

BUREAUCRATIC INFORMATION IN CONGRESS*

Luca Bellodi[†]

November 20, 2023

Abstract

The ability to cultivate expertise and produce information is a key source of legitimacy of unelected bureaucracies. However, very little is known about the extent to which bureaucratic information is actually used by legislators when making policies. In this paper, I introduce a novel measure of legislators' reliance on bureaucratic information which uses natural language processing to extract and analyze bureaucratic information used by members of Congress in 8.3 million floor and committee speeches given over the past 40 years. I find that legislators make greater use of information coming from ideologically similar bureaucracies. However, statutory features insulating agencies from political control sharply reduce the effect of ideological distance. These findings have implications for theories of bureaucratic legitimacy and for the use of evidence in policy-making. Institutional features granting independence to bureaucracy can depoliticize the role of bureaucratic information in policy-making.

Word count: 10,012

*For helpful comments, I am thankful to Alex Acs, Elliott Ash, Jack Blumenau, David Coen, Fabrizio Gilardi, Giovanna Invernizzi, Mary A. Kroeger, Benjamin Lauderdale, David Lewis, Kenneth Lowande, Massimo Morelli, Tom O'Grady, Christoph Pfeufer, Colin Provost, Jennifer Selin, Arthur Spirling, Brandon Stewart, Matia Vannoni, Hye Young You, and seminar participants at Stanford, LSE, and UC Berkeley GSPP. Usual disclaimers apply.

[†]Postdoctoral Fellow, Bocconi University.

In a statement given on 8th July 2015 during a session of the US Senate Committee on Environment and Public Works, Senator Barbara Boxer (D-CA) highlights some of the positive outcomes of President Obama's Climate Action Plan, an ambitious set of measures aimed at cutting carbon emissions.

The recent study by the Environmental Protection Agency shows us 57,000 fewer deaths per year from poor air quality, with economic benefits valued at \$930 billion, 12,000 fewer deaths per year from extreme heat and temperature changes, \$180 billion per year in avoided damages from water shortages, \$3 billion per year avoided damages from poor water quality, \$11 billion a year avoided losses in our ag sector, 40 to 59 percent fewer severe and extreme droughts and almost 8 million fewer acres burned each year from wildfires.

In the speech, the Senator cites evidence produced by the EPA to persuade the Republican-controlled committee and its chairman Senator Jim Inhofe (R-OK) to take action against climate change. These are the concluding words of her speech: "I feel stronger than ever the President is on the right path. This Committee is on the wrong path." Far from being an isolated case, my data shows that during the past 40 years the information produced by the EPA has been used in congressional speeches other 9,202 times by a total of 1,113 different members of Congress.

Due to their expertise, bureaucratic agencies produce a great wealth of information that can be used by politicians for both policy and electoral goals (Niskanen 1971; Wilson 1989; Gailmard and Patty 2013). Bureaucratic information can help lawmakers persuade political opponents, make better policies, or frame to their favor the debate around certain policy issues. At the same time, citing bureaucratic information can represent a form of position-taking, with members of Congress signalling effort and commitment to policy to their constituents and donors (Mayhew 1974; Maltzman and Sigelman 1996; Przeworski, Stokes, and Manin 1999; Grimmer 2013). Producing expertise is ultimately the main reason why unelected bureaucracies are delegated significant discretion in administering policy and there is a long tradition of work in political science that conceives bureaucracies as shaping policy through information (Weber

1922; Aberbach, Putnam, and Rockman 1981; Workman 2015). However, the scholarship has struggled to document and study the extent to which lawmakers rely on such information in the legislative arena. In this article, I remedy this lack of theory and data on the use of bureaucratic information in Congress, a question that has broad implications for evidence-based policy-making and for the legitimacy of unelected bureaucracies.

Theoretically, I show that both the preferences and the statutory design of bureaucracies matter for lawmakers' decision to use bureaucratic information. Consistent with the whole delegation literature (Miller 2005), I argue that the information asymmetries characterizing the relationship between political principals and bureaucratic agents limit legislators' ability to directly evaluate the accuracy of the information produced by bureaucracy and its compatibility with their own political goals. Therefore, legislators use the ideological leaning of bureaucracies to verify whether the information is beneficial to their political enterprises and hence use it when making policy. When exposed to information coming from ideologically distant agencies, legislators might fear that the information is partial or incorrect, as the agency can manipulate information to achieve its own political goal. Climate-change skeptics are likely to believe the EPA report mentioned by Senator Boxer is blatant propaganda. Even if they thought the information was correct, they would likely disagree with the political implications of a report perceived to be at odds with their own preferences and those of hard-line conservative voters and donors. Yet an alternative decision criterion exists which can supplement ideology when legislators verify the reliability of the information and its source. When agencies are insulated from political pressures and enjoy a high level of statutory independence, the information they produce is perceived to be more credible and less politicized by members of Congress, and the role of the ideological divide weakens.

To test this argument, I present the first attempt at studying legislators' use of the information produced by hundreds of US federal bureaucracies over the past 40 years. I introduce

a new measurement strategy that uses natural language processing to detect when MCs use quantitative evidence and statistical facts produced by bureaucratic bodies in their speeches and apply it to an original corpus of 8.3 million speeches given by US members of Congress in floor and committee sessions. First, I apply dependency parsing to the corpus of speeches and extract legislators' quotes of bureaucratic information. Second, I compute the extent to which the quote contains statistical facts and quantitative evidence, which allows me to produce continuous estimates for every sentence mentioning an agency. This measure yields a comprehensive picture of the role of bureaucratic information in Congress over a large period of time, for a wide set of agencies and legislators with different partisan affiliations.

