

Presidential Policymaking, 1877–2020*

Aaron R. Kaufman[†]
New York University, Abu Dhabi

Jon C. Rogowski[‡]
University of Chicago

June 18, 2021

Abstract

Contemporary presidents frequently make new law and create new policy outcomes through unilateral power. Empirical assessments of the unilateral presidency, however, generally draw from a single category of directives—executive orders—issued over relatively short periods. We offer the most comprehensive assessment of presidential unilateralism to date by introducing new data on more than 50,000 unilateral directives issued between 1877 and 2020 and using machine learning techniques to distinguish their substantive importance and associated issue areas. We show that the rate of significant unilateral activity has increased markedly over the last 30 years, particularly in foreign affairs and often through the substitution of memoranda for executive orders. Finally, we use our measures to formally evaluate the historical and temporal development of the unilateral presidency. Our research provides new evidence about variation in and the nature of the exercise of presidential authority and opens new avenues for empirical inquiry.

Word count: 11,729

*The Faculty of Arts and Sciences at Harvard University provided research support for this project. We thank Joshua Berry, Vivian Fernandez, Joshua Florence, Jeffrey Gu, Darwin Peng, Sapna Rampersaud, Enrique Rodriguez, Corban Ryan, Michael Scherr, Grant Schwartz, Melvin Woo, Jenny Yoon, and Andrew Zucker for excellent research assistance. We are also grateful to Avidit Acharya, Traci Burch, Brandice Canes-Wrone, Dan Carpenter, Jamie Druckman, Dan Galvin, Elisabeth Gerber, Sandy Gordon, Laurel Harbridge, Greg Huber, Jeff Jenkins, Doug Kriner, Kenneth Lowande, Greg Martin, Nolan McCarty, Julie Lee Merseeth, Pablo Montagnes, Daniel Moskowitz, Anne Joseph O’Connell, Yu Ouyang, Paul Peterson, Pia Raffler, Andrew Reeves, Andrew Rudalevige, Josh Ryan, Maya Sen, Chuck Shipan, Chris Skovron, Arthur Spirling, Sharece Thrower, Chloe Thurston, Robert Ward, Hye Young You, and participants at the 2017 Annual Meeting of the American Political Science Association, the American politics workshop at Northwestern University, the Political Institutions and Political Economy Conference at the University of Southern California, the Emory Political Economy Conference, the Presidency and the Administrative State workshop at Princeton, the Center for the Study of American Politics at Yale University, and the Political Science Speaker Series at Utah State University for generous and helpful feedback.

[†] Assistant Professor, Division of Social Sciences, Saadiyat Island, Abu Dhabi, UAE ; aaronkaufman@nyu.edu.

[‡] Associate Professor, Department of Political Science, 5828 S. University Ave, Chicago, IL 60637; jrogowski@uchicago.edu.

Modern presidents play more active and central roles in policymaking relative to their nineteenth century predecessors. Not only do contemporary presidents seek to influence legislation passed by Congress (e.g., Cameron 2000; Canes-Wrone 2006; Cohen 2012; Rudalevige 2002), but they also rely increasingly on unilateral authority to make policies on their own (Howell 2003; Krause and Cohen 2000; Mayer 2001). As gridlock and dysfunction increasingly hinder the legislative process in Congress (Binder 2015), presidential policymaking gains increased salience as a means of policy production by the national government.

In this paper, we provide the most comprehensive investigation to date on unilateral policymaking by American presidents. At its core, scholarship on unilateral action addresses a common measurement question: how much policy do presidents create through unilateral power, and in what domains? While previous research makes important contributions to our understanding of unilateral action, its evidence often draws from only a single category of unilateral directives—executive orders—and from relatively short periods of history. Empirical progress requires richer, more complete measures of unilateral activity that span longer periods of time.

Our study makes three primary contributions to empirical scholarship on unilateral action, the presidency, and policymaking. First, we introduce new data on more than 50,000 presidential directives across a comprehensive range of unilateral instruments. While existing quantitative research on unilateral action focuses largely on executive orders (e.g., Bolton and Thrower 2016; Chiou and Rothenberg 2017; Howell 2003; Warber 2006), presidents exercise unilateral powers through a variety of tools (e.g., Bailey and Rottinghaus 2013; Cooper 2014; Williams 2020). The omission of directives beyond executive orders provides an incomplete description of presidents' use of unilateral powers. Analyses based solely on one type of directive may also produce inaccurate inferences about the predictors of unilateral activity, particularly if presidents are strategic about the choice of instrument (see Lowande 2014).

Second, we study unilateral activity across more than a century, from the end of Reconstruction through the presidency of Donald Trump. The findings from existing scholarship are drawn

largely from the post-WWII era (exceptions include Bailey and Rottinghaus 2013; Bolton and Thrower 2016; Williams 2020). While a substantial literature documents historical variation in the production of legislation (e.g., Clinton and Lapinski 2006) and presidential bargaining with Congress (e.g., Cohen 2012), historical patterns of president-driven policymaking are less well understood (Lowande and Rogowski 2021). These limitations inhibit efforts to characterize the scope and nature of unilateral policymaking across time and to identify the features of the political environment associated with its use.

Third, we use advances in machine learning techniques to address measurement challenges in evaluating both the substantive importance and issue domains of presidential directives. Just as legislation (Clinton and Lapinski 2006) and government regulations (Zubek, Dasgupta, and Doyle 2021) vary in substantive importance, so too do unilateral directives. A large share of unilateral actions has trivial policy impact, and conflating important and trivial directives presents challenges for evaluating theories of unilateral action. Moreover, understanding the president's contributions to public policy requires distinguishing the substantive issues for which unilateral power is used (for related arguments, see Katznelson and Lapinski 2006). We use a supervised learning approach for estimating directive significance and policy domain using the text of the directives by building upon techniques used in related domains (e.g., Chiou and Rothenberg 2017; Christenson and Kriner 2020a; Zubek, Dasgupta, and Doyle 2021; Gerrish and Blei 2011; Spirling 2012). Our framework can be applied when studying policy-relevant activity by executives, legislatures, bureaucracies, and courts across countries and levels of government.

We document new empirical patterns and historical variation in the frequency and nature of presidential policymaking over more than a century. Notably, the last four presidents issued significantly greater numbers of unilateral directives than in virtually any other period since Reconstruction. The recent increase is driven by presidents' greater reliance on instruments other than executive orders, particularly presidential memoranda, and especially in the domains of international affairs and national defense. Using these data to study historical change in presidents'

use of unilateral powers (following Wawro and Katznelson 2014), we find mixed evidence in support of historical accounts of changes in the unilateral presidency. Both the Progressive Era and the contemporary period exhibited sharp changes in presidents' use of unilateral powers, while neither the Franklin Roosevelt administration nor the Cold War era was associated with clear increases in unilateral activity. Finally, we show that while presidents have increasingly relied on executive orders when making important policies concerning domestic affairs, they have also increasingly substituted memoranda for proclamations and executive orders when making important unilateral policies in foreign affairs. We conclude by discussing other uses for our measures and applications of our approach. Our research provides the fullest characterization to date on the exercise of presidential authority and opens new avenues for empirical inquiry.

Unilateral Power and the Presidency

All presidents have used unilateral power to make new law on their own. Presidents use unilateral powers “to implement many of their most important policy initiatives, basing them on any combination of constitutional and statutory power that is thought to be available” (Shane and Bruff 1996, 131). Unilateral directives, such as executive orders and proclamations, have wide impact: they can reorganize executive branch agencies or create new ones, enforce racial and gender equality, declare national emergencies, announce holidays and celebrations, and make new rules governing the implementation of existing policies. As Cooper (2014, ix) has observed, “There is virtually no significant policy area in which presidents operate that has not been shaped to one degree or another by the use or abuse of [unilateral] tools.”

The appeal of unilateral action for presidents is self-evident. By using unilateral power to set new policies and direct activity within the executive branch, presidents can advance their political goals without expending political capital to rally support for legislative proposals. Perhaps most importantly, unilateral action may enable presidents to achieve policy outcomes more

to their liking than a similar legislative initiative, or to enact policies that could not be passed in Congress (Howell 2003). The development and deployment of unilateral power, according to Moe and Howell (1999, 133), “virtually defines what is distinctly modern about the modern American presidency.” Establishing the merits of these claims and the consequences of unilateral power ultimately is a measurement problem: how much policy do presidents create through unilateral means and in what domains?

Empirical Approaches to Studying Unilateral Action

Most empirical research on unilateral action takes a “macro-politics”¹ approach and evaluates temporal variation in unilateral activity. Some research in this area emphasizes the growth of unilateral power across secular time (e.g., Hebe 1972; Krause and Cohen 2000). As Skowronek (1982, 175-6) argues, “The postwar consolidation of the administrative advances of the Progressive era forged a new institutional politics in American national government, a politics organized around administrative power and a stalemate of constitutional controls.” As the president’s expectations increased over time, so did “his incentive to tap the alternate source of supply deriving from his position as head of the federal bureaucracy” (Kagan 2001, 2312). These theoretical accounts make measurable claims about temporal change in presidential unilateral activity.

Other scholarship studies how unilateralism varies with institutional and political context. Separation-of-powers-based accounts argue that presidents issue unilateral directives subject to the configuration of legislators’ preferences and the courts’ willingness to uphold a president’s action (Bolton and Thrower 2016; Howell 2003; Moe and Howell 1999). To evaluate this perspective, scholars have studied how annual or biennial patterns of unilateral activity covary with congressional and judicial preferences (e.g., Bolton and Thrower 2016; Chiou and Rothenberg 2017; Howell 2003; Lowande 2014; Mayer 2001; Thrower 2017; Warber 2006). Other scholarship studies

¹This term borrows from Cameron’s (2009) discussion of research on presidential vetoes.

how the volume of unilateral activity correlates with presidential approval ratings (Christenson and Kriner 2020*b*) and issues of public concern (Rogowski 2019).² Customarily, this literature regresses the production of unilateral action in some time period on a set of institutional and political covariates.

Empirical Challenges in the Study of Unilateral Action

Empirical scholarship that studies temporal patterns of presidential unilateralism is limited in four key ways. First, though presidents exercise unilateral powers through a variety of means, studies have focused overwhelmingly on a single kind of directive: executive orders (e.g., Bolton and Thrower 2016; Chiou and Rothenberg 2017; Christenson and Kriner 2020*b*; Howell 2003; Krause and Cohen 2000; Mayer 2001; Rogowski 2019; Thrower 2017; Warber 2006).³ The omission of other directives raises two major concerns. One, if descriptive claims about presidents' use of unilateral power rely on an incomplete or unrepresentative data source, they may produce biased estimates of the frequency of unilateralism. Two, to the extent that presidents have political incentives to substitute a salient type of directive for an obscure one (or vice versa; see, e.g., Lowande 2014), studies using only executive orders ignore presidents' strategic behavior, again producing biased inferences.⁴

²In an important exception, Lowande (Forthcoming) evaluates how presidents modify status quo policies through unilateral action rather than models the rate of unilateral activity.

³Some scholars have studied other unilateral tools such as proclamations (Bailey and Rottinghaus 2013; Rottinghaus and Maier 2007) or memoranda (Lowande 2014) but considers those directives in isolation from the president's wider unilateral toolkit. Williams (2020) offers the most comprehensive comparison of executive orders and proclamations across time but does not incorporate other kinds of directives.

⁴Temporal variation in presidents' proclivities to use different directives for creating unilateral policy creates a similar problem. For instance, Bailey and Rottinghaus (2013, 189) characterize the

Because the president’s unilateral toolkit has expanded over time and the interpretation of particular forms of unilateral action has evolved along with it, these concerns may be particularly acute for studies that span several decades. According to Howell (2005, 417), for example, “almost all the trend lines point upward” in presidents’ use of unilateral powers, and he concludes that “[t]hough the total number of executive orders has declined, presidents issued almost four times as many ‘significant’ orders in the second half of the twentieth century as they did in the first.” Yet without data on directives beyond executive orders, it is difficult to know how confident one should be in accepting Howell’s conclusions. Incorporating a wider range of unilateral directives also addresses the observation from Durant (2009, 90) that “scholars have typically examined individual tools of the administrative presidency rather than the interaction of these tools over time as the implementation of presidential goals and priorities shifts to fit changing circumstances.”

Second, existing scholarship is focused largely on the contemporary—i.e., post-World War II—era.⁵ This focus has several consequences. One, without comprehensive data across long periods of time, it is difficult to evaluate claims about long-term trends and developments in the use of unilateral powers. Two, given that most studies model unilateral directives at the annual (e.g., Bolton and Thrower 2016) or biennial (e.g., Chiou and Rothenberg 2017) level, relatively short time series risk providing insufficient statistical power for evaluating the relationships of interest, such as the relationship between divided government and unilateral activity. A longer first third of the twentieth century as the “golden age of policy making by presidential proclamation” due to presidents’ reliance on proclamations to create substantive policies in that era. Studies of unilateral activity during this period that omit proclamations thus provide an incomplete portrait of presidential policymaking analogous to non-random missing data.

⁵Three important exceptions are Bailey and Rottinghaus (2013), who study proclamations from Washington through the end of the nineteenth century; Bolton and Thrower (2016), who study executive orders from 1906 to 2013; and Williams (2020), who studies proclamations and executive orders from the mid-nineteenth through the mid-twentieth centuries.

time series of presidential unilateral activity would strengthen the quality of inferences based on variables whose values change slowly or infrequently across time and allow for more in-depth investigation of historical changes in unilateral activity.

Third, not all unilateral directives are of comparable substantive import. A large share of unilateral actions makes ceremonial announcements or addresses routine administrative affairs. Just as it may be suspect to test theories of lawmaking using samples of trivial legislation (Clinton and Lapinski 2006), proper tests of theories of unilateral action should focus on substantively important directives. Existing work uses media mention of unilateral directives (Chiou and Rothenberg 2017; Christenson and Kriner 2020*a*; Howell 2003; Williams 2020), focuses on non-ceremonial orders (Bolton and Thrower 2016), or relies upon expert judgments (Warber, Ouyang, and Waterman 2018), yet the field has not reached consensus about how (or whether) to extend these approaches across directive type and time (Lowande and Rogowski 2021). More generally, the field would benefit from a measurement approach that provides directive-level measures of policy significance that avoids collapsing significant and nonsignificant directives or assuming that all significant directives are similarly important.

