

Measuring and Analyzing the Policy Agendas of American State Legislatures: 1991-2018

Alex Garlick

Assistant Professor of Political Science, The College of New Jersey, email: garlicka@tcnj.edu

The policy content of the 50 state legislatures is not consistently measured by scholars or any central agency. After reviewing extant efforts in the literature, this paper expands on an approach that uses keyword searches of LexisNexis to estimate how many bills were introduced in 22 policy areas since 1991. The measure is validated using data from the states that do code their legislation by topic, showing high levels of agreement. The keyword estimates are used to investigate agenda setting in the context of federalism. It shows an association between the number of Congressional bills introduced in a policy area and the number of state legislative bills from the prior session, which is taken as evidence of “bottom-up” federalism. There is also an asymmetric relationship between the parties, as chambers controlled by Republicans have a closer contemporaneous association with their Congressional copartisans than Democrats.

Keywords: state politics, text as data, policy, legislation, federalism

Word count (including appendix): 8,337

Notes: This paper was prepared for the (since canceled) 2020 State Politics and Policy Conference in San Diego, California. I thank Justin Kirkland, Virginia Gray and David Lowery for sharing replication data, and to several scholars for responding to my request on Twitter for examples of state legislation coded by policy area. Errors are my own. The data produced by this method is available on the Harvard Dataverse: <https://dataverse.harvard.edu/dataverse/statelegispolicy>.

What do we know about which policies are considered by legislators in the United States? It depends on where you look. There is a great deal of information available about Congress, which scholars have prepared for applied research. However, there is a glaring gap of systematic data about American state legislation in progress. There have been efforts to code *enacted* policies in all 50 states (Sorens et al., 2008; Boehmke et al., 2019), but a number of classic questions in the literature such as “Who governs?” or “How does an idea’s time come?” require a broader view that includes which policies are under consideration, so as to avoid selecting on the dependent variable of policies that had enough consensus to be passed (Binder, 2003).

This paper reviews the literature to find how scholars have measured state policy agendas. There are a great deal of works that have used inductive methods, which start with a policy and then discover how many bills address it. This approach contrasts with a deductive method that would analyze each piece of legislation, assigning them to preset categories, as the Congressional Bills Project has done using the Policy Agendas Project codebook (Adler and Wilkerson, 2015; Baumgartner et al., 2019). The first approach is useful if one is looking to study a policy area to answer a research question such as: “Where have legislators introduced bills about decriminalizing marijuana?” However, using a taxonomy allows for more general comparisons, such as the number of interest groups and bills in the same policy area over time, which can be used to answer the larger question of “What comes first, interest group attention or legislator attention to an issue?” (Lowery et al., 2004).

To address this lack of data availability, I reproduce and expand a prior effort at measuring the policy agendas of all 50 states employed by Virginia Gray, David Lowery and a number of their coauthors to measure agendas between 1995 and 2004 (Fellowes et al., 2006; Kirkland et al., 2010). The measure is a hybrid of inductive and deductive approaches, as it cobbles together 22 inductive searches, by using keyword searches of bill descriptions provided by LexisNexis. Using 22 policy areas covers all of the relevant categories, and is a more than suitable proxy of the Policy Agendas Project approach for the states.¹ The bill descriptions were exported from LexisNexis via email, and have been cleaned for analysis. The dataset is available on the Harvard Dataverse.²

Validation exercises provide confidence in the reliability of the estimates produced by the LexisNexis keyword procedure. Using the universe of legislation from 36 states that report the content of their legislation, as aggregated by the OpenStates project, I find that the keyword estimates have a good deal of precision. In other words, if the keyword procedure codes a bill as being in a policy area, there is an 83% chance that the state identifies it in that same category. However, the keyword procedure identifies an underwhelming portion of the potential documents. Comparisons with previous papers that use this method (Kirkland et al., 2010; Fellowes et al., 2006), show that previous efforts return far more bills per policy area, which is a function of coding choices necessitated by changes LexisNexis has made in the way it surfaces its data over the years.

This dataset is best suited for two types of comparisons. The first is to show the difference in the levels of activity in different policy areas. For example, it shows that state legislators write vastly more education and insurance legislation than civil rights and environmental legislation, which may belie the impression citizens would receive from media coverage of state politics that tends to focus on national policies (Garlick, 2017; McCann et al., 2015). Second, these data are helpful for research designs that use the policy area as the unit of analysis, and are interested in changes in each policy area over time.

To demonstrate the utility of these data for investigating changes within policy areas over time, I consider agenda setting in the context of federalism. Specifically, I ask if changes in the number of bills introduced in a policy area in the state legislative agenda lead to changes in that policy area in Congress as the concept of “laboratories of democracy” would suggest, or vice-versa. I find a positive relationship between the number of Congressional bills and the number of state legislative bills introduced in the *previous* session, such that state legislative activity predates Congressional activity. I do not find the opposite relationship. There is a partisan asymmetry to the contemporaneous relationship, as the number of bills introduced in state legislative chambers controlled by Republicans have a particularly strong association with the number of Congressional bills sponsored by Republicans. No such relationship exists with Democrats. One explanation for this pattern is that Democratic bill sponsorship rates do not tend to vary as much, as they nearly always introducing more bills, so there’s more variation on the Republican side.

I conclude with thoughts on the path forward in this space. Validation exercises show that these data are accurate, but limited in their coverage. Therefore, data such as these could provide

¹The Pennsylvania Policy Agendas Project has done this for the state of Pennsylvania (McLaughlin et al., 2010) and is a model project for future work beyond the legislative agenda to also observe executive branch and media institutions.

²See “State Legislative Bills by Policy Area: 1991-2018”: <https://dataverse.harvard.edu/dataverse/statelegispolicy/>.

an opportunity to serve as training data for an automated process that could be run on the corpus of state legislative text that scholars are beginning to operationalize (McCrain and Hitt, 2019; Kroeger, 2017). While important questions remain on this front — should bills be coded by their full-text or by bill descriptions? (Ragusa and Birkhead, 2020) — this dataset represents a substantial step in the direction of organizing this vital collection of political documents.

Literature on State Policy Agendas

The content of legislation has important effects on the behavior of legislators considering it. Regarding Congressional polarization, Lee (2008) has shown that the US Senate is more polarized due in part to the type of legislation US Senators are voting on, when a greater share of economic bills are on the roll-call agenda, the parties are predictably more split. Looking over the centuries, Lapinski (2013) shows that there has almost always been disagreement on domestic policies, but an important evolution has taken place on “sovereignty” policies, such as questions of citizenship. While these studies do not provide conclusive evidence of why Congress has polarized, they provide an important starting point.

The somewhat static bounds of a legislature’s policy agenda provides a useful lens to observe when an idea’s time has come (Baumgartner and Jones, 1993). This is a persistent topic for both political scientists and scholars of public policy, and seeing how much attention is paid to different policy areas over time can help pin down the *sequencing* of activity in different corners of the government, particularly if the data line up with other coding schemes, such as media attention (Wolfe et al., 2013), economic activity in a sector, or interest group density in a state (Lowery et al., 2004).

The Library of Congress maintains a useful search engine of Congressional legislation, but the best resource for scholars is the Congressional Bills Project, which sorts bills by the Policy Agendas Project codebook from 1947 to the present, using a combination of hand-coding and automated coding procedures (Purpura and Hillard, 2006). A number of other efforts have coded *enacted* legislation by its content, dating back to 1877 (Lapinski, 2013), primarily using research assistants to code bills by hand, although Ragusa and Birkhead (2020) used keyword searches of bill descriptions to code legislation according to the Policy Agendas Project codebook.

Unfortunately, such an accessible dataset is not available at the state level, despite most of the raw data being on the internet. The digital era has led to an explosion in the amount of data of American state legislatures is available to researchers. However, even though researchers have access to the full text of every bill and nearly every roll call vote taken in state legislatures, this data is not prepared for applied research. 36 states report the policy content of their legislation, albeit in an inconsistent fashion. The OpenStates project has collected and standardized these codes; however, there are no policy codes for 14 states, data collection has only begun in 2009, and the data has not been validated.

Before setting out on a coding expedition, it’s worth knowing what the destination is. There are essentially two approaches to coding a legislature’s policy agenda: inductive and deductive. An inductive scheme starts with a topic (e.g. abortion) and identifies each bill within a legislature that addresses a topic. For example, Bromley-Trujillo et al. (2019) downloaded the bill titles, sponsors, and history of approximately 527,000 bills from 2010-2016 from Legiscan. They then used keyword searches of the bill titles to identify climate change legislation, which they validated using a list of climate change legislation maintained by the National Conference of State

Table 1: Selected scholarly work that has coded state legislation by its policy content.