I present descriptive evidence alongside a rigorous test of the role of ideology in the MCs' consumption of bureaucratic information. A key stylized fact observed in the data is that Republicans engage less with bureaucratic information compared to Democrats, but only in the floor. In committees, members of the party controlling the Presidency – and, as a result, the government agenda – make greater use of information coming from bureaucratic agencies. Consistent with recent anecdotal evidence, the data shows that Republicans make greater use of the expertise of law-enforcement agencies, whereas Democrats are those almost exclusively using the information produced by the Centers for Disease Control and Prevention, perhaps an indicator of the different (ideological) approaches of Republicans and Democrats to the COVID-19 pandemic. I leverage *within*-agency changes in ideological leaning resulting from bureaucratic turnover in leadership positions across presidencies and compare the frequency of statistical facts and quantitative evidence in legislators' statements mentioning bureaucracies with a series of fixed effects models. I find strong support for the proposed ideology-driven account. However, the effects are driven by executive departments and sub-agencies under tighter control of the President, which lends support to the moderating effect of agency independence. I strengthen these results with a difference-in-differences design which exploits a 2020 Supreme

Court decision curtailing the independence of the director of the Consumer Financial Protection Bureau, and find that, after the Court’s decision, Democrats are significantly less likely to use information from the Trump-controlled Bureau.

This paper makes three contributions to the literature on the political use of information and politicians-bureaucracy interactions. First, while most of the scholarship on politicians-bureaucracy relations focus on how Congress and Presidents control drifting bureaucracies (e.g., Fiorina 1981; McCubbins, Noll, and Weingast 1987; Bolton, Potter, and Thrower 2016; Lowande 2018; but see Moe 2006), in this paper I show that bureaucracy can play an important role in congressional politics, by accounting for a prominent source of information and quantitative evidence at legislators’ disposal. Second, I present the first and largest measure of the role of bureaucratic bodies in legislative politics, presenting fine-grained data for 316 agencies and approximately 40 years of floor and committee speeches. Finally, I introduce a new, transparent, and objective way of measuring politicians’ reliance on different sources of information, which can be used to study other questions about information provision and usage across different fields in political science. The findings also have implications for the institutional design of bureaucratic agencies, suggesting that statutory features granting independence to agencies can counteract the ideological polarization underlying legislators’ decision to use bureaucratic expertise in Congress.

Bureaucratic Information in Legislative Politics

Bureaucratic agencies can have a significant impact on politics even outside their implementation domain. Krause (1996), for instance, describe the agency-political relations as a “two-way street”, where preferences of politicians towards agencies are affected by agency performance, and Carpenter (2001) shows that, during the US Progressive Era, bureaucratic agencies enjoying a good reputation across multiple audiences were able to secure their desired policies despite

the opposition of elected politicians. In fact, despite bureaucrats being legally subordinated in the hierarchy of government, bureaucrats can play major roles in determining the direction of policies (Moe 2012). Zooming in on legislative politics, Nicholson-Crotty and Miller (2012) find a positive relationship between the agency’s perceived effectiveness and politicians’ perceptions of bureaucratic influence on legislative outcomes, while Ingold and Leifeld (2016) find that vertically integrated offices with access to formal decision-making venues are on average perceived as more influential. Blom-Hansen, Bækgaard, and Serritzlew (2020) implement a series of experiments simulating the decision-making process and find that politicians are likely to rely on bureaucrats’ expertise and information, especially when it is framed more positively.

While this scholarship made important advancements in the study of the role of bureaucracies in the policy-making process, they all rely on perception measures, hence we know little about the extent to which they capture real-world phenomena. Moving to observational data, Shobe (2017) shows how federal agencies play an important role as reviewer and editor of legislative texts, either by request from Congress or as a result of the agency’s own monitoring of legislation. Kroeger (2022) reports similar findings on state legislation and finds that bureaucracy-sponsored bills are more likely to be approved by the legislature when there is unified government and when the capacity of the legislature is weaker compared to that of the bureaucracy. A similar demand-side approach to study politicians seeking information is taken by Ban, Park, and You (2022), who map the universe of witnesses testifying before Congressional committees and show that bureaucrats – on top of being the largest category of witnesses – are important providers of analytical information and they are invited to testify mostly when legislators are exploring a legislative issue and are open to acquiring new information. Similarly, when looking at bureaucrats’ incentives to supply information, Ban, Park, and You (2023) find that when appearing before committees, bureaucrats supply more analytical information to legislators who are presidential co-partisans, suggesting that ideology is key not

only to legislators' decision to use the information, but for bureaucrats' decision to supply it too.

In this paper, I build on these recent attempts at bringing evidence to the study of the role of bureaucracy in legislative politics by providing the first comprehensive picture of the extent to which the information produced by the US federal bureaucracy is used in Congressional politics.

Ideology and Independence as Verification Devices

Members of Congress are moved by a combination of re-election and policy goals, and giving speeches – arguably one of the main activities of elected officials – can be a powerful tool to communicate effectively to constituents and interest groups, and to frame a policy-problem to their favor (Mayhew 1974; Grimmer 2013; Grimmer, Westwood, and Messing 2014; Lee 2016). Bureaucracies are a one-stop-shop for MCs seeking to acquire policy information (Ban, Park, and You 2023). Members of Congress might report what said by an independent agency or a department to claim credit for the success of a program or to direct voters' attention to a specific issue. Reputable agencies can be mobilized to increase the salience of debate and change the political agenda, or to persuade other members with credible information. At the same time, members of Congress know that bureaucracies can use information to advance their own preferred policy or that of an out-partisan President, and they might be hesitant about using information coming from ideologically distant bureaus.

