Policymaking in the American States, 1787-2020*

Charles Angelucci
MIT Sloan

Elliott Ash
ETH Zurich

Nicolas Longuet-Marx
Columbia University

March 21, 2025

Abstract

This paper combines tools from natural language processing and structural estimation to analyze the adoption and diffusion of policies across U.S. states over the period 1787-2020. We employ a large language model to extract policy content from all state statutes enacted during this period (N=2.5 million). Using an unsupervised clustering algorithm applied to vector representations of policy content, we group related policies enacted across states with the goal of measuring state legislative activity on these policy clusters. After validating this measurement procedure, we use a discrete-choice model of policymaking to analyze the determinants of policy diffusion across states. States with greater geographic, economic, and political similarity implement more similar policies. While polarization has risen significantly since the 1990s, consistent with prior literature, we show that current levels of polarization are comparable to those observed in the pre-WWII period, following a U-shaped pattern over time. Polarization primarily reflects differences in policy choices within topics, rather than from states selecting different topics to address. Finally, we document an increasing nationalization of policy, with federal legislative texts exerting growing influence on state legislation since the post-war period.

*We thank Clementine Abed Meraim, Revekka Gershovich, Jairo Gudiño-Rosero, Michel Gutmann, and Xianglin Li for excellent research assistance. We are also grateful to participants at the Monash-Warwick-Zurich Text-as-Data Workshop, Columbia Political Economy Colloquium, Berkeley Economic History Workshop, Harvard Political Economy Seminar, Bocconi Political Economy Seminar, the Paris Empirical Political Economics Seminar, and the LSE Political Economy Seminar for suggestions that have improved this paper.

1 Introduction

At its inception, the U.S. federal system was designed to balance a limited central government with strong state autonomy. In Federalist No. 45, Madison famously argued that "the powers delegated by the proposed Constitution to the federal government are few and defined. Those which are to remain in the state governments are numerous and indefinite" (Madison, 1788). This vision not only preserved state independence but also encouraged a process of policy experimentation and diffusion among state legislatures — an idea embedded in the notion of states as "laboratories of democracy" (Brandeis, 1932). U.S. state legislatures remain critical engines of public policy, with their decisions affecting a broad range of issues — from education and healthcare to criminal justice and environmental regulation.

Over time, however, the federal government has assumed a growing role in national policymaking, and the boundaries between state and federal action have become increasingly blurred. Scholars such as Elazar (1966, 1987) and Shor and McCarty (2011*b*) highlight that evolving political dynamics, expanding state capacities, and shifting public expectations have transformed intergovernmental relations. In this paper, we examine how these forces have redefined state policymaking and the interplay between state and federal initiatives.

We examine three core dimensions of state policy variation over the lat 200 years. First, we assess how policies diverge across states, and how this divergence has evolved over time. Second, we analyze the factors that drive similarities and differences in policy choices, including institutional characteristics, partisan dynamics, and spillovers from neighboring states. Third, we explore the extent to which state policies drive or are driven by federal initiatives. By investigating these dimensions, our work sheds new light on the mechanisms that shape state-level innovation and policy diffusion.

To address these questions in the data, we employ a novel methodological approach using artificial intelligence and natural language processing. Our analysis is based on a newly expanded corpus comprising the universe of state laws passed by state legislatures since their creation until 2020, as well as all federal laws enacted during this period. This corpus includes approximately two million pages of state law text and an additional two million pages of federal law text.

Using state-of-the-art AI models (GPT-4o-mini) and embedding techniques (MPNet), we extract policy content from these legal texts and create clusters of closely related policies. This approach allows us to measure legal similarity between individual states over time, as well as between states and federal legislation. We provide extensive validations that the method effectively measures legislative activity across

policies.

Next, we leverage our data to estimate a dynamic discrete choice model of policymaking. Our model captures the decision process by which each state — and the federal government — chooses whether to adopt a policy within a given policy cluster. A state's decision is shaped by its own characteristics, its historical policy choices, and the actions of other states and the federal government. For instance, states may be more inclined to adopt a policy if similar jurisdictions have already done so, reflecting a dynamic process of policy diffusion.

We find that the strongest predictors of cross-state policy adoption have shifted from geography in the early twentieth century to partisanship since the 1990s, and that federal statutes increasingly shape state legislation. We don't find any evidence that state-level policy innovations prompt subsequent federal reforms. Overall, our results suggest continued independence of states in fine-grained policy decisions and a growing pull toward national party identities and legislative templates.

These results relate to several strands of the literature in political economy. First, we build on some articles that have analyzed the state session laws in the past. For example, Mulligan and Shleifer (2005) show that states with higher population also have more pages of laws in the cross section. Ash, Morelli and Vannoni (2021) show that higher legislative flows are related to higher economic output over time. Ash (2018) uses state statutes to recover the effective tax code and shows that Democrats use relatively more revenue-increasing language for income taxes but revenue-decreasing language for sales taxes, compared to Republicans.

Second, our paper relates to the literature on policy coordination and policy diffusion between U.S. states. A growing body of work has examined how states influence each other's policy choices and how that influence is shaped by factors like geography and partisanship. De Paula, Rasul and Souza (2019) document the role of geography in tax competition among U.S. states, while Mallinson (2021) finds that the effect of contiguous neighbors has waned over time. Using a comparable approach, DellaVigna and Kim (2022) show that although demographic and geographic similarities have become less relevant, political alignment has emerged as a powerful determinant of policy replication across states. Most research along these lines has relied on a relatively small, predefined set of laws, whereas our analysis uses the complete universe of enacted legislation, allowing for a more comprehensive perspective. Additionally, we extend this literature by examining the evolving relationship between federal and state-level policies.

A separate strand of scholarship has investigated the degree to which partisanship shapes policy outcomes at the subnational level. Some studies, such as Ferreira and

Gyourko (2009), argue that the mayor's party affiliation does not systematically alter urban policy, whereas Dippel (2022) report that city pensions grow faster under Democratic mayors. Focusing on states, Caughey, Xu and Warshaw (2017) show that legislative partisanship has become increasingly salient over time, a finding echoed by Carlino et al. (2023) in gubernatorial spending decisions. Parallel work by Hertel-Fernandez (2019) notes the increasing influence of lobbies and advocacy groups, including the American Legislative Exchange Council (ALEC). In an important contribution, Shor and McCarty (2011*a*) examine roll call data from state legislatures since the mid-1990s, highlighting a trend of rising polarization that mirrors developments at the federal level. Grumbach (2018) similarly argues that states, once deemed peripheral in U.S. governance, have become central arenas for partisan conflict, thus contributing to higher policy variance across the country. DellaVigna and Kim (2022) analyze a broad set of key policies adopted by U.S. states to examine diffusion patterns, finding that partisanship has become an increasingly influential factor in shaping policy adoption in recent decades.

Third, recent work also highlights the growing nationalization of local politics, with state party platforms increasingly reflecting the national agenda rather than local concerns (Hopkins, 2018; Hopkins, Schickler and Azizi, 2022). This trend coincides with the decline of local news coverage, which has encouraged voters to base decisions on national issues rather than exclusively local matters (Angelucci, Cagé and Sinkinson, 2020; Moskowitz, 2021). In addition, scholars such as Elazar (1966, 1987) have long emphasized the multifaceted relationship between state governments and federal authorities. Building on this literature, our analysis examines whether the legislative agenda and resulting policies implemented at the state level now mirror national priorities more than in earlier eras.

Fourth and finally, we add to the literature on text as data in economics (Gentzkow, Kelly and Taddy, 2017; Ash and Hansen, 2023). A large literature has used text-data methods to explore differences across political parties (Gentzkow and Shapiro, 2010; Gentzkow, Shapiro and Taddy, 2017; Ash, Morelli and Weelden, 2017; Ash, Morelli and Osnabrügge, 2019; Widmer, Galletta and Ash, 2020). While the standard approaches have done text scaling using regressions or classifiers, more recent work has taken unsupervised approaches using embeddings (Gennaro and Ash, 2021). The use of large language models to summarize legal text is quite new (Ash et al., 2024), as is the use of transformer-based document encoders (Arold et al., 2024). We show how to validate and implement these methods for large historical corpora.

The remaining of this paper is organized as follows. Section 2 provides institutional background. Section 3 develops our theoretical model. Sections 4 and 5 respectively

describe our datasets and methods. Section 6 reports the empirical results, while Section 7 provides robustness checks on the polarization aspects with an alternative ML-based approach. Section 8 concludes.

2 Institutional Background

A defining feature of the United States is the substantial autonomy exercised by its fifty state governments. Each state has a legislature that, with one exception (Nebraska's unicameral body), consists of two chambers. These bodies typically convene for shorter sessions than the U.S. Congress and often have fewer members, sometimes operating under term limits. Governors, who lead the executive branch at the state level, hold the authority to approve or veto bills passed by the legislature and to propose major legislative initiatives, most notably the annual state budget. Despite these structural similarities, considerable variation exists across states in the scope and frequency of legislative sessions, the degree of professionalization among legislators, and the extent of gubernatorial powers.

State governments handle policy areas that generally fall under their own purview, such as education, state-level taxation, and much of criminal law. At the same time, the federal government legislates on matters including defense, trade, and national taxation. Yet the boundary between state and federal responsibilities is not strictly demarcated. Many critical issues—environmental regulations, labor laws, civil rights protections—feature overlapping authority between federal and state governments (Elazar, 1966, 1987). Under the U.S. Constitution's Supremacy Clause, valid federal law can override state law in the event of a direct conflict. However, states are often free to adopt laws that go beyond federal mandates or to enact policies in areas where federal legislation is silent.

A range of examples illustrates how states can either strengthen or weaken federal-level frameworks. Where federal minimum wage legislation sets a floor ($7.25/hour), states such as Massachusetts and Washington have enacted substantially higher minimum wages in response to local conditions. Environmental protection exhibits a similar pattern: state environmental agencies often implement stricter regulations than those imposed by the Environmental Protection Agency (EPA). By contrast, some states enact policies that effectively undercut federal positions. For instance, while federal law lists marijuana as a controlled substance, multiple states have legalized its recreational use. Likewise, "sanctuary" laws in certain jurisdictions limit cooperation with federal immigration authorities, and "constitutional carry" statutes in others relax gun

regulations beyond federal requirements.

Regardless of whether a state bolsters or diminishes federal guidelines, its legislature exerts broad influence over daily life. These responsibilities range from appropriating funds for public schools and transportation infrastructure to regulating healthcare systems and shaping the criminal justice process through sentencing rules and law enforcement oversight. They also include decisions on tax structures, environmental protections, and various social policies (e.g., abortion, LGBTQ+ rights, and firearm regulations). States hence act as "laboratories of democracy,"[1] experimenting with legal frameworks that may later be adopted elsewhere. A prominent case is the legalization of recreational marijuana: after Colorado and Washington took early steps in 2012, many additional states followed suit, revealing how innovations can diffuse across state lines.

In sum, state governments not only legislate on issues of direct local impact but also play a critical part in shaping American public policy at large. Overlapping jurisdictions with the federal government create opportunities for policy coordination—or conflict—that can reshape the national political landscape. The interplay between state-level autonomy, constitutional constraints, gubernatorial authority, and federal supremacy has historically driven significant policy diversity across states. These institutional dynamics motivate our analysis. The substantial variation in state political institutions and overlapping jurisdictions with the federal government motivate our focus on policy clusters and diffusion processes. By modeling the interplay between state-level autonomy, legislative structures, and federal influence, we capture how innovations in public policy emerge and spread across jurisdictions.

3 Model

3.1 A Discrete Choice Model of Policy Adoption

We develop a framework for policy adoption and diffusion among U.S. states and the federal government. Let s index individual states — with $s = f$ representing the federal government — and let time t be discrete over an infinite horizon. Our analysis assumes a discrete policy space, where each subset of policies is referred to as a "cluster" (a terminology reflecting our data construction). For instance, one cluster pertains to wildlife conservation while another focuses on pesticide regulation violations.

Within each policy cluster c, every state faces a binary decision in each period:

[1] A term famously used by Justice Louis Brandeis in *New State Ice Co. v. Liebmann* (1932).

adopt a policy from that cluster ($n_{sct} = 1$) or not ($n_{sct} = 0$). We assume that there are no technological interdependencies across clusters, meaning that the cost and benefit of adopting a policy in cluster c are independent of a state's decisions on other clusters. However, as detailed below, the net benefit of adopting a policy within a given cluster may depend on the state's own past adoption history in that cluster, as well as on the past and current adoption decisions of other states and the federal government for policies within the same cluster.

Let $T_{-t} = \{t' \in T : t' < t\}$ denote the set of all periods preceding the current period t. We assume that payoffs are separable across policy clusters. Then, the payoff for state s from contributing to policy cluster c in year t is given by

$$U_{sct} = \underbrace{\alpha_{sct} + \delta_c \sum_{t' \in T_{-t}} n_{sct'} + \sum_{s' \neq s} \sum_{t' \in T_{-t}} \beta_{ss'ct'}\, n_{s'ct'} + \sum_{s' \neq s} \beta_{ss'ct}\, \hat{p}_{s'ct}}_{V_{sct}} + \epsilon_{sct}, \tag{1}$$

where ϵ_{sct} denotes an idiosyncratic shock.

The baseline utility of adoption is specified as

$$\alpha_{sct} = \alpha_c + \alpha_{st} + \boldsymbol{\omega}' \boldsymbol{X}_{st}, \tag{2}$$

where $\boldsymbol{X}_{st}$ is a vector of year-state level characteristics and $\boldsymbol{\omega}$ is the corresponding vector of parameters.

Next, the interaction parameter is defined by

$$\beta_{ss'ct'} = \beta_c + \beta_{s't'} + \boldsymbol{\gamma}' \boldsymbol{SimX}_{ss'}, \tag{3}$$

where $\boldsymbol{SimX}_{ss'}$ is a vector of characteristics for the state pair (s, s') and $\boldsymbol{\gamma}$ is the corresponding vector of interaction coefficients.

We assume that the error term ϵ_{sct} is i.i.d. with a logistic distribution (with a scale paramter fixed at 1), and we normalize the payoff to not contributing as zero. Further, $\hat{p}_{s'ct}$ denotes the conjectured probability that state s' contributes to cluster c in the current period t. The solution concept is a simultaneous-move Nash equilibrium. We assume that states are not forward-looking (for instance, each state legislature lives one period).

First, the term α_{sct} captures the baseline utility of adopting the policy. Here, α_c reflects cluster-specific factors that influence the inherent attractiveness or cost of the policy, while α_{st} accounts for state-specific and temporal factors common across clusters. The additional components $\boldsymbol{\omega}' \boldsymbol{X}_{st}$ incorporate state-year characteristics such

as GDP or political conditions.

Second, the parameter δ_c captures the persistence effect of policy adoption within a cluster. Specifically, it measures how a state's own history of past adoptions in cluster c (i.e., $\sum_{t' \in T_{-t}} n_{sct'}$) affects its current decision.

Third, diffusion effects are incorporated via the terms involving $\beta_{ss'ct'}$ and $\beta_{ss'ct}$. The term $\sum_{s' \neq s} \sum_{t' \in T_{-t}} \beta_{ss'ct'}\, n_{s'ct'}$ captures the influence of other states' past adoptions on state s's current utility. In contrast, the term $\sum_{s' \neq s} \beta_{ss'ct}\, \hat{p}_{s'ct}$ reflects the contemporaneous impact of the expected behavior of other states, with $\hat{p}_{s'ct}$ denoting the conjectured probability that state s' contributes in period t.

To conclude, assume that state s forms conjectures about the adoption probabilities of all other states. Then, the probability that state s contributes to policy cluster c in period t is given by the logit choice probability

$$\begin{aligned} p_{sct} &= \frac{\exp\left(V_{sct}\right)}{1+\exp\left(V_{sct}\right)} \\ &= \frac{\exp\Big\{\alpha_{sct} + \delta_c \sum_{t' \in T_{-t}} n_{sct'} + \sum_{s' \neq s}\Big[\sum_{t' \in T_{-t}} \beta_{ss'ct'}\, n_{s'ct'} + \beta_{ss'ct}\, \hat{p}_{s'ct}\Big]\Big\}}{1+\exp\Big\{\alpha_{sct} + \delta_c \sum_{t' \in T_{-t}} n_{sct'} + \sum_{s' \neq s}\Big[\sum_{t' \in T_{-t}} \beta_{ss'ct'}\, n_{s'ct'} + \beta_{ss'ct}\, \hat{p}_{s'ct}\Big]\Big\}}. \end{aligned} \tag{4}$$

Finally, in our empirical analysis we sometimes present results in terms of the log odds of adoption—a monotonic transformation of p_{sct}. Specifically,

$$\log \frac{p_{sct}}{1-p_{sct}} = \alpha_{sct} + \delta_c \sum_{t' \in T_{-t}} n_{sct'} + \sum_{s' \neq s}\left[\sum_{t' \in T_{-t}} \beta_{ss'ct'} n_{s'ct'} + \beta_{ss'ct}\hat{p}_{s'ct}\right]. \tag{5}$$

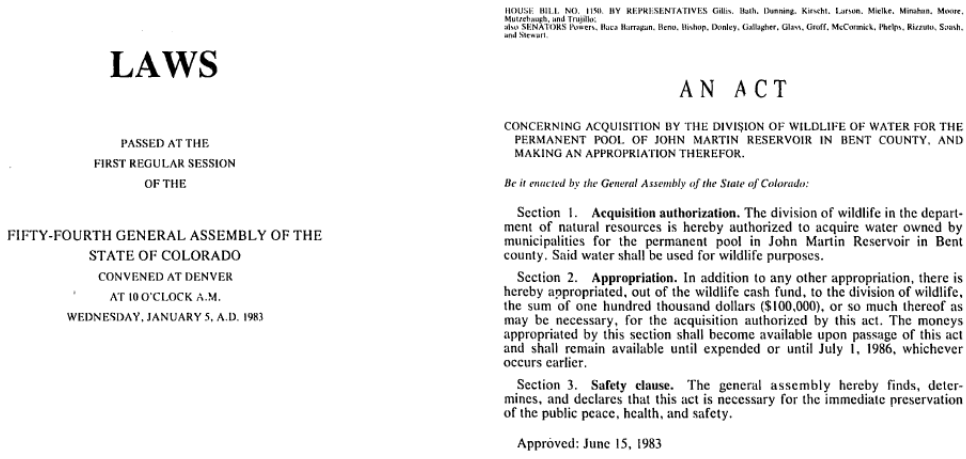
3.2 Equilibrium Uniqueness

In Appendix A.1, we show that a sufficient condition for the uniqueness of the equilibrium adoption probabilities is that the total peer influence on any state is bounded. Specifically, if

$$\max_s \sum_{s' \neq s} |\beta_{ss'ct}| < 4, \tag{6}$$

then the logit mapping is a contraction, ensuring that the probability vector $\mathbf{p}_{ct}$ is unique. This establishes that for any given vector of parameters, the equilibrium probabilities p_{sct} are uniquely defined.

