

RESEARCH NOTE OPEN ACCESS

Mapping Measurement to Theory in Policy Evolution: A Case of Net Metering Policies of the United States

Graham Ambrose¹  | Myriam Gregoire-Zawilski² 

¹School of Public and International Affairs, North Carolina State University, Raleigh, North Carolina, USA | ²Graduate School of Public and International Affairs and Institute of the Environment, University of Ottawa, Ottawa, Ontario, Canada

Correspondence: Graham Ambrose (gwambros@ncsu.edu)

Received: 17 October 2025 | **Revised:** 4 May 2026 | **Accepted:** 8 May 2026

Keywords: net metering | policy design | policy evolution | policy packaging | policy patching

关键词: 政策演变 | 政策设计 | 政策修补 | 政策组合 | 净计量

Palabras clave: Evolución de políticas | Diseño de políticas | Adaptación de políticas | Agrupación de políticas | Medición neta

ABSTRACT

The examination of policy evolution has important practical implications, but current literature often only evaluates changing policy goals and instruments at the macro-level, overlooking changes that occur at the more granular, micro-level of policy text. Furthermore, concepts such as policy patching and packaging have achieved theoretical salience but lack operationalization in the context of medium- to large-N studies. We offer a new approach for measuring dynamics of patching and packaging, proposing three dimensions of a single measure evaluating evolution in policy text: (1) depth: the extent of change, on average, across policy statements, (2) width: how these changes are distributed across the policy, and (3) the (a)symmetry of depth and width across added and terminated text. With these three dimensions, we offer a more nuanced exploration of patching and packaging dynamics. We test our approach using net metering legislations enacted and amended across 32 states since 1980. We find that when policy evolution is observed at the micro-level, both patching and packaging dynamics are prevalent, with packaging occurring almost as often as patching. Thus, we contribute to the study of the public policy process by offering a new theoretically-rooted approach to measuring policy evolution dynamics using policy text as data.

摘要

摘要政策演化的研究具有重要的实践意义,但现有文献通常仅从宏观层面评估政策目标和工具的变化,而忽略了政策文本微观层面的变化。此外,诸如政策修补和政策打包等概念虽然在理论上具有重要意义,但在中大型样本研究中缺乏实际应用。我们提出了一种衡量政策修补和政策打包动态的新方法,提出了一个包含三个维度的单一指标来评估政策文本的演化:(1)深度:政策声明中平均变化的程度;(2)宽度:这些变化在政策中的分布情况;(3)新增和终止文本的深度和宽度的(非)对称性。通过这三个维度,我们对政策修补和政策打包的动态进行了更为细致的探索。我们利用自1980年以来32个州颁布和修订的净计量法规来检验我们的方法。我们发现,在微观层面观察政策演变时,政策修补和政策组合两种动态模式都十分普遍,其中政策组合的出现频率几乎与修补一样高。因此,我们通过提供一种新的、基于理论基础的、以政策文本为数据来衡量政策演变动态的方法,为公共政策过程的研究做出了贡献。

RESUMEN

El análisis de la evolución de las políticas tiene importantes implicaciones prácticas, pero la literatura actual a menudo solo evalúa los cambios en los objetivos e instrumentos de las políticas a nivel macro, pasando por alto los cambios que ocurren a nivel

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2026 The Author(s). *Policy Studies Journal* published by Wiley Periodicals LLC on behalf of Policy Studies Organization.

micro, en el texto de las políticas. Además, conceptos como la modificación y el empaquetamiento de políticas han alcanzado relevancia teórica, pero carecen de operacionalización en el contexto de estudios con un número de muestras mediano a grande. Proponemos un nuevo enfoque para medir la dinámica de la modificación y el empaquetamiento, planteando tres dimensiones para una única medida que evalúa la evolución en el texto de las políticas: (1) profundidad: la magnitud del cambio, en promedio, en las declaraciones de la política; (2) amplitud: cómo se distribuyen estos cambios en la política; y (3) la (a)simetría de la profundidad y la amplitud en el texto añadido y eliminado. Con estas tres dimensiones, ofrecemos una exploración más matizada de la dinámica de la modificación y el empaquetamiento. Validamos nuestro enfoque utilizando las leyes de medición neta promulgadas y modificadas en 32 estados desde 1980. Observamos que, al analizar la evolución de las políticas a nivel micro, predominan tanto la adaptación como la agrupación de políticas, siendo la agrupación casi tan frecuente como la adaptación. De este modo, contribuimos al estudio del proceso de políticas públicas al ofrecer un nuevo enfoque con fundamento teórico para medir la dinámica de la evolución de las políticas utilizando el texto de las mismas como datos.

1 | Introduction

The field has a long-standing interest in understanding what precipitates and shapes policy change (Wilson 1995; Lowi 1964; Howlett and Rayner 2013; Mahoney and Thelen 2009; Howlett and Mukherjee 2014). Various streams of literature explore the antecedents and consequences of policy evolution, including in the policy design literature where understanding policy evolution has garnered substantial interest from both scholars and practitioners (Howlett and Rayner 2013; Mahoney and Thelen 2009; Howlett and Mukherjee 2014). Despite this, the theoretical constructs used to characterize dynamics of policy evolution remain difficult to operationalize in empirical research. A common gap across studies has been the lack of reliable and replicable measures for quantifying dynamics of policy design evolution. Thus, we proffer an approach to evaluate, in policy text: (1) the magnitude and concentration of policy changes, and (2) the alignment of this change with policy evolution concepts proposed in literature, that is, policy patching and policy packaging. Our approach summarizes depth, breadth and asymmetry of changes in policy text into a single composite quantitative measure, enabling us to empirically distinguish these two modes of policy evolution.

We extend the work of Ambrose et al. (2024) who hand code the text of legislative amendments, examining how policy content has evolved over time (Howlett and Rayner 2013; Mahoney and Thelen 2009; Howlett and Mukherjee 2014). While the prior work explored dynamics of individual statement change, here we propose a measurement approach that maps statement-level evolution in policy text to policy-level dynamics of policy patching and packaging. This approach can be automated across a large number of policy documents to rigorously measure these dynamics using policy text. We show that both dynamics are prevalent in our sample: policy packaging, when old language is removed and replaced with new language; and policy patching, when new language is added alongside existing language without adjustments to incumbent policy designs. Our approach adds clarity to the current discussion of these dynamics by more precisely operationalizing the concepts of patching and packaging. This, we argue, is an essential step towards fulfilling scholars' call for more work exploring the context and decision-making processes that lead to different evolution dynamics, as well as how policy evolution leads to variable implementation outcomes (Capano and Howlett 2024, 2020; Guha and Mukherjee 2025).