In practice, changes to the ideological leaning of agencies occur as a cascade: a new President is elected and, through appointments and turnover among agency leaders, she influences the ideological slant of agency communications and outputs. Let us consider a conservative member of Congress exposed to some information coming from the (liberal) Environmental Protection Agency under a new Democratic President. The preferences of both actors towards

environmental regulations are known, and the conservative politician is likely to think that the EPA is distorting the information it releases, perhaps because captured by environmentalist groups or subject to the directives of a President with a clear liberal agenda. Notice that the information advantage of bureaucracy does not mean that lawmakers are always incapable of detecting the ideological slant of agency reports and studies. Even if the conservative member had expertise on environmental policy and could isolate accurate information from biased information, she would quickly realize that the political implications of the information are at odds with her own policy preferences or with the preferences of their own constituents, and would ultimately decide to disregard it. Therefore, legislators will be skeptical towards the information produced by ideologically distant bureaucracies either because they believe the information itself is biased or because they anticipate it will not be received favorably by their constituents.¹

HYPOTHESIS 1: MCs' reliance on bureaucratic information decreases with the ideological distance between the legislator and the bureaucracy.

One important assumption of this argument – consistent with the entire principal-agent literature on information asymmetries and politicians' delegation of authority to bureaucracy (Gailmard and Patty 2012) – is that imperfectly informed legislators cannot rely on an independent system which verifies the quality and political compatibility of the information produced by agencies, and resort to ideology as a heuristic when deciding whether to use the information or not. However, suppose bureaucracies were designed in such a way as to ensure that the information they produce is not distorted for political goals or as a result of capture by interest groups or other political or economic elites. In this scenario, even though politicians might disagree ideologically with the bureaucracy, they are more likely to believe that the information is

¹Although here I consider bureaucracies' supply of information exogenous, the intuition behind cheap talk models of strategic communication yields similar predictions on legislators' decision to use information produced by bureaucracy (Crawford and Sobel 1982; Gailmard and Patty 2012). Because the receiver of the information (the member of Congress) cannot verify the veracity of information, information exchange is greater when the sender (the bureaucratic agency) of the information and the receiver have similar preferences over policy.

not biased. When members of Congress know that the source of information is an independent body, they receive an additional signal about the information agencies produce, which reduces uncertainty over the veracity of the information and makes it easier for the legislator to assess whether it is compatible with her own political strategy. Moreover, information produced by less politicized organizations would be received more favorably by MCs' constituents and donors. When agencies are granted statutory independence which insulates them from political control, members of Congress can rely on an alternative verification device. Statutory independence, acting as a credibility-enhancing device, tempers politicians' skepticism towards information produced by ideologically distant agencies.

Enhancing the credibility of policy outputs is the main goal of institutional reforms granting statutory independence to bureaucratic bodies (Keeper and Stasavage 2003). By delegating independence to agencies, elected politicians separate politics from administration, reducing the degree of control that the government exerts on bureaucracy or the risk of capture by interest groups. Another reason for granting independence is to promote agency specialization. As in Gilligan and Krehbiel (1987), principals might decide to restrict the array of procedures aimed at controlling the agent in order to promote the agent's incentive to specialize and acquire information, especially when the agent is ideologically apart from the parent body. A clear example of such commitment is the independence of central banks and regulatory agencies for the credibility of monetary policies, for controlling inflationary tendencies, and for ensuring a level-playing field for public and private businesses (Cukierman, Webb, and Neyapti 1992; Keeper and Stasavage 2003). Empirical work has shown how agency independence can improve bureaucratic policy-making, in particular the perceived and objective quality of regulation (Bertelli and Whitford 2009; Koop and Hanretty 2018). Free from undue influence, independent agencies – and the statutory provisions which define their relationship with political officials – make the ideological leaning of the agency less salient for legislators, who are more likely to use the

information produced by agencies even though they are perceived ideologically apart.

HYPOTHESIS 2: *The (negative) effect of ideological distance on MCs' use of bureaucratic information is weaker when the information comes from more independent agencies.*

This account shows how institutional features of bureaucracy can moderate the effect of the ideological divide between agencies and members of Congress.

A New Measure of Legislators' Use of Information

The role of bureaucratic information in the legislative process has generally been measured either qualitatively or through perception measures. While qualitative measures, though benefiting from “deep” observation and multiple sources of data, are limited to a few cases, answers to perception questions, like “How often do you rely on information from agency x ?”, are easily susceptible to social desirability bias.

I present a new objective, replicable, and large-scale measurement strategy that captures the extent to which bureaucratic expertise is used in Congress by applying natural language processing techniques to a large corpus of floor and committee speeches given by the universe of US lawmakers, detecting when MCs use agencies' information and extracting what type of information they use. This measurement strategy has quantitative and qualitative advantages over existing methods. First, by looking at floor and committee speeches, I am able to trace how legislators use the evidence produced by a large set of bureaucratic bodies over a long period of time and on a daily (or debate) basis. Second, by observing how MCs' use information from sources with different fixed and time-changing characteristics, this measurement strategy is uniquely suitable for theory-testing and comparative analysis. Third, I am able to measure the intensity of information usage, namely the frequency and the extent to which the information used by legislators is dense with factual information.

