PSJ Policy Studies Journal

Policy Studies Journal, Vol. 0, No. 0, 2018

Text as Policy: Measuring Policy Similarity through Bill Text Reuse

Fridolin Linder, Bruce Desmarais, Matthew Burgess, and Eugenia Giraudy

The identification of substantively similar policy proposals in legislation is important to scholars of public policy and legislative politics. Manual approaches are prohibitively costly in constructing datasets that accurately represent policymaking across policy domains, jurisdictions, or time. We propose the use of an algorithm that identifies similar sequences of text (i.e., text reuse), applied to legislative text, to measure the similarity of the policy proposals advanced by two bills. We study bills from U.S. state legislatures. We present three ground truth tests, applied to a corpus of 500,000 bills. First, we show that bills introduced by ideologically similar sponsors exhibit a high degree of text reuse, that bills classified by the National Conference of State Legislatures as covering the same policies exhibit a high degree of text reuse, and that rates of text reuse between states correlate with policy diffusion network ties between states. In an empirical application of our similarity measure, we find that Republican state legislators introduce legislation that is more similar to legislation introduced by Republicans in other states, than is legislation introduced by Democratic state legislators to legislation introduced by Democrats in other states.

KEY WORDS: policy diffusion, text analysis, state legislatures, asymmetric politics

识别在立法中非常相似的政策建议对于公共政策和立法政治领域的学者来说非常重要。我们需要建立数据集以准确反映跨越政策领域、管辖范围和时间的政策制定过程，然而以手工方法实现这一目标的成本过高。我们提出使用一种算法来识别立法文本中的相似文本序列（即文本重用）以衡量两项法案所提出的政策建议的相似性。我们以美国州议会的法案为研究对象。针对500, 000个法案的汇编，我们演算了三项真实值测试。首先，我们的结果表明，由意识形态上相似的政策倡议者引入的法案表现出高度的文本重用，由全国州立法会议列为相同类别政策的法案表现出高度的文本重用，并且各州之间的文本重用率与各州之间的政策扩散网络的联系程度相关。通过实证应用我们的相似性测量算法，我们发现，共和党州立法者引入的立法与其他州的共和党人引入的立法更相似，而且其相似程度要大于民主党州立法者提出的立法与其他州的民主党人引入的立法之间的相似程度

Text Reuse for Measuring Policy Similarity

The study of the adoption and diffusion of similar policies has been a central focus in political science research since at least Walker (1969), and is a common

doi: 10.1111/psj.12257

focus of researchers in state and local politics (e.g., Grossback, Nicholson-Crotty, & Peterson, 2004; Mooney, 2001; Shipan & Volden, 2008), comparative politics (e.g., Knill, 2005; Simmons & Elkins, 2004; Weyland, 2005), and international relations (e.g., Cao, 2010; Sharman, 2008; True & Mintrom, 2001). Research on public policy adoption and diffusion has conventionally relied on modestly sized, hand-coded datasets that record when a set of jurisdictions adopt one or a handful of similar policies (Boehmke & Skinner, 2012). In this paper, we draw upon a voluminous and comprehensive source of data with which to measure the consideration and adoption of similar policies—the text of legislation considered in U.S. state legislatures. We take an automated, text-as-data approach to the analysis of bill text. This vastly increases the volume of data that can be included in research concerned with the similarity of policy proposals in legislative text.

Through the development of an automated text analytic approach to measuring policy similarity, we consider several methodological challenges and broader conceptual hurdles. The central methodological puzzles are two-fold. First, we need to extract segments of text in pieces of legislation that communicate similar policy enactments. We refer to these segments of text in bills as "reused" text. Second, we need to develop a method for quantitatively scoring the reused text between bills. The main conceptual puzzle with which we engage is the potential ambiguity regarding whether two policies should be considered the same for the purposes of public policy research. Considering precedents in the literature, we argue for conceptualizing policy adoption in terms of a continuum of similarity rather than a dichotomy of equivalence.

We take a multi-pronged approach in evaluating the performance of extracting and scoring reused text, which we refer to as "text reuse," in measuring the consideration and adoption of similar laws. First, we investigate the relationship between the ideological distance of the legislators that proposed two bills and the text reuse scores between these bills. Second, we assess how well policy diffusion networks that have been established by previous research can be predicted using the amount of text reuse between states. And third, we use data on equivalent bills collected by the National Conference of State Legislatures to build an evaluation dataset containing true policy overlap to assess the accuracy of text reuse in predicting policy overlap. We find that text reuse scores between two bills provide a precise signal of the presence of similar policies within the bills. In an empirical application of our similarity measure, we find that Republican state legislators introduce legislation that is more similar to legislation introduced by Republicans in other states, than is legislation introduced by Democratic state legislators to legislation introduced by Democrats in other states.

The findings and open questions offered by our work contribute beyond the study of public policy diffusion or legislative politics. We add to the growing methodological literature in political science and public policy that is focused on the use of text-as-data methods for measurement tasks (see, e.g., Blair, Heikkila, & Weible, 2016; Grimmer & Stewart, 2013; Langer & Sagarzazu, 2017; Laver, Benoit, & Garry, 2003; Lowe & Benoit, 2013; Martin & Vanberg, 2008; Monroe, Colaresi, & Quinn, 2008; Nowlin, 2016). We also contribute to a broader movement in political science to measure and study political phenomena through the lens of relational data (e.g., Alemán & Calvo, 2013; Cranmer & Desmarais, 2016; McClurg, 2003; Ward, Stovel, & Sacks, 2011), moving the focus from the units to the connections between the units. Furthermore, the methodological approach we present would be applicable to any of the several lines of research in political science and public policy that draw upon text corpora to operationalize variables of interest (see, e.g., Hart, Smith-Howell, & Llewellyn, 1990; Mikhaylov, Laver, & Benoit, 2012; Mondou, Skogstad, & Houle, 2014; Van Atteveldt, Kleinnijenhuis, & Ruigrok, 2008; Witting, 2015; Young & Soroka, 2012).

Conceptualizing Policy Similarity

The measurement task on which we are focused is the identification of comparable policy actions (e.g., adoption, consideration) across jurisdictions—U.S. states in particular. Scholars of public policy, legislative politics, and related areas have used measurements of comparable policy actions in a variety of research tasks. These include the study of policy diffusion (Karch, 2007), the comparison of policies on a select legislative issue (e.g., Huber, Shipan, & Pfahler, 2001; Mycoff, Wagner, & Wilson, 2009), and influence/adoption of model legislation introduced by advocacy organizations (e.g., Burgess et al., 2016; Garrett & Jansa, 2015). In each of these areas of research, scholars make use of measurements of consistent policy actions across jurisdictions. Below we review a variety of methods, and conceptual definitions, that researchers have used in identifying similar policy actions.

We first consider the levels at which measurement has been conducted in past work. Public policy research on cross-jurisdiction comparable policy actions has conventionally involved the manual identification of related policies across states, countries or local governments in one or a handful of policy areas (e.g., Berry & Berry, 1990; Gilardi, Füglister, & Luyet, 2009; Krause, 2011; Simmons & Elkins, 2004; Walker, 1969). This is a manual approach to measurement at the monadic (i.e., governing jurisdiction) level. Recognizing the limitations of monadic levels of measurement, researchers have developed dyadic approaches to studying the diffusion relationships between states (Boehmke, 2009; Volden, 2006). In the most recent iteration of dyadic approaches, Desmarais, Harden, and Boehmke (2015) apply network inference algorithms to U.S. state adoption sequences in over 100 policy domains to empirically infer the underlying network through which policies diffuse. In another recent innovation in measurement for policy analysis, Garrett and Jansa (2015) analyze the text of U.S. state legislation through pairwise comparison with model legislation introduced by interest groups to identify the influence of model legislation.

Several variations of coding rules have been used to define sets of equivalent, or at least similar, policies. Perhaps the broadest approach—studying policies adopted within a domain, but moving the status quo in opposite directions—is represented by Glick (1992). Glick studies judicial enactments related to the "right to die"—the right of patients or their representatives to end the use of life-preserving medical technology. He considers an adoption to include any judicial enactment on this topic, whether they restrict medical providers' deference to the patients or enact broad patients' rights. Berry and Berry (1990) look at policy change in a uniform direction—states' adoption of lotteries. Lotteries, of course, vary in terms of their rules and financial models, but all of the policies studied by Berry and Berry (1990) moved the status quo from no state lottery to the existence of a state lottery. Volden (2006) considers a broad array of state laws—those implemented for the Federal Health and Human Services Children's Health Insurance Program (CHIP). CHIP laws were coded for six policy characteristics and then analyzed for diffusion dynamics. Last, Mooney and Mei-Hsien (1995) provides an example of quantitative coding of policies adopted in a given domain. They code the permissiveness of state abortion laws (pre Roe v. Wade) on a 5-point scale. These examples convey the variability with which scholars have defined the set of policies that are considered to be comparable across state borders.

Detecting Policy Similarity Through Text Reuse

The above review of the ways in which researchers have defined comparable policies offers clarity regarding the implicit concepts underlying policy comparability. We see that policies that are considered comparable exhibit two qualities. First, they often, but do not always, move the status quo in a similar direction. This could be thought of as ideological similarity. Second, they enact policy in similar domains—either narrowly or broadly conceived. Our review highlights that those who have studied the existence and/or diffusion of a particular policy across jurisdictions are typically not studying the adoption of equivalent policies. Rather, they deem policies comparable if they are highly similar along one or two dimensions—policy domain and (optionally) ideological direction. In their review and re-analysis of studies focused on morality policy—U.S. state lotteries in particular—Pierce and Miller (1999) carefully consider the potential for variation within a fairly narrow policy domain. As (Pierce & Miller, 1999, p. 702) note succinctly, "Even as students of morality policy have noted its distinctiveness from 'normal' policy, we need to appreciate the variation *within* the general category of morality policy." Before we describe the algorithm we use for detecting text reuse, and validate that algorithm, we draw upon these two concepts to define the latent variable we seek to measure through the assessment of text overlap.