Figure 1: U.S. State Session Laws Corpus



LAWS

PASSED AT THE
FIRST REGULAR SESSION
OF THE

FIFTY-FOURTH GENERAL ASSEMBLY OF THE
STATE OF COLORADO
CONVENED AT DENVER
AT 10 O'CLOCK A.M.
WEDNESDAY, JANUARY 5, A.D. 1983

HOUSE BILL NO. 1150. BY REPRESENTATIVES Gillis, Bath, Dunning, Kirscht, Larson, Mielke, Minahan, Moore, Mutzebaugh, and Trujillo;
also SENATORS Powers, Ruca Barragan, Beno, Bishop, Donley, Gallagher, Glass, Groff, McCormick, Phelps, Rizzuto, Soash, and Stewart.

AN ACT

CONCERNING ACQUISITION BY THE DIVISION OF WILDLIFE OF WATER FOR THE PERMANENT POOL OF JOHN MARTIN RESERVOIR IN BENT COUNTY, AND MAKING AN APPROPRIATION THEREFOR.

Be it enacted by the General Assembly of the State of Colorado:

Section 1. **Acquisition authorization.** The division of wildlife in the department of natural resources is hereby authorized to acquire water owned by municipalities for the permanent pool in John Martin Reservoir in Bent county. Said water shall be used for wildlife purposes.

Section 2. **Appropriation.** In addition to any other appropriation, there is hereby appropriated, out of the wildlife cash fund, to the division of wildlife, the sum of one hundred thousand dollars ($100,000), or so much thereof as may be necessary, for the acquisition authorized by this act. The moneys appropriated by this section shall become available upon passage of this act and shall remain available until expended or until July 1, 1986, whichever occurs earlier.

Section 3. **Safety clause.** The general assembly hereby finds, determines, and declares that this act is necessary for the immediate preservation of the public peace, health, and safety.

Approved: June 15, 1983

4 Data Sources

This section describes the Data sources and method for converting the statute texts to vectors for our similarity analysis.

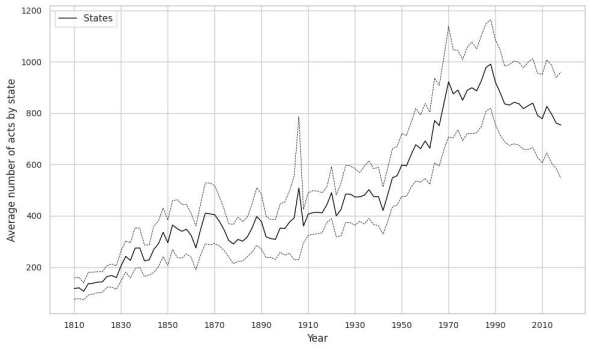
State Session Laws Corpus. Our state legislative records contain the full text of U.S. State Session Laws. The session laws contain legislative acts enacted by each state legislature during each session. This corpus has been used in several papers in economics and political science (Ash, 2018; Vannoni, Ash and Morelli, 2021; Ash, Morelli and Vannoni, 2021). The corpus come as plain-text scans of the session laws, which are published as annual or biennial journals (Figure 1).

The Session Laws may introduce new laws, make modifications (amendments) to existing laws, or repeal existing laws (deleting). Ideally, it would be possible to distinguish amendments from repeals. Unfortunately, the corpus does not provide a machine-readable approach for distinguishing these in the original text. Therefore, our text measures cover all types of legislation without distinction between amendments or repeals. However, we were able to perform qualitative tests on samples in the corpus. We found that the proportion of amendments and cancellations is relatively small. Further, we scanned the relevant text for indicators ("amend*" or "repeal*"). They are relatively rare (about 3% mention repeal and less than 1% mention amendments).

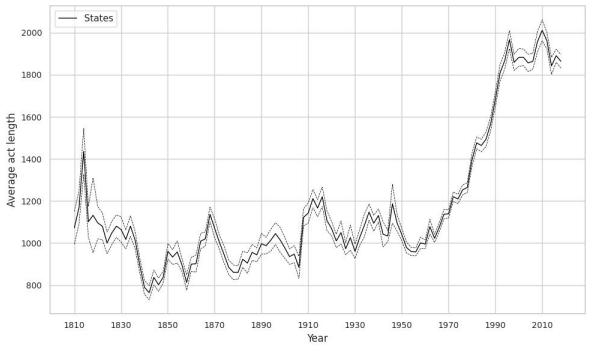
The next issue is the texts are generated by Optical Character Recognition (OCR) applied to printed legislation. When examining samples, OCR is of high quality. Regulations are old, but the quality of digitization is quite good. Nonetheless, as with any historical digitized corpus, there are a non-negligible number of OCR errors. To

Figure 2: Evolution of the quantity of acts

(a) Average number of acts per state



(b) Average number of words per act



Notes: The left panel shows the average number of acts per state per biennial legislative period. The right panel shows the average number of words per act per biennial legislative period. We report the 95% confidence interval of the mean.

investigate this, we calculated a proxy OCR error rate as the common (non-proper) noun misspelling rate. The misspelling rate is low (about 3% of common nouns) and smooth over time. These OCR errors may add measurement error to our similarity measures.

The laws are published once a year or every two years. To ensure consistency, the dataset is created every two years.

U.S. Federal Statutes at Large. We also obtain the Federal Statutes at Large, containing equivalent legislative act texts for the U.S. federal government. These come from the Library of Congress and U.S. Government Information (GovInfo) web sites.

Text Pre-processing. The first step is to merge and process this raw text. A script serves to append pages, remove headers, footers, tables of contents, indexes, and other non-statute material. Then, it segments the text into individual bills, acts, and resolutions using text markers for the start of new statutes. These include indicators for new Chapters, Articles, or Titles, such as a line with CHAPTER followed by a Roman numeral. Some states have their own standard indicators, such as P.A followed by a number to indicate a new Public Act. The script also uses common text for the beginning of a statute preamble (e.g., An act to...) and for enacting clauses (e.g., Be it enacted that...). Research assistants checked samples of the statute segmenter for each state-biennium to make sure it worked well. Figure 2 Panel A plots the average number of acts per state over time.

Next, for each statute, we normalize the text to lowercase and remove punctuation. We then apply an automated spellchecking tool to correct misspellings. Next, we further filter the text by dropping non-letter characters and dropping common *stopwords* like 'the'/'to'/'is'. We then represent statute documents as lists of words. Figure 2 Panel B plots the average number of words per act over time.

Political, economic, and geographic data. We combine several datasets in order to collect information on the demographic, economic and political conditions of each state. Economic data come from the NAICS and SIC sector shares for each year from 1940 onward, from Klarner (2015). We complement this data with NHGIS Employment by industry from decennial census files going back to 1810 (Manson, 2020).

5 Measurement of Policy Output

This section describes the AI-based approach for extraction of policies using a large language model and the subsequent encoding using a transformer-based context-sensitive encoder.

LLM extraction of policy content. First, we took the first 400 words[2] of each pre-processed statute, and summarized their policy content using GPT-3.5 (gpt-3.5-turbo-1106). We used the prompt listed in Figure 3. The prompt first provides context on the source document. Second, the assistant provides a short-phrase topic label. Third, the assistant is asked to summarize the policy being enacted. To improve quality of the outputs, the prompt includes instructions to limit the summary to 40 words and to use the entire text, and then three example outputs are provided.

The summaries produced by GPT with this prompt are extremely high-quality (see Figure 3 for an example). The summaries effectively articulate the policy content of a statute, even for historical documents and in the presence of OCR errors. From inspections of samples, we can say that almost all summaries are reasonable. It sometimes provides poor output, but that is the result of poor inputs, for example because a statute does not really enact a policy, or enacts multiple policies.

Next, we want to adjust the content of the summaries to further focus on policy content and remove information about the institutional context, such as information about time and place. We need to do this because we want the policies to be comparable across time and space. We undertake an additional AI-based processing step to

[2]We validated for a sample of 100 statutes that giving the first 400 words provided similar summaries to including the whole statute.

Figure 3: Extraction of Policies with GPT

Prompt

Here is an excerpt from a statute of a particular U.S. state government that enacts a policy. The extract, taken from a digital scan, may contain OCR errors and lacks punctuation. This excerpt includes the first 400 words from the statute. Execute two tasks: **A. Provide a brief phrase, approximately 5 words, describing the topic covered. B. Write a concise sentence, not exceeding forty words and utilizing the entire text, summarizing the policy being enacted, beginning with 'This policy ...'** For instance: 'A. Income Taxes. B. This policy introduces income taxes,' 'A. Criminal Penalties for Embezzlement. B. This policy establishes criminal penalties for embezzlement,' or 'A. Occupational Licensing for Barbers. B. This policy governs occupational licensing for barbers'.

User Prompt

public law of a of a a with congress an act to require the department of justice to develop and implement a national strategy child exploitation prevention and interdiction to improve the internet crimes against children task force (...)

Answer

Child Exploitation Prevention and Interdiction. This policy requires the development of a national strategy for child exploitation prevention, increases resources for internet crimes against children task forces, and makes other improvements to aid law enforcement in investigating and prosecuting child predators.

"anonymize" the content of the policies. We need to do this because we are going to vectorize and cluster the policy texts, and if we did not do the adjustment, the resulting similarity metrics and clusters would load substantially on time and place characteristics. For example, you would get a Texas cluster and an Idaho cluster, rather than a tax cluster and a spend cluster (see Appendix A.6.3).

The prompt for anonymizing the policy extracts is shown in Table 1. We carefully adapt the prompt through an iterative procedure to remove information about locations, but in a way that maintains the important policy information. In the included example on railroad incorporation, "Four Mile Valley Rail Road Company' becomes "a specific railroad company," "Hamilton" becomes "a specific city", "Ohio" becomes "a specific state" and "Indiana" becomes "another specific state". The policy-relevant information is maintained, but the information about the local context is excised.

Assignment of CAP topics to Policy Extracts We assign a policy topic and subtopic to each statute summary using the Comparative Agenda Project (CAP) topics and subtopics. Since a policy can address multiple topics, we allow for a primary topic and a potential secondary topic, using GPT-4o-mini in a two-step approach.

In the first step, GPT assigns a main topic (and, if applicable, a secondary topic) using the prompt in Figure A.1, which is shown along with an example. Then, conditional on the assigned primary topic, we ask GPT to select the most relevant subtopic from the list of subtopics associated with that main topic. Each CAP topic and its

Table 1: Policy Anonymization Prompt

Prompt

Your task is to fully anonymize a policy summary by removing all specific identifying information. Instructions:

- Keep the title and text as close to the original as possible.
- Replace all names, organizations, places, and time periods with generic terms (e.g., "a specific state," "a specific city," "a specific individual," "a specific year").
- Replace mentions of the state from which the statute originates with "in the state," unless it refers to the Federal Government.
- Replace mentions of locations outside that state with "in another specific state," "in another specific state's city," or "in another specific state's town," as appropriate.
- Maintain a neutral and objective tone throughout.

The final summary should be entirely anonymized, with no specific identifying information remaining.

User Prompt

Here is the state from which the statute originates: "*Ohio*"
Here is the policy summary: "*Railroad Incorporation. This policy incorporates the Four Mile Valley Rail Road Company with the power to construct a railroad from Hamilton to the state line between Ohio and Indiana, and to Richmond, Indiana.*"

Answer

Railroad Incorporation. This policy incorporates a specific railroad company with the power to construct a railroad from a specific city to the state line between a specific state and another specific state, and to another specific state's city.

corresponding list of subtopics (exactly as fed to GPT) is provided in Appendix B.1.

Appendix Figures A.2 and A.3 respectively show the proportion of assigned main topics among statutes and the top 15 topic-subtopic combinations' frequency among statutes.

MPNet Encoding of policy summaries. Next, we encode the anonymized policy summaries with a pre-trained transformer-based short document encoder, MPNet (Song et al., 2020). MPNet is a state-of-the-art encoder with good all-around performance in embedding evaluations and applications, especially document similarity and clustering. It is a BERT-based model pre-trained with a contrastive learning objective on a sample of 1.1 billion sentence pairs, to guess whether the pairs of sentences are related or unrelated.

The advantage of MPNet is that it interprets documents using the full context; for example, the word "bank" can have a different meaning for the regulation of financial institutions versus the regulation of river-ways. The result is a set of 768-dimensional vectors representing the policy content of each statute.

We compared MPNet to another popular embedding model called FastText, which

Figure 4: Topic Assignment of Policy Summaries with GPT

User Prompt

Input Example:
Here is a summary of a state statute: "*Turnpike Road Company Organization. This policy authorizes the organization and division of a specific turnpike road company into sections, and the creation of another specific turnpike road company to complete construction.*"

Prompt

Your task is to categorize this policy summary among a list of possible policy topics.
Here is the list of different policy topics:
"*Macroeconomics, Civil Rights, Health, Agriculture, Labor, Education, Environment, Energy Policy, Immigration, Transportation, Law and Crime, Social Welfare, Housing, Domestic Commerce, Defense, Technology, Foreign Trade, International Affairs, Government Operations, Public Lands, Culture*"
Respond with the primary policy topic in the format: Primary: [policy topic name].
If applicable, you may include secondary topics by adding: Secondary: [policy topic names].
Prioritize selecting only one or the fewest possible topics, focusing on relevance and listing them in order of importance.

Answer

Output Example:
"*Transportation*"

is based on averaging the word vectors in a statute. We compared the quality of clusters produced by MPNet and FastText (more detail below), and we found that the clusters produced by MPNet were of higher quality. See Appendix A.6.1 for details.

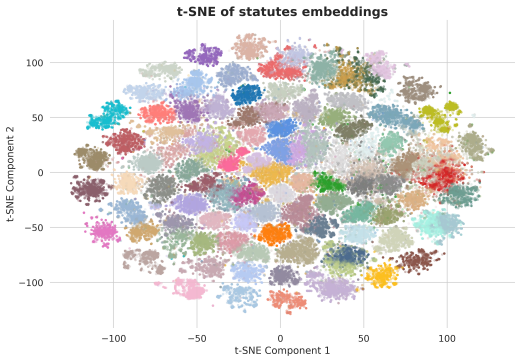
5.1 Constructing Policy Clusters

An advantage of the policy embeddings is that they are comparable across years and states. Therefore, we can identify discrete policies by extracting local clusters of policy summaries in the MPNet vector space.

Specifically, we apply k-means clustering to the policy embeddings and specify 2,048 clusters. This algorithm learns 2,048 centroids in the embedding space and assigns each policy summary to the closest centroid. We selected 2,048 clusters based in computing the silhouette score for the assigned clusters for a range of values of the number of clusters. This process is described in Appendix A.6.2.

Figure 5 provides an illustration of the k-means clustering result. The first panel shows a t-distributed Stochastic Neighbor Embedding (t-SNE) visualization of the top 100 clusters with the most statutes from the 2048-cluster configuration. t-SNE is a dimensionality reduction technique used to visualize high-dimensional data in two dimensions. In this case, the technique is applied to cluster data to reveal patterns and relationships between clusters based on their statute counts. Each color in the

Figure 5: K-Means Clustering on Policy Embeddings



Policy	Policy from same cluster
Repeal of Road Laws. This policy repeals certain road laws applicable to a specific area and replaces them with new provisions.	Repeal of Charitable Incorporation and Road Authorization. This policy repeals the incorporation of a charitable society and authorizes new commissioners to lay out a public road.
Domestic Violence Definitions. This policy amends the penal code to define domestic violence and related terms, and outlines the roles of officers and victims in cases of domestic violence.	Domestic Violence Shelter Search Warrants. This policy clarifies that requesting a search warrant before allowing law enforcement into a domestic violence shelter does not constitute a crime, and requires a search warrant before serving an arrest warrant in such shelters.
Bridge Construction Consent. This policy grants consent to a specific county and another specific county to construct, maintain, and operate a bridge across a specific river, in accordance with specified provisions.	Bridge Construction Consent. This policy grants consent to construct a bridge across a specific river between a specific borough in a specific county and another specific borough in another specific county, subject to specified conditions and limitations.

plot represents a cluster, and each point represents a statute. The positioning of the points reflects the similarity between clusters, with clusters that are closer together being more similar.

The right panel shows examples of policies in the same cluster and in different clusters. We can see that the associated policy texts are grouped by their semantic content. We provide informative labels to the clusters by sampling 10 policy summaries from each cluster and prompting GPT to label the associated policy.

We validate the cluster quality in two ways using AI comparisons. First, we evaluate whether policies in summaries within the same cluster are more similar to each other compared to policies from different clusters. For 4,000 random statute summaries, we randomly select one policy from the same cluster and another policy from a different cluster, within the same CAP topic. We assess which policy is more similar to the initial policy using GPT-4o-mini, based on the prompt presented in Appendix Figure A.13. As shown in Appendix Table A.9, statutes from the same cluster are generally preferred, demonstrating that clustering effectively captures policy similarity.

Second, we evaluate whether the policies from the same cluster pursue aligned or opposing policy goals, to check whether opposite policies (e.g. increase taxes or decreases taxes) get in the same cluster. For 10,000 sampled policy extracts, we randomly select one policy from the same cluster and another policy from a different cluster, within the same CAP topic. We assess for each pair whether the policies pursue aligned or opposite goals using the GPT-4o-mini prompt presented from Appendix Figure A.14. As shown in Appendix Figure A.15, statutes from the same cluster tend to pursue aligned policy objectives, and less so with cross-cluster pairs.