Fourth, and finally, existing research provides no comprehensive inventory of the substantive policy areas presidents have addressed through unilateral action. While the Policy Agendas Project (2021) characterizes the issue areas of each executive order issued since 1945, these issue codings are not available for a wider period of time or for directives beyond executive orders. Just as a fuller understanding of legislative representation requires evaluating the substantive outputs produced by Congress (Katznelson and Lapinski 2006), identifying presidents' contributions to substantive policy outcomes requires a more complete accounting of the domains in which use unilateral power.

We address each of these limitations by introducing new data on unilateral activity from 1877 to 2020 and a measurement strategy for characterizing the policy significance and issue area of each directive. As we describe below in greater detail, our strategy incorporates estimates from

previous work by Chiou and Rothenberg (2017) to measure the importance of unilateral directives. We leverage these measures, building a supervised machine learning model to estimate the significance of tens of thousands of new directives using their unstructured text content. We use a similar approach to estimate the substantive policy area of each directive, facilitating easy integration with the Policy Agendas Project. While our approach is focused on presidential unilateral action, it is broadly applicable across domains and could easily be extended to measure the substantive importance of other documents including legislation, bureaucratic regulations, official communications, and judicial decisions, and to classify other dimensions such as sentiment.

Data and Measurement

We introduce a new dataset that contains the most extensive inventory of presidential directives assembled to date. These data are obtained from the *CIS Index to Presidential Executive Orders & Proclamations 1987 (CIS)* and have been extended through 2020 by *ProQuest Legislative & Executive Publications*. This database contains “a complete collection of numbered and unnumbered Executive Orders and Presidential Proclamations” including approximately 100,000 presidential actions since 1789. We draw upon the full text of each of these documents.⁶

These data provide a fuller characterization of unilateral directives than existing databases. The data include documents and directives that report presidential orders or determinations but lack formal designation and thus are routinely excluded from existing studies. As the *CIS Index* notes (1987, ix), “various categories of Executive Orders and Proclamations were routinely excluded from the Numbered Series despite their similarity to materials included in the series.”⁷

⁶These documents are available as PDFs. Most are already text-extracted, but for those that are not, we transcribe them using an optical character recognition system.

⁷These documents were found in: presidential papers and collections, files of correspondence and orders maintained by Federal departments and agencies, and “other Presidential documents”

These directives functioned as “a form of executive lawmaking used in instances in which the Constitution or Congress directly or indirectly permits the President to take action” (*CIS Index to Presidential Executive Orders and Proclamations* 1987, vii). The inclusion of these documents appears conceptually appropriate; after all, no existing theory of unilateral action implicates formally-designated executive orders to the exclusion of other instances of presidentially-initiated policymaking.

Our dataset consists of 53,055 unilateral directives issued between 1877 and 2020. The documents in our dataset represent either an original document declaring a presidential action or a presidential message which contains evidence of presidential action.⁸ Documents that are not of policy consequence and/or address matters other than executive policymaking (such as pardons, nominations, routine remarks, and Statements of Administration Policy) are omitted. We classified the remaining directives into three groups: *Executive Orders*, *Proclamations*, and *Memoranda*, which correspond roughly to differences in the president’s authority, legal status, and intended audience of each.⁹ While executive orders and proclamations both carry the force of law, executive orders are generally directed to actors inside government while proclamations tend to be directed to actors outside U.S. government (Rottinghaus and Maier 2007, 339). Memoranda are similar to executive orders in that they provide instructions to officials within government; however, they are not subject to the same reporting requirements as proclamations and executive orders, and their legal status is somewhat more ambiguous (Lowande 2014, 725). As Cooper (2014, 120) writes, “[a]s a practical matter, the memorandum is now being used as the equivalent of an executive order.” Our use of the term “Memoranda” is not intended to indicate that documents were formally designated as such, but instead is in the spirit of related classifications of

contained in the Federal Register (1987, ix-x).

⁸The CIS cross-checked documents to eliminate duplication.

⁹A complete accounting of the document categories included in the *CIS Index* and our justifications for their categorization is shown in Appendix A.

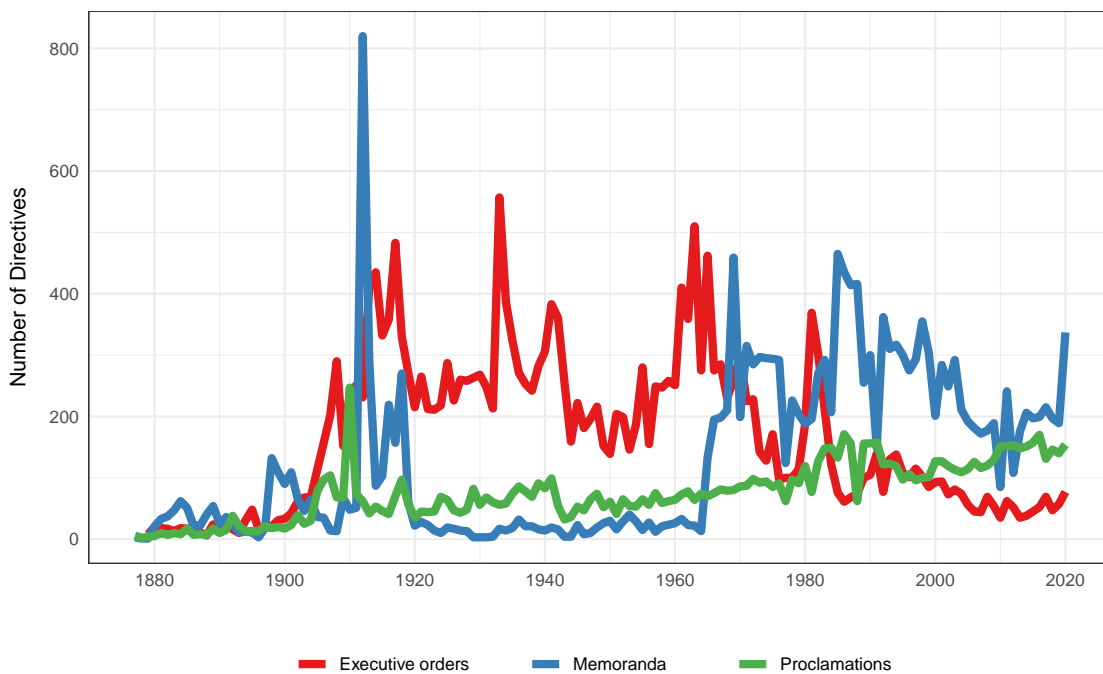
“memo orders” (Woolley and Peters 2017) which distinguish administrative actions directed by the president though means other than executive orders.

Perhaps most importantly, the distinctions between document categories are not clear-cut. As the *CIS Index* (1987, viii) reports, “Although technically, Executive Orders differ from Presidential Proclamations, this distinction has been imperfectly observed over the years. There are items in each category that seem to belong in the other.” Likewise, the Congressional Research Service (2021, 21) argues that “any distinction among these instruments—executive orders, presidential memoranda, and proclamations—is muddled by the fact that all three may be employed to direct and govern the actions of government officials and agencies.” Presidents have made use of different unilateral tools to implement similar policies. In the Trump administration, for instance, the “travel ban” restrictions on entry to the United States from citizens of certain countries were first implemented via Executive Order 13769 and later with Executive Order 13780, which was then replaced by Proclamation 9645. Moreover, while proclamations are often characterized merely as symbolic or ceremonial, this designation is more true for the modern presidency than for earlier periods of American history (Bailey and Rottinghaus 2013), and many proclamations are specifically authorized by statute, particularly for certain kinds of issues (Cooper 2014). Similarly, many high-profile executive actions have been issued in forms other than either executive orders or proclamations (Lowande 2014). For example, when President Obama took to the Rose Garden on June 15, 2012 to announce the Deferred Action for Childhood Arrivals program, his remarks did not mention a new executive order. Instead, the program was implemented at his urging via Homeland Security memorandum (and is included in our data as a memoranda). Our classification of directives is thus intended to provide a simple way for comparing patterns of presidential policymaking across unilateral tools but distinctions between them are more fluid than fixed.

Figure 1 displays the number of documents in each category from 1877 through 2020. The figure reveals several substantively interesting patterns. Prior to the twentieth century, presi-

dents issued unilateral directives infrequently. Yet since the turn of the twentieth century, the number of directives has not only increased but also has varied across time and directive type. Interestingly, while executive orders may be the most publicly salient form of unilateral actions, presidents quite often make more frequent use of other unilateral tools. Reflecting the findings reported by Howell (2003), by the mid-1980s presidents made less frequent use of executive orders than they had earlier in the century. While there was a spike in the number of proclamations early in the twentieth century, their use has slowly yet consistently trended upwards. Descriptively, the data indicate that executive orders alone do not provide a full picture of presidents' use of unilateral action. If we were to consider only executive orders, we would conclude that unilateralism has decreased since the 1980s; but if we were to consider proclamations and memoranda as well, we reach the opposite conclusion. Understanding unilateral action in the modern presidency requires considering other directives, including memoranda and proclamations, in conjunction with executive orders.

Figure 1: Annual Directives by Category, 1877 to 2020



Measuring the Significance and Policy Domains of Unilateral Action

We employ a text-based approach to estimate the substantive significance and policy domains of the unilateral actions displayed in Figure 1 (see, e.g., Kaufman 2020). Our approach extends existing approaches to measuring the significance of presidential actions in several ways. First, while previous work relies heavily on media coverage to assess substantive importance (Chiou and Rothenberg 2017; Christenson and Kriner 2020a; Howell 2003; Williams 2020), our estimates incorporate the directives' text.¹⁰ While media mentions may be correlated with policy significance, the correlation likely is noisy and variable over time, and in ways that may be difficult to discern. For instance, it is unclear whether a "significant" directive that was issued in 2013 and covered by the *New York Times* would have received the same media coverage if it had instead been issued in 1913. As we describe below, while we make use of previous media-based indicators of directive significance, our approach attempts to evaluate directive significance within a relatively fixed media environment. To a large degree, this allows us to make comparisons of important directives across time that are not simply a function of changing levels of media interest or scrutiny.¹¹

Second, our approach addresses potential measurement error that may arise from media-

¹⁰Similarly, some research distinguishes directives based on their content rather than their visibility (Fine and Warber 2012; Warber 2006).

¹¹Consider, for instance, that the *New York Times* covered as many executive actions during the first year of the Trump presidency as it did during the entire first term of the Obama presidency (Christenson and Kriner 2020b, 12). Yet President Trump dominated news coverage to a considerably greater degree than his predecessors, at least during his first 100 days in office (e.g., Patterson 2017). Thus it is not clear whether the number of Trump executive actions covered by the media reflected the substantive significance of those directives or instead the more intense focus given to the Trump administration.

based approaches. Some actions are ceremonial or symbolic in their effects but may have popular interest and subsequently receive substantial media attention. As an example, President Trump signed Executive Order 13765 on his first day in office to announce his intentions to repeal the Affordable Care Act, which received media coverage despite having no direct policy effects. Other actions, though, are substantively important but may be less salient and/or highly technical and do not receive media attention. For instance, Franklin Roosevelt’s Executive Order 8807, which created the Office of Scientific Research and Development and facilitated the Manhattan Project, went unmentioned by the *New York Times* in 1941 despite its coverage of other Roosevelt executive orders. Thus, media-based indicators of significance may not meaningfully distinguish directive importance, and they particularly may not do so along a continuous scale. We use coarsening to identify significant actions above a given threshold of significance (and without making inferences about cardinality). Under this approach, we will incorrectly identify a ceremonial document as significant only if textually-similar ceremonial executive orders received so much media attention that they move from below our significance threshold to above it, and we will fail to recognize a significant document as such only if textually-similar executive orders received far less media attention than their significance warranted. This reduces measurement error considerably vis-à-vis other media-based approaches (e.g., Chiou and Rothenberg 2017; Howell 2003). Our approach further relaxes the assumption that all equally significant unilateral actions are likely to receive equivalent media attention, as levels of popular interest may vary with the policy in question, and the nature of the media environment may shape uptake of particular actions.

Third, our approach is easily scalable and can evaluate the text of tens of thousands of presidential documents automatically and cheaply. While media-based indicators require fresh data collection on media coverage after each new directive is issued, our approach requires only that we have access to the text of the directive.¹² We can thus apply our approach backwards and

¹²As we discuss below, we consider whether changes in language over time might degrade the quality of our measure. For example, if our model learns the relationship between policy signifi-

forwards through time as new directives are issued and/or discovered. Finally, by using machine learning rather than relying solely on human coders, we avoid potential biases associated with subjective evaluations (see Katagiri and Min 2019).

We use a supervised learning approach (e.g., Gentzkow and Shapiro 2010; Hopkins and King 2010) to estimate the substantive significance and policy domains of each action.¹³ Following standard practices in the literature, our procedure consists of six steps: (1) Collect a training corpus which differs along the dimension of interest, (2) label each document in the corpus corresponding to its location along the dimension of interest, (3) convert the corpus to a data set, (4) train a supervised model on the data set and training labels, (5) use the model to predict the labels for out-of-sample documents, which are the ultimate quantity of interest, and (6) examine the results for face validity, internal validity, and external validity. We list these steps in Table 1 and briefly describe each. We provide more extensive documentation in Appendix B.

Training Data. We begin with the CIS *Index* to Presidential Executive Orders and Proclamations, as described above, which includes the text of each directive. This data set includes such important directives as the Japanese Internment executive order (EO 9066) and documents as mundane as Lyndon Johnson appointing four new members to the Committee on the National Medal of Science.

Document Labeling. We create initial training labels for both substantive significance and policy domains. To develop substantive significance labels, we take three complementary ap-

cance and 1940s policy writing, it may extrapolate less well to Trump-era orders. We discuss this concern and our approach to address it in Appendix B.

¹³Supervised learning requires labeled example documents from which to learn a relationship between words and a label. Tools like regularized regression, support vector machines, random forests, and neural networks then use those examples to estimate a complex functional form.

proaches. For numbered executive orders from 1947 to 2002, we match significance estimates from Chiou and Rothenberg (2017) to the text of the corresponding executive order.¹⁴ The estimates from Chiou and Rothenberg (2017) are continuous from roughly -1 to 3; we coarsen them to create a binary indicator of significance at the threshold of 0.5.¹⁵ Secondly, we use trained undergraduate research assistants¹⁶ to triple-code each numbered proclamation in a dichotomous fashion as either “substantive” or “ceremonial” and aggregate their results for each directive. Finally, we manually identify a set of non-numbered proclamations, executive orders, and other documents that are related to or explicitly numbered executive orders and assign to these new documents a significance score equal to the average significance of all numbered executive orders to which it is related. These 11,183 documents matched to their significance estimates constitutes the training set; the remaining 41,872 documents comprise our test set.