Information Extraction

The key assumption of the proposed measurement strategy is that MCs' use of bureaucratic information can be detected by parsing the syntactic relations of terms in segments of text (e.g., sentences). Syntactic analysis can identify the action of saying something, the subject carrying out the action, and the object of the action. Let us consider a MC saying "The Federal Reserve [*subject*] said [*action*] that higher interest rates will slow inflation [*object*]". By creating extraction rules that detect certain syntactic relationships, I can therefore match every instance in which a bureaucracy is used as a source of information in a speech and then analyze the type of information that is being used. Syntactic analysis and dependency parsing are new frontiers in political science research, but a few promising applications show the benefit of retaining dependency relationships between words when analyzing text. Atteveldt et al. (2017), for instance, shows how US and Chinese media portrayed differently the role of Hamas and Israel in the 2008-9 Gaza war, and Vannoni, Ash, and Morelli (2019) apply syntactic analysis to a corpus of US state laws to estimate delegation of powers to governors of US states.

The measurement strategy I propose consists of three steps. First, I split every speech mentioning the name of an agency into sentences, tag parts of speech (e.g., subject, verb, predicate, etc.), and detect dependency relations. Second, I extract clauses that match pre-defined syntactic frames capturing different ways in which MCs can use bureaucratic information. Third, I isolate the quote, namely the actual piece of information used by legislators, and measure the extent to which the quote reports statistical facts and quantitative evidence. Eventually, I obtain a sample of sentences where agencies are used as sources of information, and every sentence will receive a continuous measure equal to the frequency of statistical facts and quantitative evidence contained in the quote.

Table 1: Dependency Parsing.

Token ID	Token	Part-Of-Speech	Head Token ID	Dependency Relation
1	The	DETERMINER	2	determiner
2	FED	PROPER NOUN	3	nominal subject
3	said	VERB	3	ROOT
4	that	ADPOSITION	9	marker
5	higher	ADJECTIVE	7	adjectival modifier
6	interest	NOUN	7	compound
7	rates	NOUN	9	nominal subject
8	will	VERB	9	auxiliary
9	slow	VERB	3	clausal complement
10	inflation	NOUN	9	direct object
11	.	PUNCT	3	punctuation

Notes: Output of dependency parsing. Each token is assigned an ID, which is used to describe syntactic dependency relations between tokens. Token IDs in bold used as example in text.

Step 1: Parts-of-Speech Tagging and Dependency Parsing

I tag and parse the sentences with SpaCy, a supervised learning algorithm which achieves state-of-the-art performance on several NLP tasks like part-of-speech tagging and dependency parsing (Choi, Tetreault, and Stent 2015; Honnibal and Johnson 2015). After splitting speeches into sentences, the parser tags parts of speech and detects dependency relations between words.

For instance, let us consider the sentence “The FED said that higher interest rates will slow inflation”. The tokens – namely every single word – within this sentence have syntactic properties and follow specific dependency relations. For instance, “The” refers to the “FED”, which in turn is the nominal subject of the verb “to say”. The result of syntactic parsing is displayed in Table 1, which reports the token ID, the token (i.e., the word), the part-of-speech, the ID of the head token (i.e., the “parent” token), and the type of dependency relation. For instance, the head token ID of the words “higher” and “interest” is the token ID 7, “rates”.

Step 2: Extraction Rules

Once the parser has tagged each token of the sentence, I annotate the sentence based on extraction rules that detect quotes, namely instances where somebody is reporting (*i*) something said, written, or released by someone, (*ii*) the source of the information contained in the quote,

and (iii) the content of the quote. I create two comprehensive sets of extraction rules that match who-says-what syntactic structures: one that captures direct and indirect statements of agencies (“the FED said”, “as said by the FED”) and “according-to” structures (“according to the FED”); while the other captures direct or indirect outputs of agencies (“the FED’s proposal is”, “the FED’s proposal to”, the “FED’s study suggests”).

To match direct and indirect statements, I specify a vector of “say verbs” so that the parser marks the lemmatized version of the verb – therefore capturing verbs declined in every form (active or passive) or tense – and its respective subject or, in case of an indirect statement, the agent.² For “according-to” structures, the parser detects the lemmatized token “accord” and the object of the preposition, which will be the source of the information. For direct and indirect nominal outputs, I specify a vector of output-related words for the parser to detect (e.g., study, proposal, recommendation, suggestion), and their possessive determiner or the object of prepositions such as “of”, “by”, or “from” – which mark the owner of the output – will be labelled as the source of the output³ When labeling the source of the information, I also include cases where individuals affiliated with the agency are producing information. For instance, the algorithm is able to mark the following direct statement by Representative Proxmire (D-WI) as a quote from the EPA: *“EPA’s Deputy Assistant Administrator for Radiation Programs has stated that if all Americans reduced the air infiltration in their homes by 50 percent, the resulting buildup of radon gas could eventually lead to an additional 10,000 to 20,000 cases of lung cancer a year.”*

Finally, all the tokens that are dependencies of say verbs, output-related verbs, or according-to structures are labelled as quotes. Table 2 reports the precise tokens and syntactic structures used to compile the extraction rules, and the toy sentences in which a legislator

²Importantly, I exclude instances where a negation is syntactically dependent of one of these “say verbs” (e.g., “The FED did not respond to my request” will not be marked). Similarly, I do not consider questions – namely sentences terminating with a question mark – when extracting quotes.

³Say-verbs and output-type words are reported in Section A in the appendix.

Table 2: Syntactic Frames.