Table 2 presents descriptive statistics for the policy clusters. As desired, there are few clusters that are limited to a single states, and they tend to be observed over long

Table 2: Descriptive Statistics for 2048-Cluster Configuration

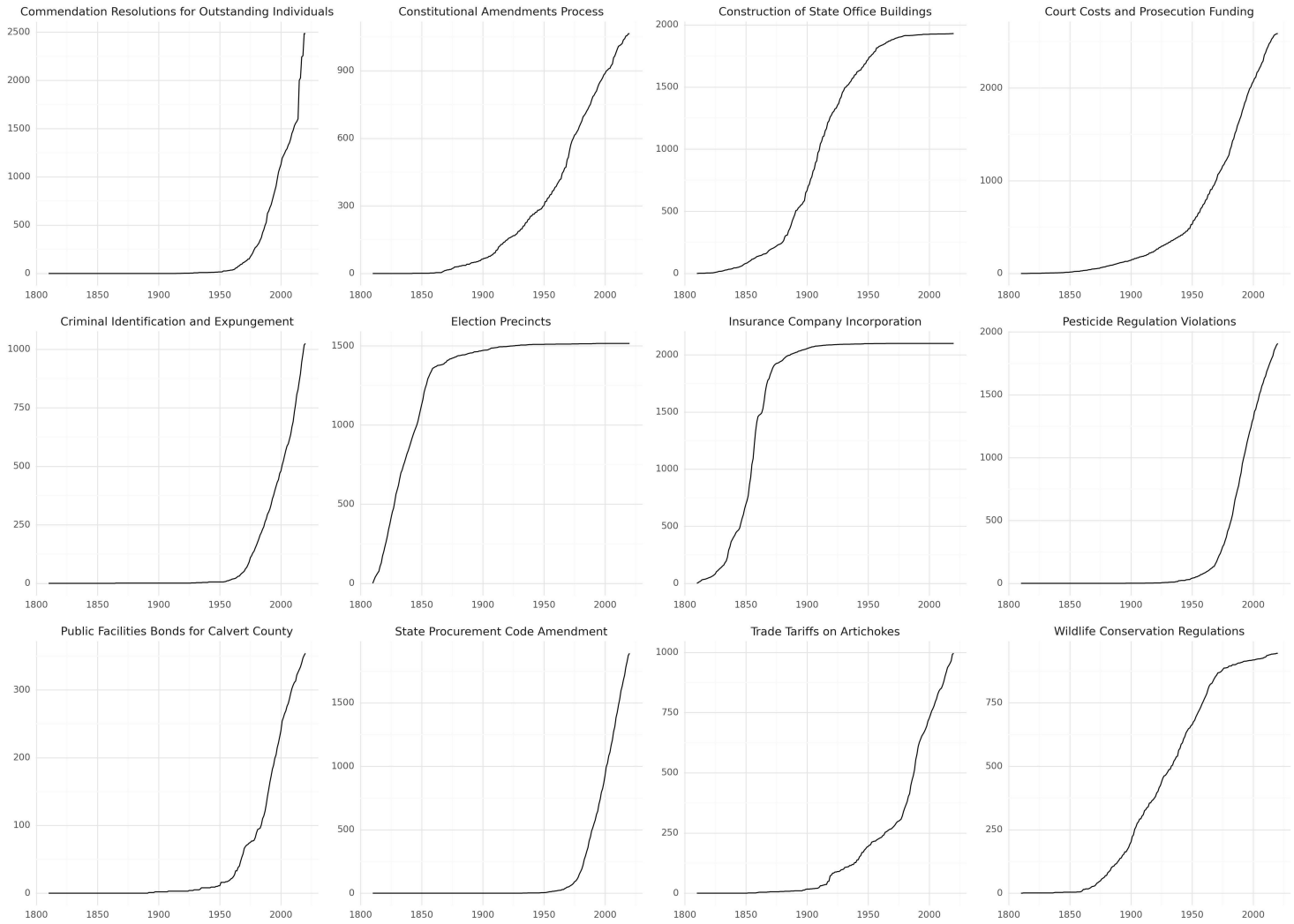
	Mean	Std dev.	5th pctile	95th pctile
Number of statutes per cluster	1219.9	723.9	338.4	2565.9
Time span per cluster	90.0	35.3	34.0	146.0
Number of states per cluster	44.7	9.3	23.0	50.0
HH index	0.43	0.16	0.24	0.73
Share of clusters with more than one state contributing	99.9%			
Share of clusters where all states are contributing	44.7%			
Share of clusters with federal government contributing	79.7%			
Share of clusters where a state contributes more than half	1.3%			
Share of clusters where federal government contributes more than half	1.7%			

Note: For statistics related to years, we removed statutes below the 10th percentile and above the 90th percentile per cluster. For statistics on states, we excluded federal government statutes from the clusters.

time periods. Hence, the method captures policies that are comparable across states and over time.

We can then get at policy priorities by state by counting the number of times a policy is enacted for each cluster. Figure 6 shows the evolution of the stock of mentions for a sample of clusters over the time period. The figure shows that different policies were enacted at different times. For example, policies around electoral precincts were introduced already in the early and mid-1800s. Regulations about pesticides were introduced in the late 1900s.

Figure 6: Evolution of the stock of 12 randomly picked clusters



Notes: Each graph shows the cumulative stock of acts adopted in each cluster over the period. Cluster labels provided by GPT prompt based on a sample of associated policy summaries.

6 Empirical Analysis

6.1 Estimation

We now take the discrete choice model from Section 3 to the data. First, we make the following additional functional-form assumptions. First, for α, we have

$$\alpha_{sct} = \alpha_c + \alpha_{st} + \omega_{1c}\,\text{Pol}_{st} + \omega_{2c}\,\text{Econ}_{st} + \omega_{3c}\,\text{Geog}_{st}, \tag{7}$$

The parameters ω_c capture the extent to which a policy is more likely to be adopted by states with specific political, economic, or geographic conditions.

Next, for the interaction parameter β we assume

$$\beta_{ss'ct'} = \beta_c + \beta_{s'\tau} + \beta^f_{s'\tau} + \gamma_{1\tau}\,\text{PolSim}_{ss't'} + \gamma_{2\tau}\,\text{EconSim}_{ss't'} + \gamma_{3\tau}\,\text{GeoSim}_{ss'}. \tag{8}$$

The parameters $\beta_{s\tau}$ capture the extent to which a policy adopted by state s tends to later be adopted by other states. We also allow this effect to differ for states' influence on the federal government, denoted by $\beta^f_{s'\tau}$. The parameters γ capture the extent to which policies tend to spread more between states that are politically, economically, or geographically similar. We allow the main coefficients of interest to vary flexibly over time, where $\tau = 10$ in the main specification.

The "Sim" measures are defined as follows. Following DellaVigna and Kim (2022), we construct a binary measure of political similarity, defined as a dummy variable equal to one if both the legislatures and governors are controlled by the same party in both states. For economic similarity, we compute the cosine similarity between states' employment sector shares and define a binary measure indicating whether the two states fall within the top third of the similarity distribution. Similarly, for geographic similarity, we calculate the geographic distance between each state's population-weighted centroid and create a dummy variable indicating whether the two states are among the top third closest pairs.

Due to well-known identification issues in discrete choice models, we impose several normalizations. In particular, we fix one cluster-specific baseline parameter, α_c, and one state-year baseline parameter, α_{st}, to arbitrary values. Similarly, we set one cluster-specific diffusion parameter, β_c, and one state-year diffusion parameter, β_{st}, to arbitrary fixed values. These normalizations do not affect the differences in the V_{sct} payoffs – and hence preserve differences in the log-odds of adoption, $\log \dfrac{p_{sct}}{1-p_{sct}}$, across states, clusters, or years – but they do affect the absolute levels of differences across probabilities p_{sct}.

Next, note that the log-likelihood function depends solely on the probabilities p_{sct}, which are determined by a logistic function. As a result, the mapping from the parameter vector Θ to the predicted probabilities p_{sct} is injective. Moreover, under standard regularity conditions (full rank condition) the log-likelihood function is strictly concave. This strict concavity guarantees that the maximum likelihood estimator is unique.

To recover the parameters (α, β, δ), we embed a fixed-point iterative algorithm for all states' probabilities p_{sct} for every cluster and year in a maximum likelihood estimation, where we maximize the log-likelihood provided below:

$$L(\theta) = \sum_{sct} \left(n_{sct} \ln(p_{sct}) + (1 - n_{sct}) \ln(1 - p_{sct}) \right) \tag{9}$$

For now, we abstract from considering contemporaneous strategic interactions and focus on the impact of previous laws, which allows us to maximize a simpler likelihood function.

Discussion. Under the imposed normalizations and the contraction condition on the β's, identification in our model relies on variation across clusters, states, and time—even though all states are fully connected and thus share the same set of peers. Variation in average adoption rates across clusters, after controlling for state–year effects, identifies the cluster-specific baseline α_c and the policy-specific coefficients ω_{1c} and ω_{2c}. Differences in adoption behavior across states within the same period reveal the state–year fixed effects α_{st}. Time-series variation within a state, where changes in its own past adoption patterns predict current behavior, identifies the persistence parameter δ_c.

Although every state considers the same set of peers, cross-sectional differences in historical and contemporaneous adoption rates provide identification for the diffusion effects. In particular, differences in past adoption behavior across states isolate the lagged interaction parameters $\beta_{ss'ct'}$, which decompose into a cluster-level baseline β_c, a state–time-specific effect $\beta_{s't'}$, and coefficients γ_{1c}, γ_{2c}, γ_{3c}, and γ_{4c} reflecting the impacts of political similarity, economic similarity, geographic proximity, and federal status, respectively. Furthermore, cross-sectional differences in current adoption—after controlling for each state's own past behavior—identify $\beta_{ss'ct}$, which measures the immediate influence of contemporaneous peer actions.

For example, suppose that in one policy cluster, after controlling for state–year fixed effects and historical adoption, there is a strong cross-sectional correlation in contemporaneous adoption decisions — when one state adopts, its peers are significantly more likely to adopt as well. This pattern leads the model to estimate a high

β_c. In contrast, if another cluster shows little or no such correlation after accounting for the same controls, the model would infer a low β_c. Alternatively, if after controlling for fixed effects and past adoption, an increase in peer adoption is associated with a lower probability of a state adopting in the current period, this negative correlation leads to a negative estimate for β_c.

Further, if in a given period t' a state s''s past adoption is strongly associated with increased subsequent adoption by other states, then $\beta_{s't'}$ would be estimated as high. Conversely, if state s''s past adoption exhibits little or even a negative association with peers' current decisions, then $\beta_{s't'}$ would be estimated as low or negative.

Next, if after controlling for other factors, states that are politically similar (i.e., with a high $\text{PolSim}_{ss't'}$) tend to exhibit stronger correlated adoption behavior, then γ_{1c} would be estimated as high and positive. The same logic holds for economic similarity and geographic proximity.

Finally, if when the federal government adopts a policy in cluster c, states are significantly more likely to follow suit, then γ_{4c} will be estimated as high and positive, reflecting a strong positive influence. Conversely, if the federal government's adoption is associated with a lower likelihood of state adoption—perhaps due to crowding-out effects—then γ_{4c} would be estimated as low or even negative.

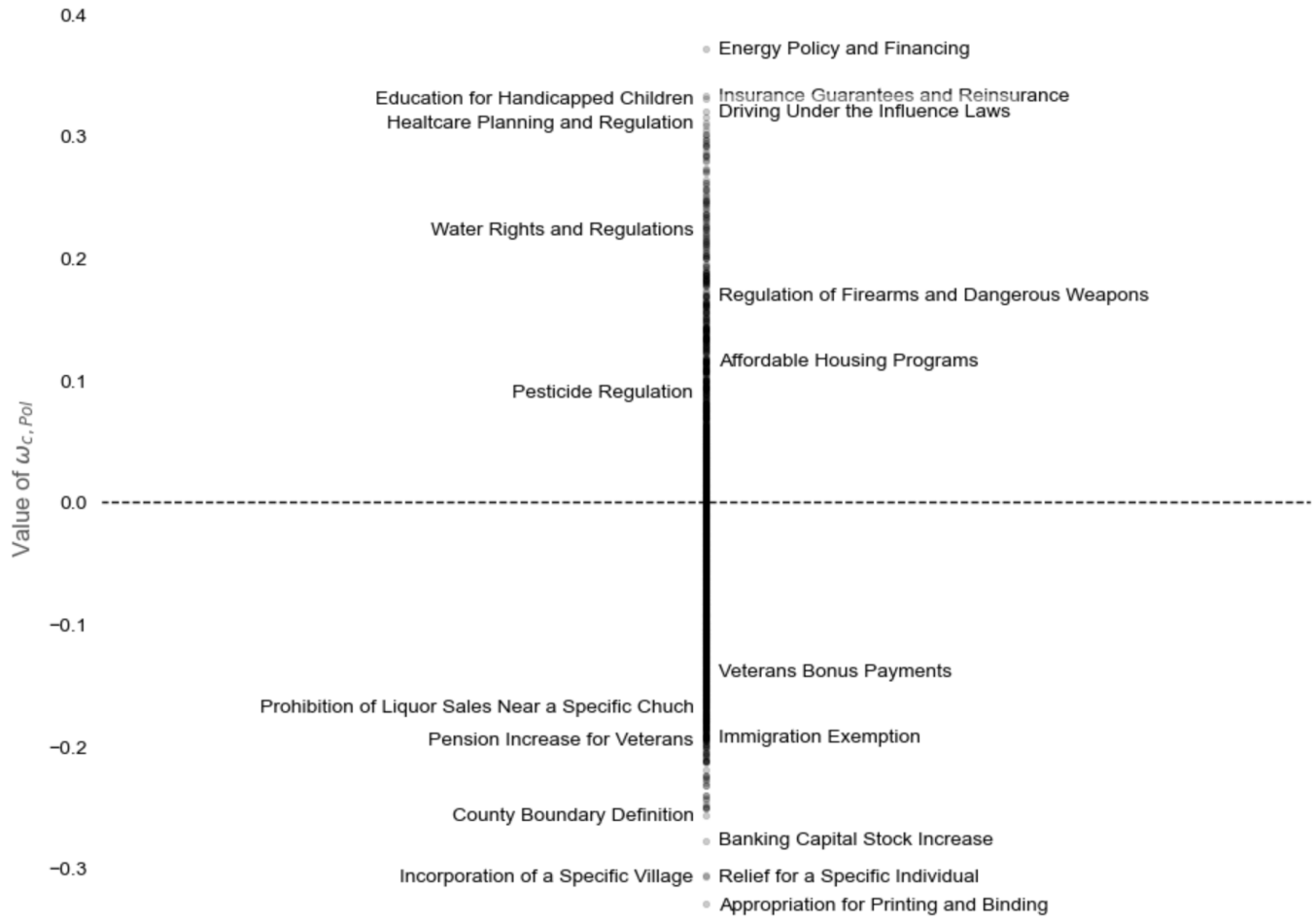
6.2 Variation in policy diffusion between states

We begin by using our structural estimates to provide a descriptive analysis of variation in the propensity to adopt laws. First, we estimate and analyze the coefficients ω_c from for the political dimension. These coefficients measure the extent to which a policy is more likely to be adopted by states with larger Democratic presidential vote shares.

Figure 7 shows these estimates, with the associated cluster labels. We see that there is substantial heterogeneity across policies, which reflect some intuition about partisan priorities. For example, Democrats emphasize energy policy, education for handicapped children, healthcare planning, regulation of firearms, and affordable housing. Republicans prioritize immigration regulations, veteran pension benefits and bonuses, and prohibition of liquor sales near churches. It is also notable how these priorities reflect relatively local priorities, compared to those one would see in the federal government.

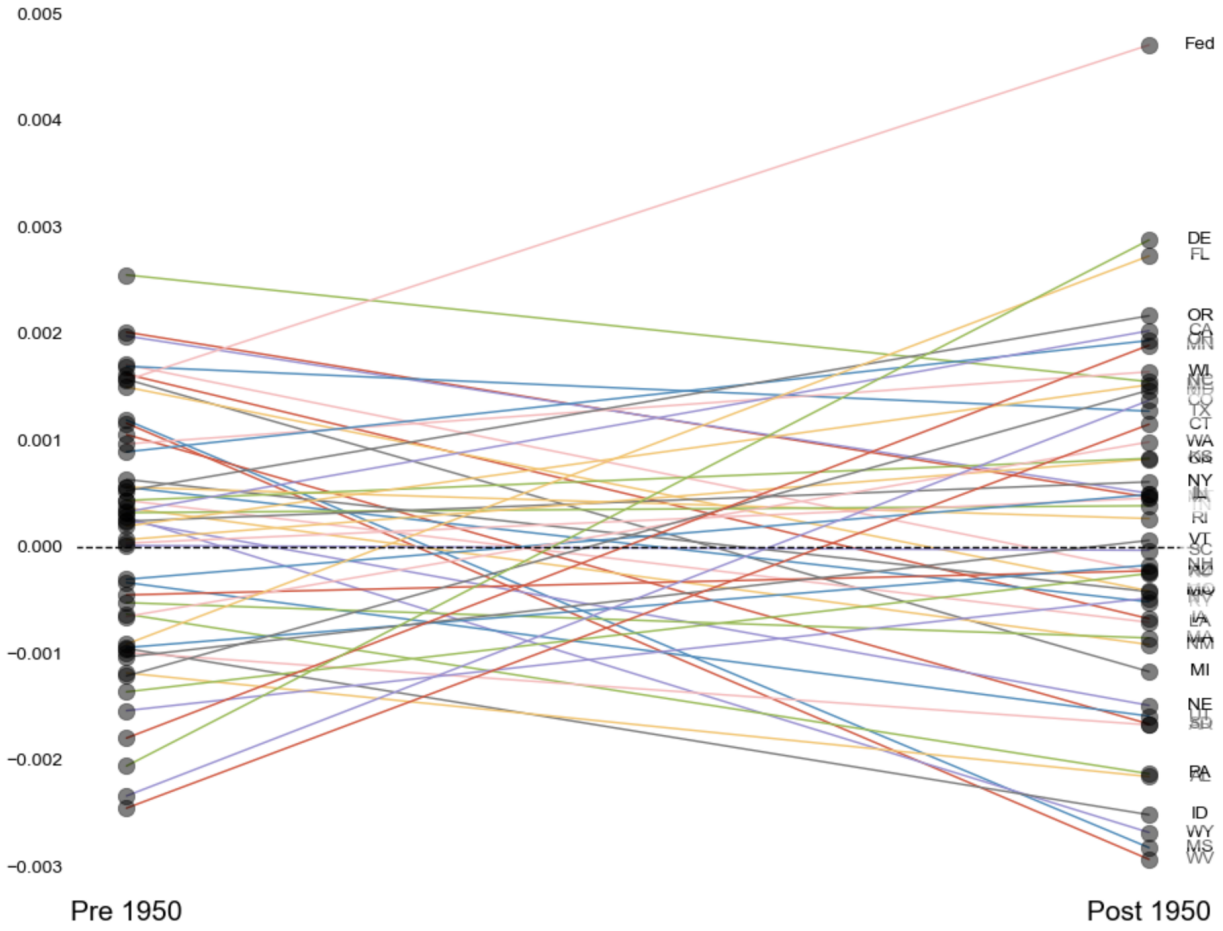
These coefficients have sizable magnitudes. For instance, the most Democratic-leaning policy, "Energy Policy and Financing," would be approximately 2 percentage points more likely to be adopted by states with a 10-point higher Democratic presidential vote share, independently of other states' behaviors.

Figure 7: Baseline Likelihood of Policy Adoption by Partisanship ($\omega_{c,Pol}$)



Notes: The figure shows the estimated parameters $\omega_{c,Pol}$ for the period 1980-2020. The coefficients measure the likelihood that each policy cluster is adopted by states with a higher Democratic presidential vote share.

Figure 8: Propensity to spread laws ($\beta_{s'}$)



Notes: This figure shows the parameters $\beta_{s'}$ from equation 8 estimated by MLE. The values with confidence intervals are reported in Appendix Table A.1.