Authors	Policies	Years	States	
<i>Deductive coding schemes</i>			<i>Taxonomy</i>	
Fellowes et al. (2006)	22	1995-1999	50	Economic sectors
Gamm and Kousser (2010)	3	1880-1997	13	Statewide, local or district
Kirkland et al. (2010)	10	2000, 2004	50	Economic sectors
McLaughlin et al. (2010)	PAP*	1979-2012	1	Policy agenda
Garlick (2017)	31	2011-2014	26	National or state
Olson (2019)	PAP	1879-1916	2	Policy agenda
<i>Inductive coding schemes</i>			<i>Policies targeted</i>	
Kreitzer (2015)	1	1973-2013	50	Pro and anti-abortion
Reingold et al. (2019)	5	1997, 2005	15	Women, black, latinx, poverty
Filindra (2019)	1	2005-2011	50	Immigration
Bromley-Trujillo and Karch (2019)	3	1993-2015	50	GMO food, HPV, tanning beds
Bromley-Trujillo et al. (2019)	1	2010-2016	50	Climate change
Reingold et al. (2020)	1	1997-2012	21	Anti-abortion

*The Policy Agendas Project has 20 major topic codes, and 229 minor topic codes.

Legislatures (NCSL).³ A deductive approach starts with a coding scheme, and assigns each bill to one or more of those categories. McLaughlin et al. (2010) has adapted the Congressional Policy Agendas Project codebook for state politics by adding a handful of categories, notably to address intergovernmental relations.

Table 1 shows a sample of papers using state legislative data coded by content. There is impressive coverage of all 50 states, as well as a variety of temporal periods. The inductive papers listed are the tip of the iceberg, as there is much more work that has focuses on a handful of policy areas. However, these works do not necessarily aggregate easily. But the decision to use an inductive or deductive approach depends on the research question of the project. For example, if one is interested in which states are considering legalizing marijuana, the appropriate coding scheme would be inductive, starting with keywords relating to marijuana. However, if one is interested in how much attention is paid to marijuana legalization in different states, the appropriate coding scheme would have to account for the other bills under consideration, as a legislator’s attention span is a zero-sum game. In this case a deductive approach is more appropriate.

Combining inductive approaches will lead to bills fitting in multiple issue areas, which could pose a conceptual challenge. The Congressional Bills Project codes each bill into a single policy area, just as the larger Comparative Agendas Project does for State of the Union speeches, newspaper articles and many other political documents (Baumgartner et al., 2019). However, this decision could be challenged for a number of reasons. One of the most consequential state legislative bills in recent decades was the 2006 Massachusetts bill “Providing Access to Affordable, Quality, Accountable Health Care” which became known as “Romneycare” and later served a template for the federal Affordable Care Act of 2010. This bill is obviously about health care, but it also instituted an individual mandate to hold health insurance, making it the most consequential insurance bill in Massachusetts history as well. It also changed the tax code to pay for a massive slice of the Commonwealth’s budget, making it a consequential tax bill. Scholars need

³The NCSL often identifies legislation across states addressing unique policy trends (e.g. the states that require restaurants to place labels with calorie counts on menus, such as <https://www.ncsl.org/research/health/trans-fat-and-menu-labeling-legislation.aspx>), but it does not maintain a central directory of legislation.

to take care when dealing with multiple issue codes, but from a substantive perspective, it is a defensible position. An emerging best practice in the automated text classification literature is for scholars not to “solve” these difficult coding decisions, but rather to estimate the uncertainty of such decisions with a bootstrap procedure (Garlick and Cluverius, 2020).

The next section describes the dataset built using LexisNexis keyword searches that has more temporal, geographic and agenda coverage than the aforementioned studies.

Estimating the subject matter of bills introduced in the states using LexisNexis

The goals of this procedure are to replicate a deductive coding procedure of state legislative agendas with as much geographic and temporal coverage as possible. While many state governments have made their legislation available online in the 21st century, LexisNexis maintains a database of legislation for all 50 states dating back to 1991. The LexisNexis *State Capital Universe* product allows for keyword searches of “bill tracking reports” that feature a short description of legislation, as well as the full-text of legislation for some years.⁴ To my knowledge, LexisNexis maintains the longest running collection of state legislation.

A notable drawback is that using LexisNexis as a source of data is that their bill tracking reports are proprietary and only available to researchers through a search function (hence the keyword-based approach). *State Capital Universe* offers several means of searching their underlying database, including a search by “synopsis”, which appears to be the bill descriptions shown in Table 3 and a search by “subject” function. However, the subject coding procedure is proprietary and not transparent.⁵ Moreover, LexisNexis redesigned their search portal in 2019 as it transitioned from *LexisNexis Academic* to *LexisNexisUni* and it appears that the collection procedure described herein is now defunct.

I followed the approach Fellowes et al. (2006) used to estimate the collect the agenda for 1995-1999 with only minor modifications. Table 2 shows the keyword that were used to search LexisNexis State Capital Universe⁶. using the “Bill Tracking by Keyword” function. For each policy, I input the keywords (e.g. “education”) in the “Synopsis only” box, as well as the session (“2011”), state (“Illinois”) and a temporal limitation. LexisNexis will only return a search with less than 1000 entries. This can be a challenge as some searches would return more than 1000 results for a year. In some extreme cases, certain searches, like “tax” in New York, would exceed the 1000 limit even if the search were limited to a single day. My understanding is that LexisNexis returns all of the bills under consideration for a given time period, so even though there wouldn’t be 1000+ bills introduced in New York on those days, there would be that many bills in the process. Therefore, I would conduct overlapping searches (sometimes on a day by day basis) to ensure temporal and geographic coverage, in the hopes of capturing bills that were may have excluded by the search limit. This can create redundant entries in the dataset. This process required tens of thousands of search queries, so I automated this process using iMacros⁷ or Kantu.⁸

LexisNexis allows users to export up to 1000 citations from a search via e-mail. Table 3 features a single citation. I imported this unformatted text into Stata, and using regular expressions, pulled out the 1) session of introduction, 2) bill prefix (which indicates which chamber the bill

⁴During data collection, the full-text search of legislation ended in 2012.

⁵During data collection, search by subject was not available after 2012.

⁶Made available via the University of Pennsylvania’s library subscription.

⁷A free browser plugin is available at <http://imacros.net/overview>.

⁸A free browser plugin is available at <https://ui.vision/>.

Table 2: Codebook for Lexis Nexis Searches derived from [Fellowes et al. \(2006, p. 52\)](#)

No.	PAP Code	Full Name	Short Name	Keywords*			Number of bills (1991-2017)
				1	2	3	
1	G0201	Civil Rights	civilr	civil rights			5,854
2	G0205	Environment	enviro	environment			18,318
3	G0207	Religion	relig	church			14,534
4	G0208	Tax Policy	tax	tax			278,419
5	G0300	Health	health	health			219,722
6	G0400	Agriculture	agric	agriculture			22,767
7	G0600	Education	educ	education			200,721
8	G0701	Utilities	util	utilities			36,963
9	G0702	Natural Resource	resourc	gas	oil	minerals	29,154
10	G1000	Transportation	trans	highways	transit	airports	67,279
11	G1200	Law	law	legal			24,495
12	G1300	Welfare	welf	social services	charities		14,700
13	G1400	Construction	const	construction			57,381
14	G1500	Bank	bank	banking	real estate		64,728
15	G1502	Small Business	smallb	retail			14,197
16	G1503	Sports	sport	sports	recreation		19,350
17	G1510	Insurance	insur	insurance			132,446
18	G1520	Manufacturing	manuf	manufacturing			6,640
19	G1600	Military	milit	military			16,583
20	G1700	Communication	comm	media	telecommunications		14,664
21	G2400	Local Government	govt	municipality	public employees		105,100
22	G2401	Police and Fire	pfire	police	fire		50,195

*Multiple keywords are separated by "OR" (e.g. the search for G0702 is: "gas OR oil OR minerals").

Table 3: Example citation drawn from an email from LexisNexis

138. 2015 Bill Tracking MN S.B. 2191, 89TH REGULAR SESSION, SENATE BILL 2191, DATE-INTRO: MAY 15, 2015, LAST-ACTION: MARCH 24, 2016; Rereferred to SENATE Committee on FINANCE., Relates to agriculture; establishes a pollinator investment grant program; appropriates money; awards a pollinator investment grant to a person who implements best management practices to protect wild and managed insect pollinators in this state., MINNESOTA BILL TRACKING Copyright 2016 LexisNexis. All Rights Reserved.

