of state legislators who ran unopposed in the general election (left) and percentage of the general election vote received if opposed (right) based on state legislators' Local Roots Index.

homegrown candidates on their side, they can afford to reroute crucial campaign dollars and other infrastructure to more competitive state legislative races.

Finally, attention should be paid to the district magnitude covariate in both models. Specifically, should local roots really matter equally in both small districts of only a few thousand constituents, versus larger ones that approach millions? Although fuller theorizing about these interactions is beyond the scope of this short

https://doi.org/10.1017/spq.2023.5 Published online by Cambridge University Press

Table 2. Effects on general election vote share

Dependent variable	GE vote share
Local Roots Index	0.70***
	(0.31)
Any undergrad education	0.63*
	(0.44)
Any postgrad education	0.29
	(0.48)
District partisan safety	46.09***
	(6.94)
Multimember district	−11.58***
	(3.22)
District magnitude	0.07
	(0.96)
SL professionalization	11.77
	(9.98)
Traditional party org	−0.43
	(0.71)
Democrat	−3.72***
	(1.34)
Prev. elected experience	−0.30
	(0.54)
Prev. party/campaign experience	−0.89
	(1.16)
Chamber seniority	4.37***
	(1.25)
Female	−0.02
	(0.41)
Nonwhite	1.69*
	(1.03)
Constant	39.97***
	(12.26)
R-squared	0.53
N	3,322

Results found using standard linear OLS regression; SEs clustered by state.
*$p < 0.1$
**$p < .05$
***$p < 0.01$

article, we note here that including logged total district population as a covariate has no tangible impact on the power of local roots – that is, results are virtually identical whether the covariate is included or not. In addition, although we do not include it in the core models, using this variable as an interaction with the LRI produced null interaction terms, indicating that the impacts of local roots on both the probability of running unopposed, as well as total vote share, are both unrelated to the magnitude of the district.

Conclusion

Local candidate roots have been shown as impactful in modern congressional elections, and in the context of measuring the effects of "place identity" as a meaningful representational connection between voters and elected leaders. We have demonstrated that these effects are substantial in state legislative races as well. In addition to proposing novel theory as to why state legislative elections are fertile

https://doi.org/10.1017/spq.2023.5 Published online by Cambridge University Press

ground for local candidate effects, we have shown that local roots produce significant disincentives for potential challengers to incumbents; and that even when they are challenged, locally rooted incumbents are in better electoral positions than those without local ties.

These results have important implications for state legislative elections. The findings for the impact of roots on running uncontested are particularly instructive in the area of candidate emergence and recruitment. Parties and outside groups looking for new candidates for a seat that is being vacated by a retiring member of their party would clearly do well to find a candidate with deep local roots, who can discourage potential challengers in future elections. On the flipside, parties looking to make headway against potentially vulnerable state legislative incumbents in the opposing party might look to deeply rooted candidates of their own to potentially attenuate the influence of partisanship and challenge these incumbents.

These findings also raise questions about whether the impact of local candidate roots is limited to elections, or plays out in more complex ways within the legislative process. It is possible that candidates with local roots (because they can achieve greater cross-party appeal) could have more moderate voting records, leading to a less ideologically polarized state legislative chamber. Future work could examine this and other potential intersections between legislator roots and their lawmaking behavior in the chamber.

Supplementary materials. To view supplementary material for this article, please visit http://doi.org/10.1017/spq.2023.5.

Data availability statement. Replication materials are available on SPPQ Dataverse at https://doi.org/10.15139/S3/FMYG8Z (Rouse 2023).

Funding statement. The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of interest. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Aspin, Larry T., and William K. Hall. 1987. "The Friends and Neighbors Effect in Judicial Retention Elections." *Western Political Quarterly* 40 (4): 703–15.

Bratton, Kathleen A., and Kerry L. Haynie. 1999. "Agenda-Setting and Legislative Success in State Legislatures: The Effects of Gender and Ethnicity. *Journal of Politics* 61 (3): 658–79.