Measuring dynamics of patching and packaging is important by themselves, but our approach can support future inquiry into these gaps.

We select the case of net metering policy to trial our measurement approach. Net metering allows utility customers to feed excess renewable energy back to the grid and is one of the most widely used state-level renewable energy policies in the United States. Since the 1980s, 32 states have adopted net metering legislation. We found a total of 206 legislative documents, initial adoptions and amendments, providing ample opportunity to study dynamics of policy evolution over several decades and across many jurisdictions.

2 | Literature Review

2.1 | Policy Evolution

2.1.1 | Defining Modes of Policy Evolution

The literature on policy design, which is more specifically interested in understanding the process of *designing* policies (policy design as a verb) and the *content* of policies (policy design as a noun¹), began engaging actively with questions of policy evolution in the mid-2000s (Howlett and Rayner 2007; Howlett and Cashore 2009). Policy design scholars broadly define evolution through different processes of layering. More specifically, *policy patching* adds new policy language without removing old language, and *policy packaging* restructures policies by replacing old language with new language rather than simply adding new language (Howlett et al. 2018; Kern et al. 2017; Howlett and Rayner 2013; Capano and Mukherjee 2020).

Yet, across the policy design literature, scholars attach additional value to the notions of patching and packaging that often goes unmeasured—particularly the dynamics of small incremental changes and large punctuations. Scholars have highlighted how intentional “smart layering” affords greater opportunity to make incremental refinements to policy. They liken this “tinkering” and intentional troubleshooting of policy designs to the way a computer programmer issues “patches” to fix code in a computer program (Guha and Mukherjee 2025; Howlett et al. 2018; Wellstead et al. 2016). In this way, changes are implied to be targeted, incrementally fixing existing subsections of text (Wellstead et al. 2016; Capano and Howlett 2024; Cashore and Howlett 2007). In other

cases, scholars discuss evolution as a substantial layering of new policy content often expanding or altering the scope of the policy, as well as the tasks and behaviors of implementers (Barnett et al. 2020; Howlett and Rayner 2007, 2013; Kern et al. 2017). In prior theoretical and qualitative empirical work, scholars often discuss the extent of layering (i.e., the depth of change) and the direction of change (additions and/or terminations) (Barnett et al. 2020; Wellstead et al. 2016; Howlett et al. 2018). To this, we add the targeted or untargeted nature of policy evolution across the policy (i.e., width of statements changed across the policy), which we argue scholars have previously implied, through notions of targeted incrementalism and broad punctuations, but have not measured directly in policy design studies.

2.1.2 | Measurement of Modes of Policy Evolution in Extant Literature: Policy Patching and Packaging

While we focus on the micro-level (individual policy statements or sentences), prior policy design literature has leveraged similar notions of width and depth at the macro- (policies within mixes) and meso-level (instruments within policies). Kern et al. (2017) evaluate dynamics of patching and packaging in policy mixes, examining the number of policy instruments added and removed (i.e., depth) within a policy domain, and the breadth of instrument types within the mix (i.e., width). Fitch-Roy et al. (2020) examine the “depth” (or degree) of novel instruments adopted, and “width” or range of policy goals when operationalizing patching and packaging. Hinterleitner et al. (2024) review metrics used in the literature on policy expansion, including indicators capturing dimensions similar to depth and width of policy targets and tools.

Current work at the micro-level of policy design is largely supported through in-depth case descriptions across one or a few cases (Capano and Howlett 2024; Capano and Howlett 2020; Wellstead et al. 2016). While this work is highly valuable, it lacks evidence of generalizability that larger samples provide. Complementing this work, our measurement approach enables medium- and large-N comparisons, which also engaging research gaps at the micro-level highlighted by Capano and Howlett (2024): (1) to better understand political conflicts and context as it relates to different patterns of policy evolution (Nair and Howlett 2020; Howlett and Cashore 2009), and (2) to better understand how evolution links to behavioral changes in implementation and tangible policy outputs and outcomes (Guha and Mukherjee 2025; Howlett and Leong 2022; Howlett et al. 2018). While we do not explore these relationships here, and our measure does not capture changes in policy content, we share the view of other scholars: dynamics of patching and packaging are in and of themselves meaningful as they can inform us on how different jurisdictions adjust and update their policies (Rayner and Howlett 2009; Kern et al. 2017) and represent important dynamics in policy process research (Capano and Howlett 2024, 2020).

Thus, in this note, we offer a generalizable approach for measuring patching and packaging to advance future medium- and large-N research in these directions. We focus on policy patching and packaging because their definitions are agnostic to and conceptually distinct from identifying the goals or intentions of the policy, which require additional examinations of the policy context that are beyond the scope of this note.

2.1.3 | Policy Evolution Dimensions: Depth, Width, and (a)symmetry of Additions and Terminations

We measure change in the text of policies along three dimensions: (1) depth (2) width, and (3) (a)symmetry of additions and terminations. *Depth* captures the extent to which policy language changes, mapping to the concept of policy *layering* developed by Howlett and Cashore (2009), Rayner and Howlett (2009), and similar work. However, these studies have used the concept of layering to characterize instances of large policy change (high depth) often through punctuation at the meso- and macro-levels. While sometimes implied in discussion such as with “smart layering,” policy design scholarship offers limited guidance for conceptualizing instances of small policy change (low depth). Yet, small incremental change is observed abundantly in other policy process theories (e.g., PET (Baumgartner and Jones 1993) and incrementalism (Lindblom 1959)). *Width* captures how change is distributed across the text of the policy. While again implied by “smart layering,” we argue that characterizations of policy evolution found in extant policy design literature do not consider this facet of change, as we can consider the concentration of change across a policy distinctly from the depth of that same change. This results in: (1) narrow change, or change across a few statements, and (2) broad change, or change across many statements. Finally, because the definitions of patching and packaging identify change in what is added and removed, we further operationalize the concepts through the *(a)symmetry* in depth and width of additions and termination. The conceptual justification of each dynamic is discussed below.

Capturing change along these three dimensions (depth, width, and (a)symmetry of additions and terminations) is necessary to credibly discern patching from packaging. Conceptually, patching occurs when individual elements are adjusted in relative isolation, whereas packaging involves more coordinated recalibration of interrelated elements. These distinct dynamics of change would be missed if we solely examined the number of words or statements changing across policy amendments. The first dimension, depth, captures the magnitude of textual change, or, when change is observed, how intensive the change is. However, regardless of their intensity, changes can be targeted, occurring in few statements, or diffuse, occurring across many statements. Hence, width is needed to capture how broadly change is distributed across the policy’s statements.