Extraction Rule	Syntactic Structure	Sentence Example
Statements		
Direct Statement	subject + say verbs	The FED said <i>that higher interest rates will slow inflation.</i>
Indirect Statement	agent + say verbs	As reported by the FED, <i>higher interest rates will slow inflation.</i>
According-to Structure	accord + object of preposition	According to the FED, <i>higher interest rates will slow inflation.</i>
Outputs		
Direct Nominal Output	output + possession modifier	The FED’s [output] is <i>to increase interest rates.</i>
Indirect Nominal Output	output + possession modifier	I fully endorse the FED’s [output] <i>to increase interest rates.</i>
Direct Output	output + say verbs	A [output] from the FED indicates <i>to to increase interest rates.</i>

Notes: Syntactic frames designed to extract quotes from sentences with examples of sentences matching each frame.

could use the information produced by the FED, with the quote in italics.

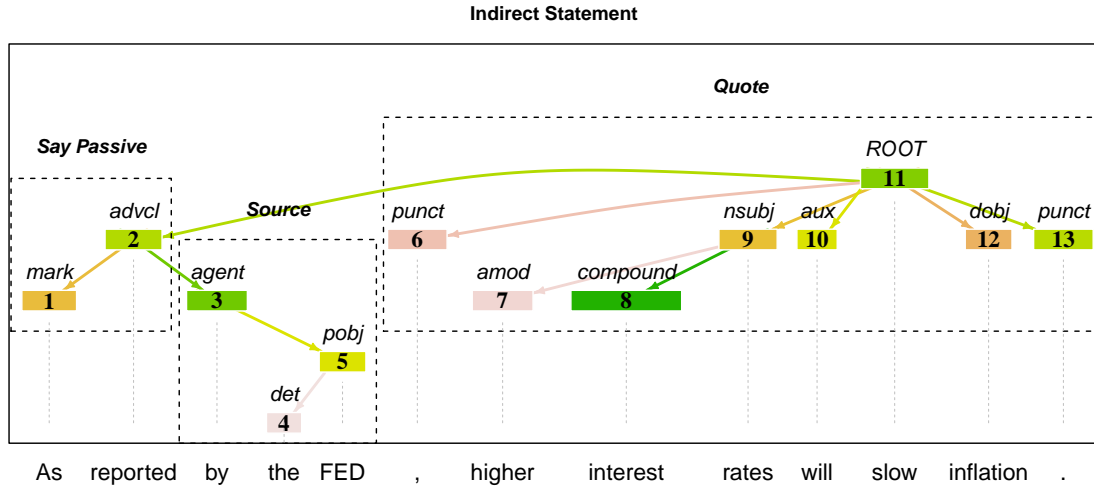
I then apply the extraction rules to the tagged sentences. Figure 1 shows the final output of the syntactic analysis for the example of the indirect statement, one that might seem particularly challenging to extract. Dependency trees of other rules are shown in Figure B.1 in the Appendix.

Step 3: Analysing Quotes

Using bureaucratic information does not occur by just citing bureaucracies. Politicians might report what is said by agencies with a negative tone (e.g., “The FED said nothing about it!”) or they could cite an agency without making any reference to policy (e.g., “The FED said that in the long term we’re all dead.”). Step three of the proposed measurement strategy addresses this issue by extracting qualitative information from the quotes, hence establishing whether the information used by politicians taps into the expertise of the bureaucracy.

For each tagged sentence identified as a quote, I extract the text of the quote and compute the frequency of statistical facts and quantitative evidence in the quote. To do that, I apply a simple dictionary-based approach, whereby every quote is assigned a score capturing the

Figure 1: Dependency Tree.



Notes: Dependency tree of an illustrative example where the FED’s information is used in a speech.

frequency of words belonging to a pre-defined dictionary of statistical facts and quantitative evidence. I use the licensed off-the-shelf LIWC dictionary (Pennebaker et al. 2015), which contains a comprehensive list of words related to quantifiers and numbers, such as “amount”, “approximately”, “average”, “entirety”, “equal”, “less”, “multiple”, “percentage”, “whole”, “twice”, “total”, as well as all digits and numbers used to express quantities, which I integrate with verbs capturing quantitative change (e.g., “increase”, “decrease”, “grow”, “drop”, etc.). In Table C.1 I report the full list of words used to build the dictionary.

The frequency measure I use is the term-frequency inverse-document-frequency (*tf-idf*), which is a weighted frequency that down-weights (up-weights) words that are in the dictionary but that appear in many (few) quotes, for they are less (more) useful at differentiating between quotes. The precise formula to build the metric is reported in the Appendix (see Section D). Sentences mentioning – without quoting – a bureaucracy will receive a score of 0, whereas when the agency is used as a quote, the score of the sentence will be its *tf-idf*.

As an example, the frequency of statistical facts and quantitative evidence in the following statement given by Senator Jim Inhofe (R-OK) equals 8, whereas the *tf-idf* equals 9.94. The tokens in italics are those matched by the dictionary.

“According to the US Department of Transportation, *every 1 billion* invested in

highway construction *creates 47,500 jobs and generates more than \$2 billion in economic activity.*”

Validation and Limitations

I present two validation tests for the measurement strategy, one focusing on the crucial step of quote extraction and one for the measurement as a whole. First, to ensure the parser can successfully detect instances in which legislators are using a bureaucratic agency as a source of information, I compare the performance of the automatic extraction of quotes to human judgement. I extract 200 random sentences which are classified as quotes by the parser and 300 which are not classified as quotes. I then take these 500 sentences and ask an independent coder to decide whether the sentence mentioning a name of an agency is using that agency as a source of information. Second, to ensure the proposed strategy is also able to capture legislators’ reliance on quantitative policy information, I rely on recent availability of large language models (LLMs) and their applications in zero-shot classification tasks. In particular, I prompt the GPT-4 language model and ask to decide whether, in the sentence, the member of Congress is citing policy information produced by the agency. As shown by Gilardi, Alizadeh, and Kubli (2023), LLMs can even outperform human annotators in text classification tasks and can hence provide new ways of producing and validating NLP pipelines. An example of the prompt and the answer given by the model is reported in the Appendix (see Table E.2). I then dichotomized the *tf-idf* of statistical facts so that quotes either contain or do not contain statistical facts and then compare the answers given by GPT and the ones given by the proposed measurement strategy.