We see past work that has identified sets of comparable policies as focusing on sets of policies that meet some threshold of similarity. As such, the latent variable that we argue below is measured effectively through text reuse is *policy similarity*. The text in legislation is an aggregate representation of the dimensions underlying the policies proposed therein. These dimensions include the domain of the policy, the ideological position enacted by the policy, the level of specificity in the policy enactment, and several other salient features of policy that are communicated through the text in legislation. We do not claim that the reuse of text can be used to measure these dimensions in a specific, disaggregated way. Rather, text reuse serves as a summary measure of the greatest overlap observed across all relevant policy dimensions represented in legislative text. Furthermore, we note that by focusing on legislative text, we are measuring the similarity of the intended policies in terms of how those policies reflect in formal enacting legislation. Once enacted, the policies that are implemented in different states, under different bureaucratic and judicial regimes, will likely differ even if they are enacted under equivalent legislative language. This divergence in policy implementation is attributable to the inevitable discretion exercised by bureaucratic officials in implementing policy (Gerber, Maestas, & Dometrius, 2005; Scott, 1997; Weingast & Moran, 1983).

We should emphasize at this point that we are not the first scholars of legislative politics, or even state legislative politics, literature to use text similarity between bills to measure policy similarity. Wilkerson, Smith, and Stramp (2015) use text similarity—indeed, the same algorithm we use—to trace the development of very similar policy proposals in legislation produced in the U.S. Congress. Garrett and Jansa (2015) and Burgess et al. (2016) both use measures of text similarity in U.S. state legislation to identify bills that incorporate similar forms of model legislation, as provided by legislative/policy advocacy groups. Our contributions to this developing literature are two-fold. First, we broaden the scope of measurement beyond tracing single ideas through bill versions or detecting model legislation—to assessing policy similarity between bills in general. Second, we present multiple validation tests to assess whether our text-reuse–based measure of policy similarity correlates with benchmarks that we deduce from theories of state policymaking.

Detecting Text Reuse

To assess the maximal similarity of policies proposed in two pieces of legislation, we seek to identify large segments of equivalent or highly similar text. Following Wilkerson et al. (2015) and Burgess et al. (2016), we use an algorithm that considers the use of similar words aligned in similar sequence—an alignment algorithm—to identify large segments of similar text in bills. Alignment algorithms stand in contrast to bag-of-words methods, which compare documents in terms of the use of similar words, but do not account for the ordering/sequence in which the words appear. Bag-of-words methods may be effective for capturing broad topical areas in legislation, but will provide results that are too coarse for precisely assessing policy overlap. For example, in modeling bills in the U.S. Congress using a very popular bag-of-words method for text analysis, probabilistic topic modeling (Nowlin, 2016), Gerrish and Blei (2011) find that topic models for legislative text perform most effectively when parameterized with 64 topics to model a corpus of 4,447 bills. While 64 topics provides a fairly detailed categorization of the domains, or "latent themes" (Nowlin, 2016, p. 315), covered in legislation, it is unrealistic to think that topic modeling can isolate individual policy proposals. If the topics identified in Gerrish and Blei's (2011) results corresponded to individual policy proposals, that would mean that the U.S. Congress considers nearly 70 bills, on average, for each specific policy proposed.

There are a variety of algorithms designed to automatically detect text reuse between documents. Most originated in the plagiarism detection literature (see, e.g., Potthast et al., 2013, for an overview). In political science, Wilkerson et al. (2015) introduced the Smith-Waterman local alignment algorithm (SW algorithm) to detect overlapping language in congressional bills in order to trace policy ideas through the legislative process of the U.S. Congress. We use the same algorithm to detect text reuse in U.S. state bills; however, our implementation of the algorithm differs slightly (see below). The SW algorithm was developed by Smith and Waterman (1981) in molecular biology, in order to match genetic sequences. Given two sequences of words it calculates the optimal alignment between these sequences. That is, it finds the longest subsequences that overlap between the two sequences. However, sequences in text that convey similar—even equivalent—policy content often do not match perfectly. If not permitted, mismatches in formatting, white space, typographical errors, and minor variations on wording would decrease the recall—the percentage of matches identified by the text reuse algorithm. The SW algorithm, therefore, returns the optimal alignment while allowing for mismatches and gaps. The extent to which such imperfections are tolerated is governed by parameters that are set by the researcher. There are three such parameters: the match score (reward for exactly matching words), the mismatch score (penalty for including words that do not match), and the gap score (penalty for including placeholder words in matching a shorter sequence to a longer sequence). The goal of the algorithm is to return the alignment with the highest score given the input sequences and the parameters. Consider the following two excerpts from Alaska State Bill 203 (28th session) and North Carolina House Bill 366 (2015), respectively:

> section 1.AS44.99 is amended by adding new sections to read: article 6. compact for a balanced budget (...)
>
> chapter 143 is amended by adding a new article to read: article 80. compact for a balanced budget (...)

These sequences could be aligned in a large number of possible ways. There is a trade-off between short sequences that match perfectly, and longer sequences with imperfections. For example, we could match just "is amended by adding new" or "compact for a balanced budget." However, by adding a gap between "adding" and "new" in the first sequence and accepting the mismatches "article"-"section" and "6."-"80.", a longer alignment can be obtained: "is amended by adding *(a)* new *(article/section)* to read: article *(6./80.)*." The SW algorithm amounts to a systematic procedure for scoring similar sequences of text, and efficiently finding the highest scoring sequences in two documents. We explain the SW algorithm, and our implementation, in detail in the Appendix.

Policy Similarity in U.S. State Legislation

Through the application of the SW algorithm to U.S. state legislation, we provide a quantitative measure of policy similarity between bills. The similarity measures between bills that we identify can be aggregated up to the legislator or state levels. The measures we provide can be used to test hypotheses regarding, among other topics, public policy diffusion, legislative politics, political parties, and interest groups.

Data

We apply the SW algorithm to a database of U.S. state bills, collected by Burgess et al. (2016) and the Sunlight Foundation.[1] This database contains approximately 500,000 bills from 2008 to 2015. Not all bills from all states are available for the whole time period. Figure 1 displays the year ranges and number of bills in the bill database. This collection of bills is based on all bills that are available through the Sunlight Foundation's openstates.org API. openstates.org is a website maintained by the Sunlight Foundation, in order to increase transparency in state politics. The Sunlight Foundation uses web scrapers (programs that automatically collect data from web sources) to access all bills that are available on the websites of state legislatures. As long as the bills can be successfully collected from the website of a legislature, there is no reason to expect bias with respect to the bills covered, as the scraper targets all available bills. There is, however, variation across states in terms of the time periods—especially going back in time—for which the bills can be successfully collected. The dataset includes enacted legislation as well as bills that are still in the legislative process, or were not enacted. Several recent research projects have used openstates.org as a data source (e.g., Burgess et al., 2016; Cook, 2017; Garcia-Olano, Arias Vicente, & Larriba Pey, 2016; Tremblay et al., 2015).

Our approach of applying the SW algorithm is similar to Wilkerson et al. (2015) but differs in several key respects. The analysis consists of three major steps: (i)

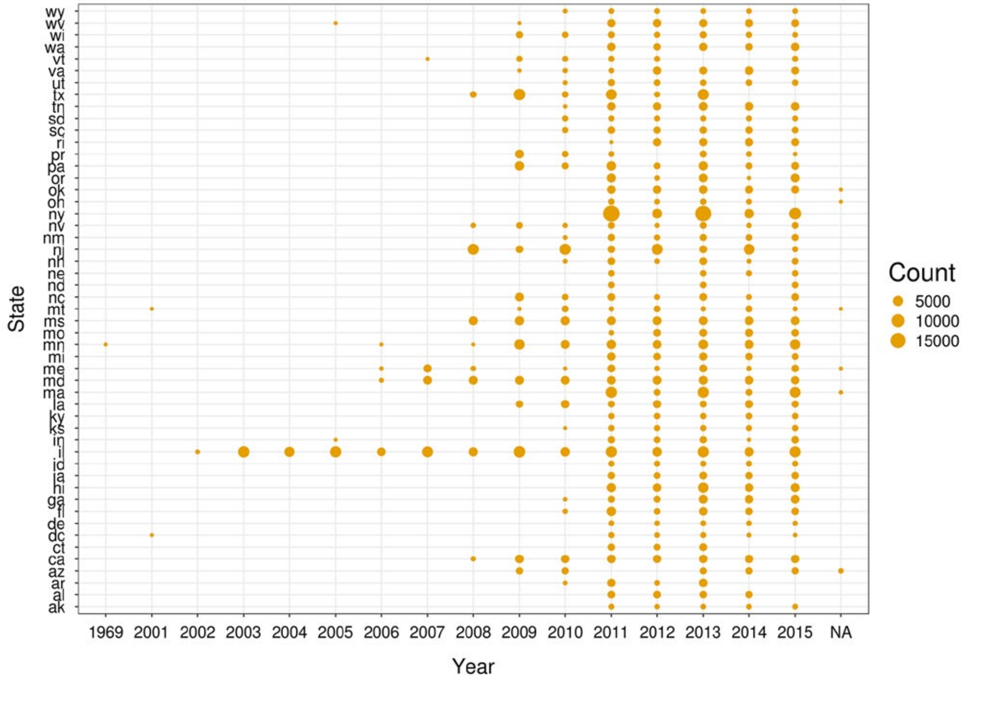


Figure 1. Availability of State Bills Over Time and By State.

preprocessing and selection of documents to compare, (ii) alignment computation, (iii) post hoc alignment adjustment. We describe each of these steps below.

Preprocessing and Selection

The SW algorithm is computationally demanding and an exhaustive comparison of all possible bill-pairs is not feasible. Wilkerson et al. (2015) approached the computational burden of the SW algorithm by excluding bill pairs that do not share at least five unique sequences of 10 words (i.e., 10-grams). Requiring several matching 10-grams is a high threshold to meet, but it aligned with Wilkerson et al.'s (2015) goal of finding major passages of legislation that had been copied exactly. Since this implies considerable exact character matches, we see this as too restrictive for our application. Because the text we are using originates in 50 different legislatures, and we are interested in identifying similar policy proposals, and not necessarily sections that were copied verbatim, exact character matches might obfuscate valuable text reuse just because of small formatting changes, changes of state names, etc. Instead, we choose to rely on a procedure that is based on finding bills with similar language.

We take a more inclusive approach than Wilkerson et al. (2015) in selecting the other bills to which a bill is compared. We use a tool from the literature on search and query in computer science—Elastic Search (Gormley & Tong, 2015)—to find the 500 bills that are most similar to each bill. We calculate the alignment scores between each bill and the 500 bills to which it is most similar based on Elastic Search. Details on Elastic Search are provided in the Appendix. Our criteria for comparing bills are more inclusive than those used by Wilkerson et al. (2015), since we are assessing the general similarity between policy provisions in the bills, not just searching for nearly replicated bill sections. In order to account for the possibility of selection bias being introduced to our analyses through the cap of 500 similar bills, we analyzed the results with different numbers of preselected bills. We find that this parameter, when set to 1,000, does not lead to such bias because the set of bills that result in modest or large text reuse scores is always smaller than 500.