Table 4: Two-year sessions

Year2	48 states	NJ + VA
1991	1991-1992	
1993	1993-1994	1992-1993
1995	1995-1996	1994-1995
1997	1997-1998	1996-1997
1999	1999-2000	1998-1999
2001	2001-2002	2000-2001
2003	2003-2004	2002-2003
2005	2005-2006	2004-2005
2007	2007-2008	2006-2007
2009	2009-2010	2008-2009
2011	2011-2012	2010-2011
2013	2013-2014	2012-2013
2015	2015-2016	2014-2015
2017	2017-2018	2016-2017

was introduced in), 3) bill number. This allows me to create a registry of all the bills introduced across sessions for each individual policy in each state legislature by a “bill code” formed by its state, session, prefix and bill number (e.g. “Colorado 2011 HB 1001”), while ignoring redundant copies. These are not exclusive categories, so a bill can be assigned to more than one subject. I only include conventional house and senate bills introduced during regular sessions. Table 4 shows how bills in states with one year sessions are aggregated into two-year sessions. In most states, I start with the odd-year, except in New Jersey and Virginia, which hold off-cycle state legislative elections, in those states a two-year session starts with the even year.

This procedure modifies the approach used by [Fellowes et al. \(2006\)](#) in two key ways. First, they include each version of a bill, whereas I collapse all of the versions down by bill code. This is necessary because of the search limit challenges mentioned in the preceding paragraphs led to untold duplicate citations. The other major difference is that they reported using LexisNexis’s “Subject” search, and I used “Synopsis” search.⁹

Internal validation

To assess the reliability of estimating the agenda using LexisNexis keyword searches, I compare my estimates to two different sets of estimates in the literature. First, I compare the measure to published estimates produced by [Fellowes et al. \(2006\)](#) and [Kirkland et al. \(2010\)](#). The search terms for my measure were drawn from these sources, so it is an appropriate comparison to make. [Kirkland et al. \(2010\)](#) collect the number of bills introduced in eight policy areas in 2000 and 2004 using the same keywords as my search,¹⁰ and there is a moderately strong relationship between the data at this level of aggregation ($r = 0.54$, $n = 412$). This relationship is shown in

⁹See footnote 5 for more detail on this discrepancy. Moreover, a validation exercise in Appendix A of [Garlick \(2016\)](#) showed that searching by “synopsis” outperformed the “subject” search by producing fewer false positives when checked with OpenStates data.

¹⁰Agriculture, Bank, Communication, Construction, Health, Manufacturing, Military, Natural Resources, Transportation.

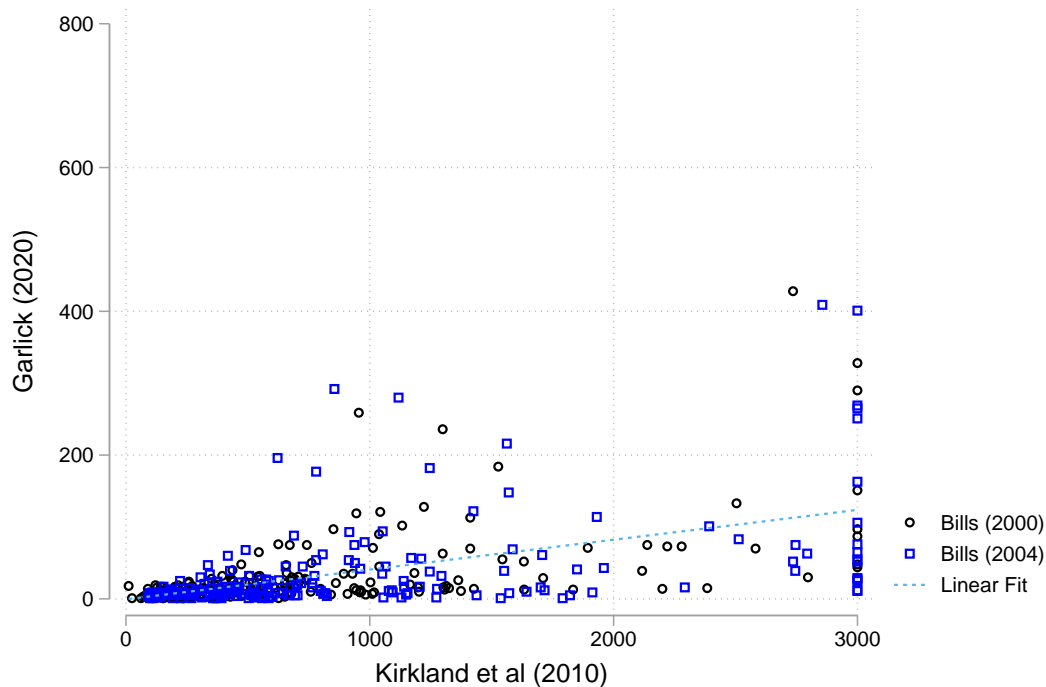


Figure 1: Correlation between the estimated number of bills per policy area and Kirkland et al. (2010) estimates, aggregated by policy/state: 2000, 2004 (n=412)

Figure 1.

Fellowes et al. (2006, p. 40) report the total number of bills across all 22 policy areas for 1995-1999,¹¹ and Figure 2 shows the relationship between these estimates and the keyword estimates. There are about 2.5 times more bills in each session in their data compared to my estimates, which is likely a function of their coding each version of a bill as a unique bill.¹² With these caveats in mind, there is again a moderate relationship between these data, whether they are aggregated by each year ($r = 0.56$, $n = 204$) or aggregated into two-year sessions ($r = 0.57$, $n = 147$).

The mixed nature of these results suggests that the LexisNexis search procedure is highly sensitive to researcher decisions. For example, it raises questions over the nature of a bill: do changes to the legislative text mean it is a new bill? Also, it is hard to answer these questions due to the opaque nature of the LexisNexis search engine: how does it differentiate between a bill synopsis and bill text search?

External validation

The more important validation exercise is to determine if this procedure resembles the way legislators and staff view their own bills. To draw this comparison, I use data from the OpenStates

¹¹The original replication data de-aggregated by policy was not available from the authors.

¹²I tried to re-run the comparison using each unique version of the LexisNexis “description” as a unique bill but it did not improve this association. However, I do not think this is the appropriate approach to this data, as changes to the underlying text may or may not be reflected in the LexisNexis description of the bill, and vice-versa. My measure codes bills as unique by their bill id (All “HB 1001” observations are coded as a single bill).

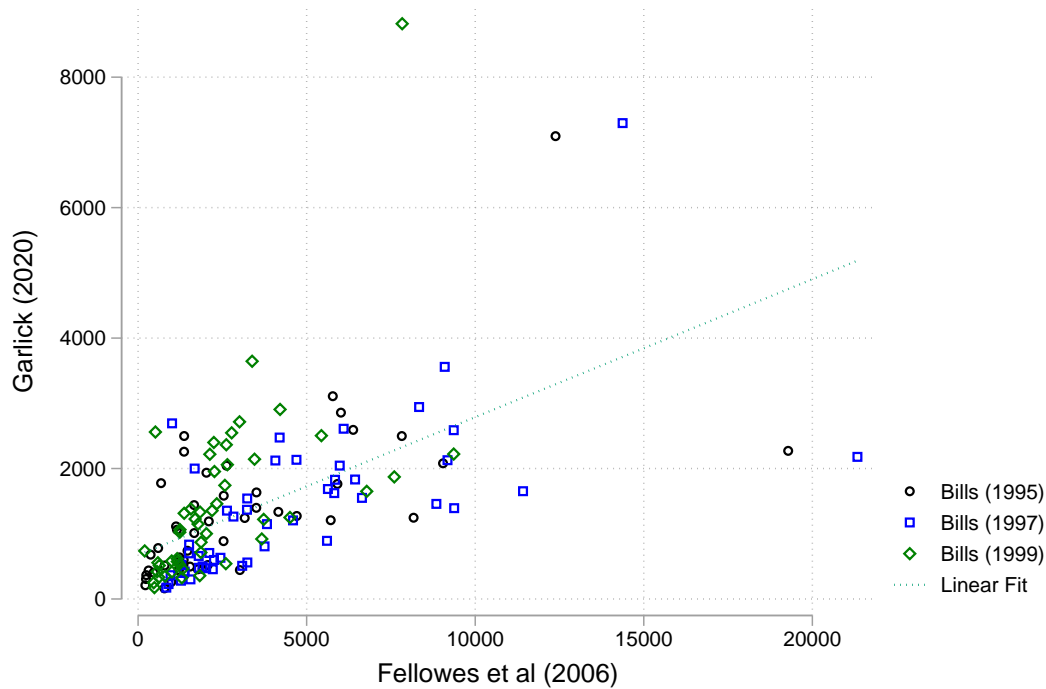


Figure 2: Correlation between the estimated number of bills per policy area and [Fellowes et al. \(2006\)](#) estimates, aggregated by state: 1995, 1997, 1999 (n=147)