Table 3 shows the performance metrics from the confusion matrices of the two classification tasks: quote extraction and statistical facts detection. The accuracy of both tasks is rather high (approximately .80) and the F1 score – assessing the balance between precision and recall – is satisfactory and above .70. These performance metrics suggest that the measurement strategy

Table 3: Performance Metrics of Validation Tasks.

	<i>Quote Extraction</i>	<i>Facts Detection</i>
Accuracy	0.80	0.79
Precision	0.70	0.64
Recall	0.79	0.81
F1	0.75	0.72

Notes: Performance metrics from the comparison of the proposed measure and manual coding (for quote extraction) and GPT output (for quote extraction and facts detection) of 500 sentences mentioning the name of a bureaucratic agency.

performs well at extracting quotes and detecting whether the quotes contain statistical facts.

Despite the validation exercises reported above, the proposed measure has limitations too. First, this measurement strategy is about implicit ways legislators could use the information produced by bureaucracies. By anchoring the quote to the name of the agency (or individuals whose affiliation with the agency appears in the text), the proposed method is only able to capture explicit ways of using bureaucratic information.⁴ Second, members of Congress could report the information produced by agencies while criticizing them in the overall speech. I rule out this possibility empirically, performing a sentiment analysis on the full text of speeches mentioning the name of a bureaucracy and estimating the effect of using the agency as a source of information (compared to simply mentioning the name) on the probability that the speech has a negative sentiment. I find a negative and precisely estimated relationship between speeches using at least one bureaucracy as a source of information and the probability that the speech is assigned a negative sentiment, even when including legislator and year fixed effects (see Section F in the Appendix).

Speeches and Bureaucracies

I apply the proposed method to a corpus of 1,634,602 floor and 6,654,065 committee speeches.

I scraped floor (1994-2022) and committee (2010-2022) speeches from the digitized version of

⁴In the Appendix, I show that the MC-agency ideological distance in a dyadic framework does not predict the probability that the MC mention the agency. This suggests that ideology matters for how – rather than whether – legislators engage with bureaucracy (see Section H3).

the Congressional Record and I obtained transcripts of congressional committee sessions for the period 1980-2009 from ProQuest.⁵ After replacing the various ways in which agencies are mentioned with a standardized name, I subset all speeches mentioning at least one agency. The list of agencies combines large samples of bureaucratic bodies from Bertelli and Grose (2011), Chen and Johnson (2014), Selin (2015), and Richardson, Clinton, and Lewis (2018) integrated with information on the type of agency directly obtained from the US government website (usa.gov/federal-agencies), for a total of 322 agencies.

A total of 285,255 floor speeches and 739,558 committee speeches mention the name of at least one agency, 17% and 11% of the total speeches, respectively. I parse these speeches into sentences and keep only the sentences which contain the name of an agency. I then apply the extraction rules described in Table 2 to each sentence. I extract the quotes from each sentence using bureaucracies as source of information and apply the dictionary analysis to the quote, measuring the *tf-idf* of words that belong to the dictionary of statistical facts and quantitative evidence. The average *tf-idf* is larger in floor speeches, accounting for 2.48, whereas it is equal to 1.94 in committee speeches. Descriptive statistics about the sample of sentences, quotes, number of agencies, and legislators are reported in Table 4.

To benchmark the frequency with which MCs rely on bureaucracy as source of information in their speeches, I compare the number of times legislators use bureaucracies as a source of information to the number of times they use competitive sources. I replicate the measurement strategy using a different set of sources including a comprehensive list of 54 newspapers and 138 think tanks. I do find that bureaucracies are used as a source of information 12 times more often than the most important newspapers and 34 times more often than the major think tanks of the country, suggesting that bureaucratic bodies are key players in providing information to Congress.⁶

⁵A note on the quality of the transcripts and the speech parsing steps are reported in Section G of the Appendix.

⁶The lists of newspapers and think tanks have been assembled from Wikipedia pages

Table 4: Descriptive Statistics.

Descriptive Statistics	Floor	Committee
Sentences with agency mentioned	917,480	1,430,393
Share of sentences with agency used as source	6.5%	5.4%
Years covered	1994-2022	1980-2022
Unique agencies used as source	248	273
Unique legislators	1,573	1,796
Average <i>tf-idf</i> of Statistical Facts		
In sentences with agency used as source	2.47	2.05
In all sentences mentioning an agency	0.16	0.11

Notes: Descriptive statistics of sample of sentences, number of unique legislators, agencies used as source of information, and average use of information both in sentences quoting agencies and mentioning agencies.