Bill Alignment Computation

Once the candidate bills are selected, we calculate alignments for all possible bill dyads. Bill pairs from the same state are available, but excluded for the analyses below. We exclude same-state comparisons in the current analysis since our validation exercises are focused on cross-state comparisons. One challenge that arises when working with legislative text is that bills contain a large amount of procedural text, or "boilerplate" (Wilkerson et al., 2015), that is unrelated to the policies enacted in the bills, but could nonetheless drive text reuse scores if there is similar boilerplate in other bills. Consider, for example, New Jersey Assembly Bill No. 2336, as introduced on February 4, 2016. This bill, which establishes limitations on health insurers' policies toward optometrists, includes three lines of moderate length that

Table 1. Alignment Example.

Left text	Right text	Score
nj_214_A1167: “the entire credit may not be taken for the taxable year in which the **renewable** energy property is placed in service but must be taken in five equal - installments beginning with the taxable year in which the **renewable** energy property is placed in service. if, in one of the years in which the installment of a credit accrues, the **renewable** energy property with respect to which the credit was claimed is disposed of, taken out of service, or **moved** out of **state,** the credit expires and the taxpayer may not take any remaining installment of the credit. the taxpayer may, however, take the portion of an installment that accrued in a previous year and was carried forward to the extent permittedunder”	**nc_2011_SB747**: ”the entire credit may not be taken for the taxable year in which the **facility** - - is placed in service but must be taken in five equal annual installments beginning with the taxable year in which the **facility** - - is placed in service. if, in one of the years in which the installment of a credit accrues, the **facility** - - with respect to which the credit was claimed is disposed - - - of - or **taken** out of **service,** the credit expires and the taxpayer may not take any remaining installment of the credit. the taxpayer may,however, take the portion of an installment that accrued in a previous year and was carried forward to the extent permitted under”	388
nj_214_A1167: “the entire credit may not be taken for the taxable year in which the renewable energy property is placed in service but must be taken in **five** equal installments - - - - - beginning with the taxable year in which the”	**ga_2011_12_HB146**: “the entire credit may not be taken for the - year in which the - - property is placed in service but must be taken in **four** equal installments over four successive taxable yearsbeginning with the taxable year in which the”	110
nj_214_A1167: “the entire credit may not be taken for the taxable year in which the **renewable energy property** is placed in service but must be taken in five equal installments beginning with the taxable year in which the”	**nc_2009_SB305**: “the entire credit may not be taken for the taxable year in which the **costs are paid** - - - -but must be taken in five equal installments beginning with the taxable year in which the”	108

Notes: Bold texts are mismatches. Dashes are gaps introduced by the alignment algorithm. Nonhighlighted text are matches.

have nothing to do with the proposed policies. One example that could result in an alignment of length 10 or more with bills from other states is, "Be It Enacted by the Senate and General Assembly of the State of New." See the next section for a discussion of how we adjust the alignment scores in order to take into account boilerplate alignments. The SW algorithm returns parts of the two bills under comparison that are determined to be aligned, which we refer to as the "left text" and "right text," as well as a score. We refer to these three pieces of information—the left text, right text, and score—as an alignment.

Table 1 displays three examples of alignments. The "Left Text" and "Right Text" columns display the bill text segments, and the alignment score is displayed in the "Score" column. This example shows the reuse of legislative text in three states (New Jersey, North Carolina, and Georgia).

For our data, we obtain from this procedure approximately 225 million individual alignments among pairs of bills. Figure 2 displays the empirical cumulative frequency distribution of the alignment scores. The alignment score is given on the x-axis, on the logarithmic scale. The y-axis gives the proportion of alignment scores less-than or equal-to the respective alignment score. The lowest score is 3, the highest 21,000 with a mean alignment score of 21.8 (median 18). From this distribution, it is clear that the majority of alignments are small with only relatively few bill pairs having large alignments. This is an intuitive finding, indicating that the overwhelming majority of bill pairs do not introduce similar policies. Of the 225 million bill pairs about 1.3 million have scores larger than 100 and approximately 43,000 have scores larger than 1,000.

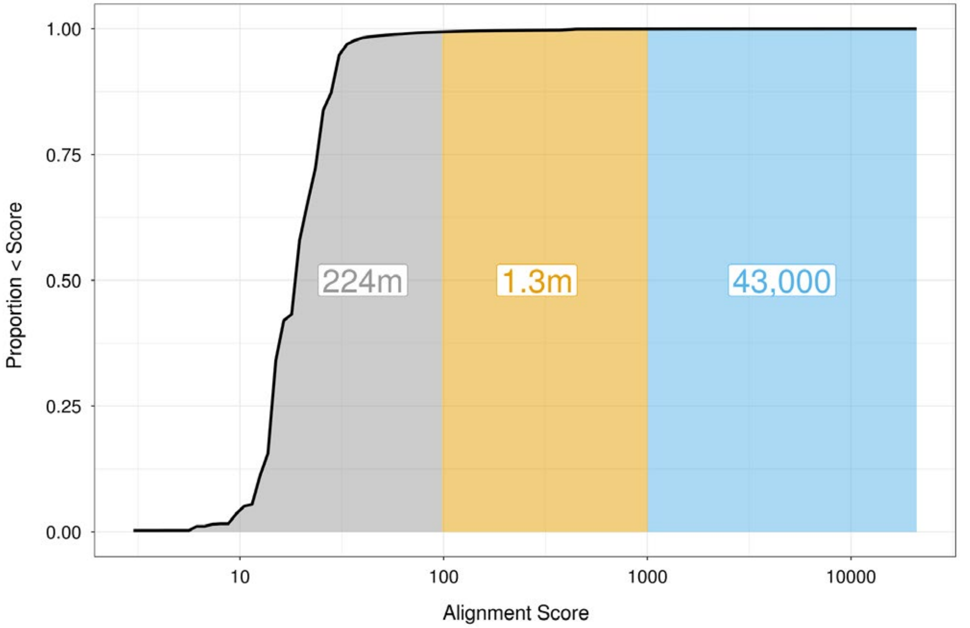


Figure 2. Cumulative Frequency Distribution of the Dyad-Level Alignment Scores.
Note: The x-axis represents the dyad-level alignment score on the log_{10} scale.

Adjusting for Boilerplate Text

Legislation is full of textual content that is not very useful in identifying the novel policy proposal(s) offered by the bill. Such text includes procedural language, headers, section titles, and common phrases such as definitions of legal terms. Text that is irrelevant to specific policy proposals represents a major—perhaps the primary—limitation on the capacity for text reuse to identify the emulation of policies across bills. Wilkerson et al. (2015) use human coders to identify alignment text that is unrelated to policy content. Given a large set of hand-labeled alignments, they estimate a model using textual content as the independent variables and the boilerplate/not boilerplate indicator as the dependent variable. That model is then used to classify all of their alignment text as either boilerplate or not. They exclude alignments classified as boilerplate. This represents one possible approach, but has two limitations. The first is that the use of human coders to label data to train the classifier is costly. The second is that the classification and exclusion approach treats legislative text as conforming to a procedural-or-not dichotomy, where language in a bill may be situated on a continuum in terms of the degree to which it contributes to the policy proposal(s) offered in the bill.

Our approach to managing boilerplate does not involve the application of a procedural-or-not dichotomy. Rather, we down-weight alignments that consist of language that is very common to all alignments. The underlying logic is that language that appears in a large number of other alignments is probably not related to specific policies. For example, alignments consisting of many stop words such as "the" or "a" or of common procedural language like "legislature" or 'section" are most likely part of phrases that are either very frequent in common language or very frequent in the legislative context. In order to quantify the similarity of an alignment's content to the contents of other alignments, we calculate the cosine similarity—a common measure of the similarity of the vocabularies in two documents (Garrett & Jansa, 2015)—of each alignment with a random sample of 1,000 alignments.[2] The adjusted score for two bills X and Y is then:

$$S^{*}_{X,Y} = S_{X,Y}\left(1 - \frac{1}{1000}\sum_{i=1}^{1000}\frac{\mathcal{A}_{X,Y}\bullet\mathcal{A}_{\mathbf{i}}}{\|\mathcal{A}_{X,Y}\|\,\|\mathcal{A}_{\mathbf{i}}\|}\right) \tag{1}$$

where $S_{X,Y}$ is the alignment score between bills X and Y, calculated as described above, $\mathcal{A}_i$, and $\frac{\mathcal{A}_{X,Y}}{\bullet}\mathcal{A}_i\|\mathcal{A}_{X,Y}\|\|\mathcal{A}_i\|$ is the cosine similarity of the alignment between bills X and Y to the i^{th} alignment in the random sample of 1,000. The cosine similarity is a score between 0 and 1, which means that the quantity in the parentheses will approach zero as the alignment between bills X and Y becomes more similar to other alignments. In this approach, we follow the common intuition in natural language processing that the informativeness of text is inversely related to the frequency with which that text occurs (Robertson, 2004). Though we cannot speak to whether or not our approach is ultimately better than that taken by Wilkerson et al. (2015), in which they train a classifier to recognize and exclude boilerplate text based on human-labeled data, we note three relative strengths of our approach. First, we do not rely on the development and implementation of a human coding scheme, which is costly and always exhibits some degree of coding error. Second, our approach results in a continuous measure of the information content of text rather than a dichotomous classification of boilerplate or not. Third, and related to the first strength, our approach assigns an informativeness score to all word types, whereas reliance on human classifiers necessarily limits the boilerplate contributions of word types to the word types that were included in the human-coded set. Our approach in adjusting SW alignment scores for boilerplate is novel, and represents a contribution to the voluminous literature on the measurement of text similarity. All this said, dealing with boilerplate/procedural text is an exciting open area of methodological inquiry for quantitative text-as-data research that is focused on legislation and other formal legal documents.