To develop labels for policy domain, we combine two approaches: for numbered executive orders issued between 1945 and 2020, we rely on the Comparative Agendas Project’s policy labels.¹⁷ These labels associate each executive order to one of 20 policy areas, ranging from agriculture and the environment to national defense and government operations. For a random sample of

¹⁴Chiou and Rothenberg (2017) do not study executive orders beyond these dates.

¹⁵Here and in one other step in our measurement strategy we coarsen a continuous score to produce a dichotomous measure of significance. While we perform a series of robustness checks on these choices (see Appendix B), we acknowledge that readers may disagree with our choices. We suspect that alternative measurement choices would have minimal effects on our substantive findings. In the interest of transparency, however, upon publication we will publish our full replication file for generating estimates of document significance on the journal Dataverse, indicating where in the code we make these choices, so that curious readers can explore alternative threshold choices.

¹⁶Research assistants all performed strongly in an undergraduate course on the U.S. presidency.

¹⁷See <https://www.comparativeagendas.net/pages/master-codebook>.

the remaining documents, we use trained research assistants to label documents according to the same coding scheme. Altogether we have policy domain training labels for 21,429 documents.

Text to Data. We convert the training set and test set into a data object that can be analyzed quantitatively by using the documents in our corpora to create term-document matrices. Each document is a row in a matrix, and each unique lexical feature is represented in a column. Entries in this matrix indicate how many times lexical object j occurs in document i . In this case, columns of the data set include unigrams, which are single words, and bigrams, which are ordered pairs of words. Consider Executive Order 13123, the Preamble of which begins: “The Federal Government, as the Nation’s largest energy consumer, shall significantly improve its energy management in order to save taxpayer dollars and reduce emissions that contribute to air pollution and global climate change.” The unigrams in this document include, among others, “energy”, “consumer”, “management”, “save”, “dollars”, “emissions”, “pollution”, and “climate”; bigrams include “energy consumer”, “energy management”, “taxpayer dollars”, “air pollution”, and “climate change.” As additional preprocessing, we remove all terms which do not occur in at least 500 documents. In total, our significance training term-document matrix (TDM) has 11,190 unique documents and 31,214 terms. Our topic area TDM has 21,429 documents and 31,214 terms.

Modeling. Next, we apply standard machine learning tools to model the relationship between lexical features and document significance or policy domain. We are methodologically and theoretically agnostic as to which model will work best, so we test a variety and measure their success using k -fold cross-validation (Mozer et al. 2020). This procedure involves partitioning a training set into a number of non-overlapping random subsamples, training a model on all but one of them, predicting the outcome measure for the omitted subsample, and comparing the model’s predictions to the true outcome labels. Stronger correlations with the known estimates generate greater confidence in its predictions for the training set.

We find that the random forest model performs best. It successfully identifies whether a doc-

ument is below or above our significance threshold 98.76% of the time, with little observable heterogeneity across significance levels and with roughly equal false-positive and false-negative rates (for more details, see Appendix B). This provides confidence that although we are estimating documents' significance with error, that error is unlikely to produce bias in our substantive results.

In producing a policy area estimation model, we confront a class imbalance problem: since we have relatively many training documents belonging to certain policy areas like Government Operations and relatively few belonging to others (e.g. Civil Rights), we risk misclassifying documents at different rates. To address this concern, we rebalance our policy area training set using Adaptive Synthetic Sampling (He et al. 2008).

Estimation. Finally, we use the random forest model to estimate the both policy significance and policy domain for the documents in our test set (those without training labels). The significance random forest model produces for each document in the test set a *probability* that it is significant;¹⁸ the policy domain random forest model produces each document's most likely policy category.

We distinguish significant unilateral actions as those whose scores are greater than 0.355, the value that equalizes the false-positive and false-negative rates. This threshold distinguishes 8,544, or roughly 16 percent, of the directives in our data as policy significant. This appears to represent an appropriate level of selectivity given related scholarship. To contextualize this figure, Howell (2005) characterized 8% of executive orders (290 of 3,749) issued between 1945 and 2001 as significant, Bolton and Thrower's (2016) study of "nonceremonial" orders included 97% of executive orders issued between 1945 and 2013 (4,045 of 4,153), and Mayer and Price (2002) classified 14% of a sample of executive orders issued between 1936 and 1999 as significant. Finally, Chiou and Rothenberg (2017) characterize 14%, 26%, and 44% of executive orders as significant

¹⁸Figure A.1 displays the distribution of significance scores across document types.

based on thresholds of 1, 0.5, and 0, respectively.¹⁹ Overall, our threshold appears to exhibit greater selectivity relative to most existing research. However, because our scores are measured with error, we do not interpret them in a cardinal way.

Relabeling. We improve the quality of our supervised learning models using Active (or Adaptive) Learning (Settles 2012; Kaufman and Klevs Forthcoming). In this procedure, we examine the predicted scores of a carefully-selected subset of the test set, relabel those documents if their predicted scores are incorrect, add them and their corrected labels to the training set, and then repeat the Model Training and Estimation steps. We follow this approach for our policy significance model by identifying 250 randomly-selected documents close to our significance threshold—the documents for which the model is least certain—and hand-label them as either significant or not. This provides additional information that the model uses to estimate the policy importance of the directives in our data and helps limit noise around the threshold we use to distinguish significant directives.

Face Validity. As an exercise in face validity, we evaluate the significance estimates and topic areas produced by our model for presidential proclamations that were issued in the last two years of the Trump administration. The vast majority of these documents are in our training set (and we omit those in our test set).²⁰ We hand-code each proclamation to indicate whether it was substantive or ceremonial in nature. Substantive proclamations tended to address immigration enforcement, national emergencies, tariffs and trade, and other foreign policy items. Ceremonial proclamations recognized national holidays, deaths of notable figures, or marked, for example,

¹⁹Moreover, if we were to establish a threshold on the basis of the significance estimate for the *least* significant executive order in Howell’s (2005) data (EO 11575, issued in 1970), we would identify more than 85% of the directives in our database as significant.

²⁰In contrast, a larger share of proclamations issued through 2018 were in the test set.

Table 1: Overview of Document Processing and Classification

Step	Significance	Policy Classifier
1. Collect Corpus	Index to Presidential Executive Orders and Proclamations	Index to Presidential Executive Orders and Proclamations
2. Training Labels	1) Coarsened Chiou & Rothenberg scores 2) Trained undergraduate coders 3) Labeled documents matched to unlabeled documents that reference them	1) Comparative Agendas Project scores 2) Trained undergraduate coders
3. Text-to-data	Standard preprocessing	Standard preprocessing
4. Modeling	1) Random Forest Regression	1) ADASYN 2) Random Forest Classification
5. Estimation	Significant if predicted score is greater than 0.355	Use the predicted category
6. Relabeling	Yes	No

National Safe Boating Week (Proclamation 10035) or Leif Erikson Day (Proclamation 10097). Second, we compare the significance estimates from our model between substantive and ceremonial directives. If our estimates do a reasonable job of distinguishing important directives, we expect that substantive proclamations will have higher significance estimates.

Table 2 shows the reports of this exercise. The top panel displays the names and significance estimates for the proclamations we identified as substantive in nature. These include emergency declarations related to the US border with Mexico (Proclamation 9844) and the coronavirus pandemic (9994), changes to entry restrictions for citizens of certain countries (e.g., Proclamations 9931, 9993, and 10043), and restrictions on imports of certain goods and materials (e.g., Proclamations 9886, 9888, and 10106). Importantly, our model classified all but one of the 37 substantive proclamations issued in 2019 and 2020 as significant, with estimates greater than 0.36.²¹ Just as importantly, *none* of the proclamations we identified as ceremonial were estimated as significant. Proclamations announcing National Safe Boating Week and Leif Erikson Day, for instance, both have significance estimates less than 0.01. The bottom panel of Table 2 compares the average significance of substantive and ceremonial directives; in both years, the average significance of substantive proclamations was greater than 0.6 while the average significance of ceremonial proclamations was near zero. The estimates suggest that our measurement approach effectively distinguishes directives on the basis of their policy significance.

Table 2 also shows that our model performs well in estimating the policy area associated with each directive. Proclamations associated with the border with Mexico (e.g., 9842 and 9844) and entry restrictions for certain residents of other countries (e.g., 9931, 9932, 9945, 9983, and 9993) are classified as immigration policy, while proclamations associated with imports and tariffs (e.g., 9886, 9888, 9893, 9894, 10060, and 10064) are classified as trade. Likewise, Trump's emergency dec-

²¹The one exception was Proclamation 10126, which recognized Morocco's claims to Western Sahara. The proclamation received some media attention, with some observers noting its symbolic importance while raising doubts about its substantive effects.

laration related to COVID-19 (9994) and revision to US marine scientific research policy (10071) appear accurately classified as government operations issues. On the whole, the model appears to do a reasonable job of associating each directive with the issue area it concerns.

To recapitulate, our approach for estimating document significance builds upon the estimates developed by Chiou and Rothenberg (2017). Those estimates reflect contemporary and retrospective judgments by media, journalists, and other observers along with exogenous variables that describe the political context and the substantive focus of the directives. Yet those estimates are available for only a subset of unilateral directives issued during the period they studied and are not available for directives issued in other years. We supplement their estimates with some original hand-coding of our own, coarsened the estimates of document significance, and identified the relationship between the text of the directives and their significance estimates. After uncovering the relationship between text and significance for the set of directives studied by Chiou and Rothenberg (2017), we generate estimates of the remaining directives in our data based on the words and phrases contained in them. As the directives in our data use terms that are associated with greater significance based on the Chiou and Rothenberg (2017) estimates, our approach will also estimate them to have greater significance. Thus, as a general matter, a proclamation in our data that was issued in 1903 and contains the same text as an executive order issued in 2003 in the Chiou and Rothenberg (2017) data will have the same significance estimate.

Our approach for estimating a document's policy domain follows much in parallel, relying on the Comparative Agendas Project's expert codings of contemporary executive orders. Since these too are available only for a subset of unilateral directives, we model the relationship between a document's text and its policy domain, then use that model to extrapolate to the full set of unilateral actions.

Table 2: Presidential Proclamations, 2019 and 2020

Year	#	Title	Score	Issue Area
2019	9842	Addressing Mass Migration Through the Southern Border of the United States	0.71	Immigration
2019	9844	Declaring a National Emergency Concerning the Southern Border of the United States	0.52	Defense
2019	9880	Addressing Mass Migration Through the Southern Border of the United States	0.70	Immigration
2019	9886	Adjusting Imports of Steel Into the United States	0.80	Trade
2019	9887	To Modify the List of Beneficiary Developing Countries Under the Trade Act of 1974	0.88	Intl Affairs
2019	9888	Adjusting Imports of Automobiles and Automobile Parts Into the United States	0.62	Trade
2019	9893	Adjusting Imports of Aluminum Into the United States	0.74	Trade
2019	9894	Adjusting Imports of Steel Into the United States	0.81	Trade
2019	9902	To Modify the List of Beneficiary Developing Countries Under the Trade Act of 1974	0.89	Intl Affairs
2019	9931	Suspension of Entry as Immigrants and Nonimmigrants of Persons Responsible for Policies or Actions That Threaten Venezuela's Democratic Institutions	0.72	Immigration
2019	9932	Suspension of Entry as Immigrants and Nonimmigrants of Senior Officials of the Government of Iran	0.72	Immigration
2019	9945	Suspension of Entry of Immigrants Who Will Financially Burden the United States Healthcare System, in Order To Protect the Availability of Healthcare Benefits for Americans	0.71	Immigration
2019	9955	To Modify Duty-Free Treatment Under the Generalized System of Preferences and for Other Purposes	0.85	Trade
2019	9974	To Take Certain Actions Under the African Growth and Opportunity Act and for Other Purposes	0.87	Trade
2020	9979	To Further Facilitate Positive Adjustment to Competition From Imports of Large Residential Washers	0.88	Trade
2020	9980	Adjusting Imports of Derivative Aluminum Articles and Derivative Steel Articles Into the United States	0.80	Trade
2020	9983	Improving Enhanced Vetting Capabilities and Processes for Detecting Attempted Entry Into the United States by Terrorists or Other Public-Safety Threats	0.71	Immigration
2020	9993	Suspension of Entry as Immigrants and Nonimmigrants of Certain Additional Persons Who Pose a Risk of Transmitting 2019 Novel Coronavirus	0.72	Immigration
2020	9994	Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID-19) Outbreak	0.60	Govt Operations
2020	9996	Suspension of Entry as Immigrants and Nonimmigrants of Certain Additional Persons Who Pose a Risk of Transmitting 2019 Novel Coronavirus	0.71	Immigration
2020	10014	Suspension of Entry of Immigrants Who Present a Risk to the United States Labor Market During the Economic Recovery Following the 2019 Novel Coronavirus Outbreak	0.72	Immigration
2020	10041	Suspension of Entry as Immigrants and Nonimmigrants of Certain Additional Persons Who Pose a Risk of Transmitting 2019 Novel Coronavirus	0.72	Immigration
2020	10042	Amendment to Proclamation of May 24, 2020	0.48	Govt Operations
2020	10043	Suspension of Entry as Nonimmigrants of Certain Students and Researchers From the People's Republic of China	0.72	Immigration
2020	10049	Modifying the Northeast Canyons and Seamounts Marine National Monument	0.61	Govt Operations
2020	10052	Suspension of Entry of Immigrants and Nonimmigrants Who Present a Risk to the United States Labor Market During the Economic Recovery Following the 2019 Novel Coronavirus Outbreak	0.73	Immigration
2020	10053	To Take Certain Actions Under the United States-Mexico-Canada Agreement Implementation Act and for Other Purposes	0.76	Trade
2020	10054	Amendment to Proclamation 10052	0.51	Govt Operations
2020	10060	Adjusting Imports of Aluminum Into the United States	0.79	Trade
2020	10064	Adjusting Imports of Steel Into the United States	0.79	Trade
2020	10071	Revision to United States Marine Scientific Research Policy	0.55	Govt Operations
2020	10101	To Further Facilitate Positive Adjustment to Competition From Imports of Certain Crystalline Silicon Photovoltaic Cells	0.88	Trade
2020	10106	Adjusting Imports of Aluminum Into the United States	0.78	Trade
2020	10107	To Modify Duty-Free Treatment Under the Generalized System of Preferences and for Other Purposes	0.90	Trade
2020	10126	Recognizing the Sovereignty of the Kingdom of Morocco Over the Western Sahara	0.08	Govt Operations
2020	10128	To Take Certain Actions Under the African Growth and Opportunity Act and for Other Purposes	0.89	Trade
2020	10131	Suspension of Entry of Immigrants and Nonimmigrants Who Continue To Present a Risk to the United States Labor Market During the Economic Recovery Following the 2019 Novel Coronavirus Outbreak	0.64	Immigration
2019		Average significance, substantive proclamations: 0.73		
2019		Average significance, ceremonial proclamations: 0.01		
2020		Average significance, substantive proclamations: 0.69		
2020		Average significance, ceremonial proclamations: 0.02		

Model Accuracy

Our approach has three key advantages over human coders: accuracy, consistency, and scalability. A statistical model will produce the same (or very similar) codings for a single document each time it is queried, while humans may not. Machine coding can also produce labels for an enormous number of documents simultaneously, while human coders may take months or years to do the same. Despite these advantages, machine coding may be less desirable if it is substantially less accurate than human coders. We evaluate the accuracy of our model through two main procedures. In the interest of space, we present the details in Appendix B but in short, we find that our model has predictive success in classifying significant documents that compares well with many of the best results applying machine learning to political text. Moreover, we compare our model's results with what might be obtained from a more conventional approach to document coding and show that the model's success rate dramatically exceeds the predictive success rate from well-trained undergraduate coders.