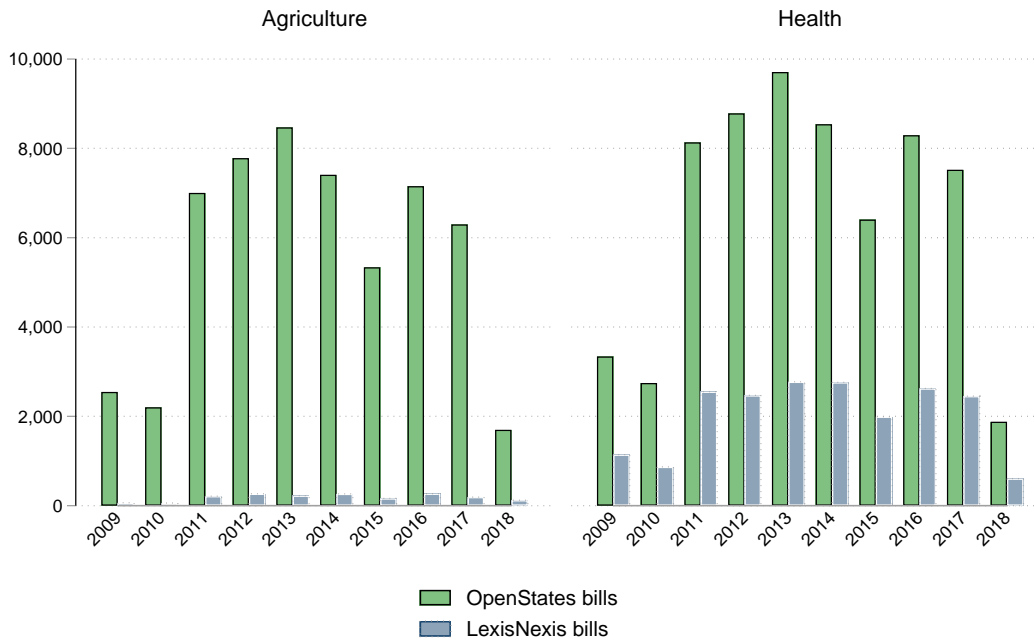
project. OpenStates scrapes state government websites to collect the universe of state legislation since about 2011. Many states have categorized their own legislation by its policy area, which OpenStates has collected for up to 36 states and aggregated to 44 policy areas (see [Garlick, 2017](#), p. 966).¹³ In addition to being a fairly direct measure of policy content of legislation, the OpenStates codes reflect a deductive coding process, which should give a more complete view of the agenda.

In comparison to the previous section, the OpenStates data allows for an individual-bill level validation of the estimates drawn from LexisNexis using an F1-score, which is the weighted harmonic mean of precision (P) and recall (R) shown in equation 1 ([Schütze et al., 2008](#)). Precision is the share of estimations that are “correct.” Since both of these procedures assign individual bills to multiple policy areas, an estimation is deemed to be “correct” if each bill’s policy area has also been recognized in the OpenStates version of the bill for the 15 policy areas shows to be matches in Table 8. Recall is the share of possibly relevant documents that are retrieved. The precision is 0.845 (n = 90,843), and the recall is 0.624 (n = 107, 219), which leads to an F1 score of 0.718. This is an impressive association considering the degree of measurement error in these comparisons. For example, a LexisNexis bill about “Women’s issues” is compared to OpenStates bills about “Sexual Orientation and Gender Issues.” Therefore a bill about discrimination against homosexuals that is not about women’s issues would be coded as a failure of a potentially relevant document in the recall exercise.

¹³The 36 states are listed in Table 9, and demonstrate a representative sample of states across geographic, ideological and professionalization dimensions.

$$F1 = \frac{2PR}{P + R} \quad (1)$$

Figure 3 shows the raw count of bills for the “Agriculture” and “Health” policy areas under these coding schemes. The deductive nature of the OpenStates coding system identifies many more bills. However, these measures do closely track each other. Aggregating the bills by year, there is an extremely strong relationship for the number of bills introduced across all 50 states for both health ($r = 0.98, n = 10$) and agriculture ($r = 0.91, n = 10$). So while this shows there to be far more coverage, the validity of the Lexis codes is high.



Number of bills introduced per year

Figure 3: Number of bills in the states in the OpenStates validation sample by 2-year period.

Descriptive Results

Figures 6 - 9 in the Appendix show the number of bills introduced with all the states aggregated to two-year sessions.¹⁴ These agendas are fairly “sticky,” in the sense that the amount of attention paid to a topic from one session to the next is relatively consistent. Table 5 shows the average biannual change in the number of bills in a policy area is less than 5 percent. There are exceptions to this rule: communication policy spiked in 1999-2002, around the period of the first “.com” boom. Natural resources had an increase after the implementation of fracking in 2013-2014. But generally this pattern fits a punctuated equilibrium model, where the number of bills about a policy topic are relatively static, with exceptions (Baumgartner and Jones, 1993).

¹⁴Figure ?? in the Appendix shows these counts for two policy areas with suspected errors that are in the process of being resolved.

Table 5: Average number of bills (and change) introduced in each two-year session

Policy Name	Bills per session	SD	Change	SD
Tax Policy	308.4	54.6	0.076	0.354
Health	210.8	39.7	0.075	0.256
Education	190.7	43.6	0.106	0.305
Local Government	147.4	25.5	0.056	0.304
Insurance	127.3	22.0	0.032	0.257
Transportation	61.6	9.9	0.061	0.229
Construction	56.6	10.7	0.069	0.309
Utilities	36.6	6.5	0.082	0.308
Natural Resources	26.7	6.0	0.060	0.233
Law	23.7	3.6	0.063	0.269
Agriculture	20.7	3.0	0.055	0.187
Sports and Recreation	17.0	2.9	0.079	0.294
Communication	13.8	3.4	0.081	0.277
Welfare	13.4	3.2	0.084	0.375
Small Business	11.4	5.5	-0.068	0.491
Environment	10.4	2.0	0.067	0.310
Manufacturing	6.2	1.2	0.086	0.295
Religion	4.5	1.3	0.140	0.385
Civil Rights	2.8	0.5	0.066	0.292
Total	63.9	15.1	0.041	0.31

Suspected coding errors: Bank ($\hat{y} = 46.0$, $sd = 69.4$), and Military ($\hat{y} = 6.6$, $sd = 8.0$).

Table 5 also shows dramatic differences in the levels of attention to the different policy areas. On average, the keyword method detects fewer than five bills about religion or civil rights, while finding about 100 times as many health and tax policy bills per two-year session. As a reminder, the relatively low recall score in the validation section casts doubt on the keyword searches capturing every possible bill, however it provides confidence in between-policy sector comparisons such as this.

Application

These data can allow us to address a persistent question in the American federalism literature; which direction does attention flow? Using the case of smoking policy, Shipan and Volden (2006) found evidence of “bottom-up” federalism, where anti-smoking laws passed in municipalities were later passed at the state level. In the other direction, McCann et al. (2015) find evidence of “top-down” federalism, where states enacted anti-smoking legislation after spikes in national attention to the matter, as measured by Congressional hearings.

Voters and the media pay far more attention to national politics than state politics (Garlick, 2017), which may lead state legislators to react to Congressional attention to issues. On the other hand, if states can serve as the “laboratories of democracy” and experiment with policies that could later be applied nationwide, their activity may later spur action from their Congressional counterparts.