Methodological Workflow Overview

In this section, we provide an overview of the entire workflow that we use to go from raw bill text to alignment scores. Starting from the set of bills collected by Burgess et al. (2016), we use the following steps to get from the bill text to the dyadic alignment scores. First all bills are imported into a database. For each bill "b" in the resulting database we then calculate the raw alignment scores on the 500 most similar bills to "b" using the SW algorighm. After obtaining the raw scores and the text associated with the respective alignment, we use the text to reweight the raw alignment score in order to decrease the scores of boilerplate alignments (as discussed above). This workflow is reiterated in the following steps.[3]

1. Scrape bills from state legislatures' websites using the Sunlight Foundation's scrapers (done by Burgess et al., 2016)
2. Import bills into elastic search database
3. Remove bills that have no text
4. For each bill b:

- Query the 500 most similar bills (S) to b, as described earlier
- For each bill pair in (b × S) use the Smith-Waterman algorithm (our implementation in Python/Cython) to find the longest aligned text sequence
- For each bill pair in (b × S) use the boilerplate algorithm to reweight its alignment score

Empirical Evaluation of Validity

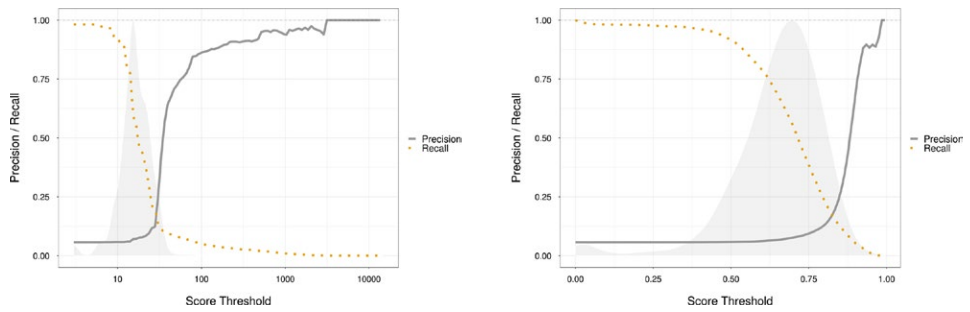
The main empirical analysis in this paper is an extensive measurement validation study. We evaluate the performance of alignment scores in empirical tests in which they should perform well if the alignment scores we introduce represent an effective measure of policy similarity. Though beyond the scope of the current paper, in the conclusion we discuss several substantive questions that could be addressed using the policy similarity scores we provide. We use three complementary tasks in order to assess the validity of text reuse as a measure of policy similarity. First, we evaluate whether there is a disproportionately high number of alignments between bills on the same policy area. Second, we analyze how well the bill similarity aggregated to the state level corresponds with policy diffusion networks identified in previous research (Desmarais et al., 2015). And third, we study the relationship between the ideological distance between the sponsors of bills and the alignment scores between those bills. If substantive policy content, and importantly, the same ideological direction of the provisions, is detected by the alignment algorithm, we will expect an inverse relationship between these two measures. In the validation experiments that follow, we make use of the quantitative, interval level, alignment scores to construct continuous measures of policy similarity. We do this to impose as little information reduction as possible. In the conclusion, we discuss the possibility of thresholding these scores in order to classify bill pairs as either containing or not containing matching policy content.

Predicting Co-Placement in NCSL Tables

In our first validation exercise, we consider whether high alignment scores provide a reliable signal of similar policy proposals. To build a ground truth dataset in terms of similar policy proposals, we rely on thematic tables published by the National Conference of State Legislatures (NCSL).[4] The NCSL is a nonpartisan organization that provides informational assistance to state legislatures, and maintains databases of state-level legislation and law (Rivera, 2014). The NCSL tables are regularly used in state politics research to identify comparable bills and policies across states (see, e.g., Brown, 2013; Commins & Wills, 2016; Marquez & Schraufnagel, 2013; Witmer, Johnson, & Boehmke, 2014; Ybarra, Sanchez, & Sanchez, 2016). We collected a sample of these tables with the following procedure:

1. Collect all URLs returned by the web search query: "site:ncsl.org 'legislation.aspx'"[5]
2. Sample 100 URLs
3. Select bills that fulfill the following three conditions:
   a. The website is a NCSL table
   b. The table refers (at least in part) to legislation introduced in 2011 or later
   c. The table contains less than 100 individual bills
4. Collect all bills from these tables

Condition (a) is necessary, since some of the URLs returned by the web query refer to blog entries or to protected websites that require NCSL membership. The rationale behind condition (b) is to maximize the number of bills we can match to our database (see Figure 1), which is most complete after 2011. We additionally constrain the number of bills per table to be below 100, in order to avoid tables that cover very broad topics. We chose 100 since this number would correspond roughly to two bills per state (often there is a House and a Senate bill in each state for a particular policy). The web query from step 1 returned 266 URLs. From the 100 sampled URLs, 34 fulfilled the three criteria. From these 34 tables, we constructed a dataset of 1,573 bills,



(a) Precision and Recall curves for classifier based on alignment scores.

(b) Precision and Recall curves for classifier based on cosine similarity.

Figure 3. Precision and Recall Curves for Classifying Two Bills being in the Same NCSL Table by Thresholding the Bill-Dyad Alignment Score.
Notes: The dotted line is recall, the solid line is precision. The x-axis displays the threshold, the y-axis represents precision and recall. The shaded area in the background displays the frequency distribution of the alignment scores.

of which we could match 1,000 to our bill database. We then calculate the complete set of pairwise alignments for these bills. After removing same-state bill pairs we obtained approximately 490,000 alignments.

Figure 3 displays the performance of the alignment score (left panel) in predicting whether two bills are in the same NCSL table compared to the performance of a simple cosine similarity measure (right panel). We present precision—the proportion of bill pairs with a given alignment score that exceed the threshold on the x-axis that are in the same NCSL table, and recall—the proportion of same-NCSL-table pairs that exceed a given alignment score threshold. We see very high precision when setting the threshold to relatively high scores (>100). Recall is low at thresholds that produce reasonable precision (above 100), but that is expected for our case, as (i) the introduction of similar policies in bills in two different states is a rare event, and relatedly, (ii) the appearance of two bills in the same NCSL table is a relatively rare event (6 percent in the current analysis); classifier recall is notoriously low in rare event data (Weiss & Hirsh, 2000)—even commonly approaching zero, as in our case (Weiss, 2004). Furthermore, using a high precision–low recall classifier in our scenario where the total number of bill pairs is extremely large produces a large number of interesting bill pairings in absolute terms.

Table 2. Area Under the Precision-Recall Curve for Classifier Based on Thresholding Alignment Scores (Alignment), Classifier Based on Thresholding Cosine Similarity Score (Cosine) and Random Classifier (Random).

	AUC	P(X<Random)	P(X<Cosine)
Random	0.06		
Cosine	0.15	0.00	
Alignment	0.17	0.00	0.00

Notes: The last three columns report one-tailed p-values for comparison between the AUCs. p-values are derived from 2000 nonparametric bootstrap iterations.

For a better understanding of how the alignment score classification performs in comparison to other methods, in Table 2 we report the area under the precision-recall curve for the alignment classifier compared to a classifier based on cosine similarity of bill pairs and a random classifier. The area under the precision-recall (PR) curve is calculated by constructing a curve of recall against precision (or vice versa), and taking the area beneath that curve. Like the area under the receiver operating characteristic (ROC) curve, the area under the PR curve is bound between 0 and 1, with higher values indicating greater predictive performance. However, the area under the PR curve is superior in cases of rare event data, such as the data with which we are working (Cranmer & Desmarais, 2017). The first column contains the AUC values for each classifier. The remaining two columns display bootstrapped p-values for the differences between the classifiers. For each bootstrapped sample, we re-sample bill pairs with replacement. We see that the alignment scores exhibit an approximately 10 percent higher AUC than cosine similarity—a difference that is statistically significant at conventional levels. Our results suggest that the presence of a large alignment score between bills is an effective indicator of policy similarity, and that alignment scores perform better than cosine similarity, a common bag-of-words approach.

Diffusion Networks and Text Reuse

To evaluate whether text reuse corresponds to the transfer of policy, we test whether the presence of a diffusion network tie between two states is a predictor of text reuse. We use the policy diffusion networks inferred in Desmarais et al. (2015). The diffusion networks were inferred using policy adoption sequences, and the network inference algorithm developed by Gomez Rodriguez, Leskovec, and Krause (2010). A tie from state i to state j in the diffusion network indicates that state j has frequently emulated state i's policies in the preceding 35 years. To calculate an aggregate alignment score for each state-pair we calculate the sum of the alignment scores associated with each pair of bills across the two states. The "Diffusion Ties" variable indicates the presence of a diffusion edge between states in the 2008 diffusion network, as measured by Desmarais et al. (2015). The diffusion network in 2008 is inferred using policy adoptions in the 35 years preceding (and excluding) 2008. There is one observation in the analysis for each of the 1,225 unique state-pairs. Since this is dyadic data, we use a matrix permutation method, quadratic assignment procedure, to calculate p-values (Krackhardt, 1988). As a robustness check, we

Table 3. Predicting Aggregate Cross-State Alignment Scores with Diffusion Ties.

	Identity Link		Log Link	
Intercept	–2.883	0.002	7.776	0.000
Diffusion tie	0.441	0.005	0.381	0.006
Coverage	0.951	0.001	1.106	0.001

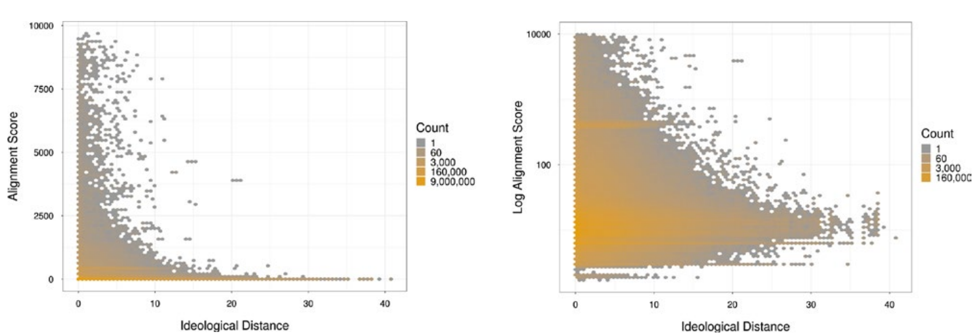
Notes: Coefficients calculated with OLS regression and normalized with standard deviation of cross-state alignment scores. p-values based on 5,000 QAP permutations.

run the model with both the identity and log link.[6] We also control for a dyadic variable (Coverage) equal to the product of the number of years for which we have data for each state in the dyad.