Table 1 shows a 2×2 matrix combining our concepts of width and depth. As a practical example, if a deadline for an agency to submit

TABLE 1 | Measuring modes of policy change.

		Depth of change	
		Low	High
Width of change	Narrow	A. Low depth & narrow width of change	B. High depth & narrow width of change
	Broad	C. Low depth & broad width of change	D. High depth & broad width of change

Note: We compute the depth of change and the width of change separately for additions and terminations in policy statements.

a report is updated, this manifests as a small change in one statement (*Low Depth & Narrow Width*). When the definition of a policy target, and subsequently their name, is revised—that is, “utilities” to “service providers”—necessitating edits to every sentence referencing the target, these edits are small but diffuse (*Low Depth & Broad Width*). When definitions are updated, such as including multi-family dwellings in the definition of residential customers, changes to this statement can be extensive but remains concentrated (*High Depth & Narrow Width*). When legislators restructure the policy’s ratemaking requirements, changing a full and lengthy policy section, these changes are large and diffuse (*High Depth & Broad Width*). These examples illustrate why both dimensions measure distinct and important information. Furthermore, while depth and width are not a direct proxy for changes in policy content, it is reasonable to expect that a policy falling into *Quadrant D* would have more substantial changes to who can do what, when, and how compared to one falling into *Quadrant A*.

An additional dimension is critical to distinguish between patching and packaging: similarities or differences between additions and terminations. Conceptually, patching involves layering new provisions, clarifications, or procedural refinements, without removing pre-existing provisions. Empirically, we should observe asymmetrical patterns between additions (large and diffuse) and terminations (small and narrow). In the high depth & narrow width example above, the definition of multi-family dwellings is *added* to the definition of eligible customers *without terminating* prior text regarding residential customers—consistent with the concept of patching. The concept of packaging, in contrast, often entails replacement, representing more symmetrical patterns between additions and terminations, suggesting more conscious recalibration of the policy. In the high depth & broad width example above, legislators restructure the policy’s ratemaking requirements by fully *terminating* a policy section and *adding* a completely new policy section in its place—a pattern indicative of packaging. The (a)symmetry between the depth and width measures of additions and terminations is therefore critical in distinguishing between patching and packaging. Our main contribution is to measure these dimensions at the micro-level using policy text as data. We present our approach in the next section.

3 | Methods

3.1 | Leveraging a Text Reuse Algorithm to Quantify Policy Evolution Through Policy Text

Text reuse methods have become more common in recent years to measure policy process phenomena, especially as a proxy for the transfer of policy ideas. Scholars have used these methods to capture the diffusion of state policies (Hinkle 2015), the reuse of model policies (Wilkerson et al. 2015; Burgess et al. 2016; Linder et al. 2020), the diffusion of advocacy documents (Pagliari and Young 2020), and the reuse of boilerplate language (Scott et al. 2022).

To the best of our knowledge, no paper has so far used these methods to track policy evolution within jurisdictions. Here, we use these methods to collect information regarding what changed, rather than what was reused. While not a substitute for expert coding of policy content, our approach sifts through policy text to clearly label what was maintained (i.e., reused), and identify what was added and removed between two iterations of a policy. Furthermore, because boilerplate language tends to be reused across iterations of the same policy, we believe the additions and terminations we observe generally point to meaningful changes in policy content.

3.2 | Introducing Text Reuse

We follow Ambrose et al.’ (2024) conceptualization of policy evolution at the policy statement-level. Policy statements can be: (1) *added*, where an entire new statement is added, (2) *terminated*, where an entire statement is removed, (3) *calibrated*, where portions of a statement are added to or terminated from a prior policy statement, and (4) *maintained*, when the statement does not undergo change. We evaluate these changes at the intra-statement-level, thus, a calibrated statement is made up of maintained words (i.e., identified as reused), as well as terminated words (i.e., in the former version but not the current

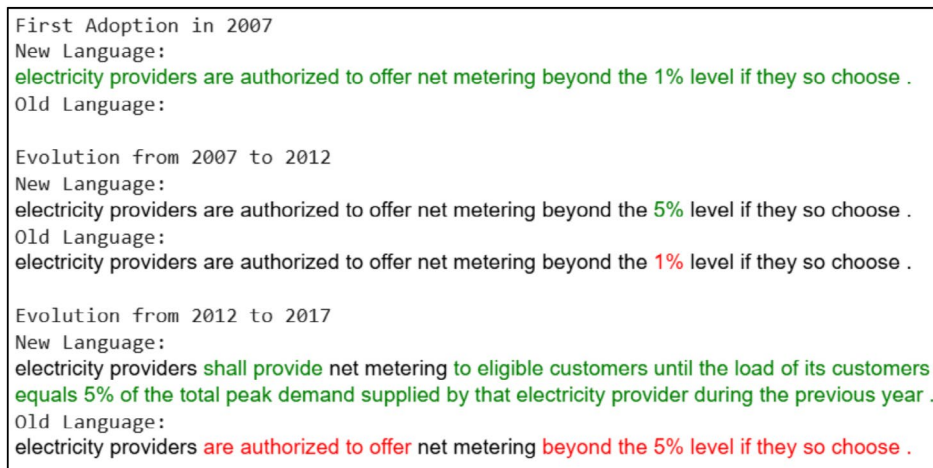


FIGURE 1 | Example of Intra-statement evolution coding across multiple amendments to the same statement. Green text: Added text from the prior version of the statement; Red text: Terminated text from the prior version of the statement; Black text: Maintained text from the prior version of the statement.

version) and/or added words (i.e., in the current but not the former).

To exemplify this, three versions of the same policy statement are offered in Figure 1. Added text is in green, terminated in red, and maintained in black. This statement was first *added* in 2007; therefore, its full text is represented in green. It was then *calibrated* in 2012, when one element—the percent cap of eligible customers—was raised from 1% (terminated text in red) to 5% (added text in green). This statement was *calibrated* again in 2017, but in a more complex way, as multiple sections were simultaneously maintained, added, and terminated. We code all statements across all policy iterations and US states in our sample in this way. The resulting maintained, added, and terminated text outputs represent the building blocks upon which our measures are constructed. We use these counts to subsequently compute our measures of change—depth and width—separately for additions and terminations.