Next we look at the parameters $\beta_{s'\tau}$, which summarize capture the extent to which state s''s policy adoption tends to lead to adoption by other states within the same policy cluster. states with higher $\beta_{s'\tau}$ are more influential in a given time period.

Figure 8 report these parameters with the associated state/federal label. We produce these estimates averaged over two periods: 1880–1950 and 1950–2020. Then we can see the ranking and broad secular changes.

The first noticeable trend is the increasing influence of the federal government on state policy adoption, indicated by the "Fed" label. In the earlier period, while the federal government was relatively influential, it was not the highest-ranked. In the later period, it was by far the highest-ranked in influence. We will return to this issue in Section 6.4 below.

Next, we observe strong heterogeneity in states' influence on one another. Some states, such as Delaware, Florida, and California, exert a strong influence on others, whereas states like Wyoming, Mississippi, and West Virginia have a much more modest impact. We note that Delaware's significant influence on other states is likely driven by its role as the primary jurisdiction for corporate law cases, given that the majority of corporations are incorporated in that state.

6.3 Determinants of policy diffusion

Next, we examine the extent to which states' influence on each other depends on their political, economic, or geographic similarity. From the discrete choice model, we estimate and report the coefficients γ by decade.

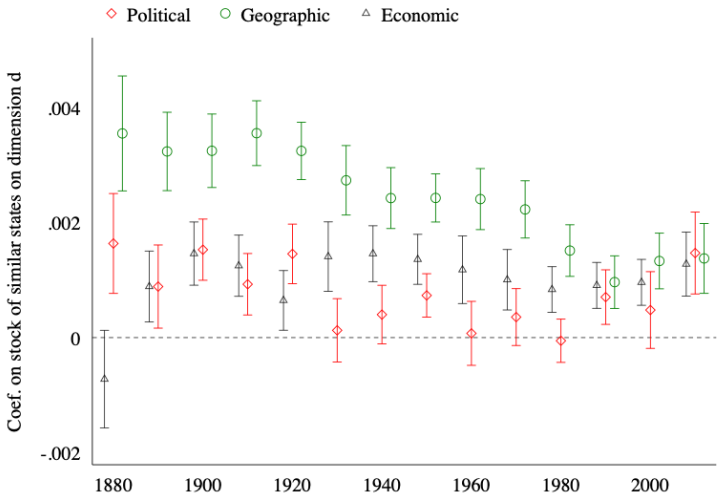
Figure 9 gives the estimates. As can be seen in green, geographic proximity once stood out as the strongest predictor of policy adoption, especially during the early to mid-20th century. Over time, however, its impact has diminished, suggesting that states now give less weight to physical adjacency when choosing policies. Despite this change, geography remains a nontrivial factor.

In gray, we assess the impact of economic similarity of occupations and sectors. It has consistently mattered. Even as the effect of geography declined, states continued to draw on shared economic attributes when shaping legislative decisions.

Finally, in red, we see partisan effects. Partisanship was a significant driver in the early 1900s but then tapered off for several decades, only to regain strength after the 1990s. This recent rise in partisan considerations is consistent with broader trends in political polarization and suggests that party alignment increasingly drives how states emulate one another's policies.

We subjected these core findings to several robustness checks. First, we estimated

Figure 9: Impact of state similarity on policy diffusion.



Notes: The figure shows the coefficients ω from equation 8 estimated by MLE. Standard errors are clustered by policy by state and by state by year.

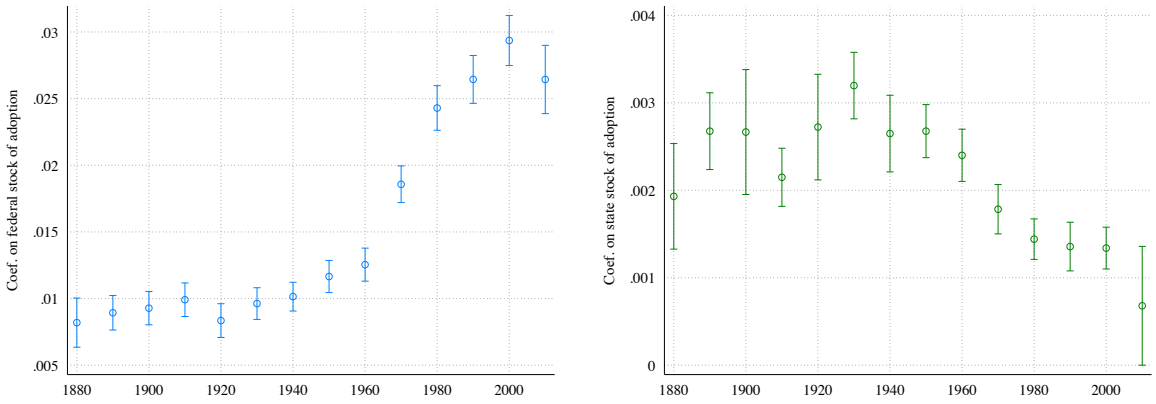
separate regressions by dimension (political, geographic, economic) to ensure that combining them in a single specification did not drive the results. Next, we varied the number of policy clusters, confirming that different levels of clustering do not undermine our conclusions. We also experimented with different normalizations of our exposure measures. Finally, we combined variations in the exposure metric with changes in the number of clusters, confirming that the main insights hold under these alternative configurations.

Appendix Figure A.16 examines whether states diverge in their choice of broad policy subtopics, aggregated into 115 categories. The exposure measure for each dimension (political, geographic, economic) is computed from these aggregated subtopics rather than fine-grained clusters. The results show that states are generally aligned in the topics they choose to address, suggesting that policy polarization arises more from how states legislate on shared issues than from disagreements over which topics merit legislative attention.

6.4 Results: Federal Influence on State Policy

We next examine how federal legislation shapes state policy choices and vice versa by analyzing the coefficients $\beta_{f\tau}$ and $\beta^s_{s\tau}$. The former captures the effect of federal policy adoption on the average state, while the latter reflects the average effect of a

Figure 10: Influence of Federal Legislation on State Policy Adoption



Notes: The first panel shows the coefficient $\beta_{f\tau}$, which captures the average impact of the federal government's adoption of a policy on state adoption of the same policy. The second panel shows the coefficients β^f_{st}, which capture the average impact of a state's adoption on the federal government's adoption of the same policy.

state's policy adoption on federal adoption.

Figure 10 (left panel) summarizes the coefficient estimates, indicating a positive association between federal adoption of policies in cluster c and subsequent state-level activity in that same cluster. The size of the coefficient $\beta_{f\tau}$ has increased over time, in line with Hopkins (2018) and others who have documented a growing "nationalization" of American political life. This suggests that states are more likely to follow the federal lead now than in earlier decades, potentially reflecting a more unified policy environment or greater policy interdependence among levels of government.

The coefficients $\beta^f_{s\tau}$, in comparison, are an order of magnitude smaller. The influence of collective state adoption on federal legislation is more limited and has been declining over time, underscoring an asymmetry in the federal–state relationship.

Taken together, these findings reinforce the idea that federal policy has become an increasingly salient reference point for states. The null or weaker result in the placebo checks supports a directional story: federal legislative decisions provide a template that states adopt, rather than states consistently prompting federal action. This convergence on federal policy frameworks affirms the notion that state-level politics have become nationalized, with broad implications for the future of policy diffusion and innovation across the United States.

7 Alternative Supervised-Learning Approach to Partisan Influence

An alternative method for measuring the partisan influence on policymaking is to use a supervised learning approach based on text and political affiliation. Drawing on prior work in political text classification (**???**), we see how well textual features of legislation predict the political party controlling a state's government. We then would like to see whether the time series of the supervised-learning measure of political influence on policymaking matches what we found earlier with the main cluster-based approach.

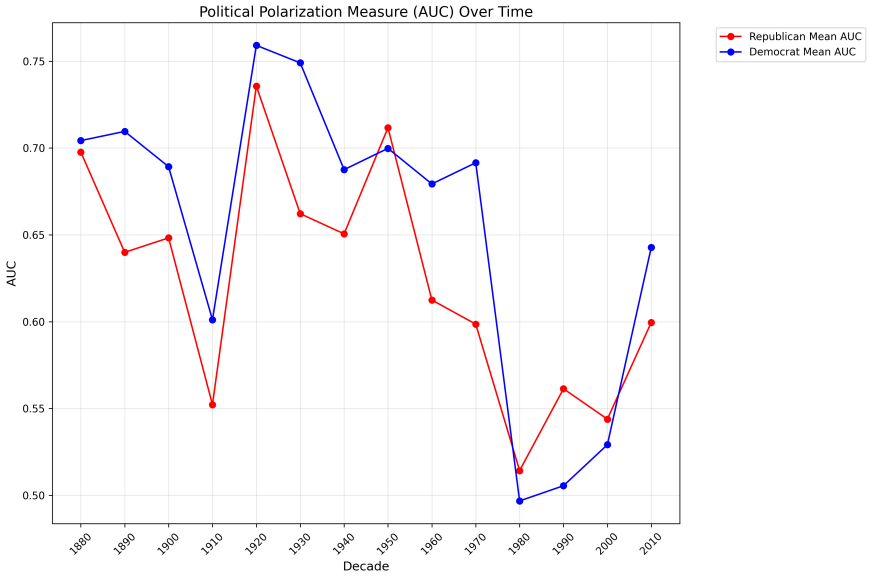
For each policy i enacted in state s at year t, we associate a 768-dimensional MP-Net embedding representation, $W_{i,s,t}$. That vectorization comes from the approach described in Section 5, where we use the LLM to extract the policy from a statute and use MPNet to assign the vector representation. The difference is that we analyze the vectors themselves rather than the subsequently constructed clusters.

The outcome label associated to i is whether s was under unified Democratic, unified Republican, or mixed-party control at time t when policy i was enacted. Our preferred machine learning algorithm is XGBoost, an optimized gradient-boosted classifier (Chen and Guestrin, 2016). XGBoost works by building an ensemble of decision trees that "vote" on the classification outcome, where different trees use different variables and different samples of data. A logistic regression model gives similar results.

Our measure of polarization is based on the performance of the ML model in predicting whether each statute is written by a Republican-controlled or a Democrat-controlled legislature. For this purpose, we compute the area under the (receiver-operating characteristic) curve (AUC) achieved by the model in held-out test data. AUC measures the overall ability of a model to distinguish between two groups, with values closer to 1 indicating stronger predictive accuracy and values around 0.5 reflecting random guessing. It can be interpreted as the probability that, for two randomly paired observations from each group, the model correctly guesses which observation is from which group. Because we have three groups (Republican, Democrat, and Mixed), we produce two AUC measures: AUC for Republican vs, not-Republican, and AUC for Democrat vs not-Democrat.

We implement a decade-by-decade training and testing procedure. For each decade, we randomly split the U.S. states into two equally sized groups, and we collect the respective samples of statutes, A and B. We then fit two XGBoost models separately for each sample, using the respective statute embeddings and party control labels. The

Figure 11: Partisan Influence on Policy: Alternative ML Approach



Notes: Performance over time of a text classifier for predicting party control based on statute embeddings. Chart shows the AUC score for Republican vs. not-Republican (red) and Democrat vs Not-Democrat (blue) based on the test-set evaluation of the XGBoost algorithm. Higher values indicate that the model can guess more accurately whether a statute is by a Republican or Democrat legislature, respectively, indicating higher partisan distinctiveness in legislating.

AUC scores for the A-fitted model are produced using the B sample, and vice versa. We then get two AUC scores for each party and for each decade, which we average to produce an AUC score for each party by decade.

The results for the supervised-ML method are shown in Figure 11. We see that partisan influence started off relatively high, and somewhat stable, with a one-decade drop in the 1910s. There was subsequently a drop in the 1960s-1980s, and then a subsequent recovery starting in the 1990s. Empirically, then, the supervised method uncovers patterns resembling those detected by our main approach. These broad patterns are robust to alternative specification choices, including using a different machine learning model and different definitions of partisan control.

8 Conclusion

In this paper, we introduced a new methodology for mapping a vast corpus of state and federal laws into coherent policy clusters. By examining how states and the federal government legislate within these clusters, we constructed measures of policy similarity

that capture both cross-state convergence and the evolving relationship between states and Washington, D.C.

Partisanship has become increasingly predictive of state-to-state policy similarity in recent decades, complementing earlier results in the literature and extending them further back in time. Contrary to common perceptions, the recent rise in polarization does not surpass historical peaks, indicating that the United States has experienced comparably polarized eras in the past.

We also documented growing federal influence on states. While geography played a substantial role in historical policy diffusion, over time it has shrunk compared to economic and partisan factors, with the federal government now emerging as a major reference point. This pattern is consistent with a nationalizing political environment, though future work is needed to clarify the precise mechanisms of federal–state policy interactions.

Future work could begin to investigate how states enact variations of federal laws, whether they expand or restrict federal provisions. Another interesting angle would be to evaluate the influence of interest groups, for example with model legislation (Burgess et al., 2016; Hertel-Fernandez, 2019). It would also be interesting to explore how innovations in communication technologies – railways, telegraphs, telephones, the internet, and AI – have and will shape the spread of policy ideas.

References

Angelucci, Charles, Julia Cagé and Michael Sinkinson. 2020. Media competition and news diets. Technical report National Bureau of Economic Research.

Arold, Benjamin W, Elliott Ash, W Bentley MacLeod and Suresh Naidu. 2024. "Do Words Matter?".

Ash, Elliott. 2018. "What drives partisan tax policy? The effective tax code." *Center for Law & Economics Working Paper Series* 4.

Ash, Elliott, Aniket Kesari, Suresh Naidu, Lena Song and Dominik Stammbach. 2024. "Translating Legalese.".

Ash, Elliott, Massimo Morelli and Matia Vannoni. 2021. "More laws, more growth? Evidence from us states.".

Ash, Elliott, Massimo Morelli and Moritz Osnabrügge. 2019. "Electoral Systems and Political Attention: Evidence from the Electoral Reform in New Zealand.".

Ash, Elliott, Massimo Morelli and Richard Van Weelden. 2017. "Elections and Divisiveness: Theory and Evidence." *The Journal of Politics* 79(4):1268–1285.

Ash, Elliott and Stephen Hansen. 2023. "Text algorithms in economics." *Annual Review of Economics* .

Brandeis, Louis. 1932. "New State Ice Co. v. Liebmann.". U.S. Supreme Court opinion. **URL:** *https://supreme.justia.com/cases/federal/us/285/262/*

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. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 57–66.

Carlino, Gerald, Thorsten Drautzburg, Robert Inman and Nicholas Zarra. 2023. "Partisanship and Fiscal Policy in Economic Unions: Evidence from US States." *American Economic Review* .

Caughey, Devin, Yiqing Xu and Christopher Warshaw. 2017. "Incremental democracy: The policy effects of partisan control of state government." *The Journal of Politics* 79(4):1342–1358.

Chen, Tianqi and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. pp. 785–794.

De Paula, Áureo, Imran Rasul and Pedro Souza. 2019. "Identifying network ties from panel data: Theory and an application to tax competition." *arXiv preprint arXiv:1910.07452* .

DellaVigna, Stefano and Woojin Kim. 2022. Policy Diffusion and Polarization across US States. Technical report National Bureau of Economic Research.

Dippel, Christian. 2022. "Political parties do matter in us cities... for their unfunded pensions." *American Economic Journal: Economic Policy* 14(3):33–54.

Elazar, Daniel J. 1966. *American Federalism: A View from the States*. New York: Thomas Y. Crowell Company.

Elazar, Daniel J. 1987. *Exploring Federalism*. Tuscaloosa, AL: University of Alabama Press.

Ferreira, Fernando and Joseph Gyourko. 2009. "Do political parties matter? Evidence from US cities." *The Quarterly journal of economics* 124(1):399–422.

Gennaro, Gloria and Elliott Ash. 2021. "Emotion and Reason in Political Language." *Economic Journal* 2021(02).

Gentzkow, Matthew, Bryan T Kelly and Matt Taddy. 2017. Text as Data. Technical Report 3 National Bureau of Economic Research.

Gentzkow, Matthew and Jesse M Shapiro. 2010. "What drives media slant? Evidence from US daily newspapers." *Econometrica* 78(1):35–71.

Gentzkow, Matthew, Jesse M Shapiro and Matt Taddy. 2017. "Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech.".

Grumbach, Jacob M. 2018. "From Backwaters to Major Policymakers: Policy Polarization in the States, 1970–2014." *Perspectives on Politics* 16(2):416–435.

Hertel-Fernandez, Alexander. 2019. *State capture: how conservative activists, big businesses, and wealthy donors reshaped the American states–and the nation*. Oxford University Press, USA.

Hopkins, Daniel J. 2018. *The increasingly United States: How and why American political behavior nationalized*. University of Chicago Press.

Hopkins, Daniel J, Eric Schickler and David L Azizi. 2022. "From Many Divides, One? The Polarization and Nationalization of American State Party Platforms, 1918–2017." *Studies in American Political Development* 36(1):1–20.

Klarner, Carl. 2015. "State Economic and Government Finance Data." *Harvard Dataverse, Version* 1.

Madison, James. 1788. *The Federalist Papers*. New York: Modern Library. Federalist No. 45.

Mallinson, Daniel J. 2021. "Who are your neighbors? The role of ideology and decline of geographic proximity in the diffusion of policy innovations." *Policy Studies Journal* 49(1):67–88.

Manson, Steven M. 2020. "IPUMS national historical geographic information system: version 15.0.".

Moskowitz, Daniel J. 2021. "Local news, information, and the nationalization of US Elections." *American Political Science Review* 115(1):114–129.

Mulligan, Casey B and Andrei Shleifer. 2005. "The Extent of the Market and the Supply of Regulation." *The Quarterly Journal of Economics* 120(4):1445–1473.

Shor, Boris and Nolan McCarty. 2011*a*. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3):530–551.

Shor, Boris and Nolan McCarty. 2011*b*. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3):530–551.

Song, Kaitao, Xu Tan, Tao Qin, Jianfeng Lu and Tie-Yan Liu. 2020. "Mpnet: Masked and permuted pre-training for language understanding." *Advances in neural information processing systems* 33:16857–16867.

Vannoni, Matia, Elliott Ash and Massimo Morelli. 2021. "Measuring discretion and delegation in legislative texts: methods and application to US states." *Political Analysis* 29(1):43–57.

Widmer, Philine, Sergio Galletta and Elliott Ash. 2020. "Media Slant is Contagious." *Center for Law & Economics Working Paper Series* 14.