Even so, we acknowledge several potential threats to our model's applicability. These threats relate to inaccurate document transcription, changes in language over time, unrepresentativeness of the training set, and heteroskedastic prediction accuracy. We detail these threats, and provide a suite of robustness checks to evaluate them, in Appendix B.

Patterns of Significant Unilateral Action, 1877–2020

Figure 2a shows the annual number of significant unilateral actions from 1877 to 2020. Across the entire time period, presidents issued an average of 59 significant directives per year. The endpoints of the time span are suggestive of the aggregate trend. President Rutherford Hayes issued zero significant directives in each of 1877, 1878, and 1879, while President Donald Trump issued

309 significant directives in 2020—the largest number observed in the data.²² Yet there is considerable variation between these dates. As other research has detailed (e.g., Howell 2003), presidents have generally made increasing use of unilateral powers to achieve significant policy outcomes over the latter half of the twentieth century. Presidents between Eisenhower and Carter (inclusive) issued an average of 48 significant directives per year, and this figure more than doubled (to 107) for presidents Reagan through Trump.

The number of significant directives has increased particularly rapidly in recent decades. Between the Reagan presidency and the second Clinton term, the annual number of significant directives increased by about a factor of three. The number of significant directives then declined during the George W. Bush presidency and the early years of the Obama administration before increasing during both Obama’s second term and Trump’s term in office. On only five occasions did Presidents from Rutherford Hayes to George H.W. Bush issue more than 100 significant directives in a year. Since 1993, however, presidents have exceeded this number in every year but four (2005–2007 and 2010).

Figure 2b compares the patterns from our data to other commonly-used measures of significant unilateral action with data from Bolton and Thrower (2016) and Howell (2005). Several patterns stand out. First, compared with Howell (2005), our measure identifies substantially more instances of unilateral action than we would observe if we focused only on executive orders. Second, it also suggests that while presidents in the first half of the twentieth century issued executive orders relatively frequently, many of them were not particularly important. Presidents between Taft and Franklin Roosevelt issued plenty of nonceremonial executive orders, yet our estimates indicate that the number of significant directives was far smaller. Third, our estimates

²²One may be concerned that our model is trained on earlier data and may perform less well when classifying more recent directives due to changes in the use of language, thereby underestimating the significance of documents from that era. To the extent this measurement error is present, the 309 significant directives in 2020 is likely to be an underestimate.

show greater temporal variation in unilateral directives when expanding the data set to include proclamations, memoranda, and other tools.

Figure 2c shows how the distribution of significance varies across various unilateral tools. Overall, between 1877 and 2020, executive orders comprised 37 percent of significant unilateral actions, while memoranda and proclamations accounted for 32 and 31 percent, respectively. These aggregate statistics obscure substantial variation across time, however. Consistent with Lowande (2014), the figure makes clear that the dramatic rise in significant unilateral action in the early 1990s was driven largely by increased use of memoranda. Before 1980, memoranda comprised an average of 14 percent of the annual number of significant actions; since then, however, memos have comprised an average of 51 percent of significant actions. The increased prominence of memoranda among significant unilateral directives has come largely at the expense of proclamations, which comprised 49 percent of significant unilateral actions before 1980 but only 12 percent since then.

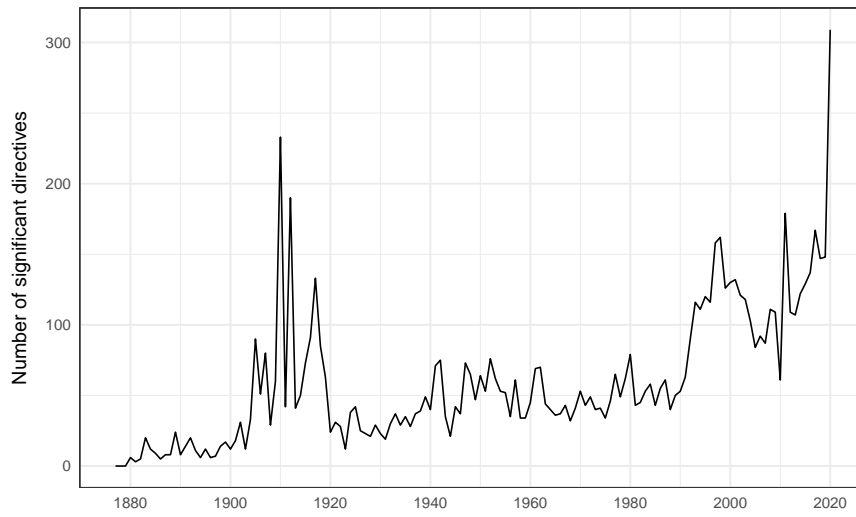
Presidential Policymaking across Issue Areas

Figure 3 shows patterns in significant unilateral action across each of the 20 issue areas. These patterns describe for what purposes presidents have deployed unilateral power. First, the figure shows that presidents address some issues more than others. Consider variation in the ranges of the y-axes. While no more than five important directives per year address civil rights, economic, education, housing, or social welfare policy, it is not uncommon for presidents to issue dozens of important directives in a given year to address defense policy, international affairs, public lands, or government operations. These patterns are striking from some perspectives, as the former group of policies typically are more politically salient than the latter group. Thus, unilateral power is most commonly exercised to address issues of lower public concern and/or on which the public and congressional lawmakers are more likely to delegate to the president.

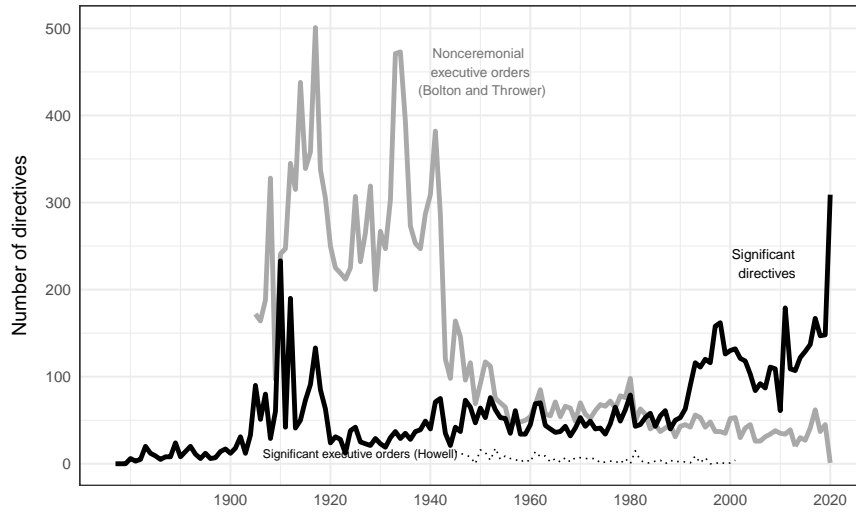
Second, the figure shows some important variation across time. Temporary spikes in an issue

Figure 2: Significant executive actions, 1877–2020

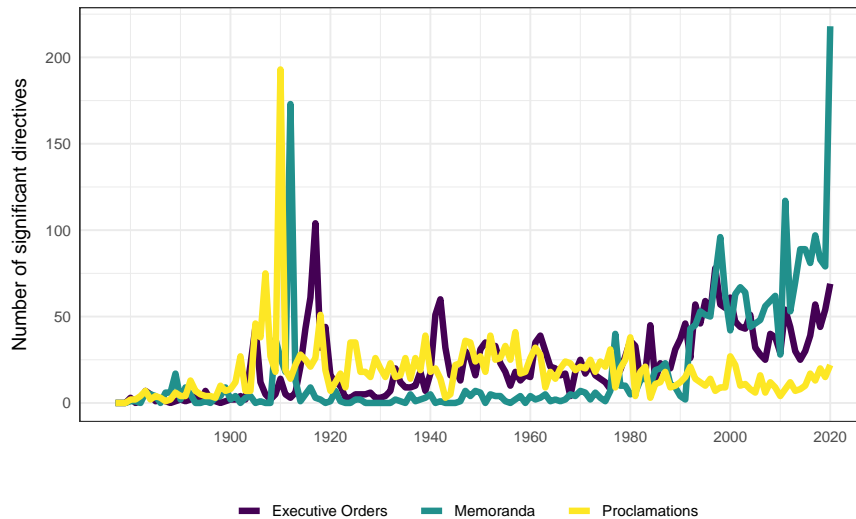
(a) Annual number of significant directives



(b) Comparison with other measures



(c) Variation across directive type



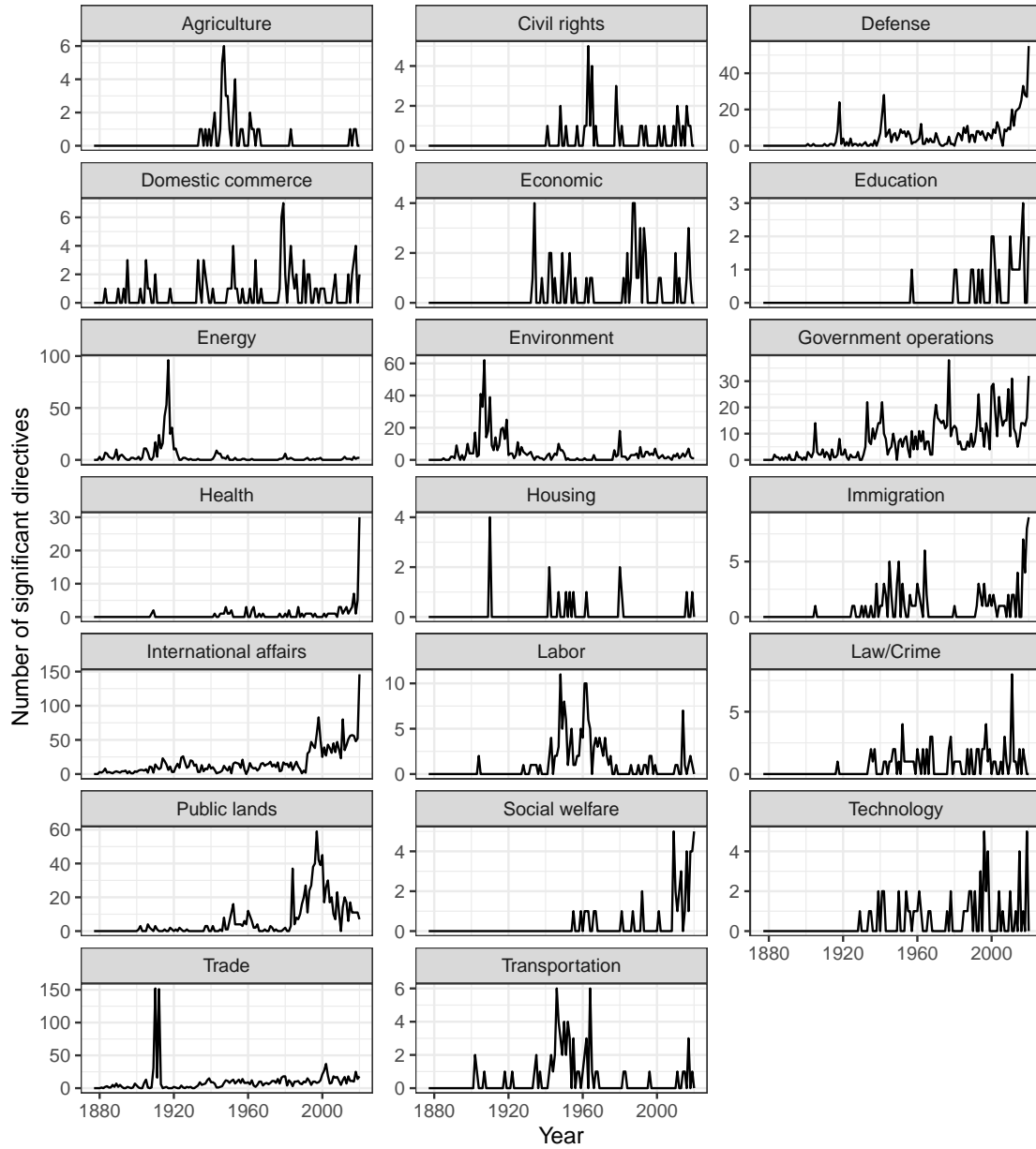
area's prominence may indicate priority areas for individual presidents; consider, for instance, the increase in civil rights directives in the early 1960s, environmental policy in the early 1900s, and labor policy in the 1940s. Perhaps most notably, over the last several decades presidents have issued increasing numbers of directives to address national defense and international affairs. These issue areas disproportionately contribute to the overall increase in unilateral activity during this period. It may also be notable that the number of immigration directives has increased over the last decade, perhaps reflecting the public salience and lack of congressional activity on this policy area.