Ideological Distance

I build a measure of ideological distance between legislators and the agency as the absolute value between the ideal points of each legislator and each agency mentioned in the sentence. Data on legislators’ ideal points are obtained from Lewis et al. (2020) and consist of the first dimension of the DW-NOMINATE score. For agency ideology, I use the dataset assembled by Chen and Johnson (2014), for it is the only one that at the same time *i*) includes both cabinet- and non-cabinet-level agencies, *ii*) allows *within*-agency variation in ideology as a result of bureaucratic turnover across different presidencies, *iii*) and is on the same scale as the DW-NOMINATE scores, thus allowing to compute meaningful ideological distance metrics.⁷

The estimates are built using federal bureaucrats’ campaign contributions to individual politicians. The key assumption – validated by Bonica (2019) – is that campaign donations are strong predictors of policy preferences. The resulting estimates are weighted averages the DW-NOMINATE scores of legislators who receive donations from bureaucrats, with weights accounting for the amount of the donation, so that larger contributions receive a larger weight.

`/List_of_think_tanks_in_the_United_States` and `/List_of_newspapers_in_the_United_States` and are available in the replication package accompanying this paper.

⁷This dataset has been widely used in political science to study the political control of the bureaucracy (Lowande 2018), strategic appointments (Moore 2018), career paths of bureaucrats (Bolton, Figueiredo, and Lewis 2019), and rule-making (Ellig and Conover 2014; Potter 2019). Another dataset with time-changing bureaucratic ideal points on the same scale as members’ of Congress is the one produced by Bertelli and Grose (2011), though it includes only cabinet-level agencies. In Section *Robustness Analysis* I show the results are robust to using this measure.

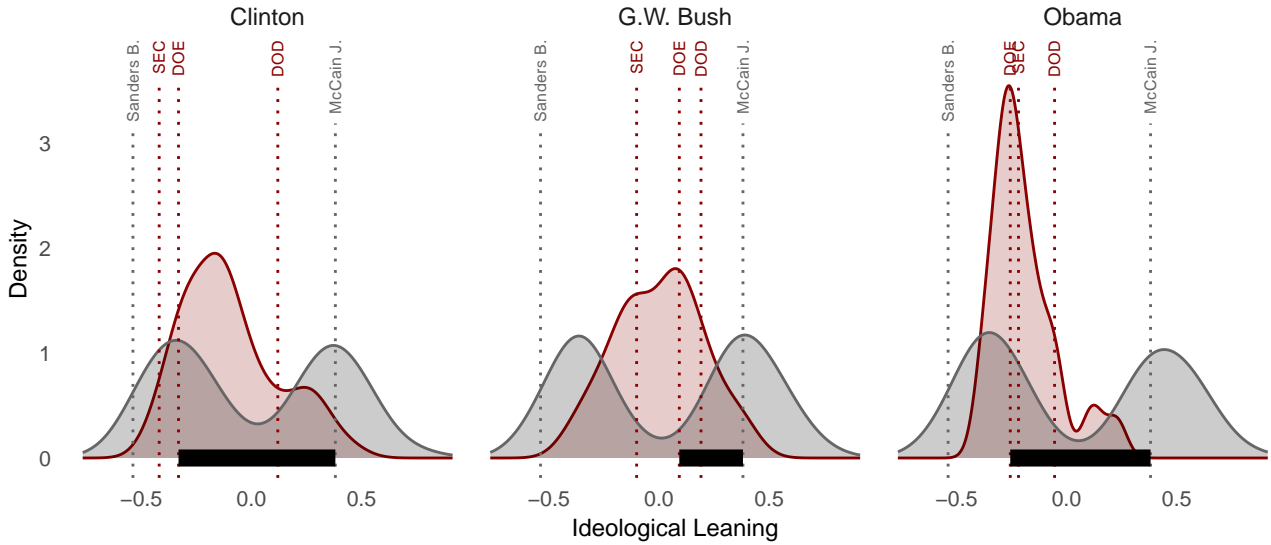
If, for instance, employees of agency i make equal donations to two legislators, the ideological score of the agency will be equal to the average of the ideal points of those two legislators. The assumption underlying these weights is twofold. First, as donations predict preferences, it is reasonable to assume that the quantity donated is correlated with the intensity of those preferences. Second, large donations are more likely to come from better-paid, upper-level bureaucrats, who in turn have more influence over agency policy-making. Clearly, all measures of agency ideology have limitations. In the case of the estimates produced by Chen and Johnson (2014), a non-trivial assumption required to obtain valid estimates is that the sample of bureaucrats who are donors is representative of the employees of the agency. To mitigate concerns about the robustness of the results, I replicate the main analysis with four additional measures of agency ideology, which I describe in Section *Robustness Analysis*.

Figure 2 shows the ideological distribution of bureaucracy and members of Congress, with the thick dark bar on the horizontal axis showing how I construct the measure of ideological distance between legislators and bureaucracy. From the plot, it is clear to see how the ideological distance between Sen. John McCain (R-AZ) and the Department of Education – an agency generally perceived very liberal – shrinking under the Bush presidency compared to the previous and following democratic presidencies.

Methods

There are three methodological concerns for identifying the effect of ideological distance on the use of bureaucratic information. First, on the legislator side, there could be many individual characteristics that are correlated with ideology, their engagement with bureaucratic agencies, and the extent to which they use bureaucratic information in speeches (e.g., education, socioeconomic background, but also legislators' level of attention to towards bureaucracy). Second, on the agency side, some agencies – e.g., executive departments or large independent agencies –

Figure 2: Agency and MCs Ideology.