3.2.1 | Depth of Change at the Statement-Level

Table 2 represents the evolution observed between 2012 to 2017 in the example from Figure 1. It identifies: (1) the word count of the new statement in 2017 (i.e., New WC), (2) the word count of the added words (i.e., Add WC), (3) the word count of the reused words (i.e., Reuse WC), (4) the word count of the terminated words (i.e., Term WC), and (5) the word count of the old statement from 2012 (i.e., Old WC).

Depth of Addition at the statement level is calculated as added word count (*Add WC*) divided by the word count of the new statements²:

$$\text{Depth of Addition} = \frac{\text{Add WC}}{\text{New Statement WC}}$$

and *Depth of Termination* is calculated as terminated word count (*Term WC*) divided by the word count of the old statements.

$$\text{Depth of Termination} = \frac{\text{Term WC}}{\text{Old Statement WC}}$$

Conceptually, our depth of evolution measures are bound between 0 and 1, where 0 would represent instances where no words are added or terminated in the statement, and 1 would represent cases where the full statement is either added or terminated.

Next, to understand micro-level dynamics, Figure 2 shows the values for Depth of Addition and Depth of Termination for each statement across eight real policy examples from our net metering dataset. This representation functionally stays at the micro-level but begins to consider multiple statements that make up a policy. This also offers a visual representation of the intra-statement evaluation offered by this approach, as statements can be fully added or terminated (i.e., Depth of Addition or Termination is equal to 1) or portions of the text are added and/or terminated (i.e., Depth of Addition or Termination is between 0 and 1). We have selected the eight state-year combinations strategically to illustrate the different dynamics of policy evolution found in our sample. Due to space constraints, Figure 2 does not show all statements for each of the policies, but to give a sense of these measures,

TABLE 2 | Depth of evolution example.

New text	New WC	Add text	Add WC	Reuse text	Reuse WC	Term text	Term WC	Old text	Old WC	Depth of addition	Depth of termination
Electricity providers shall provide net metering to eligible customers until the load of its customers equals 5% of the total peak demand supplied by that electricity provider during the previous year	31	[...] shall provide [...] to eligible customers until the load of its customers equals 5% of the total peak demand supplied by that electricity provider during the previous year	27	electricity providers [...] net metering	4	[...] are authorized to offer [...] beyond the 5% level if they so choose	12	Electricity providers are authorized to offer net metering beyond the 5% level if they so choose	16	0.871	0.750

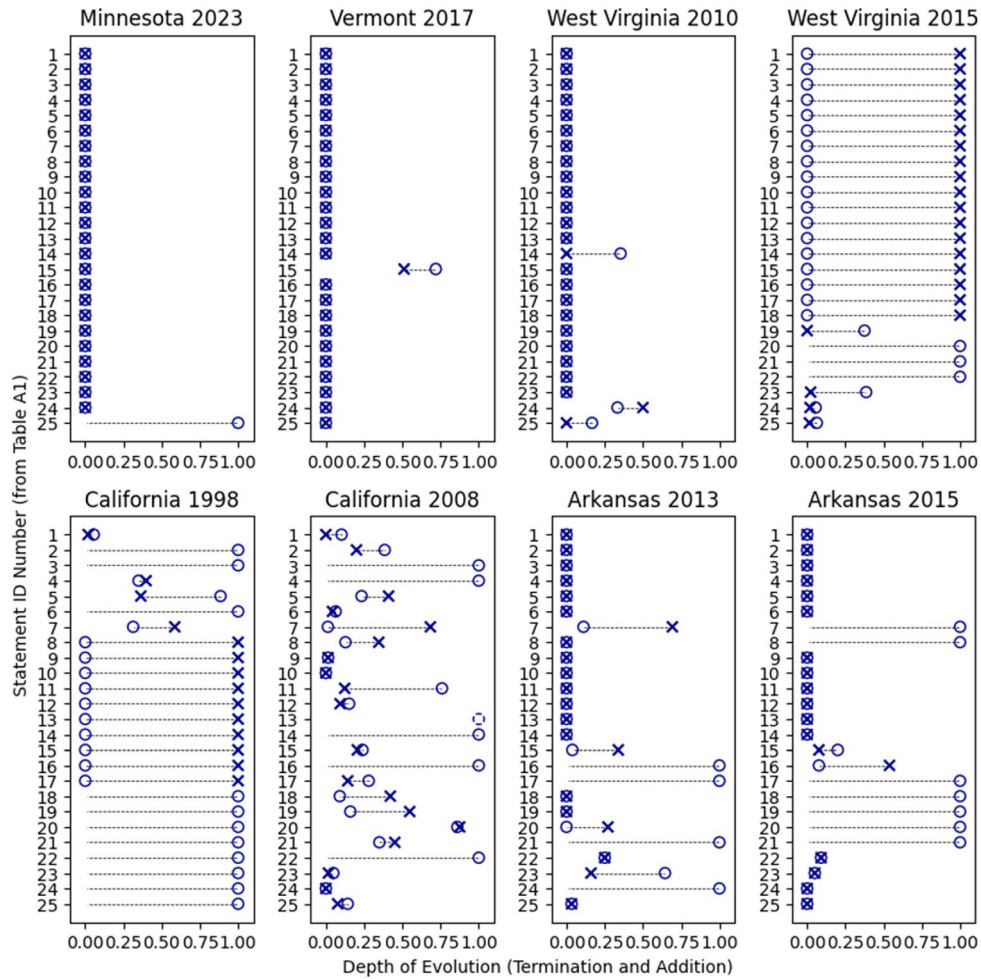


FIGURE 2 | Graphical representation of statement-level depth of addition and termination. Marker “o” indicates the “Depth of Addition” and marker “x” represents the “Depth of Termination” for each statement. When the “o” or “x” markers take a value of 1, 100% of the words were added or terminated, respectively. When 100% of the words are added, this is an entirely new statement, thus there are no words to terminate and the “x” is not displayed. When the “o” or “x” markers take a value of 0, 0 words were added or terminated, respectively. When the values of addition and/or termination are between 0 and 1, this indicates that the statement was calibrated. Titles indicate the year the amendment was enacted, and the depth of evolution represents change compared to the previous version of the policy in effect.

we show a cross section of 25 statements for each state. A full reporting across all statements for these cases is offered in Appendix S3.