Regardless of its direction, a second important question relates to the ordering of the activity of state and national actors. Lowery et al. (2010) discuss three possible patterns of action. First is a *contemporaneous* effect, where both national and state lawmakers are reacting to the same events simultaneously. The implication of this effect would be a positive relationship in the same period. Next, they discuss a *substitution* effect, where activity at one level reduces the need for the other lawmakers to act. This could be observed in a negative contemporaneous relationship, or a negative relationship in the next session. Third, there could be a *stimulation effect*, where activity at one level leads to subsequent activity in the next, perhaps as a function of policy learning, or the need to fill in details to react to another venue. For example, after states consider legalizing recreational marijuana, the federal government may be compelled to act. This could be observed with a positive relationship between the number of bills introduced and the number of bills at the other level in the next session.

The new estimates of state legislation allow to expand on this question in two substantial ways. First, instead of looking at a single policy area, we can look over the entire agenda. This is particularly important as anti-smoking is a fairly nonpartisan policy topic; one may wonder whether top-down or bottom-up federalism happens on issues that map towards partisan conflict? Second, by observing legislation under consideration, we get more of an unvarnished view of “attention” as reflected by the agenda, particularly one that reflects the input of rank-and-file legislators. Enacted legislation requires buy-in from multiple veto players, which could obscure shifts in agenda that are taking place.

Specifically in equation 2, we observe the number of bills introduced in each policy area (p) in the contemporaneous two-year period to each Congress (t), as well as the previous two-year period ($t-1$) and the subsequent two-year period ($t+1$). This parsimonious model also includes indicators for each of the 14 policy areas that overlap the LexisNexis estimates and Congressional Bills Project data. Including these fixed effects allows the intercept to vary for each policy area,

Table 6: There is a positive relationship between the number of bills introduced in a policy area in Congress and the states during the previous two-year period, as well as during the same two-year period.

DV: Congressional bills				
	(1)	(2)	(3)	(4)
Previous session in states	0.042** (0.014)			0.043** (0.015)
Concurrent session in states		0.053* (0.026)		0.062* (0.031)
Subsequent session in states			-0.005 (0.025)	-0.067* (0.030)
Constant	450.556** (67.492)	386.352** (128.961)	673.566** (127.208)	469.484** (139.740)
Observations	154	154	154	154
Absorbed indicators (policy)	14	14	14	14

Standard errors in parentheses

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

which accounts for their different levels and will detect changes in the relative quantities over time.

$$Bills_{us,p,t} = \alpha_p + \beta_1 Bills_{st,p,t-1} + \beta_2 Bills_{st,p,t} + \beta_3 Bills_{st,p,t+1} + \mu \quad (2)$$

Columns (1)-(3) of Table 6 break out each temporal relationship between the state and Congressional agendas. Column (1) shows that there is a positive relationship between the number of bills introduced the states and legislative action in the *next* Congressional session. In other words, there is evidence of “bottom-up” federalism that is consistent with a stimulation effect. It comes as no surprise that there is also a contemporaneous relationship, where national and state legislators respond to events in real time; for example, public health legislation to address the novel Coronavirus outbreak. In column (4) of this table, when all three temporal measures of the state agenda are considered, we also see a *negative* relationship between Congressional bills and the number of bills introduced in that policy area in the next session. This suggests there could be a substitution effect, such that Congressional attention to a policy area could negate the need for state activity on the matter in later years.

Lowery et al. (2010) used Congressional hearings as their dependent variable and had more muted results.¹⁵ This implies that Congressional bill introductions and hearings could have a different relationship to state policy agendas. Hearings are mostly driven by Committee chairs and bills can be introduced by rank-and-file members, so the divergence in these results could be a result of rank-and-file members being more reactive to activity in the states.

An important linkage to consider with agenda setting is partisanship; Lovett et al. (2015) found that a president’s ability to set the agenda with the State of the Union is conditional on their party holding the US House, and a similar pattern could exist across the federal system.

¹⁵Also my estimates allow for much more statistical power.

Table 7: There is a contemporaneous relationship driven by the number of bills introduced by Congressional Republicans and GOP controlled state chambers, but not Democrats.

DV: Congressional bills introduced by:			
	All	Dem.	GOP
	(1)	(2)	(3)
State bills	0.053* (0.026)		
St. Bills (Dem.)		0.040 (0.024)	0.009 (0.015)
St. Bills (GOP)		0.011 (0.030)	0.056** (0.019)
Constant	386.352** (128.961)	207.853 (116.159)	166.197* (71.780)
Observations	154	154	154
Absorbed indicators (policy)	14	14	14

Standard errors in parentheses

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

Figure 5 shows the number of bills introduced by Members of Congress and across the state legislatures, broken out by the party in control of the chamber. It shows that there are consistently more education bills being introduced in chambers controlled by Democrats than Republicans, but an uneven temporal pattern in relation to Congress. There appears to be an increase in state level attention to education in the mid 1990s while Congressional attention is at its nadir, before a spike. In line with Table 6, this could be considered evidence for a “bottom-up” effect. Figure 4 shows a similar pattern for health care policy, which appears to be on the rise in Democratic and Republican held chambers, during a period when Congressional attention was declining.

Table 7 breaks out the contemporaneous relationship from Equation 2 by the party of the member of Congress who introduced each bill, and the party in control of the chamber. It shows that this association is driven by chambers controlled by Republicans and Republican members of Congress. A possible explanation for this asymmetry is that Republican legislators are more uniformly ideological (Grossmann and Hopkins, 2015), which affects their attention to issues. A second possible explanation is that Democratic bill introductions may be more inelastic to events; in other words, because Democrats are always introducing more legislation, their behavior is less responsive to events or activity elsewhere in the federal system. This could particularly be the case in states that have bill introduction limits.

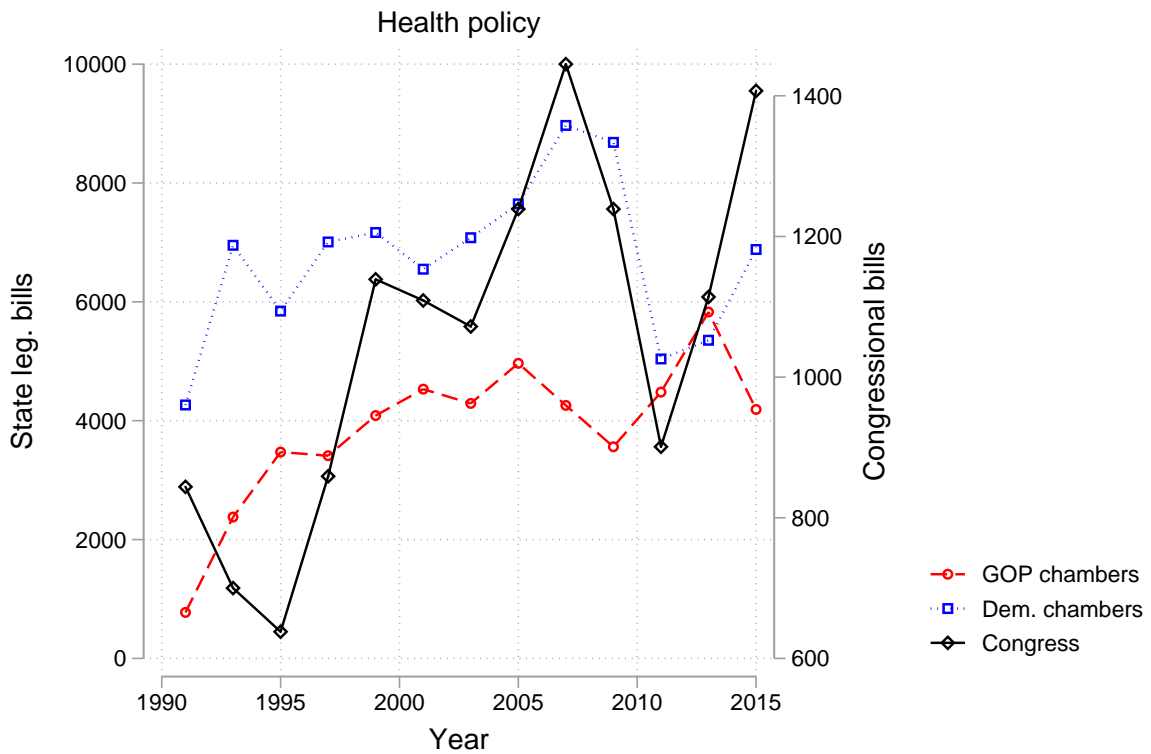


Figure 4: The relationship between Congressional bills and the number of state legislative bills introduced by chamber: Health