Cain, Bruce, John Ferejohn, and Morris Fiorina. 1987. *The Personal Vote: Constituency Service and Electoral Independence*. Cambridge, MA: Harvard University.

Campbell, Rosie, Philip Cowley, Nick Vivyan, and Markus Wagner. 2019. "Why Friends and Neighbors? Explaining the Electoral Appeal of Local Roots." *Journal of Politics* 81 (3): 937–51.

Evans, Jocelyn, Kai Arzheimer, Rosie Campbell, and Philip Cowley. 2017. "Candidate Localness and Voter Choice in the 2015 General Election in England." *Political Geography* 59: 61–71.

Fenno, Richard. 1978. *Home Style: House Members in Their Districts*. Boston: Little, Brown.

Fiorina, Morris P., and David W. Rohde. 1991. *Home Style and Washington Work: Studies of Congressional Politics*. Ann Arbor: University of Michigan.

Galdieri, Christopher J. 2019. *Stranger in a Strange State: The Politics of Carpetbagging from Robert Kennedy to Scott Brown*. Albany: SUNY.

Garand, James C. 1988. "Localism and Regionalism in Presidential Elections: Is There a Home State or Regional Advantage?" *Western Political Quarterly* 41 (1): 85–103.

Germany, Julie Barko. 2008. "Constituent Relationship Management for State Legislators." *Institute for Politics, Democracy & the Internet.* Accessed March 18, 2022. Available at: https://leg.mt.gov/content/For-Legislators/orientation/constituent-relations-ipdi.pdf.

Hinchliffe, Kelsey L., and Frances E. Lee. 2016. "Party Competition and Conflict in State Legislatures." *State Politics & Policy Quarterly* 16 (2):172–97.

Hopkins, Daniel J. 2018. *The Increasingly United States*. Chicago: The University of Chicago.

Hunt, Charles R. 2021a. "Beyond Partisanship: Outperforming the Party Label with Local Roots in Congressional Elections." *Congress & the Presidency* 49: 1–30.

Hunt, Charles R. 2021b. "Expanding Constituency Support Through Shared Local Roots in US House Primaries." *American Politics Research* 49 (2): 233–44.

Hunt, Charles R. 2022. *Home Field Advantage: Roots, Reelection, and Representation in the Modern Congress*. Ann Arbor: University of Michigan.

Jacobs, Nicholas F., and B. Kal Munis. 2018. "Place-Based Imagery and Voter Evaluations: Experimental Evidence on the Politics of Place." *Political Research Quarterly* 72 (2): 263–77.

Jewell, Malcolm. 1982. *Representation in State Legislatures*. Lexington: University Press of Kentucky.

Key, V.O. 1949. *Southern Politics in State and Nation*. Knoxville: University of Tennessee.

Lewis-Beck, Michael S., and Tom W. Rice. 1983. "Localism in Presidential Elections: The Home State Advantage." *American Journal of Political Science* 27 (3): 548–56.

Maestas, Cherie. 2003. "The Incentive to Listen: Progressive Ambition, Resources, and Opinion Monitoring Among State Legislators. *Journal of Politics* 65 (2): 439–56.

Munis, B. Kal. 2021. "Place, Candidate Roots, and Voter Preferences in an Age of Partisan Polarization: Observational and Experimental Evidence." *Political Geography* 85: 102345.

Panagopoulos, Costas, Jan E. Leighley, and Brian T. Hamel. 2017. "Are Voters Mobilized by a 'Friend-and-Neighbor' on the Ballot? Evidence from a Field Experiment." *Political Behavior* 39 (4): 865–82.

Parker, Anthony J. 1982. "The 'Friends and Neighbours' Voting Effect in the Galway West Constituency." *Political Geography Quarterly* 1 (3): 243–62.

Put, Gert-Jan, Åsa Von Schoultz, and Veikko Isotalo. 2020. "Fighting over Friends and Neighbors: The Effect of Inter-Candidate Geographic Distance on Intra-Party Competition." *Political Geography* 81: 102219.