Within this representation, additions (represented by marker “o”) and terminations (represented by marker “x”) are plotted separately on the X-axis, for each statement listed on the Y-axis. Statements that were fully added or removed are discernable by the polarization between “o” and “x” markers. For example, Statement #25 in Minnesota was fully added in 2023: the Depth of Addition and the “o” marker takes a value of 1, meaning 100% of the words were added, and, because this is an entirely new statement, there are no words to terminate and the “x” is not displayed. In contrast, Statements #1-#18 in West Virginia were fully terminated in 2015. When the values of addition and/or termination is between 0 and 1, that is, Statement #15 in Vermont 2017, this indicates that the statement was calibrated. Finally, Figure 2 also clearly identifies cases where statement-level calibration is driven by termination or addition only, that is, Statement #20 in Arkansas 2013 and Statement #14 in West Virginia 2010, respectively.

3.3 | Width and Depth of Change at the Policy-Level

While Figure 2 identifies individual depths of change across statements, it also begins to convey the importance of capturing both depth and width of change at the whole policy-level. For additions and terminations, we aggregate our statement-level Depth of Addition and Termination scores as follows:

$$Depth_{js} = \left(\frac{\sum_{i=1}^n X_{ijs}}{n} \right), X_{ijs} > 0$$

Here, depth is computed as the average statement-level Depth of Addition or Termination score X for the j th amendment of the policy in state s , in which there are n statements i that underwent change (i.e., the statement-level Depth of Addition or Termination score was > 0). Furthermore, we compute our measure of depth separately for additions and terminations. This measure is bound by 0 and 1, where a score of 0 signifies the policy did not go through any addition or termination, and 1 indicates, in statements where change is observed, all text was added or terminated. Thus, our

policy-level depth measures appropriately scales the statement-level depth measures by capturing the average extent of change across statements that underwent change in a policy.³

Second, as we are now examining multiple policy statements, we can explore how policy change is targeted to a few statements (i.e., narrow width of evolution) or diffuse across many statements (i.e., broad width of evolution). Returning to Figure 2, we can visually see that Minnesota 2023, Vermont 2017, and West Virginia 2010 represent narrow widths of evolution, while West Virginia 2015 and California 1998 and 2008, represent broader widths of evolution. To measure this empirically, we compute the inverse of the Gini coefficients,⁴ capturing the width of additions and terminations. Gini coefficients are commonly used to explore the distribution or concentration of a variable. In our application, this measure captures how the total volume of words that have undergone change is distributed across a policy's different statements. Importantly, unlike our measure of depth (computed at the statement-level and aggregated to the policy-level), width is inherently defined at the policy-level:

$$Width_{js} = 1 - \frac{\sum_{i=1}^n (2i - n_{js} - 1)X_{ijs}}{n \sum_{i=1}^n X_{ijs}}$$

More specifically, width is computed for the j th amendment of the net metering policy in state s , in which we count n statements i . The number of words that undergo change (addition or termination) within each statement i is denoted as X_{ijs} . This measure is also, in most cases,⁵ bound by 0 and 1, where values approaching 1 indicate that changes are more equally distributed

across a policy's statements (i.e., broad width), and small values approaching 0 indicate that changes are concentrated in few statements (i.e., narrow width). Furthermore, when there is no addition or termination in a policy, the policy takes on a Gini value of 0.⁶ This results in the depth and width of evolution scores identified in Table 3.⁷

We integrate our measures with the modes of policy change in Table 1 to capture the dynamics of patching and packaging at the full policy level. This operationalization focuses on the (a)symmetry between additions and terminations in policy text. Figure 3 illustrates the approach using California 2008. *Panel 1* calculates statement-level depth of evolution. *Panel 2* aggregates these measures to the policy-level, producing separate depth and width scores for additions and terminations (*Panel 3b*). Population thresholds for the modes of evolution are defined as the mean policy-level depth and width—across both additions and terminations—within the full sample (*Panel 3a*; Table 1). Applying these thresholds to the California 2008 scores assigns the evolution modes (*Panel 3c*). *Panel 4* then identifies symmetry (packaging) or asymmetry (patching). Because California 2008 shows symmetric C-to-C additions and terminations, the case is classified as packaging—specifically low-depth, broad-width packaging. Results are discussed in the next section.

4 | Validation: Integrating Measurement and Concepts of Policy Evolution

Finally, we include references and discussions from qualitative literature as well as a qualitative discussion of the policy text itself to offer validation of our approach—with additional

TABLE 3 | Depth and Width of Evolution Scores Using Figure 2 Examples.

State and year of policy	Evolution type	Depth of evolution	Width of evolution	Modes of change
Minnesota 2023	Addition	1.000	0.012	B
	Termination	0.000	0.000	A
Vermont 2017	Addition	0.859	0.039	B
	Termination	0.756	0.037	B
West Virginia 2010	Addition	0.174	0.013	A
	Termination	0.165	0.029	A
West Virginia 2015	Addition	0.599	0.032	B
	Termination	0.979	0.517	D
California 1998	Addition	0.859	0.397	D
	Termination	0.812	0.346	D
California 2008	Addition	0.313	0.391	C
	Termination	0.256	0.296	C
Arkansas 2013	Addition	0.564	0.124	B
	Termination	0.392	0.108	A
Arkansas 2015	Addition	0.806	0.295	D
	Termination	0.301	0.053	A

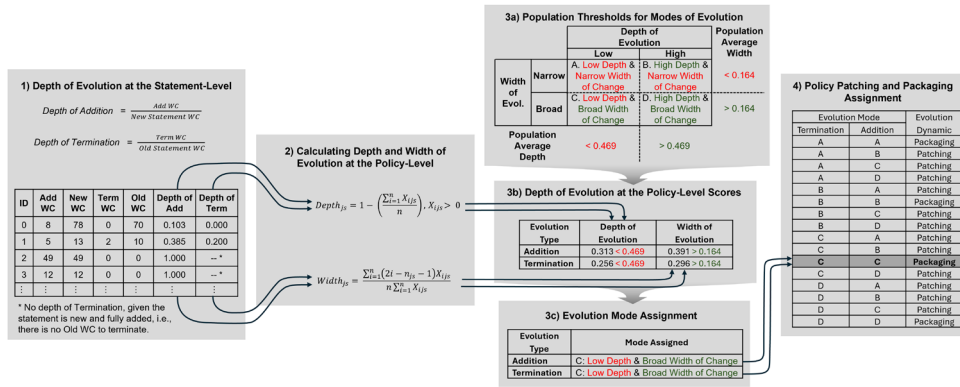


FIGURE 3 | Measurement flow diagram using California 2008 as an example.