Results of the dyadic regression are presented in Table 3. In both specifications there is a positive relationship between the number of diffusion ties and the number of alignments. The relationship is statistically significant at the 0.05 level in each analysis. Furthermore, the magnitudes of the relationships are substantively significant. With the identity link, the presence of a diffusion tie leads to a 0.44 standard deviation increase in the expected aggregate alignment score. Based on the log link, the addition of a diffusion tie corresponds to a 46 percent increase in the expected alignment score.[7] These results offer further evidence of the validity of quantifying text reuse as a measure of policy similarity, as the aggregate volume of text reuse between states is positively associated with previously identified diffusion ties between states.

Text Reuse and the Ideological Distance Between Sponsors

In the previous two validity tests, we focused mainly on the policy domain aspect of policy similarity, rather than ideological similarity. In the third and final validity test, we ask whether we observe a greater volume of text reuse between bills introduced by legislators who are ideologically similar. For calculating ideological distance, we rely on latent ideology scores measured by Shor and McCarty (2011). The dataset contains scores for 20,738 legislators from 50 state legislatures. The `openstates` API, which we can match to our bills data, contains data and identifiers on 12,000 legislators. Of these, we were able to uniquely match about 8,000 legislators to their records in the ideal point data. This allowed us to obtain ideal points for the sponsors of 60 percent of bills and $88^{*}10^{6}$ (40 percent) pairs of bills. In order to assess the validity of text reuse as a measure of substantive policy overlap, we expect the quality of the alignments to be inversely correlated to the distance between the bill sponsors' ideal points. The ideal points are located on a common scale across all states; the distance between sponsors from different states is, therefore, meaningful.



(a) Hexbin plot with alignments on natural scale. **(b)** Hexbin plot with alignments on log_{10} scale.

Figure 4. Hexbin Plot of Alignment Score and Ideological Distance of All 90 million Bill Dyads. *Note:* The shading of the bins indicates the density of data in the bin.

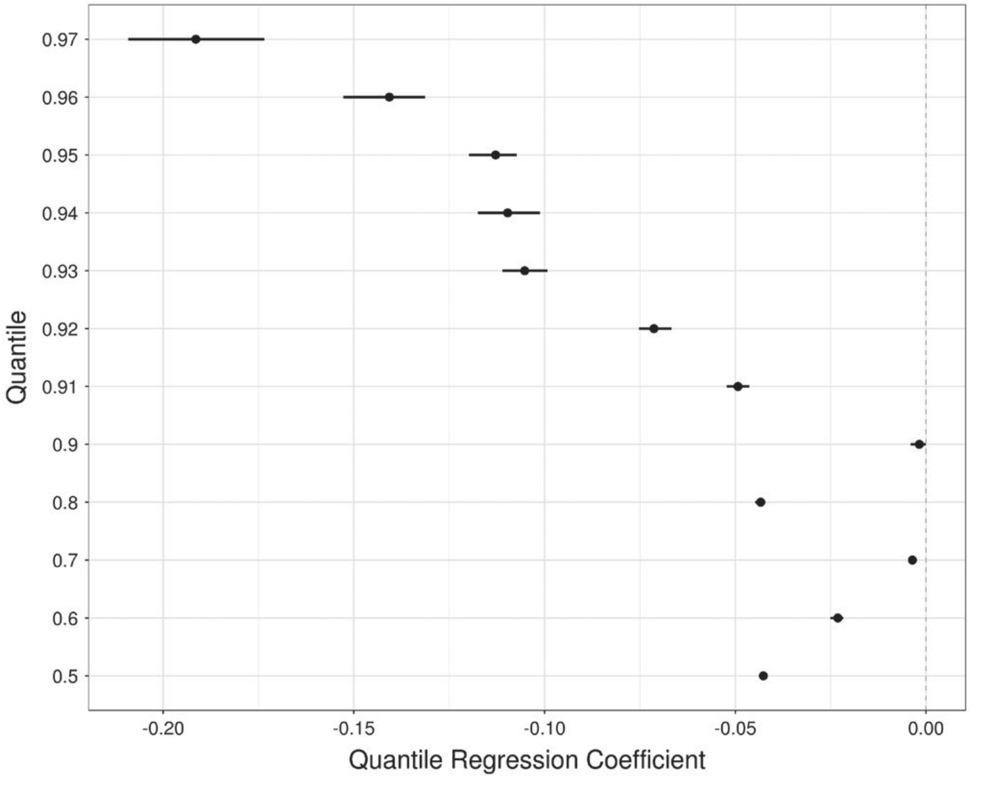


Figure 5. Coefficients of Quantile Regression for Alignment Score on Ideological Distance. *Notes:* Dots indicate the point estimate. Horizontal bar represent 95 percent confidence intervals obtained with clustered bootstrap. Note that the y-axis is not in a natural scale.

In the following sections, we present several analyses to assess this correlation. The ideal point of the bill is derived from the ideal point of the primary bill sponsor. For bills with several primary sponsors, we use the average of those sponsors' ideal points to obtain a single measure. Figure 4 displays the distribution of ideological distance and log-alignment score for the roughly 90 million dyads in the dataset. The left panel displays the alignment scores on the natural scale. In order to get a better view of the distribution for small alignments, the right panel is on the logarithmic scale. Each dot is a hexagonal tile, the shading of the tile represents the density in the area of the tile. Two major observations can be made in this plot. First, the vast majority of alignment scores are close to zero. Second, the triangular shape of the distribution indicates the high precision–low recall character of the alignment score discussed above: few bills exhibit meaningful policy overlap with other bills, but in the region of lower ideological distance we observe a greater occurrence of positive outliers, which arise from significant policy overlap.

In order to investigate this pattern more closely, we additionally calculate quantile regressions for the relationship between the alignment score and the ideological distance. In a quantile regression, the dependent variable is the value of the predicted quantile of the distribution of the dependent variable conditional on the

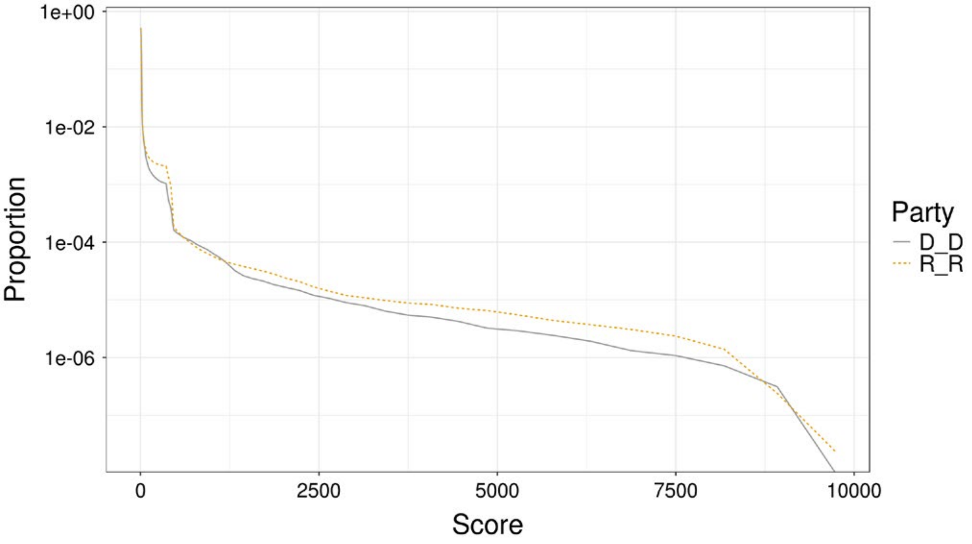


Figure 6. Cumulative Distribution (Proportion Greater Than) of Alignment Scores for Republican and Democratic (Primary Sponsor) Bill Dyads.

independent variables (e.g., the predicted median for median regression with the quantile set to 0.50 [Harden & Desmarais, 2011]). Figure 5 displays the coefficients from quantile regressions on a sequence of quantiles from the median to the 0.97th quantile.[8] Confidence intervals are calculated by resampling bills, which represents an implementation of the clustered nonparametric bootstrap (Harden, 2011). We see that there is only a very weak relationship between the ideological similarity of bill sponsors and the median alignment scores between bills. However, as the quantile we model increases we see that the relationship grows strong and negative in magnitude, and is statistically significant. These results both validate the use of text reuse as a measure of policy similarity, and reinforce the high precision and low recall nature of text reuse as a measure of policy similarity.

Asymmetric Politics and Policy Similarity Across the U.S. States

In this section, we illustrate at least one way in which the similarity scores provided by our method can be used to advance our understanding of politics and public policy. We adapt and test a theory regarding structural differences between the Democratic and Republican parties, which has been developed through the study of public opinion and electoral politics, to the context of policymaking in state legislatures. Grossmann and Hopkins (2015) presents extensive analyses of public opinion data, through which they demonstrate that those who support the Republican party make up a coherent ideological movement that rewards "doctrinal purity." On the other hand, they find that Democratic supporters form a loose coalition of supporters of various progressive interests who reward concrete achievements that are relevant to their specific policy goals. Grossmann and Hopkins (2016) term this difference between the parties "Asymmetric Politics." Grossmann and Hopkins (2016) consider the implications of this difference for legislative productivity, and show that the U.S. Congress tends to be more productive when under Democratic control. In further research on Congress, Russell (2017) finds that this pattern of asymmetric politics characterizes the rhetoric of U.S. Senators on Twitter.

We build upon the recent work by Grossman and Hopkins to analyze the implications of the Asymmetric Politics theory for the similarity of policies proposed by Republicans and those proposed by Democrats in U.S. state legislatures. Grossmann and Hopkins (2015) paint a picture of Republican supporters in which the most salient feature of a public policy proposal is whether or not it fits within the broad conservative agenda, whereas individual Democratic supporters are more likely to be motivated by interest in specific policy issues. We deduce from this difference that policies proposed by Republican legislators in one state are more likely to be attractive to Republican legislators in other states than are policies proposed by Democrats in one state likely to be attractive to Democratic legislators in other states. Since Republican legislators face the criterion of passing policy that fits within a broad philosophical agenda, legislation proposed by Republicans is likely to include policies that would be supported by Republican constituents in other states. However, since policy-relevant variables such as demographics, industry composition, and economic conditions vary considerably across states, we expect the specific policy goals of the voters that make up the Democratic coalition to also vary considerably across states. In this section we test the hypothesis that policies proposed by Republican state legislators are more similar to policies proposed by Republican legislators in other states than policies proposed by Democratic state legislators are to policies proposed by Democratic legislators in other states.