Our data and measures generate several new descriptive findings about patterns of unilateral activity by American presidents over nearly a century and a half of US history. We show that presidents have made significantly more substantive policies via unilateral power than existing data sources have documented, particularly over the last several decades. We also show that, despite the attention commonly paid to executive orders that address domestic policy, more commonly presidents have used unilateral power on issues of foreign policy. These patterns suggest that analyses focused on executive orders or any other single unilateral tool miss important information about unilateral policymaking.

Structural Change in Unilateral Activity

Using our measures of presidential unilateralism, we investigate more formally how the use of unilateral power developed across historical periods. Our interest in is identifying distinct regimes corresponding to patterns of unilateral activity. The development and deployment of unilateral power looms large in accounts of the institutional evolution of the presidency (see, e.g., Dodds 2013; Ellis 2015; Greenstein 1978; Moe and Howell 1999; Schlesinger 1973). In the main, these accounts argue that using unilateral power more frequently, and for larger purposes, accompanied the institutional ascendance of the presidency. For instance, according to Green-

Figure 3: Significant Actions by Issue Area, 1877–2020



stein (1978, 45-46): “From a presidency that normally exercised few unilateral powers, there has been a shift to one that is provided... with many more occasions for direct policy making through executive orders and other actions not normally ratified by Congress.”

Though scholars have raised questions about the modern presidency paradigm (Rogowski 2016; Skowronek 1993), there has been relatively little focus on identifying historical variation in presidents’ use of unilateral powers (important exceptions include Bolton and Thrower 2016; Dodds 2013; Williams 2020). There has been even less research dedicated to systematically identifying distinctive period of unilateral activity. Dodds (2013, 24) goes furthest in this direction, positing that

Four areas of this picture of the development of executive orders seem to stand out: (1) a period of very low usage of executive orders from the founding of the country through the nineteenth century, (2) a sharp rise under Theodore Roosevelt, (3) a less dramatic but steady increase from Taft through FDR, and (4) a marked decline in usage after World War II.

However, he does not formally test whether these periods correspond to distinctive regimes of unilateral activity, and his observations are based on executive orders rather than the larger pool of unilateral directives.

We use methods implemented by Kleiber et al. (2002) to identify structural breaks (or change points) in presidential unilateral actions. If the use of unilateral action coincided with distinct eras of presidential behavior, then we would expect to observe noticeable changes in patterns of unilateral activity that coincide with these eras. Our investigation is similar to Park (2010), who evaluates long-term changes in presidents’ use of military force, Spirling (2012), who studies the evolution of the terms of treaties between the US government and American Indians, and Wawro and Katznelson (2014), who examine the relationship between party realignments and states’ partisan representation in the Senate.

Following this work and the approach introduced by Bai and Perron (2003), we use a linear regression framework to identify the stability of annual patterns of unilateral action across time. This allows us to evaluate how annual patterns of unilateral activity have systematically varied across time and estimates the points in our time series where any changes occurred. When fitting the model, we require a minimum segment length of four years. Substantively, this allows us to identify whether patterns in unilateral action vary systematically across individual presidents. Following Bai and Perron (2003), we identify the number and locations of structural breaks using BIC as an indicator of model fit.

Figure 4 shows the results of this analysis. The points show the annual number of significant unilateral actions, the dashed vertical lines show the dates corresponding to the change points identified by the model, and the solid horizontal lines show the average number of directives issued in each period. The top panel shows results for all unilateral actions (as depicted in Figure 4). The best-fitting model identifies five breaks, corresponding to 1904, 1918, 1991, and 2016.²³ The results in this plot provide several new findings about patterns of unilateral power. First, the location of the first two estimated breaks suggest that the number of unilateral actions issued by Progressive Era presidents sharply contrasted with their predecessors. The location of the first break corresponds with Theodore Roosevelt's time as president. According to Dodds (2013, 123), "Earlier presidents issued some very important unilateral directives, but [Roosevelt]'s presidency marked a significant change in both the number and the nature of such directives." Our data are consistent with this claim. The second break (1909) matches the beginning of the Taft presidency, who, despite his apparent philosophical differences with Roosevelt (see Howell 2013, 102-3), issued unilateral actions at an even greater clip. This pattern continued under Woodrow Wilson's presidency. These findings, too, support Dodd's (2013, 152) interpretation that Roo-

²³We do not display the confidence intervals around these breaks for the purposes of visual clarity, but they are relatively small in magnitude. The 95% confidence interval for each break is: [1902, 1906], [1914, 1922], [1989, 1993], and [2014, 2018].

sevelt's 'practice of frequently issuing unilateral presidential directives stuck and became the new norm' among his immediate successors.

Second, we find no evidence that the presidency of Franklin Roosevelt was a turning point in the use of unilateral powers. This finding is striking given the centrality of Roosevelt in accounts that describe the emergence of the modern presidency (Greenstein 1978). Instead, we find that the period from 1918 to 1991 suggested a long mid-century of presidential consistency in unilateral activity. During this period, presidents issued an annual average of 45 significant directives. While this is four times the annual rate of significant directives issued by presidents Hayes through McKinley, it suggests greater continuity with the rate established by Theodore Roosevelt (43 per year) rather than a clear break in the use of unilateral power.

Third, the last two breaks, 1991 and 2016, suggest a more sharply upward trajectory in unilateral action over the last three decades. These breaks correspond roughly to the beginnings of the Clinton and Trump presidencies (though we make these claims cautiously given the uncertainty associated with the estimated breaks). Nevertheless, the data are consistent with Howell's (2005) claim about presidents' increased use of unilateral power in recent decades, with the change points suggesting that presidents Clinton, Bush, and Obama made significantly greater use of unilateral action on average than their twentieth century predecessors, which increased further still under President Trump.

The bottom three plots in Figure 4 show the results of a similar exercise applied separately to each class of significant directives. The patterns of executive order use reveal more short-term changes in their use and are somewhat more supportive of Dodds' account. It may be noteworthy that the second and third breaks correspond to 1914 and 1919, roughly bracketing World War I, while the fourth and fifth breaks correspond to 1939 and 1943, bracketing the time in which the US entered World War II. Though only suggestive, these findings may reflect observations from Schlesinger (1973), who noted that the two world wars significantly augmented presidential authority, particularly through unilateral powers. Perhaps just as interestingly, though the annual

number of executive orders decreased after the wartime periods, they returned to a higher base rate relative to the period that preceded war.

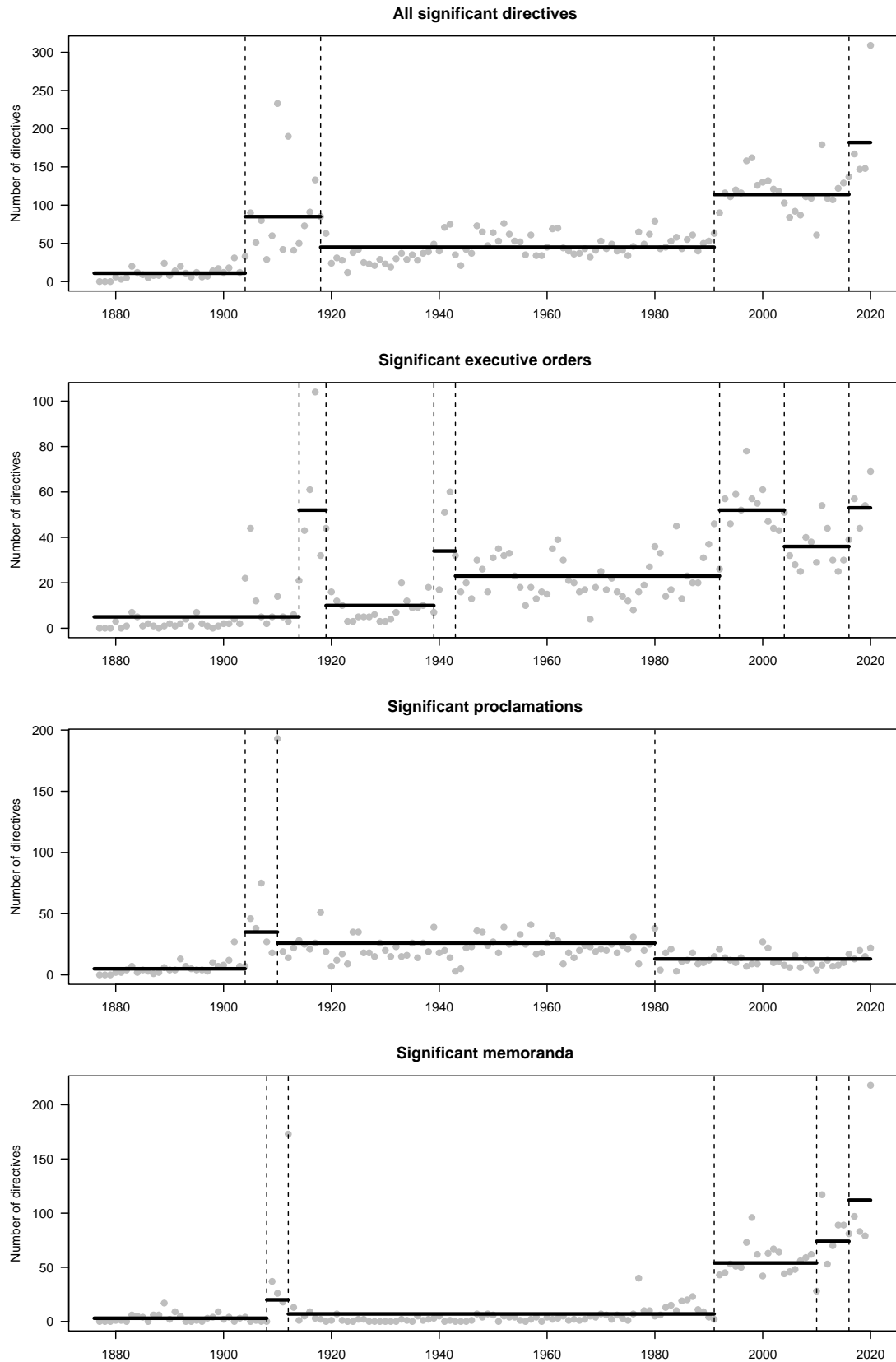
Changes in the use of proclamations were muted by comparison. While there was a significant uptick in the number of proclamations between 1904 and 1910 (consistent with Williams 2020), for the remainder of the time period their annual use was relatively even. Finally, the bottom plot details the growth of memoranda over time. The model estimates breaks in 1908 and 1912, again demonstrating the growth of unilateral power by a president (Taft) who expressed reticence to deploy it. The more significant breaks occur in 1991 and again in 2010 and 2016, as the Clinton and George W. Bush administrations, and later the Obama and Trump administrations, increased their reliance on memoranda for issuing unilateral directives.

Using our measures of significant unilateral directions and methods for examining change over time, we provide new evidence about historical patterns in presidential behavior. We find evidence to support accounts that emphasize the contributions of Theodore Roosevelt and other Progressive Era presidents to the development and deployment of unilateral power. Yet we also find, contrary to scholarship that periodizes the modern presidency coincident with Franklin Roosevelt's administration, that presidents from roughly 1920 through 1990 were more similar than different in their use of unilateral power. To be sure, we occasionally find that major wars coincided with higher rates of certain unilateral directives. But we do not find that these wars led to durable shifts in the exercise of unilateral action. The patterns reported above also show temporal complementarities across unilateral tools. The increased use of one tool does not necessarily coincide with greater use of another. Therefore, we obtain a fuller portrait of the unilateral presidency when aggregating across them.

Directive Substitution in Presidential Policymaking

Our final set of results concerns substitution in how presidents conduct unilateral policymaking. Because presidents exercise unilateral powers through a variety of tools, recent scholarship

Figure 4: Structural Change in Presidential Unilateralism, 1877–2020



Note: Vertical dashed lines are the estimated change point locations. Solid horizontal lines are the average annual number of directives issued within each period.

suggests the possibility of strategic substitution between directives (see, e.g., Lowande 2014, 739). That is, directives vary not only in what they can functionally achieve but also in the political costs and benefits their use offers to presidents. If these costs and benefits vary across time, presidents may have incentives to exercise unilateral powers through different means.

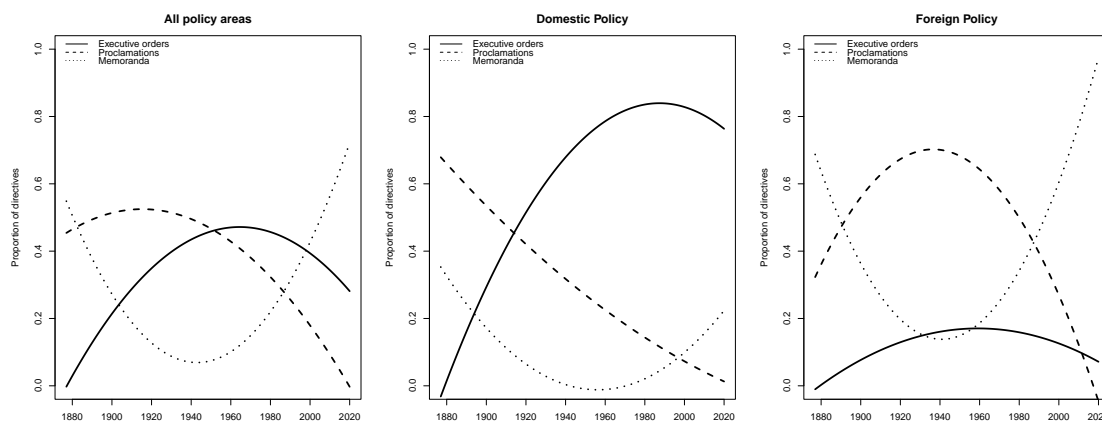
We study patterns of directive substitution using our estimates of directive significance and policy areas. Rather than study the rate of unilateral action, as in our analyses above, we study presidents' choice of unilateral tool conditional on their decision to issue a substantively important directive. To do so, we model the share of important directives issued via each unilateral tool as a quadratic expression of time. We report results using a second-degree polynomial, which we believe represents an appropriately simple yet flexible specification, but Appendix C shows similar patterns when using linear specifications and third-, fourth-, and fifth-degree polynomials.