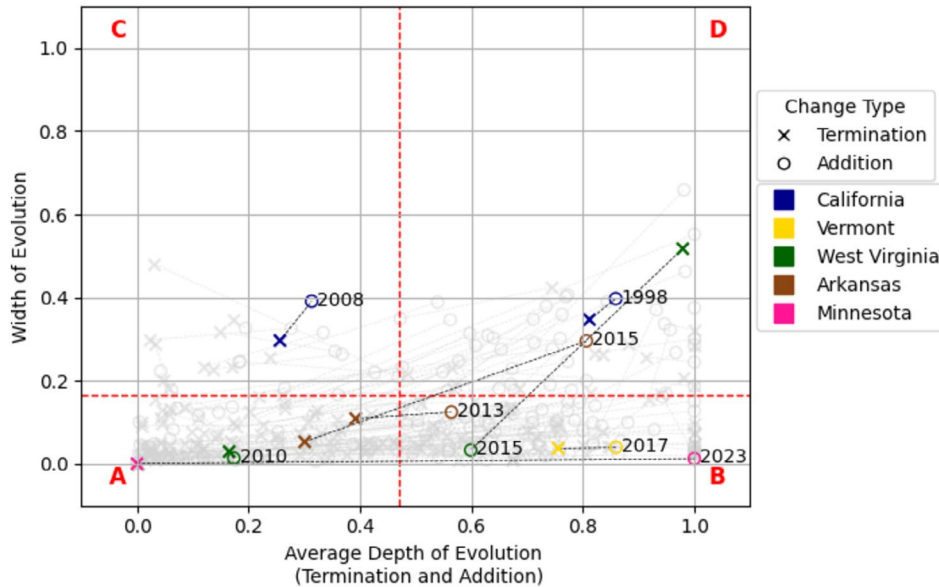


FIGURE 4 | Patterns of evolution using Figure 2 examples.

discussion in Appendix S5. In Figure 4, the quadrant letters from Table 1 are superimposed onto the graph. The dashed, red lines dividing quadrants are identified by calculating the mean of policy-level depth and width—both termination and additions—across our full sample. The colors of the markers correspond to the state, and the year is offered to the right of the addition (“o”) marker. Finally, while we focus here on the same eight examples as before, which have been strategically selected to illustrate a diversity of dynamics. We show values for the full sample in gray in the background to give readers a sense of these examples’ representativeness of our broader dataset.

First, as Table 3 and Figure 4 illustrate, different combinations of our measures of *depth* and *width of evolution* across terminations and additions signal different modes of policy change. Starting with the amendment passed in California in 1998, the high depth and broad width of changes occurring to both additions and terminations (D-to-D) are indicative of macro-level policy packaging commonly discussed in the literature. In the case of California 1998, Stokes (2015) discusses a systematic shift in policy design, as this packaging responded to two changes: (1) the “first policy expansion” (p. 207) requiring restructuring in the policy to account for newly eligible commercial customers and technology, and (2) the deregulation of the California

electricity market, requiring generalizable policy structures rather than utility-specific requirements (Taylor et al. 2007). This dynamic of full termination and full addition is supported by the statement-level representation of the policy in Figure 2, as well. When policymakers design sweeping overhauls, they add entire bundles of policy instruments, rules, and other policy design elements, often removing and replacing pre-existing provisions as well. These changes are not only important in terms of their magnitude (depth), but they also manifest across large swaths of the policy design (broad width). This signals deliberate attempts to amend the policy’s design, compared to the piecemeal adjustments we observe when policies are patched.

Policy packaging can also occur when changes in the text of policies have lower depth and width, but again, are represented by similar depth and width magnitudes across additions and terminations. For example, in 2008, Stokes (2015) discusses California’s change as targeted but diffuse, “expand[ing] net metering to allow local governments to participate” (p. 227). Yet, instead of adding a new section, the legislature restructured the definition and use of the policy’s target, moving from “distribution utility” to “electric service provider” to be inclusive of “local publicly owned electric utility” (Hess and Lee 2020). This meant the underlying design largely did not change in the net

metering context (low depth), but the concomitant change in reference to the updated target was diffuse (broad width) (C-to-C). This form of policy packaging occurs when a policy is altered given a new actor, entity, technology, etc., but the overall mechanism of the policy—and its individual statements—generally goes unaltered. In such cases, we expect to observe changes to specific names or phrases within current policy statements. Those changes are small, but occur in a diffuse way, everywhere these actors, entities, or technologies were previously mentioned throughout the policy.

In our dataset, almost a third of the amendments represent dynamics of policy packaging.⁸ The most common were combinations of low depth and narrow width of additions and terminations (A-to-A) and high depth and narrow width of additions and terminations (B-to-B)—akin to the “smart layering” referenced in literature. To illustrate, West Virginia 2010 represents the A-to-A dynamic as policymakers made minor adjustments to two statements regarding the definitions of “Advanced coal technology” and “Biologically derived fuel”. Vermont 2017 represents the B-to-B dynamic as policy makers update the reporting requirement of the Department of Energy to better determine “the current state of net metering in Vermont.” While representing different depths of change, both cases are defined by narrow widths of change. While rarely discussed in policy evolution literature that engages policy design—or broader net metering literature given its small and incremental nature—this concentrated evolution is common to policy process theories rooted in incrementalism, such as MSF and PET.

In contrast, mis-matched changes between additions and terminations, we suggest, indicate a different form of policy evolution: policy patching. Large changes concentrated in select areas of the policy—especially when these mainly consist of additions—may indicate that a new rule or instrument was incorporated without altering other sections of the policy. Arkansas is dominated by patching dynamics across its two example amendments: (1) low depth and narrow width termination paired with high depth and narrow width addition (A-to-B) in 2013; and (2) low depth and narrow width termination paired with high depth and broad width addition (A-to-D) in 2015. Both highlight an evolutionary design tendency to add new language without removing or altering older provisions (Ambrose et al. 2024; Stanton 2019)—that is, patching. While a more extreme version of A-to-B patching, the Minnesota 2023 example highlights targeted nature of policy evolution. In this case, one statement is added in full to the policy requiring utilities to provide usage data to customers upon request. While this changes the relationship between customers and utilities, the other statements remain non- or minimally-impacted. Furthermore, while these examples were chosen to represent these dynamics, A-to-B and A-to-D patching was common throughout our dataset. Practically, this represented the addition of new provisions for new technology or arrangements—for example, the addition of community solar required new infrastructure, actors, and relationships—to be regulated without substantial changes to other parts of the policy.

While the most extreme example of broad width and high depth termination to narrow and high depth addition (D-to-B) in our dataset, West Virginia 2015 represents an important patching

dynamic defined by policy termination. More specifically, 11 of the 12 policy subsections were fully terminated in 2015. This is captured by the high width of termination as well as the depth of termination near one, representing retrenchment qualitatively discussed elsewhere (Stokes 2015; Stokes and Breetz 2018). While prior literature discusses policy patching as the addition of new policy language without reference to current policy language, we argue that patching can also be driven by termination where large portions of the policy are terminated without the addition of new language regulating the policy environment.