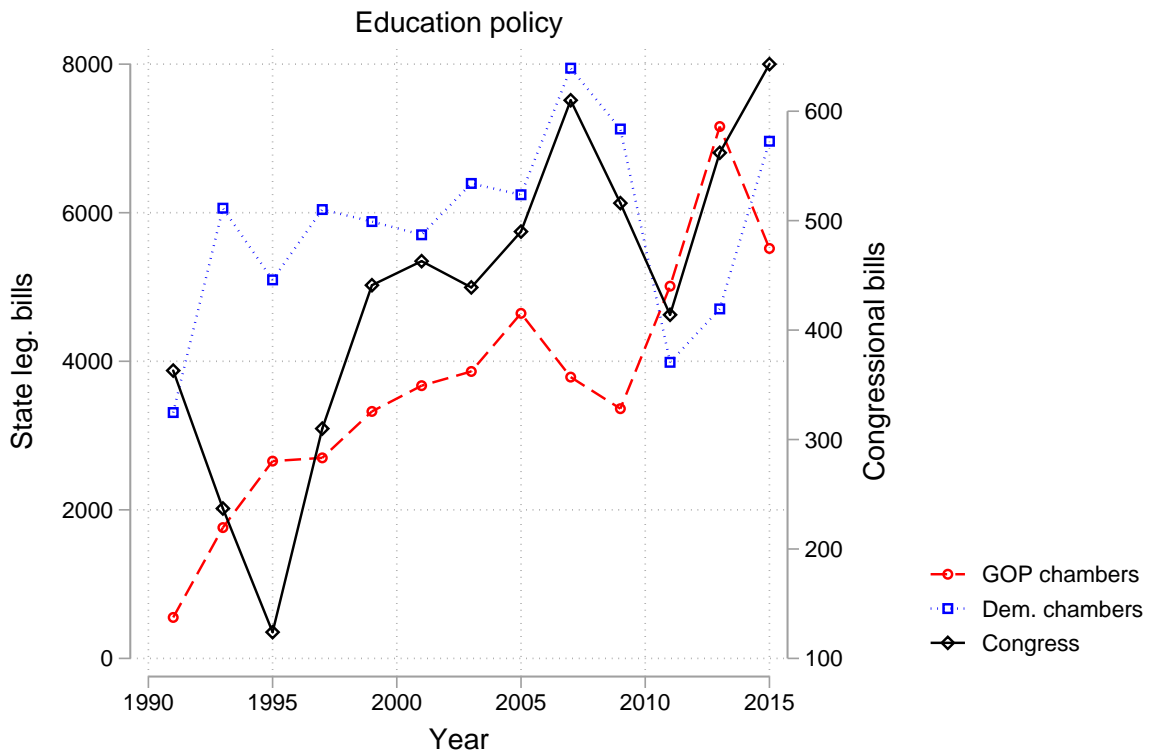


Figure 5: The relationship between Congressional bills and the number of state legislative bills introduced by chamber: Education

Discussion

Despite a boom in the amount of information about subnational politics in the United States being made available to scholars by the digital age, there is no standardized measure of the policy agenda of state legislatures. This is unfortunate as the policy agenda is a useful tool to measure the power of interest groups, legislative branch and executive branch actors. A review of the literature finds pockets of quality data, but it is limited by the portion of the agenda they observe (many are inductive approaches that only cover one or two policy areas) or geography (the Pennsylvania Policy Agendas Project only covers the agenda of that one state). This paper expands an ambitious approach to observe the policy agendas of all 50 states for several years using keyword searches in LexisNexis (Fellowes et al., 2006; Kirkland et al., 2010). This is a hybrid approach that combines 22 inductive searches in an effort to cover the whole agenda. It produces data on all 50 states from 1991-2018, and is best for applied research on changes in policy areas over time.

Validation exercises show the positives and negatives of this approach. Compared to the state legislatures that code the policy content of legislation, and which have been standardized by the OpenStates project, the LexisNexis keyword approach is only coding a portion of the potential legislation. However, these measures correlate very closely, which suggests the keyword estimates are accurate, but struggle with their recall of potentially relevant documents.

Using these data, I revisit an important agenda setting question. Within the American federal system, does attention flow from the top-down or from the bottom-up? I find a relationship between the number of bills introduced at the states level and the *subsequent* Congressional agenda, but not the inverse, which is evidence of bottom-up federalism. There is also a contemporaneous relationship between these quantities, and the association is concentrated in the relationship between the number of bills introduced by Congressional Republicans and the bills introduced in state legislative chambers controlled by Republicans. There is no such relationship with Democratic lawmakers.

In an ideal scenario, a portion of the policy agenda that has been accurately coded could be used to build an automated procedure to deductively code the universe of state legislation. Similar efforts have been taken to code an annual census of the interest group population by their economic sector (Garlick and Cluverius, 2020). A number of political scientists have drawn on the entire corpus of state legislative text (McCrain and Hitt, 2019; Kroeger, 2017), that Legiscan has made available since about 2010,¹⁶ so this scenario could be feasible. Although, building their database of enacted Congressional laws, Ragusa and Birkhead (2020) argue that descriptions are sufficient. As there is a great deal of repetitive text in legislation, this method could prove useful.

Garlick and Cluverius (2020) offer a few principles that could guide such a future work. First, to allow for the least possible measurement error in validation exercises and the ability to revisit previous findings, codebooks should accords with previous literature, as this paper does with Virginia Gray and David Lowery's many coauthored works in this area. Second, it should be externally valid to the modern context. This paper attempts to do make such a comparison using OpenStates's measurement of the agenda, however, the generation of those data is also opaque. Third, it should be reproducible. LexisNexis is extremely limited in this regard, as its subject ratings of bills by policy area are proprietary and it has made material changes to its search engine function. The search procedure described herein is defunct, and initial testing of

¹⁶See <https://legiscan.com/datasets>.

LexisNexisUni suggests that significant changes may have to be made to replicate what has been described in this paper.

Observing the policy content of state legislation is a worthy endeavor. This paper uses a legacy data source to break ground and produces a dataset that will be useful to evaluate certain research questions. However, more work could be done that would put state legislative scholars on equal footing with their Congressional counterparts.

References

- Adler, E. S. and Wilkerson, J. (2015). Congressional Bills Project: 1998-2014. *NSF 00880066 and 00880061*, (<http://www.congressionalbills.org/>).
- Baumgartner, F. R., Breunig, C., and Grossman, E. (2019). *Comparative Policy Agendas: Theory, Tools, Data*. Oxford University Press.
- Baumgartner, F. R. and Jones, B. D. (1993). *Agendas and instability in American politics*. University of Chicago Press.
- Binder, S. A. (2003). *Stalemate: Causes and consequences of legislative gridlock*. Brookings Institution Press.
- Boehmke, F. J., Brockway, M., Desmarais, B. A., Harden, J. J., LaCombe, S., Linder, F., and Wallach, H. (2019). SPID: A New Database for Inferring Public Policy Innovativeness and Diffusion Networks. *Policy Studies Journal*.
- Bromley-Trujillo, R., Holman, M., and Sandoval, A. (2019). Hot Districts, Cool Legislation: Evaluating Agenda Setting in Climate Change Bill Sponsorship in US States. *State Politics & Policy Quarterly*, 19(3):375-395.
- Bromley-Trujillo, R. and Karch, A. (2019). Saliency, Scientific Uncertainty, and the Agenda-Setting Power of Science. *Policy Studies Journal*.
- Fellowes, M., Gray, V., and Lowery, D. (2006). What's on the table? The content of state policy agendas. *Party Politics*, 12(1):35-55.
- Filindra, A. (2019). Is Threat in the Eye of the Researcher? Theory and Measurement in the Study of State-Level Immigration Policymaking. *Policy Studies Journal*, 47(3):517-543.
- Gamm, G. and Kousser, T. (2010). Broad bills or particularistic policy? Historical patterns in American state legislatures. *American Political Science Review*, 104(01):151-170.
- Garlick, A. (2016). *Interest Groups, Lobbying and Polarization in the United States*. PhD thesis, University of Pennsylvania.
- Garlick, A. (2017). National Policies, Agendas, and Polarization in American State Legislatures: 2011 to 2014. *American Politics Research*, 45(6):939-979.
- Garlick, A. and Cluverius, J. (2020). Automated Estimates of State Interest Group Lobbying Populations. *Interest Groups and Advocacy*.
- Grossmann, M. and Hopkins, D. A. (2015). Ideological Republicans and group interest Democrats: The asymmetry of American party politics. *Perspectives on Politics*, 13(01):119-139.