The empirical investigation of our Asymmetric Politics hypothesis is focused on the differences between the distributions of alignment scores calculated on pairs of bills sponsored by Republicans in different states and pairs of bills sponsored by Democrats in different states. These distributions are depicted, as empirical cumulative distribution functions, in Figure 6. In Figure 6, the alignment score is given on the x-axis, and the y-axis gives the proportion of scores greater than or equal to the respective score value. The y-axis is depicted on the log scale. We see from this figure that, especially among the large scores that strongly indicate the presence of similar policies, the proportions for Republican bill dyads are higher than those for Democratic bill dyads. Basic summary statistics for the distributions of bill similarity scores are given in Table 4. These statistics fit with our theoretical claim—the Republican dyad distribution has a larger mean, median, 95th percentile, 99th percentile, and maximum than the Democratic dyad distribution.

Table 4. Summary Statistics for Alignment Scores in Dyad Types.

Party Dyad	Count	Mean	Median	Maximum	95th percentile	99th percentile
D-D	21,411,237	13.16	10.27	9,501.88	20.40	50.17
R-R	20,275,515	14.20	10.45	9,729.71	20.46	58.82

Table 5. Distribution of Bill Scores by Party Dyad.

	Democratic	Republican	Ratio D/R	CI low	CI high
% > 10	57.7605	59.2424	0.9750	0.9722	0.9777
% > 100	0.4330	0.6557	0.6604	0.6233	0.6962
% > 1000	0.0130	0.0119	1.0874	0.9007	1.2898
% > 5000	0.0006	0.0013	0.4641	0.2556	0.7942
% > 7000	0.0002	0.0006	0.4070	0.1603	0.8614

Notes: Ninety-five percent confidence intervals based on 1,000 bootstrap iteration clustered on left bill ID.

In Table 5, we present tests of whether the differences between distributions can be considered statistically significant. We look at five rounded score thresholds that correspond to approximately the 50, 99.50, 99.90, 99.9990, and 99.9995 percentiles—focusing on the extreme right tail that provides the strongest indication of similar policies. We test whether the proportion of scores that exceed the threshold among Democratic dyads is greater than the proportion of scores that exceed the threshold among Republican dyads. Specifically, we calculate the D/R ratio of these two proportions, and construct 95-percent block-bootstrap confidence intervals by resampling left bills of dyads with replacement and calculating the ratio from the resulting sample of dyads (using 1,000 bootstrap iterations). For all but one threshold (1,000), the ratio of D/R is statistically significantly lower than 1, indicating that Republican dyad scores exceed the thresholds at greater rates than do Democratic dyad scores. The results in this section support the hypothesis that Republican state legislators introduce legislation that is more similar to legislation introduced by Republicans in other states than is legislation introduced by Democratic state legislators to legislation introduced by Republicans in other states.

Conclusion

Scholarship on public policy and legislative politics relies heavily on measuring the content of legislation, especially with a relative or comparative approach. Manual comparison of bills is prohibitively time-consuming when it comes to covering a large proportion of legislation introduced in one or more legislatures. We show that the automated detection of similar text strings in bills provides an effective approach to comparing bill contents. The log alignment scores serve as a highly valid summary measure of the similarity of two bills. We show high validity of this measure in three complementary tests. First, it correlates with the presence of a diffusion tie inferred from patterns of policy adoption. Second, it correlates with the ideological similarity between bill sponsors. Third, it exhibits high precision in predicting whether two bills are listed in the NCSL policy area tables.

The alignment scores we derive provide a resource for scholars to test hypotheses of the causes and consequences of the introduction and adoption of similar policies—a measure which crosses different policy areas, states, and time. In a validation experiment, we show that a wide span of quantiles of alignment scores between bills—from the median up through and above the 95th percentile—are negatively correlated with the ideological distance between the legislators who introduced the two bills. This result opens the door to many other hypotheses that could be tested using alignment scores for bill pairs. Where and when do similar policies cross state borders? Which legislators introduce similar legislation? Does lobbying by pressure groups result in the adoption of similar policies across states? Do similar patterns of campaign contributions to legislators predict similar policies in proposed legislation? When states adopt similar policy regimes, do they see similar societal outcomes? What societal outcomes/indicators, if any, are predictive of one state following the policies introduced by another state? Can bill text alignment be used to automatically discover tables of similar bills, like those compiled by NCSL? Questions such as these can be directly investigated using the policy similarity measures we introduce. The alignment scores for all bill pairs—a novel database which we contribute—are available in the supplemental materials with this article. The validation exercises ensure that researchers can use the alignment scores that we derive for our bill pairs database to study the policy similarity between bills at the level of bill pairs, or aggregated up to higher levels.

We present an illustrative application of the alignment scores to investigate the way in which asymmetric party politics shapes cross-state policy similarity. The asymmetric politics theory, which holds that Republican Party supporters represent an ideologically cohesive group, in which voters and politicians are focused on advancing a relatively universal conservative agenda, whereas the Democratic Party's support base is best characterized as a coalition of separate interests, each in pursuit of its own specific progressive policy goals. We deduce from this the expectation that policies introduced by Republican legislators in different states will be more similar to each other than those introduced by Democrats in different states. This expectation is robustly supported through the comparison of dyadic alignment scores for Republican and Democratic cross-state dyads. This substantive application represents a contribution to the literature on asymmetric politics in that we deduce and show support for the policy similarity implications of this relatively new theoretical paradigm.

The nature of our measure of policy similarity based on text reuse is that it is high precision and low recall. As such, in future research we advise scholars to focus on upper quantiles of the distribution of alignments when the inquiry requires the identification of bills that introduce equivalent policies. For predictive modeling, this can be done through the use of quantile regression at high quantiles. For searching out bill pairs, this can be done through thresholding on high alignment scores (e.g., 50+) to indicate the presence of highly similar policy proposals in legislation. Such thresholding may discard informative variation on policy similarity in terms of bill pairs that exhibit borderline overlap when it comes to containing equivalent policy content. On the other hand, high-score thresholding would focus the analysis on bill pairs that reliably contain very similar policy enactments, and pose little risk of a false positive.

Fridolin Linder is a postdoctoral researcher at the Social Media and Political Participation Lab at New York University. He received his Ph.D. in political science from Pennsylvania State University and was a 2015 Data Science for Social Good Fellow. His research focuses on the development of computational social science methodology to explain political phenomena.

Bruce Desmarais is an associate professor in the Department of Political Science and the Institute for CyberScience at Pennsylvania State University. His research focuses on the development and application of research methods to analyze the complex interdependence underlying the formation and implementation of law and public policy.

Matthew Burgess is a Software Engineer at Google Research. His research focuses on building data powered systems with high social impact in a range of disciplines ranging from government transparency to open data. He has a Doctorate of Computer Science from University of Michigan and was a 2015 Data Science for Social Good fellow at the University of Chicago.

Eugenia Giraudy co-authored this paper as Director of Projects and Analytics at YouGov US. She has recently joined Facebook's Core Data Science Team as a Research Scientist. Her research uses computational social science tools to understand party coalitions, electoral strategies, and interest groups influence. She has a PhD in Political Science from the University of California Berkeley and was a 2015 Data Science for Social Good fellow at the University of Chicago.

Notes

This work was supported by National Science Foundation [grant number SES-1558661, SES-1619644, CISE-1320219 and DGE-1144860]. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect those of the sponsor.

1. http://sunlightfoundation.com/
2. We rely on a random sample for computational reasons. Calculating the complete similarity matrix is computationally infeasible and unnecessary.
3. All of the code required to replicate our measurements of similarity through alignment computation, and the analyses that follow, is provided in a public repository at https://github.com/desmarais-lab/text_reuse. The bill similarity scores are available on The Dataverse at https://doi.org/10.7910/DVN/CZ25GF.
4. www.ncsl.org
5. We used the bing websearch API to collect the URLs.
6. The p-values were calculated using 5,000 random matrix permutations.
7. Calculated as $100\times[\exp(0.38)-1]=46.2$.
8. When we go above this quantile, the quantile regression estimator becomes unstable (e.g., returning bootstrap confidence intervals in which the point estimates are far from the center).
9. For details on the exact algorithm we refer the reader to the Elasticsearch documentation (https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-mlt-query.html as well as the lucene engine documentation (https://lucene.apache.org/core/4_9_0/core/org/apache/lucene/search/similarities/TFIDFSimilarity.html).

References

Alemán, Eduardo, and Ernesto Calvo. 2013. "Explaining Policy Ties in the Argentine and Chilean Congresses: A Network Analysis of Bill Initiation Data." *Political Studies* 61 (2): 356–77.

Berry, Frances Stokes, and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *American Political Science Review* 84 (2): 395–415.

Blair, Benjamin, Tanya Heikkila, and Christopher M. Weible. 2016. "National Media Coverage of Hydraulic Fracturing in the United States: Evaluation Using Human and Automated Coding Techniques." *Risk, Hazards & Crisis in Public Policy* 7 (3): 114–28.

Boehmke, Frederick J. 2009. "Policy Emulation or Policy Convergence? Potential Ambiguities in the Dyadic Event History Approach to State Policy Emulation." *Journal of Politics* 71 (3): 1125–40.

Boehmke, Frederick J., and Paul Skinner. 2012. "State Policy Innovativeness Revisited." *State Politics & Policy Quarterly* 12 (3): 303–29.

Brown, Elizabeth K. 2013. "Foreclosing on Incarceration? State Correctional Policy Enactments and the Great Recession." *Criminal Justice Policy Review* 24 (3): 317–37.

Burgess, Matthew, Eugenia Giraudy, Julian Katz-Samuels, Joe Walsh, Derek Willis, Lauren Haynes, and Rayid Ghani. 2016. "The Legislative Influence Detector: Finding Text Reuse in State Legislation." KDD 2016, San Francisco, CA.

Cao, Xun. 2010. "Networks as Channels of Policy Diffusion: Explaining Worldwide Changes in Capital Taxation, 1998–2006." *International Studies Quarterly* 54 (3): 823–54.

Commins, Margaret M., and Jeremiah B. Wills. 2016. "Reappraising and Extending the Predictors of States? Immigrant Policies: Industry Influences and the Moderating Effect of Political Ideology." *Social Science Quarterly* 98 (1): 212–29.

Cook, James M. 2017. "Twitter Adoption and Activity in U.S. Legislatures: A 50-State Study." *American Behavioral Scientist* 61 (7): 724–40.

Cranmer, Skyler J., and Bruce A. Desmarais. 2016. "A Critique of Dyadic Design." *International Studies Quarterly* 60 (2): 355–62.

———. 2017. "What Can We Learn from Predictive Modeling?" *Political Analysis* 25 (2): 145–66.