5 | Discussion and Conclusion

Our measurement approach quantifies the depth and width of changes in policy text, and the (a)symmetry in these patterns between additions and terminations, advancing the field's efforts to understand and classify different dynamics of policy evolution and offering significant progress in empirically measuring these phenomena. Our proposed measurement approach was designed to be replicable across a wide range of policy domains, jurisdictions, and temporal periods. This could enable the policy process community to accumulate consistent evidence about phenomena of policy evolution and test theoretical propositions on a much larger scale. Furthermore, our approach adds nuance to traditional frames of policy patching and packaging, which tend to focus on A-to-D (patching—large diffuse additions and small concentrated terminations) and D-to-D (packaging—large diffuse additions and terminations) modes of evolution. While nuance in measurement is important, our net metering case shows that much of the policy evolution in our sample is incremental—that is, A-to-A, A-to-B, and B-to-B, see *Table B2 in Appendix S2*—evolution that may not be captured by traditional frames of policy patching and packaging. We nevertheless find instances of large punctuations, which we discuss in the qualitative validation of our measures (section 4). These correspond to D-to-D, A-to-D, C-to-C, and D-to-B patterns. Irrespective of the mode or dynamic, the proposed approach can help researchers highlight dynamics of change.

However, our approach does not interpret the meaning, significance, or practical implications of these changes. Thus, we offer a tool to aid research, but suggest our approach could be paired with qualitative analysis, as (1) changes in language do not always mean changes in practical responsibilities, relationships, or implementation, and (2) very small changes (e.g., changing the cap of net metering from 1% to 5%) can have large ramifications despite low depth or widths of change. Thus, while we argue our approach is important in and of itself for advancing policy evolution research (i.e., Figure 4 and Table 3), our statement-level text reuse algorithm can also help identifying which policy content has changed (i.e., Figure 1). For example, future research could combine our measurement approach with additional manual or computational coding of policy text to further examine how patching and packaging relate to the substantive evolution of policy content.

Finally, through this approach, future research can examine the broader research gaps of policy design at the micro-level highlighted by Capano and Howlett (2024) through medium- and larger-N studies, such as: (1) how does political conflicts and context relate to different patterns of policy evolution, and (2)

how do patterns of policy evolution link to behavioral changes in implementation?

Funding

The authors have nothing to report.

Ethics Statement

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The datasets were derived from sources in the public domain.

Endnotes

¹We use Howlett and Rayner's definition of policy design: "how specific types of policy tools or instruments are bundled or combined in a principled manner into policy 'portfolios' or 'mixes' to attain policy goals" (2013, p.172).

²The denominator of Depth of Addition and Termination are different for two reasons. First, the calculations as defined represents the desired conceptual representation. More specifically, if a statement is fully terminated, the depth of termination would be equal to 1 as the word count of the old statement is the same as the word count of the terminated text. Other specifications of the denominator e.g., word count of the new text or union of old and new statements (i.e., $Add\ WC + Reuse\ WC + Term\ WC$) would result in a specification that does not match this conceptualization. Second, the numerator can never be more than the denominator, as you cannot add more words than the new text nor terminate more words than the old text. This results in a specification that constrains the ratio between 0 and 1.

³We argue that averaging only statements that underwent change is conceptually important. Taking Minnesota 2023 as an example, one statement out of 88 total statements is observed to change. This statement is fully added (statement-level depth of addition is equal to 1) and is the only statement that underwent change, thus the policy-level depth of addition is also 1. If we take the depth of change at the policy-level as the average of all statements, the alternative score would be 0.011. While this might offer a representation of the total textual change in the policy, the depth of change score is (1) conflating extent of change (depth) with our concentration of change (width) and (2) as a result of this conflation, the depth score no longer narrowly measures the extent of change as it is now also influenced by how many statements are or are not undergoing change (i.e., width).

⁴To improve computational efficiency, the standard formula for the Gini coefficient was algebraically transformed into this mathematically equivalent representation. Appendix S4 provides the full derivation, alongside an example from our sample.

⁵This transformed unadjusted Gini equation yields a maximum of $1 - (1/n)$, approaching 1 only as $n \rightarrow \infty$. We therefore recommend that researchers use it on a sample of policies that have been parsed into a large n number of statements. When comparing width values across policies, researchers should ensure the policies contain a similar n number of statements. For example, when $n=100$, the maximum width value is 0.99. When using these measures on a small number of statements, researchers should normalize the width scores by dividing the unadjusted Gini by $1 - (1/n)$.

⁶While this would result in a calculation error as there is no X_{ijs} to sum, replacing these values with 0 is consistent with the logic of the width

measure. More specifically, when all statements go through the same amount of change, the width measure approaches 1. When change is concentrated to one statement, the width measure approaches but does not reach 0. Thus, as even less change is observed, i.e., no change, a value of 0 is offered. As constructed, when the width measure moves from 1 to 0, fewer statements are changed. Thus, when no statements are changed, to remain logically consistent, a Gini coefficient of 0 is assigned. Furthermore, because we rely on assessments of the (a)symmetry between additions and terminations, we cannot drop observations where either no language was added or terminated from the analysis, since those are needed for the subsequent analysis of patching and packaging.

⁷Summary statistics for our full sample are available in Appendix S2.

⁸Table B2 in the Appendix S2 shows the full distribution of combinations of quadrants in our data and maps them to concepts of patching and packaging.