Rouse, Stella M. 2013. *Latinos in the Legislative Process: Interests and Influence*. New York: Cambridge University.

Rouse, Stella and Charles Hunt. 2023. "Replication Data for: Local Candidate Roots and Electoral Advantages in U.S. State Legislatures." https://doi.org/10.15139/S3/FMYG8Z, UNC Dataverse, V1, UNF:6:3KVzySrjMLe0wiEsVdQlJg==[fileUNF].

Schulte-Cloos, Julia, and Paul C. Bauer. 2021. "Local Candidates, Place-Based Identities, and Electoral Success." *Political Behavior.* doi: 10.1007/s11109-021-09712-y.

Squire, Peverill. 2000. "Uncontested Seats in State Legislative Elections." *Legislative Studies Quarterly* 25 (1): 131–46.

Squire, Peverill. 2007. "Measuring State Legislative Professionalism: The Squire Index Revisited." *State Politics & Policy Quarterly* 7 (2): 211–27.

Stevens, Benjamin Allen, Md Mujahedul Islam, Roosmarijn de Geus, Jonah Goldberg, John R. McAndrews, Alex Mierke-Zatwarnicki, Peter John Loewen, and Daniel Rubenson. 2018. "Local Candidate Effects in Canadian Elections." *Canadian Journal of Political Science* 52 (1): 83–96.

Tatalovich, Raymond. 1975. "'Friends and Neighbors' Voting: Mississippi, 1943–73." *Journal of Politics* 37 (3): 807–14.

Cite this article: Hunt, Charles Russell, and Stella Rouse. 2023. Local Candidate Roots and Electoral Advantages in US State Legislatures. *State Politics & Policy Quarterly* **23** (2): 233–243, doi:10.1017/spq.2023.5

State Politics & Policy Quarterly

State Politics & Policy Quarterly (SPPQ), peer-reviewed and published quarterly, is the premier publication outlet for original research on state politics and policy. The official publication of the State Politics and Policy Section of the American Political Science Association, SPPQ publishes high quality academic studies that develop and test general hypotheses of political behavior and policymaking, exploiting the unique advantages of the states.

Manuscript submission

All submissions to SPPQ should be made online through ScholarOne at <https://mc.manuscriptcentral.com/sppq>. For complete instructions and submissions guidelines please visit <cambridge.org/sppq/ifc>.

Subscription Information

State Politics & Policy Quarterly (ISSN: 1532-4400) is published quarterly, in March, June, September, and December, by Cambridge University Press, One Liberty Plaza, New York, NY 10006, USA.

The subscription price of Volume 23 (2023) including delivery by air where appropriate (but excluding VAT), is $499.00 (£293.00) for institutions print and online; $398.00 (£235.00) for institutions online only. 2023 subscription rates: Members of the State Politics and Policy Section of APSA receive a hard copy of *State Politics & Policy Quarterly* and have free access to the journal online for an annual charge of $20.00.

Orders, which must be accompanied by payment, may be sent to a bookseller, subscription agent or direct to the publisher: Cambridge University Press, Journals Fulfillment Department, One Liberty Plaza, New York, NY 10006, USA; or Cambridge University Press, University Printing House, Shaftesbury Road, Cambridge CB2 8BS, England. Alternatively, you can place an order online at <cambridge.org/spq>.