Desmarais, Bruce A., Jeffrey J. Harden, and Frederick J. Boehmke. 2015. "Persistent Policy Pathways: Inferring Diffusion Networks in the American States." *American Political Science Review* 109: 392–406.

Garcia-Olano, Diego, Marta Arias Vicente, and Josep Larriba Pey. 2016. "Automated Construction and Analysis of Political Networks via Open Government and Media Sources." Proceedings of the First Workshop on Data Science for Social Good, ECML PKDD 2016 September 19, Riva del Garda, Italy.

Garrett, Kristin N., and Joshua M. Jansa. 2015. "Interest Group Influence in Policy Diffusion Networks." *State Politics & Policy Quarterly* 15 (3): 387–417.

Gerber, Brian J., Cherie Maestas, and Nelson C. Dometrius. 2005. "State Legislative Influence Over Agency Rulemaking: The Utility of Ex Ante Review." *State Politics & Policy Quarterly* 5 (1): 24–46.

Gerrish, Sean, and David M. Blei. 2011. "Predicting Legislative Roll Calls from Text." 28th International Conference on Machine Learning, June 28–July 2, Bellevue, WA.

Gilardi, Fabrizio, Katharina Füglister, and Stephane Luyet. 2009. "Learning from Others: The Diffusion of Hospital Financing Reforms in OECD Countries." *Comparative Political Studies* 42 (4): 549–73.

Glick, Henry R. 1992. "Judicial Innovation and Policy Re-invention: State Supreme Courts and the Right to Die." *The Western Political Quarterly* 45 (1): 71–92.

Gomez Rodriguez, Manuel, Jure Leskovec, and Andreas Krause. 2010. "Inferring Networks of Diffusion and Influence." *ACM Transactions on Knowledge Discovery from Data (TKDD) 5* (4): Art. 21.

Gormley, Clinton, and Zachary Tong. 2015. *Elasticsearch: The Definitive Guide*. Newton, MA: O'Reilly Media, Inc.

Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*: 267–97.

Grossback, Lawrence J., Sean Nicholson-Crotty, and David A. M. Peterson. 2004. "Ideology and Learning in Policy Diffusion." *American Politics Research* 32 (5): 521–45.

Grossmann, Matt, and David A. Hopkins. 2015. "Ideological Republicans and Group Interest Democrats: The Asymmetry of American Party Politics." *Perspectives on Politics* 13 (1): 119–39.

———. 2016. *Asymmetric Politics: Ideological Republicans and Group Interest Democrats*. Oxford: Oxford University Press.

Harden, Jeffrey J. 2011. "A Bootstrap Method for Conducting Statistical Inference with Clustered Data." *State Politics & Policy Quarterly* 11 (2): 223–46.

Harden, Jeffrey J., and Bruce A. Desmarais. 2011. "Linear Models with Outliers Choosing between Conditional-Mean and Conditional-Median Methods." *State Politics & Policy Quarterly* 11 (4): 371–89.

Hart, Roderick P., Deborah Smith-Howell, and John Llewellyn. 1990. "Evolution of Presidential News Coverage." *Political Communication* 7 (4): 213–30.

Huber, John D., Charles R. Shipan, and Madelaine Pfahler. 2001. "Legislatures and Statutory Control of Bureaucracy." *American Journal of Political Science* 45 (2): 330–45.

Karch, Andrew. 2007. "Emerging Issues and Future Directions in State Policy Diffusion Research." *State Politics & Policy Quarterly* 7 (1): 54–80.

Knill, Christoph. 2005. "Introduction: Cross-National Policy Convergence: Concepts, Approaches and Explanatory Factors." *Journal of European Public Policy* 12 (5): 764–74.

Krackhardt, David. 1988. "Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data." *Social Networks* 10 (4): 359–81.

Krause, Rachel M. 2011. "Policy Innovation, Intergovernmental Relations, and the Adoption of Climate Protection Initiatives by U.S. Cities." *Journal of Urban Affairs* 33 (1): 45–60.

Langer, Ana Ines, and Iñaki Sagarzazu. 2017. "Are All Policy Decisions Equal? Explaining the Variation in Media Coverage of the UK Budget." *Policy Studies Journal* 45 (2): 337–58.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97 (2): 311–31.

Lowe, Will, and Kenneth Benoit. 2013. "Validating Estimates of Latent Traits from Textual Data Using Human Judgment as a Benchmark." *Political Analysis* 21 (3): 298–313.

Marquez, Timothy, and Scot Schraufnagel. 2013. "Hispanic Population Growth and State Immigration Policy: An Analysis of Restriction (2008–12)." *Publius: The Journal of Federalism* 43 (3): 347–67.

Martin, Lanny W., and Georg Vanberg. 2008. "A Robust Transformation Procedure for Interpreting Political Text." *Political Analysis* 16 (1): 93–100.

McClurg, Scott D. 2003. "Social Networks and Political Participation: The Role of Social Interaction in Explaining Political Participation." *Political Research Quarterly* 56 (4): 449–64.

Mikhaylov, Slava, Michael Laver, and Kenneth R. Benoit. 2012. "Coder Reliability and Misclassification in the Human Coding of Party Manifestos." *Political Analysis* 20 (1): 78–91.

Mondou, Matthieu, Grace Skogstad, and David Houle. 2014. "Policy Image Resilience, Multidimensionality, and Policy Image Management: A Study of U.S. Biofuel Policy." *Journal of Public Policy* 34 (1): 155.

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn. 2008. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis* 16 (4): 372–403.

Mooney, Christopher Z. 2001. "Modeling Regional Effects on State Policy Diffusion." *Political Research Quarterly* 54 (1): 103–24.

Mooney, Christopher Z., and Mei-Hsien Lee. 1995. "Legislative Morality in the American States: The Case of pre-Roe Abortion Regulation Reform." *American Journal of Political Science* 39 (3): 599–627.

Mycoff, Jason D., Michael W. Wagner, and David C. Wilson. 2009. "The Empirical Effects of Voter-ID Laws: Present or Absent?" *PS: Political Science & Politics* 42 (1): 121–26.

Nowlin, Matthew C. 2016. "Modeling Issue Definitions Using Quantitative Text Analysis." *Policy Studies Journal* 44 (3): 309–31.

Pierce, Patrick A., and Donald E. Miller. 1999. "Variations in the Diffusion of State Lottery Adoptions: How Revenue Dedication Changes Morality Politics." *Policy Studies Journal* 27 (4): 696–706.

Potthast, Martin, Matthias Hagen, Tim Gollub, Martin Tippmann, Johannes Kiesel, Paolo Rosso, Efstathios Stamatatos, and Benno Stein. 2013. "Overview of the 5th International Competition on Plagiarism Detection." In *CLEF Conference on Multilingual and Multimodal Information Access Evaluation. CELCT,* 301–31.

Ramos, Juan. 2003. "Using tf-idf to Determine Word Relevance in Document Queries." In *Proceedings of the First Instructional Conference on Machine Learning* 242: 133–42.

Rivera, Michael U. 2014. "Immigration, Public Opinion, and State Policy (Un) Responsiveness: A Case for Analyzing Policies that Expand the Rights of Immigrants." In *Annual State Politics and Policy Conference*, Bloomington, IN.

Robertson, Stephen. 2004. "Understanding Inverse Document Frequency: On Theoretical Arguments for IDF." *Journal of Documentation* 60 (5): 503–20.

Russell, Annelise. 2017. "U.S. Senators on Twitter: Asymmetric Party Rhetoric in 140 Characters." *American Politics Research* 46 (4): 695–723.

Scott, Patrick G. 1997. "Assessing Determinants of Bureaucratic Discretion: An Experiment in Street-Level Decision Making." *Journal of Public dministration Research and Theory* 7 (1): 35–58.

Sharman, Jason C. 2008. "Power and Discourse in Policy Diffusion: Anti-Money Laundering in Developing States." *International Studies Quarterly* 52 (3): 635–56.

Shipan, Charles R., and Craig Volden. 2008. "The Mechanisms of Policy Diffusion." *American Journal of Political Science* 52 (4): 840–57.

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

Simmons, Beth A., and Zachary Elkins. 2004. "The Globalization of Liberalization: Policy Diffusion in the International Political Economy." *American Political Science Review* 98 (1): 171–89.

Smith, Temple F., and Michael S. Waterman. 1981. "Identification of Common Molecular Subsequences." *Journal of Molecular Biology* 147 (1): 195–97.

Tremblay, Marie-Claude, Pierre Pluye, Genevieve Gore, Vera Granikov, Kristian B. Filion, and Mark J. Eisenberg. 2015. "Regulation Profiles of e-Cigarettes in the United States: A Critical Review with Qualitative Synthesis." *BMC Medicine* 13: 130.

True, Jacqui, and Michael Mintrom. 2001. "Transnational Networks and Policy Diffusion: The Case of Gender Mainstreaming." *International Studies Quarterly* 45 (1): 27–57.

Van Atteveldt, Wouter, Jan Kleinnijenhuis, and Nel Ruigrok. 2008. "Parsing, Semantic Networks, and Political Authority Using Syntactic Analysis to Extract Semantic Relations from Dutch Newspaper Articles." *Political Analysis* 16 (4): 428–46.

Volden, Craig. 2006. "States as Policy Laboratories: Emulating Success in the Children's Health Insurance Program." *American Journal of Political Science* 50 (2): 294–312.

Walker, Jack L. 1969. "The Diffusion of Innovations Among the American States." *American Political Science Review* 63 (3): 880–99.

Ward, Michael D., Katherine Stovel, and Audrey Sacks. 2011. "Network Analysis and Political Science." *Annual Review of Political Science* 14: 245–64.

Weingast, Barry R., and Mark J. Moran. 1983. "Bureaucratic Discretion or Congressional Control? Regulatory Policymaking by the Federal Trade Commission." *Journal of Political Economy* 91 (5): 765–800.

Weiss, Gary M. 2004. "Mining with Rarity: A Unifying Framework." *ACM SIGKDD Explorations Newsletter* 6 (1): 7–19.

Weiss, Gary M., and Haym Hirsh. 2000. "Learning to Predict Extremely Rare Events." AAAI Technical Report WS-00-05. Menlo Park, CA: Association for the Advancement of Artificial Intelligence.

Weyland, Kurt. 2005. "Theories of Policy Diffusion Lessons from Latin American Pension Reform." *World Politics* 57 (2): 262–95.