- Kirkland, J. H., Gray, V., and Lowery, D. (2010). Policy Agendas, Party Control, and PAC Contributions in the American States. *Business and Politics*, 12(4):1–24.
- Kreitzer, R. J. (2015). Politics and morality in state abortion policy. *State Politics & Policy Quarterly*, 15(1):41–66.
- Kroeger, M. A. (2017). *How Groups Write the Law: An Empirical Analysis of Group Influence in American State Legislatures*. PhD thesis, Princeton University.
- Lapinski, J. S. (2013). *The Substance of Representation: Congress, American Political Development, and Lawmaking*. Princeton University Press.
- Lee, F. E. (2008). Agreeing to disagree: Agenda content and Senate partisanship, 1981–2004. *Legislative Studies Quarterly*, 33(2):199–222.
- Lovett, J., Bevan, S., and Baumgartner, F. R. (2015). Popular Presidents Can Affect Congressional Attention, for a Little While. *Policy Studies Journal*, 43(1):22–43.
- Lowery, D., Gray, V., and Baumgartner, F. R. (2010). Policy Attention in State and Nation: Is Anyone Listening to the Laboratories of Democracy? *Publius: The Journal of Federalism*, pages 1–25.
- Lowery, D., Gray, V., Fellowes, M., and Anderson, J. (2004). Living in the moment: Lags, leads, and the link between legislative agendas and interest advocacy. *Social Science Quarterly*, 85(2):463–477.
- McCann, P. J. C., Shipan, C. R., and Volden, C. (2015). Top-Down Federalism: State Policy Responses to National Government Discussions. *Publius: The Journal of Federalism*, pages 1–31.
- McCrain, J. and Hitt, M. (2019). Measuring Legislative Valence in the U.S. States. *State Politics and Policy Conference at College Park, Maryland*.
- McLaughlin, J. P., Wolfgang, P., Leckrone, J. W., Gollob, J., Bossie, J., Jennings, J., and Atherton, M. J. (2010). The Pennsylvania policy database project: A model for comparative analysis. *State Politics and Policy Quarterly*, 10(3):320–336.
- Olson, M. (2019). Restoration and Representation: Legislative Consequences of Black Disfranchisement in the American South, 1879–1916. *State Politics and Policy Conference in College Park, Maryland*.
- Purpura, S. and Hillard, D. (2006). Automated classification of congressional legislation. In *Proceedings of the 2006 international conference on Digital government research*, pages 219–225.
- Ragusa, J. M. and Birkhead, N. A. (2020). *Congress in Reverse: Repeals from Reconstruction to the Present*. University of Chicago Press.
- Reingold, B., Kreitzer, R. J., Osborn, T., and Swers, M. L. (2020). Anti-abortion Policymaking and Women’s Representation. *Political Research Quarterly*, page 1065912920903381.
- Reingold, B., Widner, K., and Harmon, R. (2019). Legislating at the Intersections: Race, Gender, and Representation. *Political Research Quarterly*, page 1065912919858405.
- Schütze, H., Manning, C. D., and Raghavan, P. (2008). *Introduction to Information Retrieval*. Cambridge Univ Press.

- Shipan, C. R. and Volden, C. (2006). Bottom-up Federalism: the diffusion of antismoking policies from US cities to states. *American journal of political science*, 50(4):825–843.
- Sorens, J., Muedini, F., and Ruger, W. P. (2008). US state and local public policies in 2006: A new database. *State Politics & Policy Quarterly*, 8(3):309–326.
- Wolfe, M., Jones, B. D., and Baumgartner, F. R. (2013). A failure to communicate: Agenda setting in media and policy studies. *Political Communication*, 30(2):175–192.

Appendix

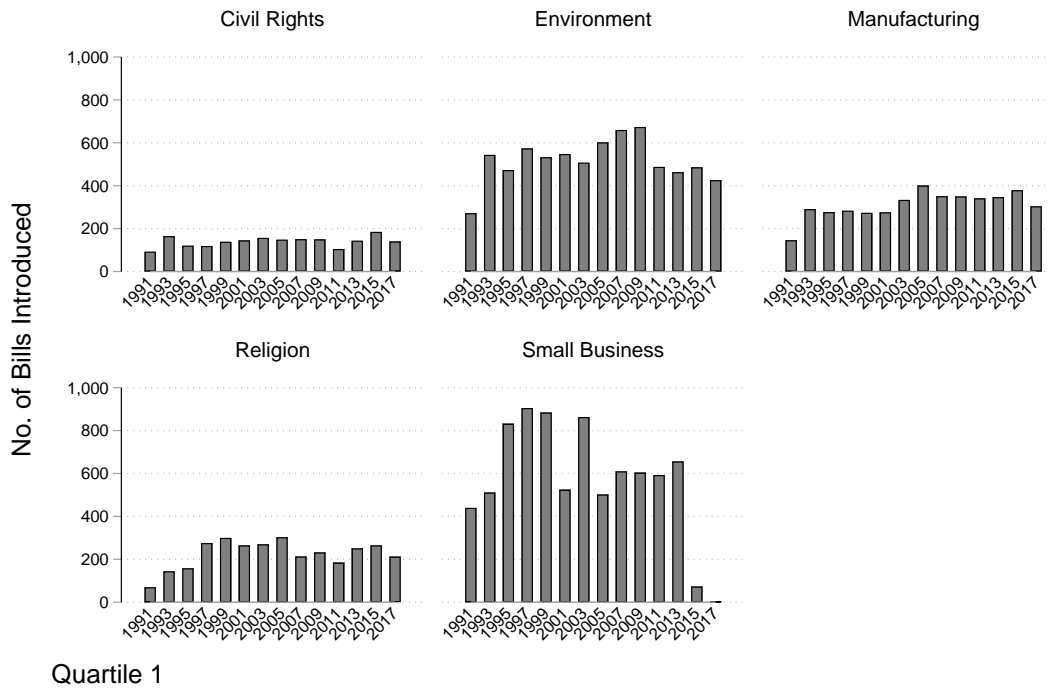
Descriptive Results

Figures 6-9 show the number of bills introduced in 21 policy areas over 1991-2018. In reference to Table 2, it combines G2400 and G2401 into a single category. Also, two policy areas have suspected coding errors and are presented separately: post-2003 for Military, and pre-2005 for Bank. There is no reason to believe these estimates are corrupted outside those temporal periods. The data collection was done in two major parts (approximating those temporal periods), so future drafts of this paper will address that discrepancy.

Supporting material for the OpenStates validation

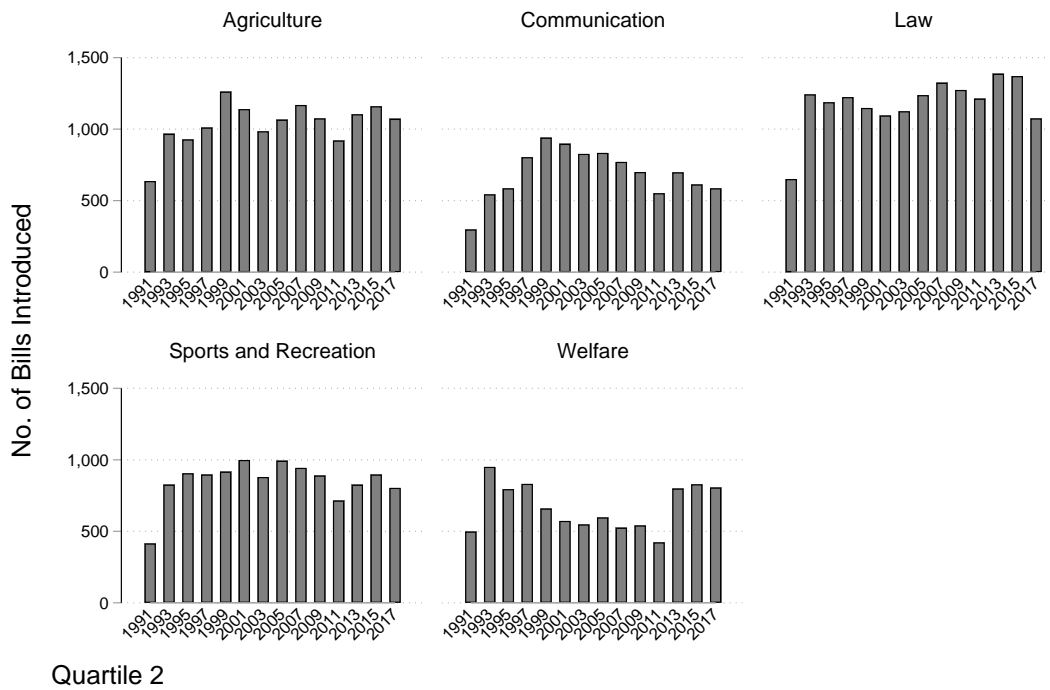
Table 8 relates the LexisNexis search terms from Table 2 with the 44 Open States codes. This alignment is meant to fit the LexisNexis terms within corresponding OpenStates subject, in order to minimize Type I errors. For example, see G0201, where a bill being coded as Civil Rights should be a necessary condition to fit “Civil Liberties and Civil Rights”, even though the inverse would not be true.