Figure 5 shows the fitted proportion of important directives in each policy area that was issued as an executive order, proclamation, or memoranda for each year in the data. The plot on the left shows results when using directives from all issue areas. The results suggest that presidents have created important unilateral policies through different means across time. The predicted share of important directives issued as executive orders increased through much of the twentieth century to a peak of 47 percent in the mid 1960s; in 2020, by contrast, about 28 percent of important directives were estimated to be executive orders. The plot also suggests that memoranda may substitute for executive orders, as the temporal patterns are inverted. About 55 percent of significant directives in the late 1870s were memoranda, which declined to about 7 percent by the FDR presidency; today, however, memoranda comprise nearly 70 percent of significant unilateral directives. Proclamations, however, have generally declined as a share of significant directives. About half of significant directives were issued as proclamations in the 1910s (consistent with Bailey and Rottinghaus 2013), while in recent years this has declined to the low single digits.

The middle and right plots in Figure 5 show how these patterns vary by issue area. For sim-

plicity, we distinguished domestic and foreign policy issues.²⁴ The figures show that presidents make unilateral policy through different means across issue areas, and that these patterns have also changed across time. Executive orders, and to some degree memoranda, have increasingly displaced proclamations as means of unilateral policymaking on domestic issues. Proclamations have also been displaced as a means of unilateral policymaking on foreign policy issues; instead, memoranda account for increasingly large shares of important unilateral directives in this policy domain. This development may be particularly notable in the context of claims that Congress has increasingly abdicated its role in foreign policy making to the president (see, e.g., Fisher 1999). Not only are contemporary presidents issuing substantially more foreign policy directives than their predecessors, but they are doing so through directives that are subject to the weakest transparency and reporting requirements. While a fuller examination of these phenomena is beyond the scope of this paper, these developments invite further scrutiny for assessing positive claims about presidential influence over foreign affairs and normative claims about the president’s foreign policy making power.

Figure 5: Directive Substitution across Policy Areas



²⁴We classified immigration, defense, trade, and international affairs as foreign policy; all other issues from Figure 3 were classified as domestic policy.

Conclusion

Unilateral powers are perhaps more publicly salient than ever. Presidential candidates routinely promise flurries of policy changes on their first day in office, and recent presidents have issued dozens of executive orders in the days following their inaugurations. While presidents can affect policy change through a number of different unilateral directives, researchers have focused primarily on studying executive orders to the exclusion of other directives. We present new data on presidential unilateralism over nearly a century and a half and introduce a new methodological for analyzing them. As Andrew Rudalevige has written, “taking a full inventory of the toolbox of directives available to presidents helps us better understand the scope of executive authority more generally.”²⁵ By accounting for a wider range of unilateral directives, we have uncovered new empirical patterns of unilateral policymaking and identified the domains in presidents have used these powers.

The patterns we identified contribute to contemporary debates about the normative bounds of presidential power. First, we show that presidents have made increased use of unilateral power in recent decades. While this could be interpreted as evidence of an “imperial” presidency that threatens the separation of powers, it could also represent increased delegation of authority to presidents in an era where Congress has been fragmented and gridlocked. This latter interpretation may be supported by our finding that the increase in unilateral action is driven largely by directives issued to address foreign affairs. To the extent that Congress defers to presidents in foreign policy, the increase in unilateral activity may not represent a usurpation of legislative power. Second, to the extent that increased unilateralism has tended to occur through more

²⁵“Most of Trump’s executive orders aren’t actually executive orders. Here’s why that matters.” *Washington Post*, Monkey Cage, January 30, 2017; <https://www.washingtonpost.com/news/monkey-cage/wp/2017/01/30/most-of-trumps-executive-orders-arent-actually-executive-orders-heres-why-that-matters/>.

obscure forms, this may raise questions about transparency and accountability. For example, if memoranda are increasingly used for important unilateral policies but are less available to and/or covered by the media, then there may be fewer opportunities to hold presidents accountable for the policies they create through this tool.

Our data, which are freely available upon publication, permit new empirical investigations into the presidency and political institutions more generally. Not only can scholars revisit existing accounts of unilateral action, but these accounts can also be tested in individual issue areas, across longer periods of American history, and among specific types of directives. For example, while previous work studies the predictors of the annual number of executive orders, memoranda, and proclamations (Bolton and Thrower 2016; Howell 2003; Lowande 2014; Williams 2020), future work could use the data introduced here to study how the constraints on presidential unilateralism vary across directives, issues, and time. These analyses could point toward new explanations for the use of unilateral power and its contributions to presidential policy influence. The text corresponding to the directives can also be useful for new and innovative analyses about the nature and impact of unilateral activity, measuring for example the liberalism or conservatism of individual directives.

Finally, the methods we use for analyzing unilateral directives could be productively extended to other institutions and contexts. Text-based documents are issued by virtually every government entity at the national, state, and local levels. Future research could use strategies like those used here to characterize the policy outputs from presidents, bureaucracies, courts, and legislatures across time and levels of government. And rather than study institutional outputs in isolation, future research could generate comparable estimates of policy significance and/or issue area for policies produced by multiple institutions. Moreover, these documents could be classified according to a host of criteria beyond policy significance or issue area. Extending the guidance offered by Wawro and Katznelson (2014), combining methodological advancements (e.g., Erlich et al. (2021)) in text-based analysis with increased availability of government documents may be

a particularly fruitful approach for scholars interested in historical research.

References

- Archer, Kellie J., and Ryan V. Kimes. 2008. "Empirical Characterization of Random Forest Variable Importance Measures." *Computational Statistics & Data Analysis* 52: 2249–2260.
- Bai, Jushan, and Pierre Perron. 2003. "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics* 18(1): 1–22.
- Bailey, Jeremy D., and Brandon Rottinghaus. 2013. "The Development of Unilateral Power and the Problem of the Power to Warn: Washington through McKinley." *Presidential Studies Quarterly* 43(1): 186–204.
- Binder, Sarah. 2015. "The Dysfunctional Congress." *Annual Review of Political Science* 18: 85–101.
- Bolton, Alexander, and Sharece Thrower. 2016. "Legislative Capacity and Executive Unilateralism." *American Journal of Political Science* 60: 649–663.
- Cameron, Charles M. 2000. *Veto bargaining: Presidents and the politics of negative power*. Cambridge University Press.
- Canes-Wrone, Brandice. 2006. *Who Leads Whom? Presidents, Policy, and the Public*. Chicago: University of Chicago Press.
- Chiou, Fang-Yi, and Lawrence S. Rothenberg. 2014. "The Elusive Search for Presidential Power." *American Journal of Political Science* 58: 653–668.
- Chiou, Fang-Yi, and Lawrence S. Rothenberg. 2017. *The Enigma of Presidential Power: Parties, Policies and Strategic Uses of Unilateral Action*. New York: Cambridge University Press.

- Christenson, Dino, and Douglas Kriner. 2020a. "Beyond the Base: Presidents, Partisan Approval, and the Political Economy of Unilateral Action." *Journal of Political Institutions and Political Economy* 1: 79–103.
- Christenson, Dino P., and Douglas L. Kriner. 2020b. *The Myth of the Imperial Presidency: How Public Opinion Checks the Unilateral Executive*. University of Chicago Press.
- CIS Index to Presidential Executive Orders and Proclamations*. 1987. Bethesda, MD: Congressional Information Service, Inc.
- Clinton, Joshua D., and John S. Lapinski. 2006. "Measuring Legislative Accomplishment, 1877-1994." *American Journal of Political Science* 50(1): 232–249.
- Cohen, Jeffrey E. 2012. *The President's Legislative Policy Agenda, 1789-2002*. Cambridge University Press.
- Congressional Research Service. 2021. "Executive Orders: An Introduction." CRS Report R46738; available at <https://crsreports.congress.gov/product/pdf/R/R46738>.
- Cooper, Phillip J. 2014. *By Order of the President: The Use and Abuse of Executive Direct Action*. University Press of Kansas.
- Dodds, Graham G. 2013. *Take Up Your Pen: Unilateral Presidential Directives in American Politics*. Philadelphia: University of Pennsylvania Press.
- Durant, Robert. 2009. "Back to the Future? Toward Revitalizing the Study of the Administrative Presidency." 39: 89–110.
- Ellis, Richard J. 2015. *The Development of the American Presidency*. New York: Routledge.
- Executive Orders*. 2021. "The Policy Agendas Project at the University of Texas at Austin." www.comparativeagendas.net. Accessed June 5, 2021.

- Erlich, Aaron, Stefano G. Dantas, Benjamin E. Bagozzi, Daniel Berliner, and Brian Palmer-Rubin. 2021. "Multi-Label Prediction for Political Text-as-Data." *Political Analysis* p. 1–18.
- Fine, Jeffrey A., and Adam L. Warber. 2012. "Circumventing Adversity: Executive Orders and Divided Government." *Presidential Studies Quarterly* 42: 256–274.
- Fisher, Louis. 1999. "Congressional Abdication: War and Spending Powers." *Saint Louis University Law Review* 43: 931–1012.
- Gentzkow, Matthew, and Jesse M Shapiro. 2010. "What Drives Media Slant? Evidence from US Daily Newspapers." *Econometrica* 78(1).
- Gerrish, Sean M., and David M. Blei. 2011. "Predicting Legislative Roll Calls from Text." *Proceedings of the 28th International Conference on Machine Learning* .
- Greenstein, Fred I. 1978. "Change and Continuity in the Modern Presidency." In *The New American Political System*, ed. Anthony King. Washington, D.C.: American Enterprise Institute.
- He, Haibo, Yang Bai, Edwardo A Garcia, and Shutao Li. 2008. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. IEEE pp. 1322–1328.
- Hebe, William. 1972. "Executive Orders and the Growth of Presidential Power." *Villanova Law Review* 17: 688–712.
- Hopkins, Daniel J, and Gary King. 2010. "A method of automated nonparametric content analysis for social science." *American Journal of Political Science* 54(1): 229–247.
- Howell, William G. 2003. *Power without Persuasion: The Politics of Direct Presidential Action*. Princeton: Princeton University Press.

- Howell, William G. 2005. "Unilateral Powers: A Brief Overview." *Presidential Studies Quarterly* 35(3).
- Howell, William G. 2013. *Thinking about the Presidency*. Princeton: Princeton University Press.
- Huang, Jin, and Charles X Ling. 2005. "Using AUC and accuracy in evaluating learning algorithms." *IEEE Transactions on knowledge and Data Engineering* 17(3): 299–310.
- Kagan, Elena. 2001. "Presidential Administration." *Harvard Law Review* 114(8): 2245–2385.
- Katagiri, Azusa, and Eric Min. 2019. "The Credibility of Public and Private Signals: A Document-Based Approach." *American Political Science Review* 113(1): 156–172.
- Katznelson, Ira, and John S. Lapinski. 2006. "The Substance of Representation: Studying Policy Content and Legislative Behavior." In *The Macropolitics of Congress*, ed. E. Scott Adler, and John S. Lapinski. Princeton: Princeton University Press pp. 96–128.
- Kaufman, Aaron, and Aja Klevs. Forthcoming. "Adaptive Fuzzy String Matching: How to Merge Data Sets with Only One (Messy) Identifying Field." *Political Analysis* .
- Kaufman, Aaron R. 2020. "Measuring the Content of Presidential Policy Making: Applying Text Analysis to Executive Branch Directives." *Presidential Studies Quarterly* .
- Kleiber, Christian, Kurt Hornik, Friedrich Leisch, and Achim Zeileis. 2002. "strucchange: An R Package for Testing for Structural Change in Linear Regression Models." 7(2): 1–38.
- Krause, George A., and Jeffrey E. Cohen. 2000. "Opportunity, Constraints, and the Development of the Institutional Presidency: The Issuance of Executive Orders, 1939–96." *Journal of Politics* 62(1): 88–114.
- Ling, Charles X, Jin Huang, Harry Zhang et al. 2003. AUC: a statistically consistent and more discriminating measure than accuracy. In *IJCAI*. Vol. 3 pp. 519–524.

- Lowande, Kenneth. Forthcoming. "Presidents and the Status Quo." *Quarterly Journal of Political Science* .
- Lowande, Kenneth, and Jon C. Rogowski. 2021. "Presidential Unilateral Power." *Annual Review of Political Science* .
- Lowande, Kenneth S. 2014. "After the Orders: Presidential Memoranda and Unilateral Action." *Presidential Studies Quarterly* 44: 724–741.
- Mayer, Kenneth R. 2001. *With the Stroke of a Pen*. Princeton: Princeton University Press.
- Mayer, Kenneth R., and Kevin Price. 2002. "Unilateral Presidential Powers: Significant Executive Orders, 1949-99." *Presidential Studies Quarterly* 32(2): 367–386.
- Moe, Terry, and William Howell. 1999. "The Presidential Power of Unilateral Action." *Journal of Law, Economics, and Organization* 15: 132–179.
- Mozer, Reagan, Luke Miratrix, Aaron Russell Kaufman, and L Jason Anastasopoulos. 2020. "Matching with text data: An experimental evaluation of methods for matching documents and of measuring match quality." *Political Analysis* 28(4): 445–468.
- Park, Jong Hee. 2010. "Structural Change in U.S. Presidents' Use of Force." *American Journal of Political Science* 54(3): 766–782.
- Patterson, Thomas E. 2017. "News Coverage of Donald Trump's First 100 Days." HKS Working Paper No. RWP17-040; available at <http://dx.doi.org/10.2139/ssrn.3040911>.
- Rogowski, Jon C. 2016. "Presidential Influence in an Era of Congressional Dominance." *American Political Science Review* 110: 325–341.
- Rogowski, Jon C. 2019. "Unilateral Action, Public Opinion, and Presidential Responsiveness." Working paper.

- Rottinghaus, Brandon, and Jason Maier. 2007. "The Power of Decree: Presidential Use of Executive Proclamations, 1977-2005." *Political Research Quarterly* 60(2): 338–343.
- Rudalevige, Andrew. 2002. *Managing the President's Program: Presidential Leadership and Legislative Policy Formulation*. Princeton: Princeton University Press.
- Schlesinger, Arthur. 1973. *The Imperial Presidency*. Houghton Mifflin.
- Settles, Burr. 2012. "Active learning." *Synthesis lectures on artificial intelligence and machine learning* 6(1): 1–114.
- Shane, Peter M., and Harold H. Bruff. 1996. *Separation of Powers Law*. Durham, NC: Carolina Academic Press.
- Simmons, Joseph P, Leif D Nelson, and Uri Simonsohn. 2011. "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant." *Psychological science* 22(11): 1359–1366.
- Skowronek, Stephen. 1982. *Building a New American State: The Expansion of National Administrative Capacities, 1877-1920*. New York: Cambridge University Press.
- Skowronek, Stephen. 1993. *The Politics Presidents Make: Leadership from John Adams to Bill Clinton*. Harvard University Press.
- Spirling, Arthur. 2012. "US treaty making with American Indians: Institutional change and relative power, 1784–1911." *American Journal of Political Science* 56(1): 84–97.
- Thrower, Sharece. 2017. "The President, the Court, and Policy Implementation." *Presidential Studies Quarterly* 47(1): 122–145.
- Warber, Adam L. 2006. *Executive Orders and the Modern Presidency: Legislating from the Oval Office*. Boulder, CO: Lynne Rienner Publishers.