Wilkerson, John, David Smith, and Nicholas Stramp. 2015. "Tracing the Flow of Policy Ideas in Legislatures: A Text Reuse Approach." *American Journal of Political Science* 59 (4): 943–56. https://doi.org/10.1111/ajps.12175

Witmer, Richard C., Joshua Johnson, and Frederick J. Boehmke. 2014. "American Indian Policy in the States." *Social Science Quarterly* 95 (4): 1043–63.

Witting, Antje. 2015. "Measuring the Use of Knowledge in Policy Development." *Central European Journal of Public Policy* 9 (2): 54–62.

Ybarra, Vickie D., Lisa M. Sanchez, and Gabriel R. Sanchez. 2016. "Anti-Immigrant Anxieties in State Policy: The Great Recession and Punitive Immigration Policy in the American States, 2005–2012." *State Politics & Policy Quarterly* 16 (3): 313–39.

Young, Lori, and Stuart Soroka. 2012. "Affective News: The Automated Coding of Sentiment in Political Texts." *Political Communication* 29 (2): 205–31.

Appendix

In this appendix, we provide extended discussions of several of the technical aspects of our use of the Smith-Waterman algorithm

Description of Smith-Waterman Algorithm

Consider again these two example sections of text from bills in North Carolina and Alaska, respectively.

section 1.AS44.99 is amended by adding new sections to read: article 6. compact for a balanced budget (...)

chapter 143 is amended by adding a new article to read: article 80. compact for a balanced budget (...)

We illustrate the design of the Smith-Waterman algorithm in Figure A1. A matrix is created where each combination of elements of the two sequences is assigned a cell. And an additional row and column of 0 is appended to the beginning of each sequence (to allow for a gap in the beginning of one sequence). The quality or alignment score of every possible alignment is then calculated cumulatively with the following algorithm. Denote the two sequences as $\mathcal{A}=(a_1, a_2, ..., a_n)$ and $\mathcal{B}=(b_1, b_2, ..., b_k)$. Additionally let δ, ε, and γ be the match, mismatch, and gap scores. Define the scoring function:

$$S(a_i,b_j)=\delta^{\mathbb{I}(a_i=b_j)}+\epsilon^{\mathbb{I}(a_i\neq b_j)} \tag{2}$$

where $\mathbb{I}(.)$ is the indicator function. Then the entry for each cell $M_{i,j}$, i=1,2,...,n and j=1,2,...,k of the matrix is filled by the following rule:

$$M_{i,j}=\max\left(M_{i-1,j-1}+S(a_i,b_j),M_{i-1,j}+\gamma,M_{i,j-1}+\gamma,0\right) \tag{3}$$

In Figure A1, each row corresponds to a word in the sentence from the Alaska bill and each column to the sentence from the North Carolina bill. On the right side of the matrix, the optimal alignment is displayed.

The four values of which the maximum is chosen in Equation 3 describe four steps in this dynamic programming array. The first value corresponds to a diagonal step which means in the resulting alignment sequence elements a_i and b_j are matched and the score of the alignment is either increased by δ if they match or decreased by ε

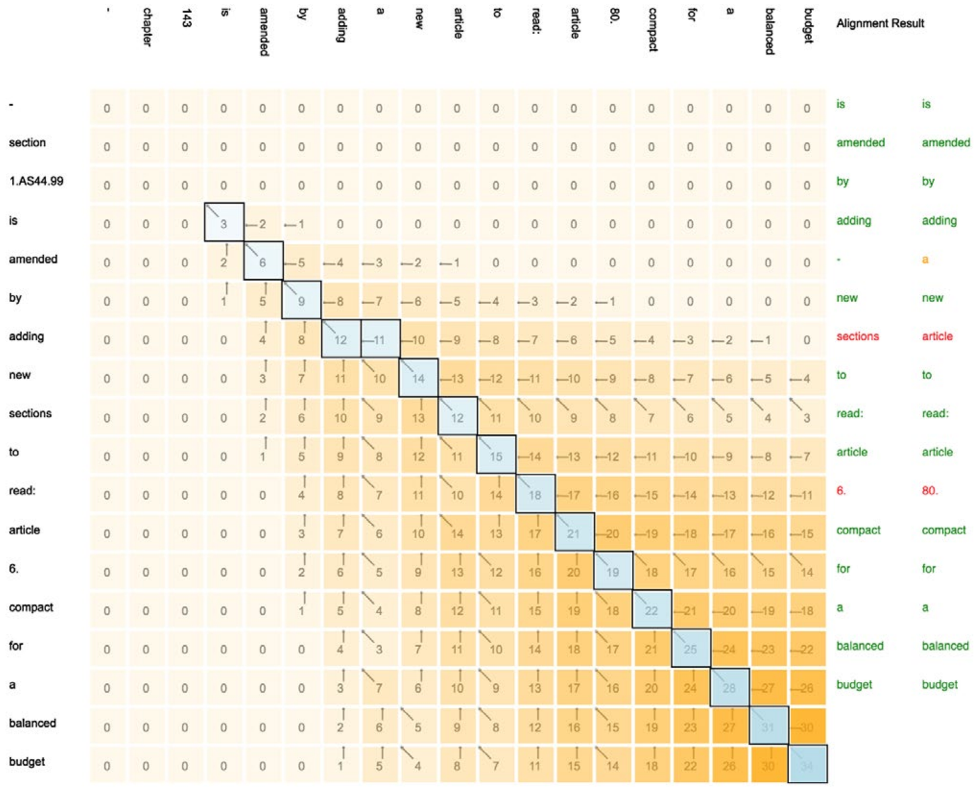


Figure A1. Dynamic Programming Array of Local Alignment Algorithm.
Note: The parameters in this instance are: match score: 3, mismatch score: –2, gap score: –1. The optimal alignment is displayed on the right side of the matrix. The left column corresponds to the text sequence in the rows of the matrix, the right column to the sequence in the columns. Mismatches are colored in red, gaps are indicated by "–" and orange color. The highlighted (blue) cells in the matrix display the path of the optimal alignment.

if they don't match. The second and third value describe a gap step either in horizontal or vertical direction. That is, if we move from cell (*i*,*j*) to cell (*i*+1,*j*), in the resulting alignment sequence element a_{i+1} is matched with a gap. On the other hand, if we move to cell (*i*,*j*+1) sequence element b_{j+1} is matched with a gap in the resulting alignment. If the first three values are all <=0, the score in the cell is 0 meaning that no step is taken and every potential alignment stops at this point. This can be observed in Figure A1. The highlighted path describes the optimal alignment. The first two elements in both sequences don't match, therefore the score remains zero. Therefore, the resulting alignment starts at the fourth element of each sequence with four diagonal steps, each adding points (δ=5) to the final score which reaches 20 at the alignment ending in "adding"-"adding." After that, a step to the right is taken, meaning that the additional "a" in the column sequence is matched with a gap in the row sequence (indicated by "-" and orange coloring in the result on the right-hand side). The cumulative score of the alignment is reduced by γ=−1 to 19. After that only diagonal steps are taken. Note that there are two mismatches each of which reduces the score by ε=−2.

The optimal alignment is identified by finding the highest score in the matrix, at which the alignment ends, and back-tracing the path of the alignment. The steps in the path are chosen using Equation 3, and are indicated in Figure A1 by arrows pointing to the left, the upper left, up, or no arrow. The back-tracing continues until a zero is encountered.

This algorithm is computationally intensive ($\mathcal{O}(nk)$ in computation time and memory demand), but returns the globally optimal local alignment given the scoring parameters. We provide a proof of this in the Appendix. The property follows from the recursive relation used to fill the dynamic programming matrix. An alignment that is constructed by adding one element to an already optimal alignment is optimal itself. Starting by aligning the first two elements of the sequence optimally (which is trivial), a globally optimal can be found by iteratively adding to this alignment.

In our analyses, we use a slightly modified version of the local alignment algorithm, in which the first gap in a series of multiple gaps receives a higher penalty then the following ones (Wilkerson et al. [2015], use the same modification). This penalizes many small gaps and makes the algorithm produce longer gaps. The idea behind this modification is that someone who changes text in a bill might insert several words into an existing piece of text. The fact that something was inserted should weigh heavier than the length of the insertion which is achieved by down-weighting gap extensions.

In their application of the SW algorithm to the dataset that we use, Burgess et al. (2016) optimize the SW algorithm parameters through a predictive experiment. They collected 165 bills that were hand-labeled by experts to reuse text from one or more bills within the 165 bill corpus. Looking at all 13,530 pairs of bills that can be formed from the corpus of 165, they used the alignment score to predict whether or not there was text reuse between the bills. They found the optimal parameters, in terms of area under the ROC curve, were a match score of 3, mismatch score of −2, and gap score of −3. We followed Burgess et al. (2016) in setting the parameters of the SW algorithm since they use the same data and have already optimized the parameter values. Future work should expand upon treatment of these parameter values in at least two directions. First, it would be valuable to develop a criterion function such that one does not need hand-labeled data, which is costly and may exhibit its own biases, to select parameter values. Second, it would be valuable to develop an approach to representing uncertainty in the parameter values as well as in the optimal alignments and scores.

Describing Elasticsearch

We store all available bills in an Elasticsearch database. Elasticsearch is an open source web search engine which is designed to find documents that match a search query (Gormley & Tong, 2015). We utilize the "more like this" query, which is designed to find documents that are similar to a given document. It works in the following way: From the focus document, the k n-grams with the highest tf-idf (Ramos, 2003) scores are selected and transformed into a search vector that is representative of the query document. Then, for each document in the collection, a form of the cosine similarity is calculated and the m bills with the highest scores are selected for further analysis. For this analysis we set k to 25, n to 5 and m to 500.[9]

Proof that SW Alignment Is Globally Optimal

F_i is the alignment of size i with i=1,2,...,(n+k−1). $S(F_i)$ is the score of the alignment.

1. There are four options for F_1 (Gap in $\mathcal{A}$, Gap in $\mathcal{B}$, Mismatch, Match). This alignment is optimal when $S(F_1)$ is maximized.
2. Every alignment's score $S(F_i)$ with i>1 when maximized, is composed of the score of the "one-shorter" alignment $S(F_{i-1})$ and the score of the i-th element. If not the score could be increased by choosing a different F_{i-1} that has a higher score.
3. By induction, therefore, follows that every alignment represented in the dynamic programming matrix (Figure A1) and calculated according to Equations 2 and 3 is an optimal alignment ending in the locations specified by the row and column index in the matrix.
4. The dynamic programming matrix contains all possible alignments. Since every cell is optimal, the highest score in the matrix must be globally optimal.