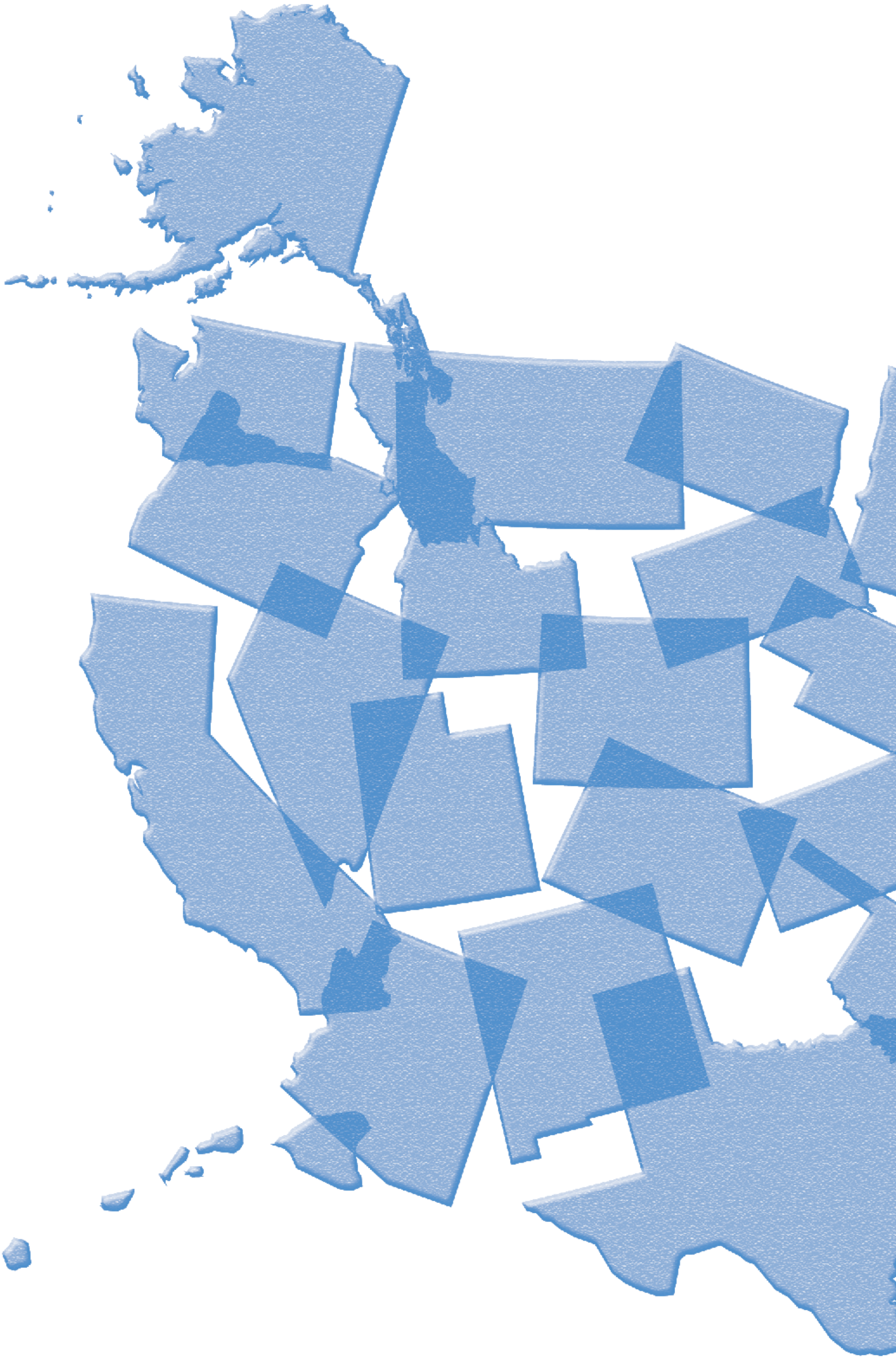
For single issues, please contact customer_service@cambridge.org.

Advertising

For information on display ad sizes, rates, and deadlines for copy, please visit the journal homepage at cambridge.org/spq or contact ad_sales@cambridge.org.

ISSN: 1532-4400
EISSN: 1946-1607
Bundle-code SPQB-0000

Postmaster: Send address changes to *State Politics & Policy Quarterly*, Cambridge University Press, One Liberty Plaza, NewYork, NY 10006, USA.



https://doi.org/10.1017/spq.2023.10 Published online by Cambridge University Press