A Appendix

A.1 Proof of Uniqueness

Let $\mathbf{p}_c = (p_{1ct}, p_{2ct}, \ldots, p_{Nct})$ be the vector of adoption probabilities for all N states in cluster c. Define the mapping $\mathbf{T}\colon [0,1]^N \to [0,1]^N$ as

$$\Big[\mathbf{T}(\mathbf{p}_c)\Big]_s = \frac{\exp\Big\{\alpha_{sct} + \delta_c \sum_{t' \in T_{-t}} n_{sct'} + \sum_{s' \neq s}\Big[\sum_{t' \in T_{-t}} \beta_{ss'ct'}\, n_{s'ct'} + \beta_{ss'ct}\, p_{s'ct}\Big]\Big\}}{1 + \exp\Big\{\alpha_{sct} + \delta_c \sum_{t' \in T_{-t}} n_{sct'} + \sum_{s' \neq s}\Big[\sum_{t' \in T_{-t}} \beta_{ss'ct'}\, n_{s'ct'} + \beta_{ss'ct}\, p_{s'ct}\Big]\Big\}}. \tag{10}$$

A fixed point of this mapping, $\mathbf{p}_c^*$, satisfies $\mathbf{T}(\mathbf{p}_c^*) = \mathbf{p}_c^*$.

Step 1 (Existence): Since $[0,1]^N$ is a nonempty, compact, and convex subset of $\mathbb{R}^N$ and the mapping $\mathbf{T}$ is continuous, Brouwer's Fixed Point Theorem guarantees that there exists at least one fixed point $\mathbf{p}_c^*$ in $[0,1]^N$.

Step 2 (Uniqueness): We now show that $\mathbf{T}$ is a contraction mapping under suitable conditions. Since the derivative of the logistic function $f(x) = \dfrac{\exp(x)}{1+\exp(x)}$ attains a maximum value of $\dfrac{1}{4}$ (when $x = 0$). Thus, f is Lipschitz continuous with a Lipschitz constant no larger than $\dfrac{1}{4}$.

The only dependence of V_{sct} on $\mathbf{p}_c$ appears through the linear peer effect term $\displaystyle\sum_{s' \neq s} \beta_{ss'ct}\, p_{s'ct}$. Let $\mathbf{p}_c, \mathbf{q}_c \in [0,1]^N$. Then, by the mean value theorem applied to the logistic function, for each state s we have

$$\Big|\Big[\mathbf{T}(\mathbf{p}_c)\Big]_s - \Big[\mathbf{T}(\mathbf{q}_c)\Big]_s\Big| \leq \frac{1}{4} \sum_{s' \neq s} |\beta_{ss'ct}|\; |p_{s'ct} - q_{s'ct}|\,. \tag{11}$$

Taking the supremum (i.e., the infinity norm) over s, we obtain

$$\|\mathbf{T}(\mathbf{p}_c) - \mathbf{T}(\mathbf{q}_c)\|_\infty \leq \frac{1}{4}\Big(\max_s \sum_{s' \neq s} |\beta_{ss'ct}|\Big) \|\mathbf{p}_c - \mathbf{q}_c\|_\infty. \tag{12}$$

Define

$$L = \frac{1}{4} \max_s \sum_{s' \neq s} |\beta_{ss'ct}|\,. \tag{13}$$

If we assume that $L < 1$, then $\mathbf{T}$ is a contraction mapping. By the Banach Fixed Point Theorem, $\mathbf{T}$ then has a unique fixed point in $[0,1]^N$, i.e., there exists a unique vector $\mathbf{p}_c^*$ such that $\mathbf{T}(\mathbf{p}_c^*) = \mathbf{p}_c^*$.

□

Figure A.1: Subtopic Assignment of Policy Summaries with GPT

User Prompt

Input Example:
Here is a summary of a state statute: "*Turnpike Road Company Organization. This policy authorizes the organization and division of a specific turnpike road company into sections, and the creation of another specific turnpike road company to complete construction.*" Here is the main topic of the policy summary: [primary topic]

Prompt

Your task is to categorize this policy summary among a list of possible policy subtopics.
Here is the list of different policy subtopics along with their definition: [subtopics of the primary topic]
Respond with the primary policy subtopic in the format: Primary: [policy subtopic name].
If applicable, you may include secondary subtopics by adding: Secondary: [policy subtopic names].
Prioritize selecting only one or the fewest possible subtopics, focusing on relevance and listing them in order of importance.

Answer

Output Example:
"*Highways*"

A.2 Additional Tables and Figures not shown in the main paper

A.3 Statute Summaries

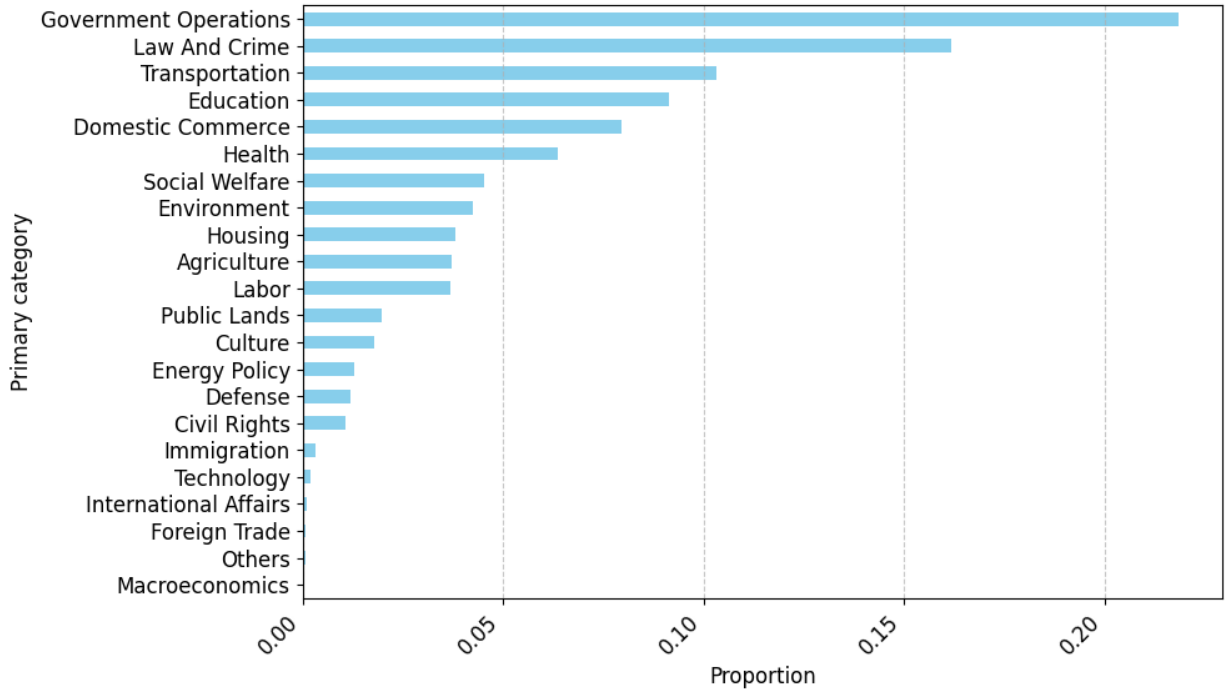
A.4 Summary Validation

A.4.1 Statutes reflects unique policies

Our analysis relies on the idea that most statutes reflect a single policy. This allows us to comprehensively summarize statutes and serves as the foundation for our clustering approach, where we group similar unique policies together. To validate this assumption, we start with the raw statutes we summarized and draw a 1% random sample (approximately 25,000 raw statutes) to test its validity. Then, we use GPT-4o-mini to identify all distinct policies mentioned in each statute. Figure A.4 presents the prompt used, along with an example.

Figures A.5 and A.6 respectively illustrate the average number of policies per statute, as identified by GPT, and the share of statutes where GPT detected more than one policy, by decade.

Figure A.2: Proportion of Each Primary Topic Category in Statute Summaries



Notes: Topics are based on the Comparative Agenda Project (see Appendix B.1).

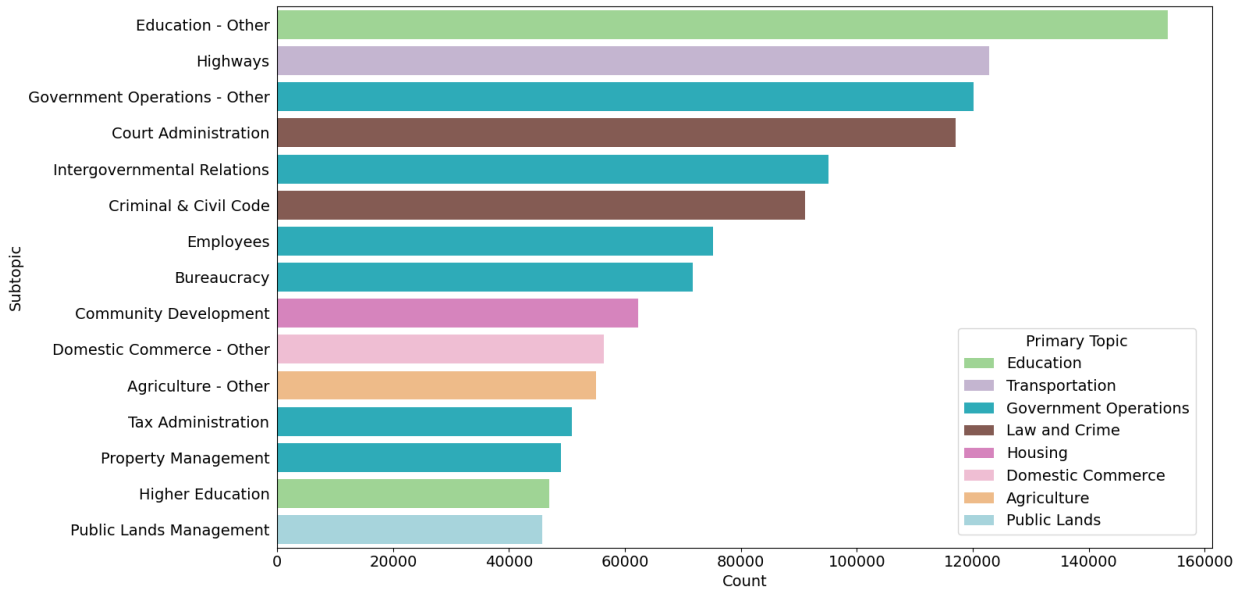
A.4.2 Fitting Raw Statutes into Clusters and Topics

Besides examining their plurality, the comprehensive list of policies extracted from raw statutes by GPT can be used to assess whether they align with the summaries in terms of policy content.

We follow the methodology outlined in Appendices **??** and 5 for statute summaries to anonymize and assign CAP topics to the listed policies. Then, we fit each listed policy into the clustering model built with the summaries. The underlying assumption is as follows: if at least one of the distinct policies identified by GPT for a given raw statute is assigned to the same cluster (i.e., policy group) as the summary of that statute, then the summary plausibly reflects the content of the raw statute. Furthermore, the number of distinct policies from the raw statute that are assigned to the same cluster as its summary indicates the comprehensiveness of the summary in capturing the full scope of the statute. We also compare the CAP topic of the listed policies to that of the statute summary for the same raw statutes.

Table A.2 presents descriptive statistics on this clustering assignment, including the share of summaries that have at least one GPT-listed policy in their assigned cluster,

Figure A.3: Top 15 Topic-Subtopic Category by Count in Statute Summaries



Notes: Topics and subtopics are based on the Comparative Agenda Project (see Appendix B.1).

as well as the share of statutes whose summaries are assigned to the same cluster as at least one of their distinct policies, and a comparison of CAP topics.

A.5 Federal Statutes

We provide a more comprehensive validation of the analysis conducted in the previous section, focusing specifically on federal statutes. These statutes, passed by the federal government, may differ from state statutes in terms of their length, the breadth of policies they cover, and the topics they address.

Figure A.7 presents the distribution of primary topics for federal statutes. Additionally, we analyze the full list of distinct policies from all raw statutes, as in Appendix A.4.1, covering nearly 90,000 statutes. Figures A.8 and A.9 summarize the average number of policies per statute, as identified by GPT, and the share of statutes where GPT detected more than one policy, by decade, for the entire corpus of raw federal statutes. Table A.3 presents the same descriptive statistics shown in Appendix A.4.2 for all federal statutes.

Figure A.4: Identification of Distinct Policies in Raw Statutes Using GPT

Prompt

Here is a statute from a U.S. state or federal government that enacts one or more policies. The statute text is derived from a digital scan and may contain OCR errors or lack punctuation.
Carefully analyze the text to identify and summarize the enacted policy or policies:

1. **If the statute enacts only one policy**, summarize it as:
   - Start with: "This policy..."
   - Combine all sub-policies under one overarching description.
2. **If the statute enacts multiple distinct policies**, list them as:
   - "1. This policy...; 2. This policy..."
   - Ensure each is distinct, non-overlapping, and less than 20 words.

Avoid treating minor sub-policies as distinct if they logically fall under a broader policy.

User Prompt

Input Example:
Here is the statute:

> "*An Act relating to all counties having a population of not less than 56,500 nor more than 59,000 inhabitants, according to the 1970 or any subsequent federal decennial census, to provide an additional expense allowance for members of the county commission.*
>
> Be it enacted by the legislature of Alabama:
>
> **Section 1:** The provisions of this act shall apply to all counties having a population of not less than 56,500 nor more than 59,000 inhabitants, according to the 1970 or any subsequent federal decennial census.
>
> **Section 2:** In addition to any and all other salary, compensation, and expense allowances provided for by law, the members of the county commission in counties to which this act applies shall be paid an additional expense allowance in the amount of $150 per month from the county general fund.
>
> **Section 3:** The provisions of this act shall become effective on the first day of the month next following the date this act becomes law. This act became a law under Section 125 of the Constitution on May 17, 1977, without approval by the governor.
>
> Act No. 622, H. 968, Crawford Sasser."

Answer

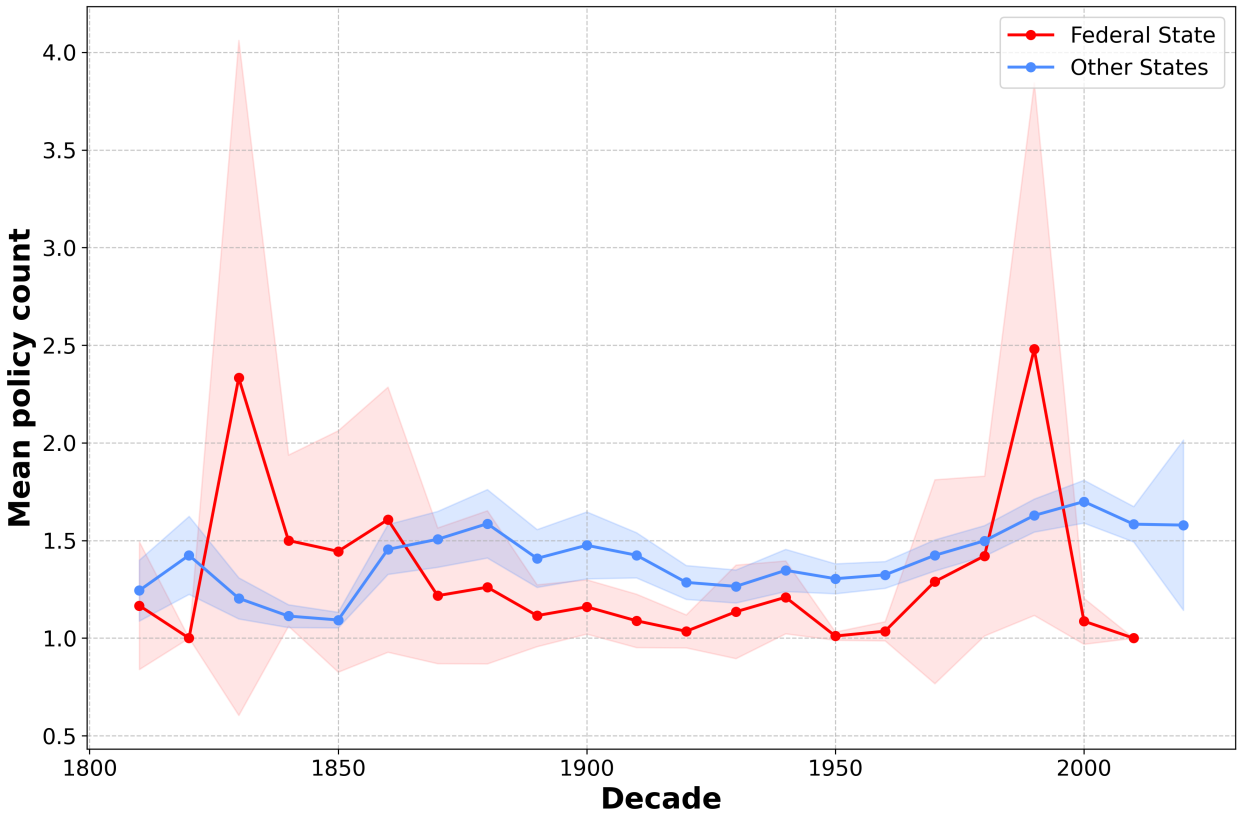
Output Example:
" *1. This policy provides an additional monthly expense allowance of $150 for county commission members in specific Alabama counties.* "

Remark

For comparison, here is the statute summary computed with GPT-3.5-turbo:
"*County Commission Expense Allowance. This policy provides an additional expense allowance for members of the county commission in specific population range counties, effective the month following enactment.*"

Notes: The raw statute was corrected from OCR errors using ChatGPT.

Figure A.5: Average Number of Policies per Raw Statute



Notes: Based on a 1% random sample of statutes. The figure distinguishes federal statutes (in red) from other state statutes (in blue).