Table 9 shows that states that report the subject of their legislation, which has been standardized by the OpenStates project. These 36 states demonstrate a useful amount of variation on geographic, ideological and professionalization dimensions who provides confidence that the validation sample is not biased by the states which choose to code their legislation. However, the subject matter of legislation in these states is taken at face value. Future work may look to validate these codes as well.



Quartile 1

Figure 6: Number of groups in each 2-year period



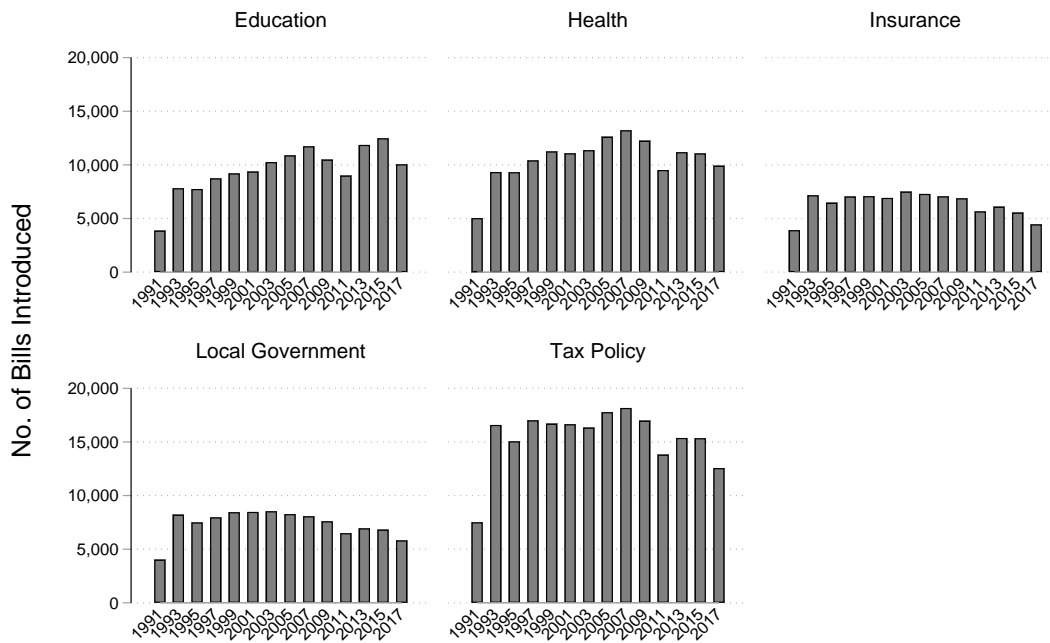
Quartile 2

Figure 7: Number of groups in each 2-year period



Quartile 3

Figure 8: Number of groups in each 2-year period



Quartile 4

Figure 9: Number of groups in each 2-year period

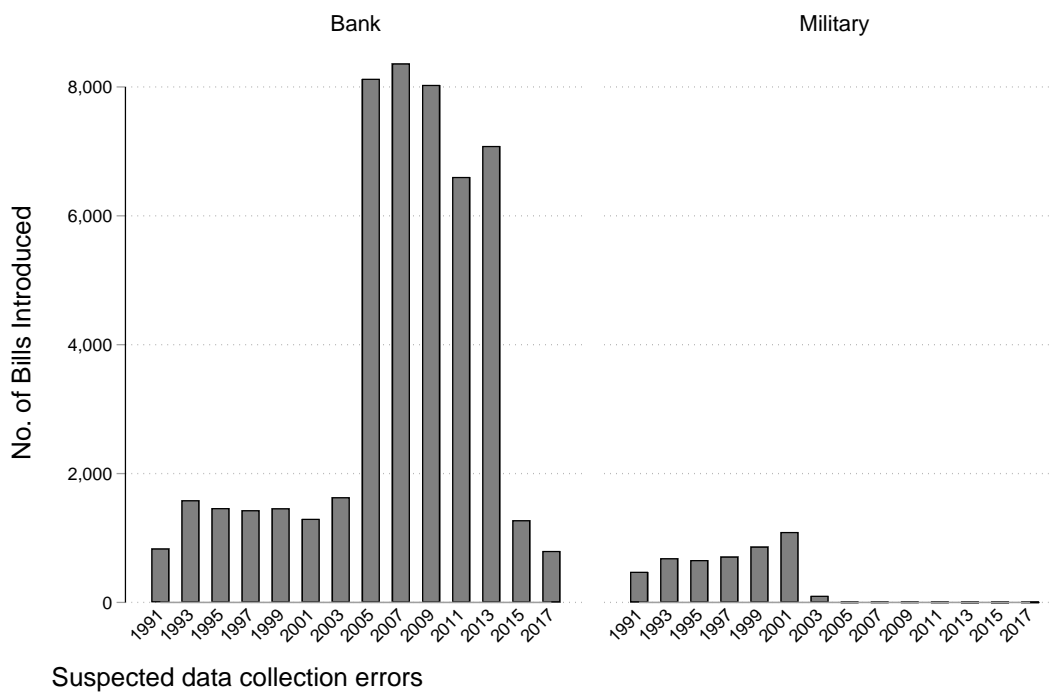


Figure 10: Number of groups in each 2-year period

Table 8: Subject coding scheme from LexisNexis to OpenStates

No.	PAP code	LexisNexis Subject	OpenStates Subject
1	G0201	Civil Rights	Civil Liberties and Civil Rights
2	G0202	Women	Sexual Orientation and Gender Issues
3	G0205	Environment	Environmental
4	G0209	Good Government	Campaign Finance and Election Issues
5	G0300	Health	Health
6	G0400	Agriculture	Agriculture and Food
7	G0600	Education	Education
8	G1000	Transportation	Transportation
9	G1200	Law	Legal Issues
10	G1500	Bank	Housing and Property
11	G1503	Sports	Recreation
12	G1510	Insurance	Insurance
13	G1600	Military	Military
14	G2400	Local Government	Municipal and County Issues
15	G2401	Police and Fire	Public Services
Unmatched codes			
x1	G0207	Religion	Animal Rights and Wildlife Issues
x2	G0208	Tax Policy	Arts and Humanities
x3	G0701	Utilities	Budget, Spending, and Taxes
x4	G0702	Natural Resource	Business and Consumers
x5	G1300	Welfare	Commerce
x6	G1400	Construction	Crime
x7	G1501	Hotel	Drugs
x8	G1502	Small Business	Energy
x9	G1504	Business Services	Executive Branch
x10	G1520	Manufacturing	Family and Children Issues
x11	G1700	Communication	Federal, State, and Local Relations
x12			Gambling and Gaming
x13			Government Reform
x14			Guns
x15			Immigration
x16			Indigenous Peoples
x17			Judiciary
x18			Labor and Employment
x19			Legislative Affairs
x20			Nominations
x21			Other
x22			Reproductive Issues
x23			Resolutions
x24			Senior Issues
x25			Social Issues
x26			State Agencies
x27			Technology and Communication
x28			Trade
x29			Welfare and Poverty

Table 9: States that report the subject of their data, as aggregated by OpenStates

State	First report	Last report
AK	2011	2018
AL	2011	2018
CA	2010	2018
CT	2011	2018
HI	2011	2018
IA	2011	2012
ID	2011	2018
IN	2011	2018
KY	2011	2015
LA	2010	2012
MD	2010	2018
ME	2011	2018
MI	2011	2018
MN	2010	2018
MO	2012	2018
MS	2011	2018
MT	2012	2017
NC	2011	2018
ND	2011	2017
NJ	2010	2018
NM	2011	2018
NV	2011	2018
NY	2011	2013
OK	2012	2016
OR	2011	2012
RI	2012	2018
SC	2011	2016
SD	2011	2018
TN	2012	2018
TX	2010	2017
UT	2011	2018
VA	2010	2018
VT	2013	2014
WA	2011	2018
WI	2011	2018
WV	2014	2018