References

- Ambrose, G., M. Gregoire-Zawilski, S. Siddiki, and N. Oesterling. 2024. "Understanding Policy Evolution Using Institutional Grammar: Net Metering Policies in the United States." *Policy Design and Practice* 7, no. 2: 227–249.
- Barnett, B., A. M. Wellstead, and M. Howlett. 2020. "The Evolution of Wisconsin's Woody Biofuel Policy: Policy Layering and Dismantling Through Dilution." *Energy Research & Social Science* 67: 101514.
- Baumgartner, F. R., and B. D. Jones. 1993. *Agendas and Instability in American Politics*. University of Chicago Press.
- Burgess, M., E. Giraudy, J. Katz-Samuels, et al. 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*, 57–66.
- Capano, G., and M. Howlett. 2020. "The Knowns and Unknowns of Policy Instrument Analysis: Policy Tools and the Current Research Agenda on Policy Mixes." *SAGE Open* 10, no. 1: 2158244019900568.
- Capano, G., and M. Howlett. 2024. "Calibration and Specification in Policy Practice: Micro-Dimensions of Policy Design." *Policy Design and Practice* 7, no. 2: 115–128.
- Capano, G., and I. Mukherjee. 2020. "Policy Design and Non-Design: Discerning the Content of Policy Packaging, Patching, Stretching and Layering." In *A Modern Guide to Public Policy*, 204–221. Edward Elgar Publishing.
- Cashore, B., and M. Howlett. 2007. "Punctuating Which Equilibrium? Understanding Thermostatic Policy Dynamics in Pacific Northwest Forestry." *American Journal of Political Science* 51, no. 3: 532–551.
- Fitch-Roy, O., D. Benson, and D. Monciardini. 2020. "Going Around in Circles? Conceptual Recycling, Patching and Policy Layering in the EU Circular Economy Package." *Environmental Politics* 29, no. 6: 983–1003.
- Guha, P., and I. Mukherjee. 2025. "Calibrations and Compliance: The Role of Motivations in Policy Instrument Design." *Policy and Society*: puaf028. <https://doi.org/10.1093/polsoc/puaf028>.
- Hess, D. J., and D. Lee. 2020. "Energy Decentralization in California and New York: Conflicts in the Politics of Shared Solar and Community Choice." *Renewable and Sustainable Energy Reviews* 121: 109716.
- Hinkle, R. K. 2015. "Into the Words: Using Statutory Text to Explore the Impact of Federal Courts on State Policy Diffusion." *American Journal of Political Science* 59, no. 4: 1002–1021.
- Hinterleitner, M., C. Knill, and Y. Steinebach. 2024. "The Growth of Policies, Rules, and Regulations: A Review of the Literature and Research Agenda." *Regulation & Governance* 18, no. 2: 637–654.
- Howlett, M., and B. Cashore. 2009. "The Dependent Variable Problem in the Study of Policy Change: Understanding Policy Change as a

- Methodological Problem.” *Journal of Comparative Policy Analysis* 11, no. 1: 33–46.
- Howlett, M., and C. Leong. 2022. “What Is Behavioral in Policy Studies?: How Far Has the Discipline Moved Beyond Traditional Utilitarianism?” *Journal of Behavioral Public Administration* 5, no. 1.
- Howlett, M., I. Mukherjee, and J. Rayner. 2018. “Understanding Policy Designs Over Time: Layering, Stretching, Patching and Packaging.” In *Routledge Handbook of Policy Design*, 136–144. Routledge.
- Howlett, M., and J. Rayner. 2007. “Design Principles for Policy Mixes: Cohesion and Coherence in ‘New Governance Arrangements.’” *Policy and Society* 26, no. 4: 1–18.
- Howlett, M., and J. Rayner. 2013. “Patching vs Packaging in Policy Formulation: Assessing Policy Portfolio Design.” *Politics and Governance* 1, no. 2: 170–182.
- Howlett, M. P., and I. Mukherjee. 2014. “Policy Design and Non-Design: Towards a Spectrum of Policy Formulation Types.” *Politics and Governance* 2, no. 2: 57–71.
- Kern, F., P. Kivimaa, and M. Martiskainen. 2017. “Policy Packaging or Policy Patching? The Development of Complex Energy Efficiency Policy Mixes.” *Energy Research & Social Science* 23: 11–25.
- Lindblom, C. E. 1959. “The Science of ‘Muddling Through.’” *Public Administration Review* 19, no. 2: 79–88.
- Linder, F., B. Desmarais, M. Burgess, and E. Giraudy. 2020. “Text as Policy: Measuring Policy Similarity Through Bill Text Reuse.” *Policy Studies Journal* 48, no. 2: 546–574.
- Lowi, T. J. 1964. “American Business, Public Policy, Case-Studies, and Political Theory.” *World Politics* 16, no. 4: 677–715.
- Mahoney, J., and K. Thelen. 2009. *Explaining Institutional Change: Ambiguity, Agency and Power*. Cambridge University Press.
- Nair, S., and M. Howlett. 2020. “Policy Myopia as a Source of Policy Failure: Adaptation and Policy Learning Under Deep Uncertainty.” In *Policy Learning and Policy Failure*, 133–154. Policy Press.
- Pagliari, S., and K. L. Young. 2020. “Exploring Information Exchange Among Interest Groups: A Text-Reuse Approach.” *Journal of European Public Policy* 27, no. 11: 1698–1717.
- Rayner, J., and M. Howlett. 2009. “Introduction: Understanding Integrated Policy Strategies and Their Evolution.” *Policy and Society* 28, no. 2: 99–109.
- Scott, T. A., N. Marantz, and N. Ulibarri. 2022. “Use of Boilerplate Language in Regulatory Documents: Evidence From Environmental Impact Statements.” *Journal of Public Administration Research and Theory* 32, no. 3: 576–590.
- Stanton, T. 2019. *Review of State Net Energy Metering and Successor Rate Designs*. National Regulatory Research Institute.
- Stokes, L. C. 2015. “Power Politics: Renewable Energy Policy Change in US States.” Doctoral dissertation, Massachusetts Institute of Technology.
- Stokes, L. C., and H. L. Breetz. 2018. “Politics in the US Energy Transition: Case Studies of Solar, Wind, Biofuels and Electric Vehicles Policy.” *Energy Policy* 113: 76–86.
- Taylor, M., G. Nemet, M. Colvin, L. Begley, C. Wadia, and T. Dillavou. 2007. *Government Actions and Innovation in Clean Energy Technologies: The Cases of Photovoltaic Cells, Solar Thermal Electric Power, and Solar Water Heating: PIER Project Report*. California Energy Commission.
- Wellstead, A., J. Rayner, and M. Howlett. 2016. “Alberta’s Oil Sands Reclamation Policy Trajectory: The Role of Tense Layering, Policy Stretching, and Policy Patching in Long-Term Policy Dynamics.” *Journal of Environmental Planning and Management* 59, no. 10: 1873–1890.
- Wilkerson, J., D. Smith, and N. Stramp. 2015. “Tracing the Flow of Policy Ideas in Legislatures: A Text Reuse Approach.” *American Journal of Political Science* 59, no. 4: 943–956.
- Wilson, J. Q. 1995. *Political Organizations*. Princeton University Press.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Appendix S1:** Data construction. **Appendix S2:** Summary statistics. **Appendix S3:** Figure 2 Underlying data. **Appendix S4:** Gini coefficient. **Appendix S5:** Qualitative validation of the measure.