A.6 Policy Cluster Validation

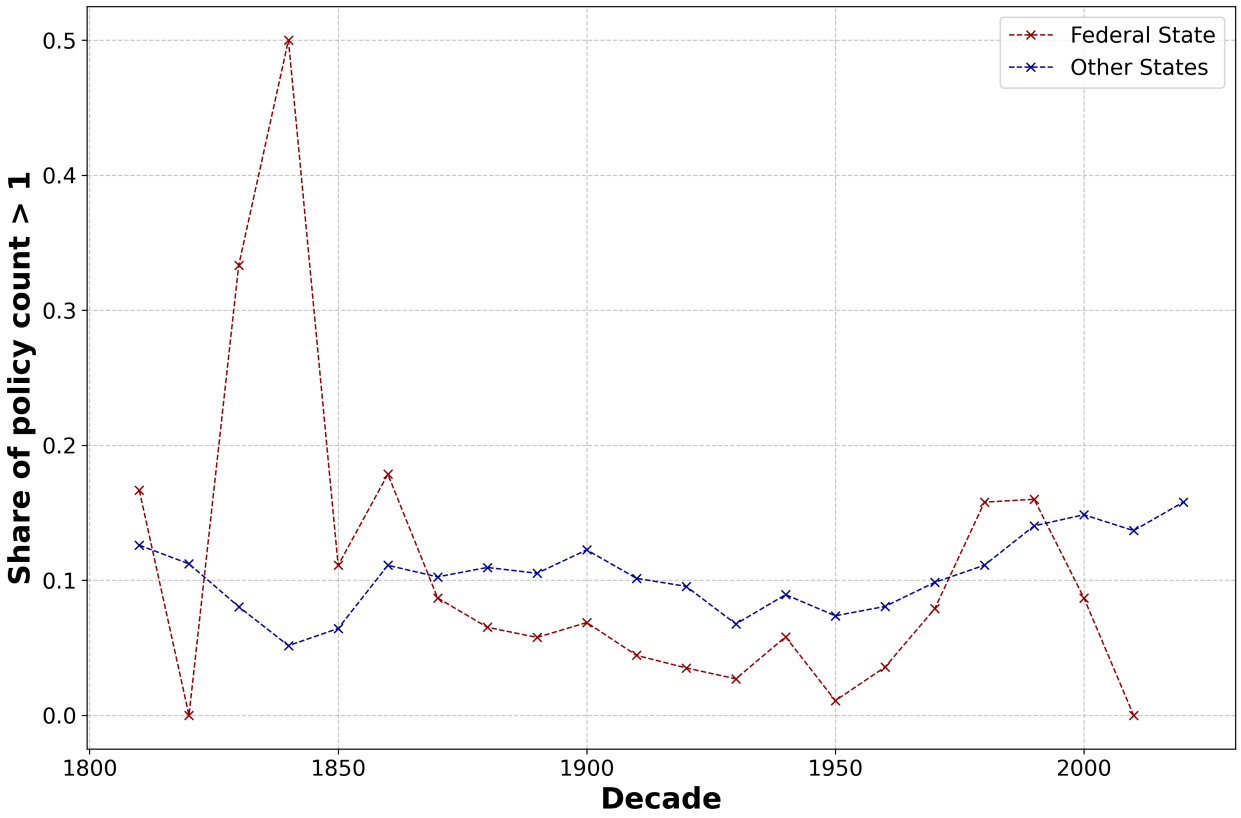
This section validates the approach outlined in Sections **??** and 5.1, focusing on the vectorization and construction of policy clusters. Specifically, we aim to demonstrate that our chosen model is the most robust among existing classical encoders and that the number of policy clusters effectively groups statutes with similar policy goals.

A.6.1 Model Validation: MPNet vs. FastText

To ensure that our clustering approach is optimal, we compare MPNet embeddings, with an alternative method, FastText embeddings. Developed by Facebook AI, FastText represents words as vectors in a continuous vector space, typically using 300-dimensional embeddings.

To evaluate their relative performance, we replicate the clustering technique described in Section **??** using both MPNet and FastText embeddings, testing across

Figure A.6: Proportion of Raw Statutes with More Than One Identified Policy

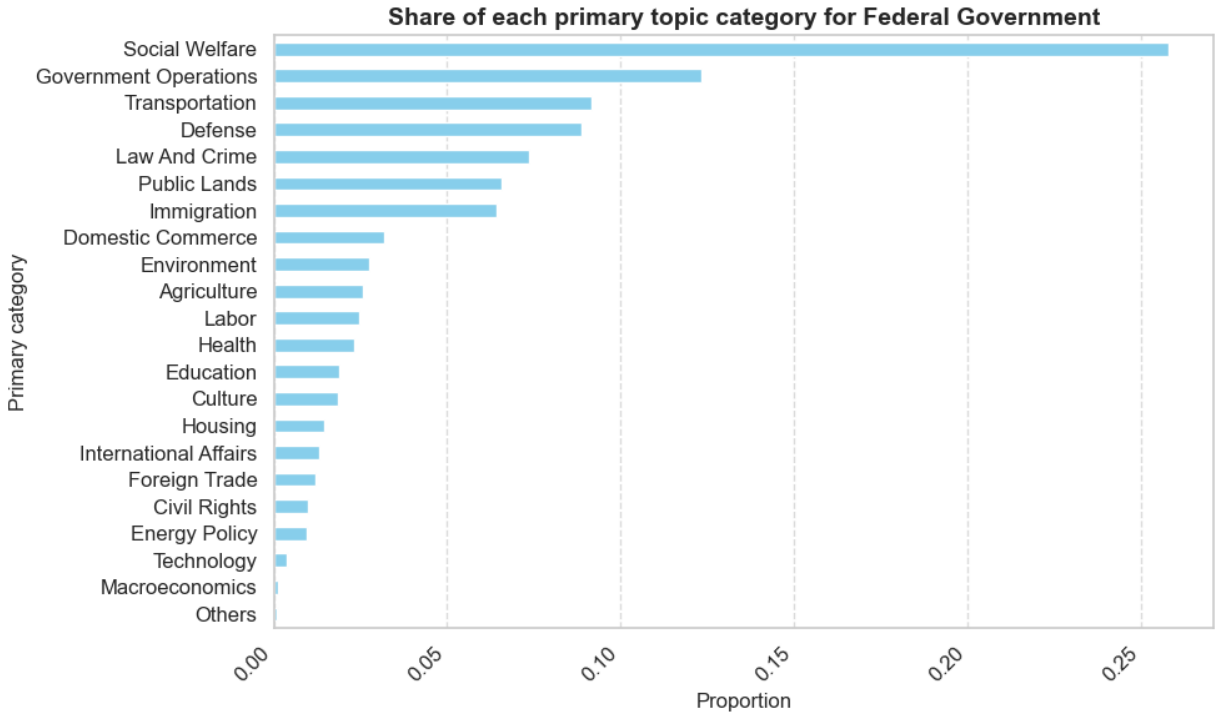


Notes: Based on a 1% random sample of statutes. The figure distinguishes federal statutes (in red) from other state statutes (in blue).

different cluster sizes: 1024, 1536, 2048, and 2560. We then assess clustering quality using the following methodology: A random sample of 4,000 statute summaries is selected. For each cluster size, each statute is compared with a randomly selected statute from the same cluster, as defined by clusters constructed using MPNet embeddings, and another randomly selected statute from the same cluster, as defined by clusters constructed using FastText embeddings.

Figure A.10 presents the prompt used for GPT-4o-mini, along with an example. The results of this evaluation, summarized in Table A.4, indicate that MPNet consistently outperforms FastText, supporting our choice of MPNet embeddings for policy clustering.

Figure A.7: Proportion of Each Primary Topic Category in Federal Statute Summaries



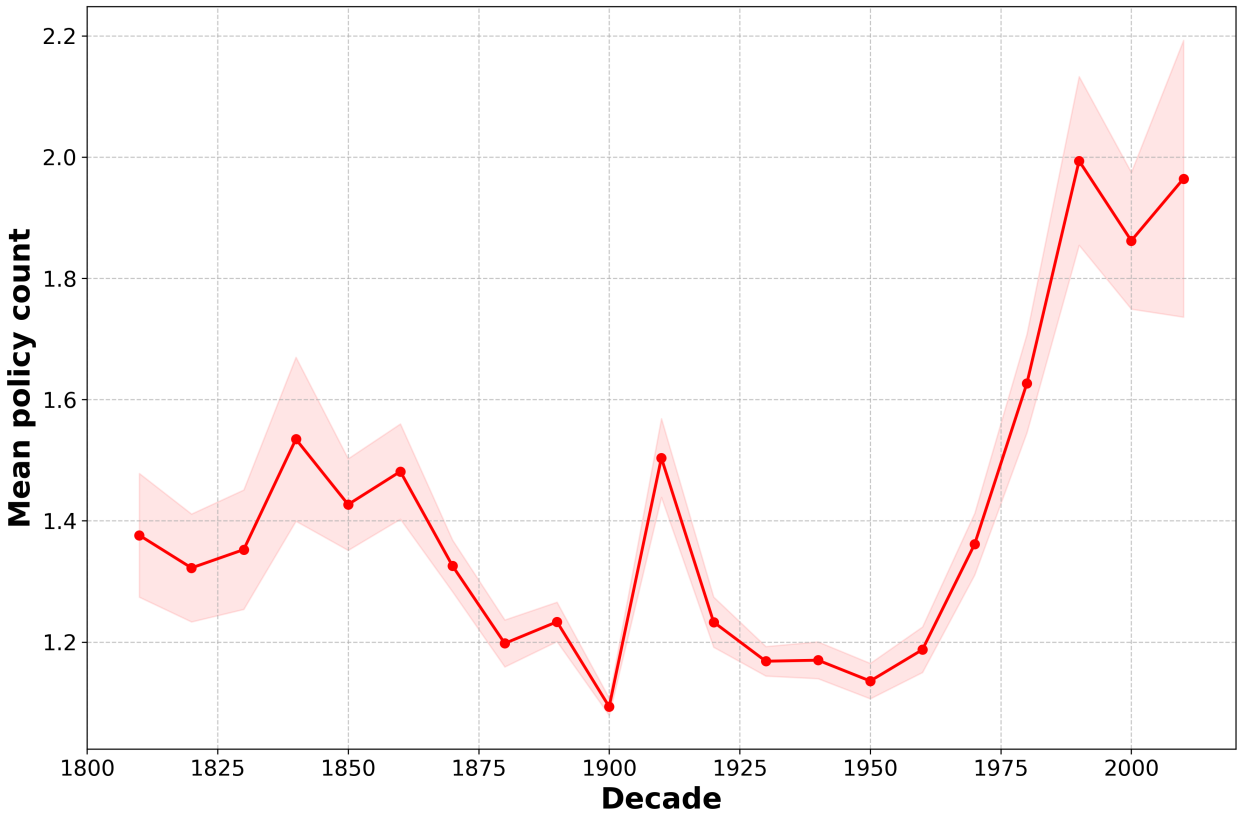
Notes: Topics are based on the Comparative Agenda Project (see Appendix B.1).

A.6.2 Cluster Count Validation: Silhouette Score

Section 5.1 describes our K-means clustering methodology, which requires the pre-specification of the number of clusters, denoted as k. Determining the best k is a complex problem, as no universally accepted method exists. Still, there are various metrics that can help in making this choice. It is important to note that there is no "optimal" value of k in clustering, as the best k depends on the specific context: it is contingent upon the data's distribution, inherent structure, and the specific objectives of the analysis. For example, in our case, we are interested in policy-level clustering rather than topic-level clustering.

To assess the validity of our chosen cluster size, we employ the silhouette score, a robust metric for evaluating clustering quality. The silhouette score measures the quality of clustering by assessing both the cohesion and separation of the clusters. A higher silhouette score indicates that the data points are well matched to their own cluster and poorly matched to neighboring clusters, suggesting well-defined and distinct clusters. For each data point i, the silhouette coefficient $s(i)$ ranges from -1 to 1 and

Figure A.8: Average Number of Policies per Raw Federal Statute



Notes: Based on Federal Statutes (nearly 90,000).

is computed as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

- $a(i)$ represents the average distance between i and all other points within the same cluster (cohesion).
- $b(i)$ denotes the minimum average distance from i to points in a different cluster, minimized over clusters (separation).

The final silhouette score averages over all sampled points. Figure A.11 presents the silhouette score of our clustering for k ranging from 10 to 3000, with increments of 10.[3]

[3]For computational efficiency, the silhouette scores the silhouette scores were calculated using a random 5% sample of the dataset.

Figure A.9: Proportion of Raw Federal Statutes with More Than One Identified Policy



Notes: Based on Federal Statutes (nearly 90,000).

A.6.3 Cluster Count Validation: Descriptive Statistics

Table 2 and A.7 present descriptive statistics for clusters of sizes 2048 and 2560, respectively. Additionally, Tables A.5 and A.6 display the proportion of clusters for which a given CAP topic is the most frequently assigned topic.

To highlight the importance of anonymization in our approach, as a comparative baseline, Table A.8 provides descriptive statistics for a non-anonymized 2048-cluster configuration.

A.6.4 Cluster Validation: Coherent Policy Goals

To validate the cluster structures and assess their ability to capture coherent policy goals, we conduct a series of robustness tests using a GPT-approach.

The first test evaluates whether policies in summaries within the same cluster are more similar to each other compared to policies from different clusters, conditional on

Figure A.10: Comparison of MPNet and FastText Policy Clusters

User Prompt

Input Example: Here is an initial statement:
"*School Bus Stoparms Requirement. This policy mandates the use of school bus stoparms on old and new buses, with state reimbursement for their purchase, and outlines approved transportation expenses for school districts.*"
You are provided with two additional statements, each identified by an ID (formatted as 'ID: text'):
[randomly ordered]
"*MN_1973_0090_347: School Bus Traffic Regulations. This policy amends highway traffic regulations related to school buses, including the use of equipment, color, and conduct of other vehicles, such as stopping and loading/unloading procedures.*" [MPNET]
"*SC_1946_0060_10: School District Financing and Infrastructure Development. This policy authorizes school districts to borrow money, issue bonds, and levy taxes for school purposes, as well as for the construction and improvement of school buildings and facilities.*" [FastText]

Prompt

Determine which statement discusses a policy most similar to the initial statement. Respond only with the ID of the most similar statement.
If both are equally similar, respond with 'Both'.

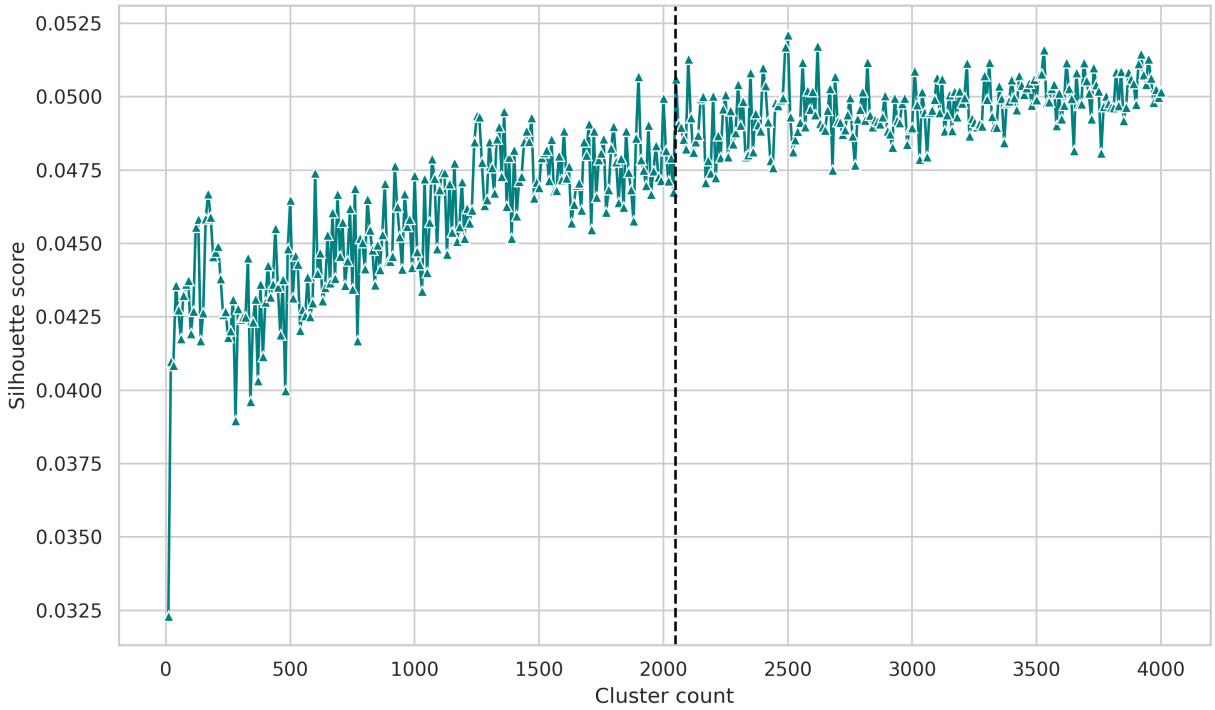
Answer

Output Example:
" *MN_1973_0090_347* " [MPNET]

the same CAP topic. For 4,000 random statute summaries, we randomly select one policy from the same cluster and another policy from a different cluster, within the same CAP topic. We assess which policy is more similar to the initial policy using GPT-4o-mini, based on the prompt presented in Figure A.13, along with an example. As shown in Table A.9, statutes from the same cluster are generally preferred, demonstrating that clustering effectively captures policy similarity. As the cluster size increases, the model shows diminishing returns in similarity improvements.

The second test evaluates whether the policies in summaries within the same cluster pursue aligned or opposing policy goals, and compares them to policies from different clusters, conditional on the same CAP topic. For 10,000 random statute summaries, we randomly select one policy from the same cluster and another policy from a different cluster, within the same CAP topic. We assess for each pair whether the policies pursue aligned or opposite goals based on their political ideologies and contexts rather than the policy topic itself, using GPT-4o-mini, with the prompt presented in Figure A.14. As shown in Figure A.15, statutes from the same cluster often pursue similar policy objectives, comparison between within-cluster and cross-cluster pairs shows that the alignment of policy goals is stronger within the same cluster, demonstrating that clustering effectively captures policy goals.

Figure A.11: Silhouette Score for Different Cluster Sizes

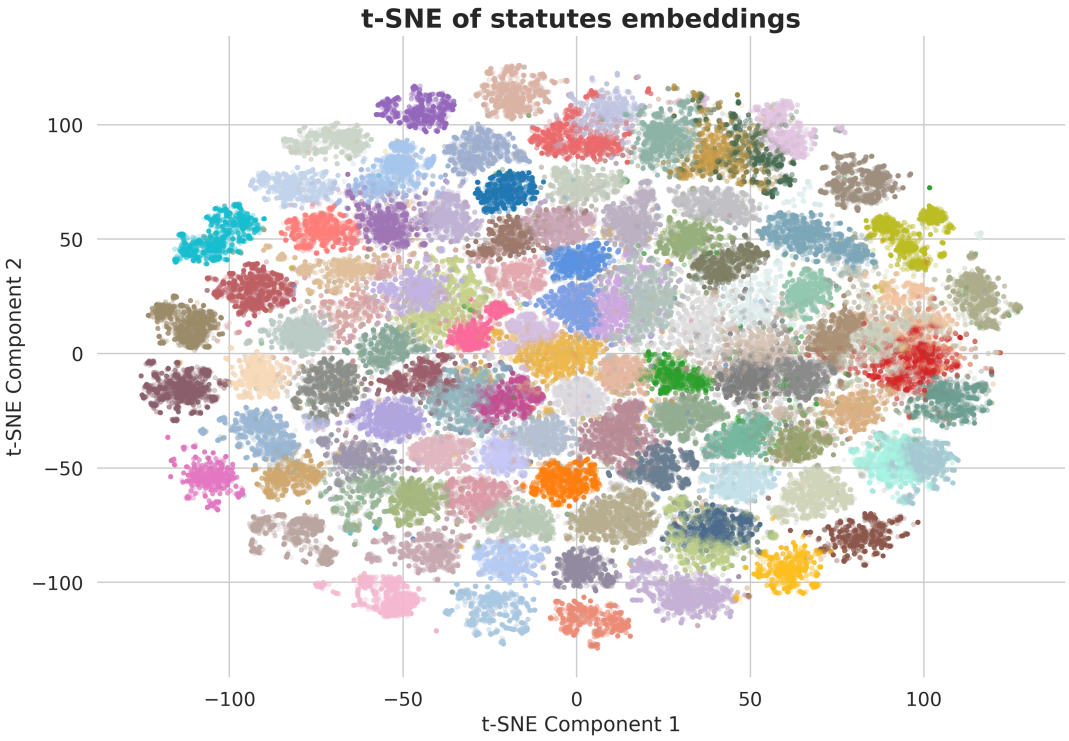


Notes: The silhouette scores were computed on a random sample comprising 5% of the data points (statute summary embeddings), with k ranging from 10 to 3000 in increments of 10.