- Warber, Adam L., Yu Ouyang, and Richard W Waterman. 2018. "Landmark Executive Orders: Presidential Leadership Through Unilateral Action." *Presidential Studies Quarterly* 48(1): 110–126.
- Wawro, Gregory J., and Ira Katznelson. 2014. "Designing Historical Social Scientific Inquiry: How Parameter Heterogeneity Can Bridge the Methodological Divide between Quantitative and Qualitative Approaches." *American Journal of Political Science* 58(2): 526–546.
- Williams, Raymond T. 2020. "Unilateral Politics in the Traditional Era: Significant Executive Orders and Proclamations, 1861-1944." *Presidential Studies Quarterly* 50(1): 146–162.
- Woolley, John T., and Gerhard Peters. 2017. "Do Presidential Memo Orders Substitute for Executive Orders? New Data." *Presidential Studies Quarterly* 47: 378–393.
- Zubek, Radoslaw, Abhishek Dasgupta, and David Doyle. 2021. "Measuring the Significance of Policy Outputs with Positive Unlabeled Learning." *American Political Science Review* 115(1): 339–346.

ONLINE APPENDIX

Robustness Checks and Supplementary Analyses for
Presidential Policymaking, 1877–2020

Contents

A	Supplementary Appendix: Source Record Groups	1
A.1	Executive Orders	1
A.2	Memoranda	1
A.3	Proclamations	2
A.4	Distribution of significance estimates	3
B	Supplementary Appendix: Model Validity Robustness Checks	4
B.0.1	AUC and Multi-class classification	6
B.1	Document Transcription	6
B.2	Language Heterogeneity & Unrepresentative Training Sets	7
B.3	Heteroskedastic Predictive Accuracy	8
B.4	Temporal Variation in Modeling Accuracy	8
B.5	Feature Importance	10
B.6	Robustness to Threshold Choices	11
C	Supplemental Appendix: Alternative Polynomial Specifications	13

A Supplementary Appendix: Source Record Groups

In this appendix we indicate which unilateral action “Source Record Groups” we group into each larger category of unilateral action.

A.1 Executive Orders

This category contains documents which are numbered and unnumbered executive orders.

EO - Numbered Executive Orders 1862-present

03 - Public Land Orders 1942-present

06 - Secretary of Interior Orders 1920-1950

22 - Executive Orders Relating to the Panama Canal 1902-1934

33 - Executive Orders Relating to Public Lands 1841-1935

41 - Executive Orders Relating to Public Lands 1820-1913

A.2 Memoranda

This category contains Executive Memoranda or other such memoranda from collections of presidential documents.

04 - Presidential Documents 1936-present

05 - White House Records 1869-present

08 - Manuscript collections 1790-1929

12 - Treasury and Justice Dept Records 1789-1908

17 - Navy and War Dept Records 1789-1884

20 - Messages and Papers of the President 1789-1899

21 - Public Papers of the Presidents 1789-present

37 - Abandoned Military Lands 1826-1905

52 - Miscellaneous Printed Sources 1789-1936

53 - Weekly Compilation of Presidential Documents 1965-present

56 - Presidential Policy Directives & National Security Decision Memoranda

A.3 Proclamations

This category includes only documents clearly noted as proclamations.

PR - Numbered Proclamations 1789-present

29 - Treaty Proclamations 1789-present

35 - Proclamations Relating to Public Lands 1834-1907

A.4 Distribution of significance estimates

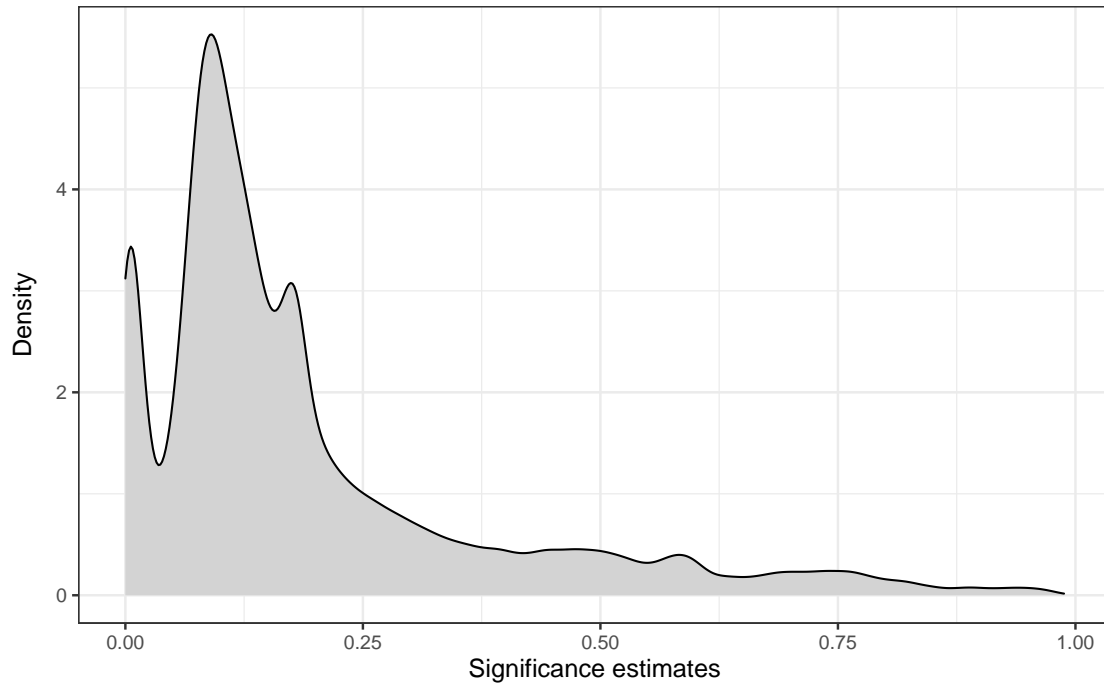


Figure A.1: Distribution of significance estimates across documents.

B Supplementary Appendix: Model Validity Robustness Checks

In interpreting the results of our significance model, we note that all documents are scored on the same scale: an estimated significance of 1 means the same thing for executive orders, proclamations, and memoranda. However, we expect that our measurement strategy induces *heterogeneous measurement error*. Some of our documents have their significance scores hand-coded or assumed; these have minimal measurement error if any. Some documents, on the other hand, have measurement error derived from any number of challenges: poor document transcription, language heterogeneity, the rarity of highly significant documents, and other problems generally associated with transfer learning.

We assess model accuracy through two means. The first is through cross-validated AUC, or the area under the precision-recall curve – we do this for both our significance model and our policy classifier. The second method is through comparisons to human coders, which we perform for our significance model only. Cross-validated AUC measures how well a model measures the relationship between covariates and outcomes in the training data. This is a difficult task: text-as-data methods are best suited to measuring concrete and measurement error-free concepts, while unilateral action significance is anything but concrete. Regardless, we achieve notable success in cross-validation accuracy.

When examining binary outcomes, as is the case in our significance modeling approach, the most common accuracy measures are precision and recall (Ling et al. 2003; Huang and Ling 2005). Precision is the number of correct positive identifications divided by the total number of identifications; recall is the number of correct positive identifications divided by the total number of true positive cases. Taken together, these measures produce a Receiver operating characteristic (ROC).²⁶ The area under the ROC curve, called AUC, is the gold standard standard measure of

²⁶Consider a model which produces probabilities that an observation is either a 1 or a 0. To measure the accuracy of this model, we must first specify a predictive cutoff. Perhaps we de-

predictive accuracy for binary classification tasks. In Figure B.1, we present the ROC and AUC for our model with a significance threshold of 0.5. The AUC is 0.904, on par with many of the best results in the application of machine learning to the social sciences.

This result is difficult to interpret without a relevant benchmark. Ideally, that benchmark would be the best alternative to using a machine learning model. To establish that benchmark, we trained three undergraduate research assistants to manually code unilateral actions as significant (1) or ceremonial (0) and compared those human coders' accuracy to that of our significance model.²⁷ We presented the research assistants with 100 executive orders that have significance scores from Chiou and Rothenberg (2014), as well as 100 other unilateral actions from our data set that did not have Chiou and Rothenberg (2014) estimates, and asked the students to code the significance of those documents. We then performed two analyses on these hand-coded significance scores. The first measures inter-coder reliability. An important advantage of machine learning models for coding documents is consistency: the model will yield a similar or identical result every time it is queried. Human coders, however, are often inconsistent. The research assistants' hand-coded executive order significance scores were not highly correlated with each other. Taking the undergraduates in pairs and measuring their percent agreement at coding unilateral actions as significant or ceremonial, the three undergraduates agree with each others at rates of 65%, 71%, and 63%.

In the second analysis, we calculate AUC scores for the three sets of hand-coded documents

termine that any observation predicted to be 1 with $p > 0.5$ is a 1, and otherwise a 0, then we can measure precision and recall. However, as we vary our predictive cutoff, precision and recall change. The ROC curve captures precision and recall for all values of the predictive cutoff from 0 to 1.

²⁷The undergraduate coders were asked to research the unilateral actions and assess their policy significance using their own best judgment and knowledge of the relevant historical/political context.

compared to with Chiou and Rothenberg’s scores for the same documents, thresholded at 0.5. If the research assistants’ AUC scores, individually or aggregated, are lower than the machine learning model, then we can be confident that the machine learning model is an improvement over the current state of the art. We find that the research assistants’ codings produce AUCs of 0.68, 0.67, and 0.65, each of which is substantially lower than the AUC of 0.90 produced by the machine learning model. In practice, when using research assistants to hand code noisy data, it is common to average hand codes to produce a more reliable measure. We take the elementwise average of the three hand-coded significance codes and calculate that the AUC for that aggregated coding is 0.71. These exercises suggest that our machine learning model is substantially more accurate than trained undergraduate research assistants, and provides a dramatic improvement as a consistent and scalable approach for measuring document significance.

B.0.1 AUC and Multi-class classification

A variant of AUC is applicable to our policy classification model as well: for each of 20 policy categories, we can calculate an AUC by dichotomizing the outcome as related to Policy X or not; by averaging each of these 20 AUC curves we can construct an average AUC. Our policy classifier receives an average AUC of 0.86, a strong score for a 20-class classification problem.

B.1 Document Transcription

Many of the documents we analyze are simply scanned images of printed pages in PDF format. We extracted text from these PDFs using Google’s Tesseract 4 optical character recognition (OCR) system. For documents with typed text, this OCR procedure produces high quality text. However, for many earlier and hand-written documents, the OCR-derived text is of poor quality. To improve the data quality in these cases, as well as in cases where more than 10% of the words are not found in a dictionary, we transcribed these documents by hand. Together, these two samples account for 5% of our total corpus. As a validity check, we transcribed 20% of this sample

twice; concordance between the doubly-transcribed documents ensures us that our transcriptions are satisfactory.

B.2 Language Heterogeneity & Unrepresentative Training Sets

A critical assumption for our analysis is that the language and word choice indicative of significant executive orders is sufficiently similar to that of other types of significant unilateral directives. For example, the tone and style of significant executive orders and proclamations may be very legalistic, while important memoranda may be more rhetorical; if this is the case, then many of the textual features which contribute to a document's significance may be legalistic, biasing downward a memorandum's estimated significance. This problem may be especially severe in cases where the training set comes from a more limited set of years than the test set.

To fortify both of our model against this weakness, we expand our training set to include more representative documents. However, since we do not have significance scores or policy labels for documents other than executive orders, we infer them using a manual matching procedure. We first select a random 500 executive orders from our training set. Then, using the ProQuest Executive Actions database, we manually search for documents which reference one and only one executive order in our random sample. If we find a document which is substantively related to a single executive order, we assign that document the same significance, either 0 or 1, as the executive order it mentions. By assigning equal significance and identical policy labels to those two documents, we teach our model to recognize the significance and policy classifications of a wider variety of rhetorical styles. We find matches for 86 of the 500 executive orders in our random sample. Many of those executive orders have multiple matching documents; as a result, our matching procedure adds 287 observations to our training data for both models.

B.3 Heteroskedastic Predictive Accuracy

A third challenge is ensuring that our models' predictions, which we aggregate into dependent variables for regressions, are not systematically biased. If the predictions are unbiased but measured with error, that measurement error will force our regression coefficients toward zero. If, however, the predictions are biased, then the regression coefficients may be artificially extreme. We can observe whether our predictions are systematically biased by examining closely our cross-validation accuracy for heteroskedasticity.

Importantly, there is little observed heteroskedasticity: our significance model's residuals are only very weakly correlated with the true significance labels ($\rho = 0.017$). However, insofar as there is heteroskedasticity, it is among the predicted high-significance documents. Documents which Chiou and Rothenberg estimate to be of high significance our model often overestimate as being highly significant, further justifying our coarsening. This is critical for performing additional analysis, as any systematic bias in our model's accuracy would subsequently bias any regression results for which we use our model's predictions.

B.4 Temporal Variation in Modeling Accuracy

A potential criticism of this significance modeling approach is that our model may underestimate the significance of documents whose text is unlike the text of numbered Executive Orders or proclamations in our training set. Consider, for example, a model trained only on data from the 1940s, used to evaluate the significance of documents from the 2010s. Due to changes in language over time, that model may be unlikely to perform well. The same result holds, though, if there are *more* documents from the 1940s than there are from the 2010s.

To test this, we perform a similar cross-validation procedure as in our main results, except instead of each fold consisting of random subsamples, each subsample is a decade of text. This allows us to test whether our model fails to accurately estimate the significance of documents

from time periods outside the training set. As Figure B.1 shows, the temporal cross-validation accuracy is not substantially lower than the randomly partitioned cross-validation accuracy, providing evidence that the lexical cues indicating document significance do change over time. This gives us confidence that our model is robust to the relatively mild changes in language usage we observe, though we still acknowledge that much earlier documents pose a significant estimation challenge. However, as we discuss in the Results section, any measurement error induced by this estimation challenge should serve only to reduce the absolute magnitude of our regression coefficients.