B Supplementary

B.1 Comparative Agenda Project

Figure A.12: t-SNE Visualization of the Top 100 Clusters with the Most Statutes (2048-Cluster Configuration)



Notes: Computed for the top 100 clusters with the most statutes from the 2048-cluster configuration.

Main Topic	Subtopics
Macroeconomics	• **Interest Rates:** includes issues related to inflation, cost of living, prices, and interest rates; • **Unemployment Rate:** includes issues related to the unemployment rate and the impact of unemployment; • **Monetary Policy:** includes issues related to the monetary policy, central bank, and the treasury; • **National Budget:** includes issues related to public debt, budgeting, and efforts to reduce deficits; • **Tax Code:** includes issues related to tax policy, the impact of taxes, and tax enforcement; • **Industrial Policy:** includes issues related to manufacturing policy, industrial revitalization and growth; • **Price Control:** includes issues related to wage or price control and emergency price controls; • **Other:** includes issues related to other macroeconomics subtopics.

Main Topic	Subtopics
Civil Rights	• **Minority Discrimination:** includes issues related to minority, ethnic, and racial group discrimination; • **Gender Discrimination:** includes issues related to sex, gender, and sexual orientation discrimination; • **Age Discrimination:** includes issues related to age discrimination, including mandatory retirement age policies; • **Handicap Discrimination:** includes issues related to handicap and disease discrimination; • **Voting Rights:** includes issues related to voting rights, expanding or contracting the franchise, participation, and related issues; • **Freedom of Speech:** includes issues related to freedom of speech, religious freedoms, and other types of freedom of expression; • **Right to Privacy:** includes issues related to privacy rights, including privacy of records, access to government information, and abortion rights; • **Anti-Government:** includes issues related to anti-government activity groups, such as the communist party and local insurgency groups; • **Other:** includes issues related to other civil rights subtopics.
Health	• **Health Care Reform:** includes issues related to broad, comprehensive changes in the health care system; • **Insurance:** includes issues related to health insurance reform, regulation, availability, and cost; • **Drug Industry:** includes issues related to the regulation and promotion of pharmaceuticals, medical devices, and clinical labs; • **Medical Facilities:** includes issues related to facilities construction, regulation, and payments, including waitlists and ambulance services; • **Insurance Providers:** includes issues related to provider and insurer payments and regulation, including other types of benefits or multiple benefits; • **Medical Liability:** includes issues related to medical liability, malpractice issues, medical fraud and abuse, and unfair practices; • **Manpower:** includes issues related to the supply and quantity of labor in the health care industry, training, and licensing; • **Disease Prevention:** includes issues related to disease prevention, treatment, and health promotion, including specific diseases not covered in other subtopics; • **Infants and Children:** includes issues related to infants and children, including coverage and quality of care, health promotion, and school health programs; • **Mental Health:** includes issues related to mental health care and mental health disease; • **Long-term Care:** includes issues related to long-term care, home health care, the terminally ill, and rehabilitation services; • **Drug Coverage and Cost:** includes issues related to prescription drug coverage, programs to pay for prescription drugs, and policy to reduce the cost of prescription drugs; • **Tobacco Abuse:** includes issues related to tobacco abuse, treatment, education, and health effects; • **Drug and Alcohol Abuse:** includes issues related to alcohol and illegal drug abuse, treatment, education, and health effects; • **R&D:** includes issues related to health care research and development; • **Other:** includes issues related to other health care topics.

Main Topic	Subtopics
Agriculture	• **Trade:** includes issues related to the regulation and impact of agricultural foreign trade; • **Subsidies to Farmers:** includes issues related to government subsidies to farmers and ranchers, including agricultural disaster insurance; • **Food Inspection & Safety:** includes issues related to food inspection and safety, including seafood and labeling requirements; • **Food Marketing & Promotion:** includes issues related to the regulation of agricultural marketing and providing information to consumers regarding a healthy diet; • **Animal and Crop Disease:** includes issues related to animal and crop disease, pest control, pesticide regulation, and welfare for domesticated animals; • **Fisheries & Fishing:** includes issues related to fishing, commercial fishery regulation, and conservation; • **R&D:** includes issues related to agricultural research and development; • **Other:** includes issues related to other agricultural subtopics.
Labor	• **Worker Safety:** includes issues related to worker safety and protection and compensation for work-related injury and disease; • **Employment Training:** includes issues related to job training for adult workers, workforce development, and efforts to retrain displaced workers; • **Employee Benefits:** includes issues related to all employee benefits, pensions, and retirement accounts, including government-provided unemployment insurance; • **Labor Unions:** includes issues related to labor unions, collective bargaining, and employer-employee relations; • **Fair Labor Standards:** includes issues related to fair labor standards such as the minimum wage and overtime compensation, and labor law; • **Youth Employment:** includes issues related to youth employment, child labor, and job training for youths; • **Migrant and Seasonal:** includes issues related to migrant, guest, and seasonal workers; • **Other:** includes issues related to other labor policy.

Main Topic	Subtopics
Education	• **Higher:** includes issues related to higher education, student loans and education finance, and the regulation of colleges and universities; • **Elementary & Secondary:** includes issues related to elementary and primary schools, school reform, safety in schools, and efforts to generally improve educational standards and outcomes; • **Underprivileged:** includes issues related to education of underprivileged students, including adult literacy programs, bilingual education needs, and rural education initiatives; • **Vocational:** includes issues related to vocational education for children and adults and their impact; • **Special:** includes issues related to special education and education for the physically or mentally handicapped; • **Excellence:** includes issues related to education excellence, including efforts to increase the quality of specific areas, such as math, science, or foreign language skills; • **R&D:** includes issues related to research and development in education; • **Other:** includes issues related to other subtopics in education policy.
Environment	• **Drinking Water:** includes issues related to domestic drinking water safety, supply, pollution, and additives; • **Waste Disposal:** includes issues related to the disposal and treatment of wastewater, solid waste, and runoff; • **Hazardous Waste:** includes issues related to hazardous waste and toxic chemical regulation, treatment, and disposal; • **Air Pollution:** includes issues related to air pollution, climate change, and noise pollution; • **Recycling:** includes issues related to recycling, reuse, and resource conservation; • **Indoor Hazards:** includes issues related to indoor environmental hazards, indoor air contamination, and indoor hazardous substances; • **Species & Forest:** includes issues related to species and forest protection, endangered species, and regulation of laboratory or performance animals; • **Land and Water Conservation:** includes issues related to land and water resource conservation; • **R&D:** includes issues related to research and development in environmental technology; • **Other:** includes issues related to other environmental subtopics.

Main Topic	Subtopics
Energy Policy	• **Nuclear:** includes issues related to nuclear energy, safety and security, and disposal of nuclear waste; • **Electricity:** includes issues related to general electricity, hydropower, and regulation of electrical utilities; • **Natural Gas & Oil:** includes issues related to natural gas and oil, drilling, oil spills and flaring, oil and gas prices, shortages, and gasoline regulation; • **Coal:** includes issues related to coal production, use, trade, and regulation, including coal gasification and clean coal technologies; • **Alternative & Renewable:** includes issues related to alternative and renewable energy, biofuels, hydrogen, and geothermal power; • **Conservation:** includes issues related to energy conservation and energy efficiency, including vehicles, homes, commercial use, and government; • **R&D:** includes issues related to energy research and development; • **Other:** includes issues related to other energy subtopics.
Immigration	• **Immigration:** includes issues related to immigration, refugees, and citizenship.
Transportation	• **Mass:** includes issues related to mass transportation construction, regulation, safety, and availability; • **Highways:** includes issues related to public highway construction, maintenance, and safety; • **Air Travel:** includes issues related to air travel, regulation, and safety of aviation, airports, air traffic control, pilot training, and aviation technology; • **Railroad Travel:** includes issues related to railroads, rail travel, rail freight, and the development and deployment of new rail technologies; • **Maritime:** includes issues related to maritime transportation, including maritime freight and shipping, safety and security, and inland waterways and channels; • **Infrastructure:** includes issues related to infrastructure and public works, including employment initiatives; • **R&D:** includes issues related to transportation research and development; • **Other:** includes issues related to other transportation subtopics.

Main Topic	Subtopics
Law and Crime	• **Agencies:** includes issues related to all law enforcement agencies, including border customs and other specialized enforcement agencies and their appropriations; • **White Collar Crime:** includes issues related to white-collar crime, organized crime, counterfeiting and fraud, cyber-crime, and money laundering; • **Illegal Drugs:** includes issues related to illegal drug crime and enforcement, criminal penalties for drug crimes, including international efforts to combat drug trafficking; • **Court Administration:** includes issues related to court administration, judiciary appropriations, guidelines for bail, pre-release, fines, and legal representation; • **Prisons:** includes issues related to prisons and jails, parole systems, and appropriations; • **Juvenile Crime:** includes issues related to juvenile crime and justice, juvenile prisons and jails, and efforts to reduce juvenile crime and recidivism; • **Child Abuse:** includes issues related to child abuse, child pornography, sexual exploitation of children, and parental kidnapping; • **Family Issues:** includes issues related to family issues, domestic violence, child welfare, and family law; • **Criminal & Civil Code:** includes issues related to domestic criminal and civil codes, including crimes not mentioned in other subtopics; • **Crime Control:** includes issues related to the control, prevention, and impact of crime; • **Police:** includes issues related to police and other general domestic security responses to terrorism, such as special police; • **Other:** includes issues related to other law, crime, and family subtopics.
Social Welfare	• **Low-Income Assistance:** includes issues related to poverty assistance for low-income families, including food assistance programs, programs to assess or alleviate welfare dependency, and tax credits directed at low-income families; • **Elderly Assistance:** includes issues related to elderly issues and elderly assistance, including government pensions; • **Disabled Assistance:** includes issues related to aid for people with physical or mental disabilities; • **Volunteer Associations:** includes issues related to domestic volunteer associations, charities, and youth organizations; • **Child Care:** includes issues related to parental leave and child care; • **Other:** includes issues related to other social welfare policy subtopics.

Main Topic	Subtopics
Housing	• **Community Development:** includes issues related to housing and community development, neighborhood development, and national housing policy; • **Urban Development:** includes issues related to urban development and general urban issues; • **Rural Housing:** includes issues related to rural housing; • **Rural Development:** includes non-housing issues focused on economic, infrastructure, and other developments in non-urban areas; • **Low-Income Assistance:** includes issues related to housing for low-income individuals and families, including public housing projects and housing affordability programs; • **Veterans:** includes issues related to housing for military veterans and their families, including subsidies for veterans; • **Elderly:** includes issues related to housing for the elderly, including housing facilities for the handicapped elderly; • **Homeless:** includes issues related to housing for the homeless and efforts to reduce homelessness; • **R&D:** includes issues related to housing and community development research and development; • **Other:** includes issues related to other housing and community development subtopics.
Domestic Commerce	• **Banking:** includes issues related to the regulation of national banking systems and other non-bank financial institutions; • **Securities & Commodities:** includes issues related to the regulation and facilitation of securities and commodities trading, regulation of investments and related industries, and exchanges; • **Consumer Finance:** includes issues related to consumer finance, mortgages, credit cards, access to credit records, and consumer credit fraud; • **Insurance Regulation:** includes issues related to insurance regulation, fraud and abuse in the insurance industry, the financial health of the insurance industry, and insurance availability and cost; • **Bankruptcy:** includes issues related to personal, commercial, and municipal bankruptcies; • **Corporate Management:** includes issues related to corporate mergers, antitrust regulation, corporate accounting and governance, and corporate management; • **Small Businesses:** includes issues related to small businesses, including programs to promote and subsidize small businesses; • **Copyrights and Patents:** includes issues related to copyrights and patents, patent reform, and intellectual property; • **Disaster Relief:** includes issues related to domestic natural disaster relief, disaster or flood insurance, and natural disaster preparedness; • **Tourism:** includes issues related to tourism regulation, promotion, and impact; • **Consumer Safety:** includes issues related to consumer fraud and safety in domestic commerce; • **Sports Regulation:** includes issues related to the regulation and promotion of sports, gambling, and personal fitness; • **R&D:** includes issues related to domestic commerce research and development; • **Other:** includes issues related to other domestic commerce policy subtopics.

Main Topic	Subtopics
Defense	• **Alliances:** includes issues related to defense alliances and agreements, security assistance, and UN peacekeeping activities; • **Intelligence:** includes issues related to military intelligence, espionage, and covert operations; • **Readiness:** includes issues related to military readiness, coordination of armed services, air support and sealift capabilities, and national stockpiles of strategic materials; • **Nuclear Arms:** includes issues related to nuclear weapons, nuclear proliferation, and modernization of nuclear equipment; • **Military Aid:** includes issues related to military aid to other countries and the control of arms sales to other countries; • **Personnel Issues:** includes issues related to military manpower, military personnel and their dependents, military courts, and general veterans issues; • **Procurement:** includes issues related to military procurement, conversion of old equipment, and weapons systems evaluation; • **Installations & Land:** includes issues related to military installations, construction, and land transfers; • **Reserve Forces:** includes issues related to military reserves and reserve affairs; • **Hazardous Waste:** includes issues related to military nuclear and hazardous waste disposal and military environmental compliance; • **Civil:** includes issues related to domestic civil defense, national security responses to terrorism, and other issues related to homeland security; • **Civilian Personnel:** includes issues related to non-contractor civilian personnel, civilian employment in the defense industry, and military base closings; • **Contractors:** includes issues related to military contractors and contracting, oversight of military contractors, and fraud by military contractors; • **Foreign Operations:** includes issues related to direct war-related foreign military operations, prisoners of war, and collateral damage to civilian populations; • **Claims against Military:** includes issues related to claims against the military, settlements for military dependents, and compensation for civilians injured in military operations; • **R&D:** includes issues related to defense research and development; • **Other:** includes issues related to other defense policy subtopics.
Technology	• **Space:** includes issues related to government use of space, space resource exploitation agreements, and space exploration; • **Commercial Use of Space:** includes issues related to the regulation and promotion of commercial space use, commercial satellite technology, and related government efforts; • **Science Transfer:** includes issues related to science and technology transfer and international science cooperation; • **Telecommunications:** includes issues related to telecommunications infrastructure, internet access, and regulation; • **Other:** includes additional technology subtopics.

Main Topic	Subtopics
Foreign Trade	• **Trade Agreements:** includes issues related to trade negotiations, disputes, and agreements, including tax treaties; • **Exports:** includes issues related to export regulation, subsidies, promotion, and control; • **Private Investments:** includes issues related to international private business investment and corporate development; • **Competitiveness:** includes issues related to productivity and competitiveness of domestic businesses and balance of payments issues; • **Tariff & Imports:** includes issues related to tariffs and other barriers to imports, import regulation, and impact of imports on domestic industries; • **Exchange Rates:** includes issues related to exchange rates and related issues; • **Other:** includes issues related to other foreign trade policy subtopics.
International Affairs	• **Foreign Aid:** includes issues related to foreign aid not directly targeting international development; • **Resources Exploitation:** includes issues related to international resource exploitation and agreements, law of the sea, and international ocean conservation efforts; • **Developing Countries:** includes issues related specifically to developing countries; • **International Finance:** includes issues related to international finance and economic development, the World Bank, International Monetary Fund, regional development banks, sovereign debt, and implications for international lending institutions; • **Western Europe:** includes issues related to Western Europe and the European Union; • **Specific Country:** includes issues related specifically to a foreign country or region not codable using other codes, including political assessments and relations between countries; • **Human Rights:** includes issues related to human rights, human rights violations, treaties, conventions, UN reports, and crimes such as genocide or crimes against humanity; • **Organizations:** includes issues related to international organizations, NGOs, the United Nations, International Red Cross, UNESCO, International Olympic Committee, and the International Criminal Court; • **Terrorism:** includes issues related to international terrorism, hijacking, piracy, anti-terrorism efforts, and international legal mechanisms against terrorism; • **Diplomats:** includes issues related to diplomats, diplomacy, embassies, citizens abroad, foreign diplomats in the country, visas, and passports; • **Other:** includes issues related to other international affairs policy subtopics.

Main Topic	Subtopics
Government Operations	• **Intergovernmental Relations:** includes issues related to intergovernmental relations and local government issues; • **Bureaucracy:** includes issues related to general government efficiencies and bureaucratic oversight; • **Postal Service:** includes issues related to postal services, regulation of mail, and post-civil service; • **Employees:** includes issues related to civil employees, government pensions, and general civil service issues; • **Appointments:** includes issues related to nominations and appointments; • **Currency:** includes issues related to currency, national mints, medals, and commemorative coins; • **Procurement & Contractors:** includes issues related to government procurement, government contractors, contractor and procurement fraud, and procurement processes; • **Property Management:** includes issues related to government property management, construction, and regulation; • **Tax Administration:** includes issues related to tax administration, enforcement, and auditing for individuals and corporations; • **Scandals:** includes issues related to public scandals and impeachment; • **Branch Relations:** includes issues related to government branch relations, administrative issues, and constitutional reforms; • **Political Campaigns:** includes issues related to regulation of political campaigns, campaign finance, political advertising, and voter registration; • **Census & Statistics:** includes issues related to census and statistics collection by the government; • **Capital City:** includes issues related to the capital city; • **Claims against the Government:** includes issues related to claims against the government and compensation for victims of terrorist attacks; • **National Holidays:** includes issues related to national holidays and their observation; • **Other:** includes issues related to other government operations subtopics.
Public Lands	• **National Parks:** includes issues related to national parks, memorials, historic sites, and recreation, including the management and staffing of cultural sites; • **Indigenous Affairs:** includes issues related to indigenous affairs, indigenous lands, and assistance to indigenous people; • **Public Lands:** includes issues related to natural resources, public lands, and forest management, including forest fires and livestock grazing; • **Water Resources:** includes issues related to water resources, water resource development and civil works, flood control, and research; • **Dependencies & Territories:** includes issues related to territorial and dependency issues and devolution; • **Other:** includes issues related to other public lands policy subtopics.
Culture	• **Culture:** includes issues related to cultural policy and its impact on society.

Figure A.13: Comparison of Policy Similarity Within and Across Clusters

User Prompt

Input Example: Here is an initial statement:
"*Taxpayer Protest of Property Situs. This policy amends the tax code to allow taxpayers to protest the inclusion of their property on the appraisal records based on taxable situs, with provisions for notification and effective date.*"
You are provided with two additional statements, each identified by an ID (formatted as 'ID: text'):
[randomly ordered]
"*VA_1901_0121_727: Property Tax Assessment Correction. This policy allows a specific organization to request a correction of the assessment of its lands in a specific county before a specific date, as per the general law..*" [Within]
"*OR_2007_0010_324: Cooperative Governance. This policy outlines the governance structure for cooperatives, including the qualifications, election, and removal of directors, as well as the powers and management of the cooperative's affairs.*" [Across]

Prompt

Determine which statement discusses a policy most similar to the initial statement. Respond only with the ID of the most similar statement.
If both are equally similar, respond with 'Both'.

Answer

Output Example:
" *VA_1901_0121_727* " [Within]

Table A.1: Propensity to spread laws by state ($\beta_{s'}$) for the pre vs. post 1950 periods

State	Pre 1950	Post 1950
AL	-0.0022	0.0017
	(-0.0029, -0.0014)	(0.0008, 0.0026)
AR	-0.0017	0.0009
	(-0.0025, -0.0009)	(-0.0001, 0.0018)
AZ	-0.0002	-0.0009
	(-0.0011, 0.0006)	(-0.0022, 0.0004)
CA	0.0020	0.0004
	(0.0013, 0.0027)	(-0.0005, 0.0012)
CO	0.0014	0.0016
	(0.0005, 0.0022)	(0.0005, 0.0027)
CT	0.0011	-0.0025
	(0.0004, 0.0019)	(-0.0033, -0.0016)
DE	0.0029	-0.0006
	(0.0020, 0.0038)	(-0.0016, 0.0003)
FL	0.0027	0.0003
	(0.0020, 0.0035)	(-0.0006, 0.0012)
Fed	0.0047	0.0016
	(0.0037, 0.0057)	(0.0010, 0.0028)
GA	-0.0004	0.0006
	(-0.0012, 0.0003)	(-0.0003, 0.0014)
IA	-0.0007	0.0002
	(-0.0015, 0.0002)	(-0.0007, 0.0012)
ID	-0.0025	0.0011
	(-0.0034, -0.0017)	(-0.0001, 0.0022)
IL	0.0005	0.0020
	(-0.0002, 0.0012)	(0.0011, 0.0030)
IN	0.0005	0.0003
	(-0.0004, 0.0013)	(-0.0006, 0.0012)
KS	0.0008	0.0020
	(0.0000, 0.0016)	(0.0011, 0.0029)
KY	-0.0005	-0.0005
	(-0.0015, 0.0004)	(-0.0015, 0.0005)
LA	-0.0007	-0.0023
	(-0.0015, 0.0001)	(-0.0033, -0.0014)
MA	-0.0009	0.0003
	(-0.0017, 0.0000)	(-0.0006, 0.0012)
MD	0.0015	0.0004
	(0.0007, 0.0022)	(-0.0004, 0.0012)
ME	0.0015	-0.0003
	(0.0008, 0.0023)	(-0.0012, 0.0006)
MI	-0.0012	0.0004
	(-0.0020, -0.0003)	(-0.0004, 0.0013)
MN	0.0019	0.0000
	(0.0011, 0.0027)	(-0.0008, 0.0008)
MO	-0.0004	0.0017
	(-0.0013, 0.0005)	(0.0007, 0.0027)
MS	-0.0028	-0.0005
	(-0.0036, -0.0020)	(-0.0014, 0.0005)
MT	0.0005	0.0002
	(-0.0004, 0.0013)	(-0.0009, 0.0013)
NC	0.0015	0.0015
	(0.0008, 0.0023)	(0.0007, 0.0023)
ND	-0.0002	0.0006
	(-0.0010, 0.0006)	(-0.0005, 0.0018)
NE	-0.0015	0.0016
	(-0.0024, -0.0006)	(0.0005, 0.0027)
NH	-0.0002	-0.0018
	(-0.0010, 0.0007)	(-0.0027, -0.0009)
NJ	-0.0003	-0.0009
	(-0.0011, 0.0006)	(-0.0017, -0.0002)
NM	-0.0009	-0.0007
	(-0.0018, -0.0000)	(-0.0018, 0.0005)
NV	-0.0005	-0.0010
	(-0.0012, 0.0003)	(-0.0020, 0.0000)
NY	0.0006	-0.0012
	(-0.0001, 0.0014)	(-0.0019, -0.0005)
OH	0.0019	-0.0014
	(0.0010, 0.0029)	(-0.0022, -0.0005)
OK	0.0008	0.0000
	(-0.0000, 0.0017)	(-0.0011, 0.0012)
OR	0.0022	0.0012
	(0.0014, 0.0029)	(0.0003, 0.0021)
PA	-0.0021	0.0002
	(-0.0030, -0.0013)	(-0.0006, 0.0010)
RI	0.0003	-0.0021
	(-0.0006, 0.0011)	(-0.0030, -0.0011)
SC	-0.0000	0.0001
	(-0.0008, 0.0008)	(-0.0008, 0.0009)
SD	-0.0017	-0.0010
	(-0.0025, -0.0008)	(-0.0021, 0.0002)
TN	0.0004	-0.0003
	(-0.0004, 0.0011)	(-0.0013, 0.0006)
TX	0.0013	0.0006
	(0.0005, 0.0021)	(-0.0004, 0.0015)
UT	-0.0016	0.0010
	(-0.0026, -0.0006)	(-0.0004, 0.0023)
VA	0.0005	0.0012
	(-0.0003, 0.0012)	(0.0004, 0.0020)
VT	0.0001	-0.0012
	(-0.0010, 0.0011)	(-0.0022, -0.0002)
WA	0.0010	0.0025
	(0.0001, 0.0019)	(0.0015, 0.0036)
WI	0.0016	0.0005
	(0.0007, 0.0025)	(-0.0003, 0.0013)
WV	-0.0029	-0.0015
	(-0.0039, -0.0020)	(-0.0026, -0.0005)
WY	-0.0027	-0.0010
	(-0.0037, -0.0017)	(-0.0023, 0.0002)

Table A.2: Descriptive Statistics of Distinct Policies – 1% Random Statutes

Category	**Value**
Total statutes	24,977
Total distinct policies	35,928
Single-policy statutes	22,346
Multiple-policy statutes	2,631
Total policies in multiple-policy statutes	13,582
CAP Topic	
All: cluster match (each policy & corresponding summary)	38.35%
All: cluster or neighboring cluster match (each policy & corresponding summary)	44.19%
All: cluster match (at least one policy per statute & corresponding summary)	48.94%
Single-policy statutes: cluster match (each policy & corresponding summary)	47.84%
Multiple-policy statutes: cluster match (each policy & corresponding summary)	22.73%
Multiple-policy statutes: cluster match (at least one policy per statute & corresponding summary)	58.30%
Clustering	
All: cluster match (each policy & corresponding summary)	38.35%
All: cluster or neighboring cluster match (each policy & corresponding summary)	44.19%
All: cluster match (at least one policy per statute & corresponding summary)	48.94%
Single-policy statutes: cluster match (each policy & corresponding summary)	47.84%
Multiple-policy statutes: cluster match (each policy & corresponding summary)	22.73%
Multiple-policy statutes: cluster match (at least one policy per statute & corresponding summary)	58.30%

Notes: Cluster size = 2,048. The true number of 1% random statutes is 24,981, but our approach, using GPT for listing distinct policies combined with an extraction method to list each distinct policy, misses some.

Table A.3: Descriptive Statistics of Distinct Policies – Federal Statutes

Category	**Value**
Total statutes	89,050
Total distinct policies	114,126
Single-policy statutes	82,853
Multiple-policy statutes	6,197
Total policies in multiple-policy statutes	31,273
CAP Topic	
All: cluster match (each policy & corresponding summary)	33.69%
All: cluster or neighboring cluster match (each policy & corresponding summary)	38.50%
All: cluster match (at least one policy per statute & corresponding summary)	39.56%
Single-policy statutes: cluster match (each policy & corresponding summary)	38.47%
Multiple-policy statutes: cluster match (each policy & corresponding summary)	21.03%
Multiple-policy statutes: cluster match (at least one policy per statute & corresponding summary)	54.11%
Clustering	
All: cluster match (each policy & corresponding summary)	33.69%
All: cluster or neighboring cluster match (each policy & corresponding summary)	38.50%
All: cluster match (at least one policy per statute & corresponding summary)	39.56%
Single-policy statutes: cluster match (each policy & corresponding summary)	38.47%
Multiple-policy statutes: cluster match (each policy & corresponding summary)	21.03%
Multiple-policy statutes: cluster match (at least one policy per statute & corresponding summary)	54.11%

Notes: Cluster size = 2048. The true number of federal statutes is 89,316, but our approach,using GPT for listing distinct policies combined with an extraction method to list each distinct policy, misses some.

Table A.4: Comparison of MPNet and FastText Clusters

Cluster Size	**MPNet**	**FastText**	**Both**	**N/A**
1024	54.6	24.7	20.7	0.0
1536	58.0	26.0	16.0	0.1
2048	58.6	26.7	14.7	0.0
2560	60.0	26.8	13.0	0.1

Notes: Based on a random sample of 4,000 statute summaries, each statute is compared with another from the same cluster as defined by clusters constructed using MPNet embeddings and with a statute from the same cluster as defined by clusters constructed using FastText embeddings. GPT-4-o-mini then selects the most similar policy. Values represent percentages.

Table A.5: Proportion of Most Frequent CAP Topic (2048-Cluster)

Topic	Share of clusters (%)	Topic	Share of clusters
Government Operations	27.0	Culture	2.0
Law And Crime	15.6	Public Lands	1.8
Transportation	10.4	Defense	1.1
Domestic Commerce	8.3	Energy Policy	0.9
Education	8.2	Civil Rights	0.8
Health	6.2	Immigration	0.3
Social Welfare	4.2	Foreign Trade	0.1
Environment	3.5	Technology	0.1
Labor	3.4	International Affairs	0.0
Housing	3.3	Macroeconomics	0.0
Agriculture	2.6	Others	0.0

Notes: This table reports the proportion of clusters for which a given CAP topic is the most frequently assigned topic. Topics are based on the Comparative Agenda Project (see Appendix B.1). Values represent percentages.

Table A.6: Proportion of Most Frequent CAP Topic (2048-Cluster)

Topic	Share of clusters	Topic	Share of clusters
Government Operations	25.9	Culture	1.8
Law And Crime	17.3	Public Lands	1.6
Transportation	10.2	Defense	1.1
Domestic Commerce	8.5	Energy Policy	0.9
Education	7.6	Civil Rights	0.8
Health	6.2	Immigration	0.4
Environment	3.9	Technology	0.2
Social Welfare	3.8	Foreign Trade	0.1
Labor	3.7	International Affairs	0.0
Housing	3.2	Macroeconomics	0.0
Agriculture	2.9	Other	0.0

Notes: This table reports the proportion of clusters for which a given CAP topic is the most frequently assigned topic. Topics are based on the Comparative Agenda Project (see Appendix B.1). Values represent percentages.

Table A.7: Descriptive Statistics for 2560-Cluster Configuration

	Mean	Std dev.	5th pctile	95th pctile
Number of statutes per cluster	975.9	604.1	271.0	2076.7
Time span per cluster	87.4	35.5	33.0	145.0
Number of states per cluster	43.7	10.1	20.0	50.0
HH index	0.46	0.16	0.25	0.78
Share of clusters with more than one state contributing	99.8%			
Share of clusters where all states are contributing	38.3%			
Share of clusters with federal government contributing	76.1%			
Share of clusters where a state contributes more than half	1.9%			
Share of clusters where federal government contributes more than half	1.8%			

Note: For statistics related to years, we removed statutes below the 10th percentile and above the 90th percentile per cluster. For statistics on states, we excluded federal government statutes from the clusters.

Table A.8: Descriptive Statistics for 2048-Cluster Configuration Without Statutes Anonymization

	Mean	Std dev.	5th pctile	95th pctile
Number of statutes per cluster	1219.9	814.2	305.4	2578.3
Time span per cluster	89.5	35.1	35.0	145.6
Number of states per cluster	45.7	8.4	28.0	50.0
HH index	0.42	0.15	0.23	0.70
Share of clusters with more than one state contributing	99.9%			
Share of clusters where all states are contributing	46.1%			
Share of clusters with federal government contributing	81.2%			
Share of clusters where a state contributes more than half	1.8%			
Share of clusters where federal government contributes more than half	0.3%			

Note: For statistics related to years, we removed statutes below the 10th percentile and above the 90th percentile per cluster. For statistics on states, we excluded federal government statutes from the clusters.

Table A.9: Comparison of Policy Similarity Within and Across Clusters

Cluster Size	**Same Cluster**	**Different Cluster**	**Both**	**N/A**
1024	69.5	5.2	25.1	0.2
1536	72.5	6.0	21.5	0.1
2048	**75.5**	**4.8**	**19.6**	**0.1**
2560	76.6	4.7	18.6	0.1

Notes: Values represent percentages.

Figure A.14: Comparison of Policy Goal Alignment Within and Across Clusters

Prompt

You are given summaries of two statutes enacted by a U.S. state or the federal government in [year].
Assess whether these statutes pursue opposite or aligned policy goals, considering the political ideologies and contexts rather than the policy topic itself.
Important: Answer using one of the following options only (no extra explanation or phrases):

- Definitely opposing goals
- Probably opposing goals
- Probably aligned goals
- Definitely aligned goals
- Undetermined

Examples:

- Opposing: One statute increases environmental regulations to reduce industrial pollution, while another relaxes such regulations to promote economic growth.
- Aligned: Both statutes aim to improve education, with one funding early childhood education and the other providing incentives for teacher training programs.

User Prompt

Input Example - Same Cluster
Here are the summaries:
" *Probation Officer Appointments and Compensation. This policy establishes the appointment, duties, and compensation of probation officers, assistant probation officers, and other personnel for the juvenile court in a specific county in the state.* "
" *Division of Probation. This policy establishes the Division of Probation as a division in the executive department, outlining the duties, functions, and powers of the division, and repealing certain sections of the correction law.* "

Answer

Output Example - Same Cluster
" *Definitely aligned goals* "

User Prompt

Input Example - Different Cluster
Here are the summaries:
" *Probation Officer Appointments and Compensation. This policy establishes the appointment, duties, and compensation of probation officers, assistant probation officers, and other personnel for the juvenile court in a specific county in the state.* "
" *Small Claims Court Procedures. This policy amends small claims court procedures, allowing for motions to vacate judgments and appeals to the superior court under specific conditions, including proper service and appearance.* "

Answer

Output Example - Different Cluster
" *Undetermined* "

Figure A.15: Comparison of Policy Goal Alignment Within and Across Clusters

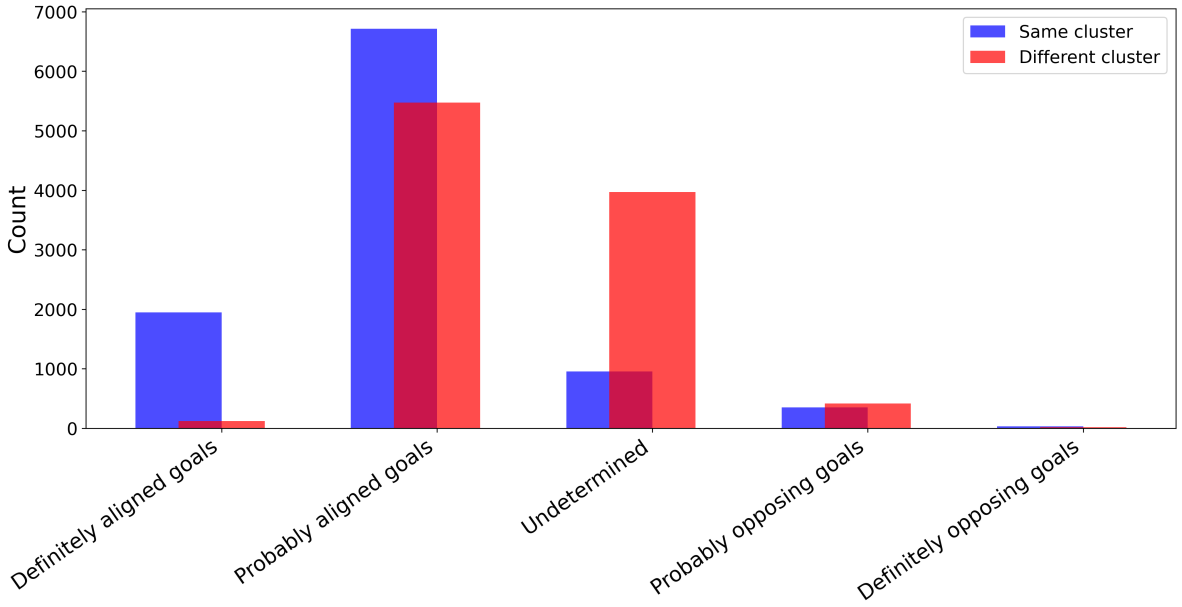
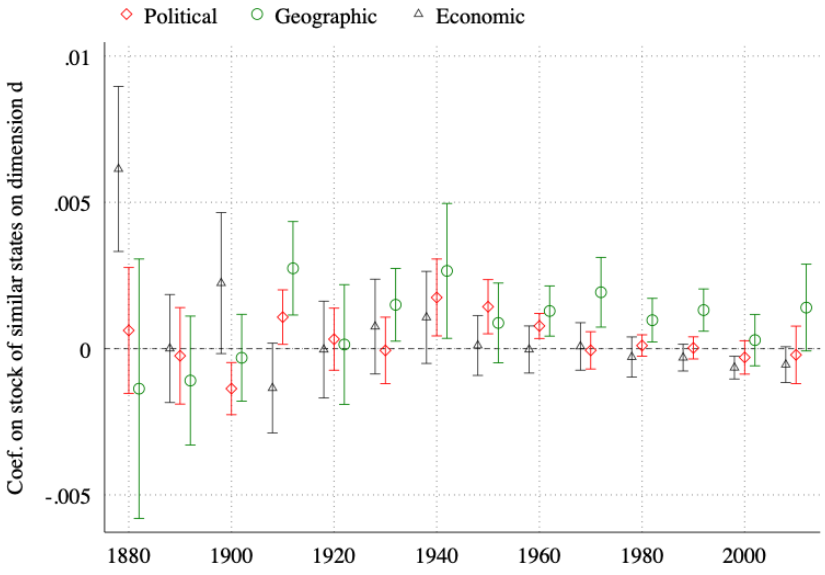


Figure A.16: Topic Polarization



$Z^d_{c,s,t}$ computed from aggregated subtopics (k=115).