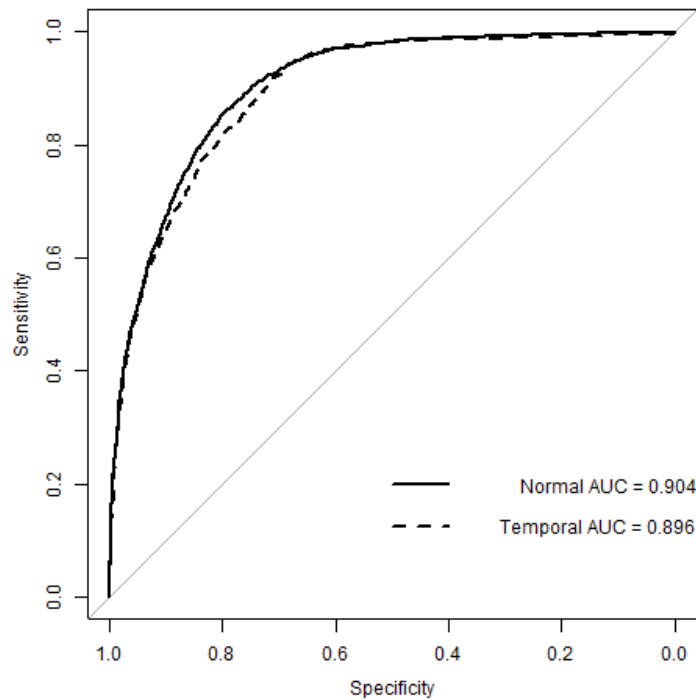


Figure B.1: The cross-validated AUC for the randomly partitioned analysis is not substantially lower than the temporally partitioned analysis, indicating that the model has significant capacity to estimate document significance outside its temporal support.

B.5 Feature Importance

As a face validity check on our significance model, we perform a descriptive feature importance analysis. Since random forest models are largely black boxes where features enter and predictions are returned, determining which covariates contribute most to the model's success can be difficult. One commonly used method to extract feature importances from tree-based models involves "feature depth" (Archer and Kimes 2008). Since random forests consist of decision trees that are ordered variable splits, features that systematically appear earlier in the decision tree are more important to the model. A covariate's feature importance, then, is proportional to the average number of times that feature appears in the decision tree, weighted by how early in the tree it appears; more simply, higher values indicate more strongly predictive features.

Note that this is purely a descriptive exercise; if our model performs as we intend, we expect that the most important features it identifies relating to policy significance will be ones which intuitively discriminate significant orders, which discuss tariffs, military conflict, and industry, from ceremonial ones, which memorialize the dead or declare National Ice Cream Day. We calculate feature importance for random forests model and present the 30 most important terms below in columns 1 and 2 of Table B.1, then 20 largely insignificant terms in columns 3 and 4. Among the most significant words are section, pursuant, provision, necessary, and articles, which generally indicate appeals to either constitutional or statutory authority. Low-significance terms include liberty, anniversary, Thanksgiving, bravery, and victims. A feature importance table derived only from executive orders follows in Table B.1.

An initial inspection of our approach to coding presidential documents, therefore, suggests that we have uncovered a meaningful dimension that distinguishes actions based on whether they address consequential policy issues or more are ceremonial in nature.

Similarly, we expect a different but related set of important features for our policy classifier: terms related to the policies in question. We find that housing, education, land, labor, defense, agricultural, and discrimination are all important policy-related key words; advancing, stronger,

undermine, desired, and everywhere are policy-unrelated ones (Table B.2).

B.6 Robustness to Threshold Choices

In two places in our measurement strategy we identify thresholds for indicating significant documents. First, we select a threshold of 0.50 in coarsening Chiou & Rothenberg’s significance scores into a dichotomous indicator. Second, after generating predictions from our random forest model, we identify documents with a probability of ≥ 0.355 as being significant.

The first threshold of 0.50 is the result of our qualitative reading of a large number of these documents. As we argue in Section , our decision to threshold here reduces measurement error vis-à-vis Chiou and Rothenberg (2014, 2017), but also turns our modeling problem from a regression problem on a scale from -1 to 3 with difficult interpretability to a much easier and more reliable classification problem. The outcome of our model can be interpreted as the probability that a document would receive a Chiou and Rothenberg score of greater or less than our threshold of 0.50. This interpretation likewise allows us to bootstrap standard errors. The choice of the threshold itself at 0.50, rather than at 0 or at 1, is the result of extensive qualitative readings of these documents.

The second threshold is of the modeling results. Our model produces probabilities that a document would receive a Chiou and Rothenberg score of greater or less than 0.50. To convert these probabilities into a prediction of whether a document is significant, we identify a second threshold.²⁸ We identify 0.355 as the estimated probability at which we declare a document “significant.” While we could reasonably choose 0.50, or 0.95, or many other thresholds, we select 0.355 since it is the value which equalizes the false-positive rate and the false-negative rate. This is a desirable property since it mitigates bias and heteroskedasticity in our regression analyses.

While we perform robustness checks to show that our substantive results are not sensitive to

²⁸Note that measuring AUC as above does not require selecting a threshold here.

our choices of these thresholds, we acknowledge that these choices induce “researcher degrees of freedom” (Simmons, Nelson, and Simonsohn 2011). In the interests of transparency and reproducibility, our complete replication file will be made public on the Harvard Dataverse upon publication. Importantly, the replication code will include clearly demarcated points where we perform our coarsening allowing interested researchers to experiment with alternative thresholding decisions.

C Supplemental Appendix: Alternative Polynomial Specifications

Figure C.1: Directive Substitution across Policy Areas (Linear specification)

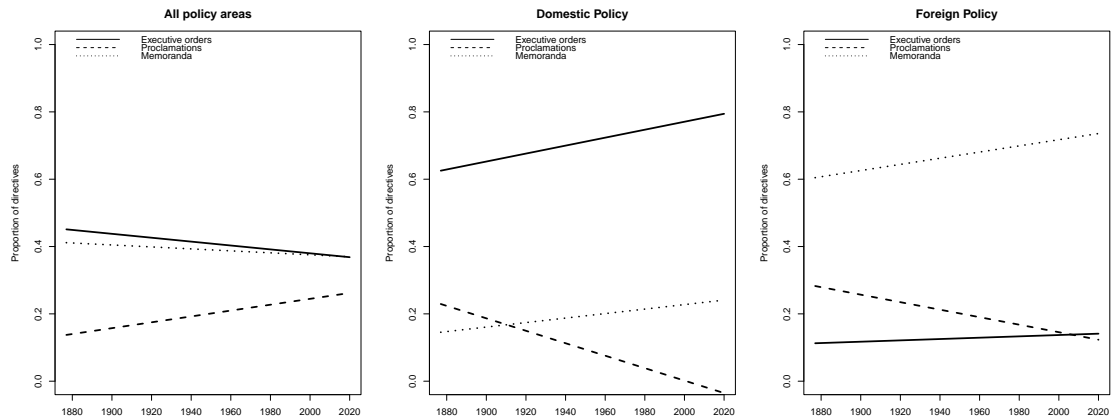


Figure C.2: Directive Substitution across Policy Areas (Third degree polynomial)

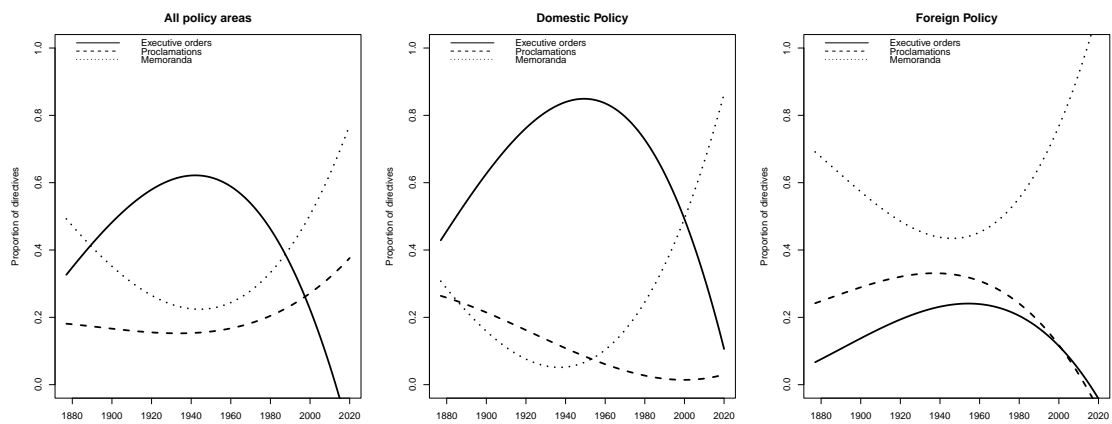


Figure C.3: Directive Substitution across Policy Areas (Fourth degree polynomial)

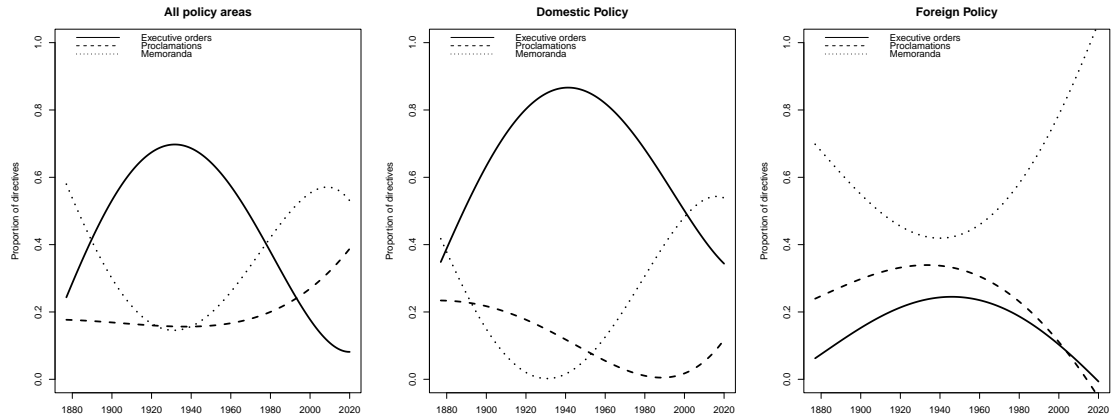
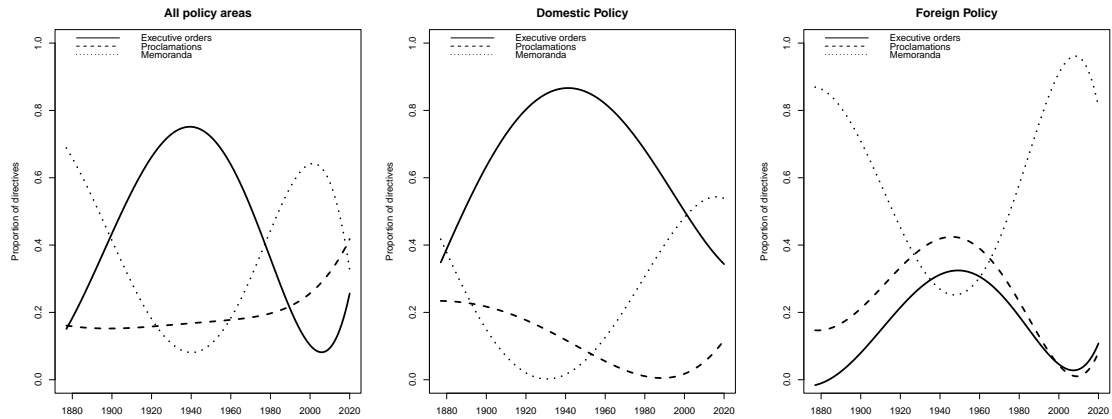


Figure C.4: Directive Substitution across Policy Areas (Fifth degree polynomial)



Term	Importance	Term	Importance
section	6.456	billion	-0.067
pursuant	5.692	liberty	-0.068
provisions	4.422	working	-0.068
necessary	3.946	remember	-0.069
proclamation	3.616	anniversary	-0.069
including	3.561	sacrifice	-0.07
withdrawn	3.342	choices	-0.07
entered	3.302	countless	-0.07
warehouse	2.976	college	-0.07
consumption	2.85	greatest	-0.07
follows	2.771	patients	-0.07
purposes	2.545	victims	-0.07
modified	2.545	mothers	-0.071
respect	2.196	productive	-0.072
effective	2.176	courage	-0.073
americans	2.081	generations	-0.074
secretary	2.043	generation	-0.075
authority	1.999	substances	-0.075
provided	1.961	disabled	-0.077
determined	1.955	helping	-0.077
actions	1.955	challenges	-0.078
articles	1.915	thanksgiving	-0.079
certain	1.862	safeguard	-0.08
register	1.832	equality	-0.082
executive	1.823	millions	-0.084
modifications	1.811	inspiration	-0.084
limited	1.757	success	-0.085
schedule	1.739	transition	-0.086
amended	1.729	bravery	-0.086

Term	Importance	Term	Importance
housing	100.944	preclude	0.113
recreational	47.765	attended	0.113
education	46.919	discipline	0.112
revokes	42.538	realized	0.112
reserved	41.118	presente	0.112
refugees	35.577	pittsburgh	0.11
interior	34.484	everywhere	0.109
executive	33.678	advancing	0.109
agriculture	33.131	ciation	0.108
educational	31.274	package	0.107
defense	31.223	artictc	0.107
section	30.958	disclosed	0.106
agricultural	30.232	benjamin	0.105
president	29.819	complaint	0.104
appointment	29.419	managing	0.102
discrimination	27.994	proximately	0.102
federal	27.564	madison	0.102
secretary	27.555	christopher	0.102
meridian	26.809	providers	0.101
announced	26.03	conveyed	0.101
approved	24.888	shipped	0.1
roosevelt	24.791	stronger	0.1
franklin	23.716	desiring	0.099
ordered	23.6	empower	0.098
dispute	21.912	specification	0.098
national	21.504	remarkable	0.097
manager	21.491	attempted	0.096
amended	21.414	sanction	0.096
sections	21.35	undermine	0.096