Notes: Distribution of the ideology scores of bureaucratic agencies (in red) and of members of Congress (in grey) across the first presidencies of Clinton, G.W. Bush, and Obama. Thick black bars represent the changing ideological distance between two actors: Senator John McCain and the Department of Education.

might be more salient than others, and their salience might be correlated with their ideological leaning and their level of politicization. Third, the attention to bureaucracy and policy for each legislator can change over time, and the fact that legislators use frequently the information produced by certain bureaucracies might be the result of a legislator following closely a specific policy sector in that particular period of time or holding specific roles in committees, rather than being the result of more similar ideological positions.

To address these sources of omitted variable bias, I exploit within-agency variation in ideological leaning to account for agency-level heterogeneity in MCs' reliance on information. By estimating agency fixed effects I account for all time-constant differences between agencies, their statutory features (i.e., independent agencies, executive departments, executive sub-agencies, agencies within the office of the president, and government-owned corporations), their level of informal politicization, and their own policy domain. In addition, I include a set of MC-by-congress fixed effects to hold constant both time-invariant and time-changing characteristics at the legislator-level (e.g., their legislative activism, committee roles, seniority, etc.), as well

as other shocks that could affect the use of the information produced by agencies for every legislator in any given congress.

The sample on which I perform the analysis consists of all the sentences mentioning a bureaucratic agency in floor and committee speeches, so that I can compare politicians using bureaucratic information supplied by agencies that are ideologically close or far.⁸

In particular, I estimate the following model via OLS:

$$tf\text{-idf of Statistical Facts}_{ijat} = \gamma_{jt} + \delta_a + \beta \text{Ideological Distance}_{jat} + X_{ijat}\zeta' + u_{ijat} \quad (1)$$

where the outcome variable is the *tf-idf* of the statistical facts in sentence i , congress t , and given by legislator j , mentioning agency a . γ_{jt} are legislator-by-congress fixed effects, δ_a are agency fixed effects, and $\text{Ideological Distance}_{jta}$ is the time-changing ideological distance between legislator j and agency a in congress t . Variation in treatment status is given by the different ideological leaning of bureaucratic agencies (which changes every presidency) and the ideology of legislators (fixed over time). X_{ijat} is a vector of covariates, namely the length of the sentence, whether the sentence was given in a floor or committee speech and, in specifications without legislator-by-congress dummies, it includes legislator-level covariates: whether the legislator is a subcommittee chair, majority-party leader, minority-party leader, and legislative effectiveness score, which synthesizes several indicators about the proven ability of a legislator to advance her agenda items through the legislative process and into law (data from Volden and Wiseman 2020). β , the marginal effect of ideological distance on the use of statistical facts when mentioning bureaucratic agency, is identified by exploiting shocks to ideological leaning of bureaucracies resulting from changes to personnel happening with presidential transitions.

⁸In the Appendix, I show that the effects are similar if we look at the intensive margin of the use of information, using as analysis dataset only the sentences quoting bureaucracies (see Section H2), whereas the results from MC-agency dyadic regressions show that ideology does not predict whether MC mention agencies, hence the analysis does not suffer from a selection effect whereby MCs do not mention ideologically distant agencies in the first place (see Section H3).

Standard errors are clustered at the legislator level.

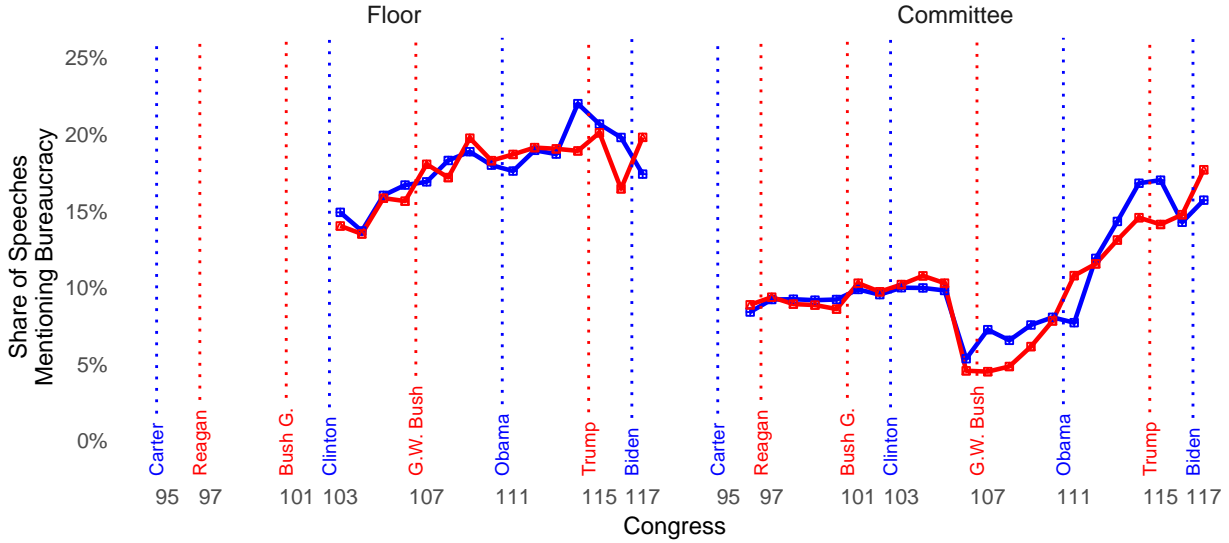
Results

Out of the 1,914 unique legislators who cite bureaucratic agencies in the period under study, Christopher H. Shays (R-CT) is the one citing bureaucracies the most during the 14 Congresses he served, for a total of 8,889 quotes. Representative Shays cited information produced by bureaucracy in 5.1% of the universe of statements he gave mentioning the name of a bureaucratic agency. Together with Daniel K. Akaka (D-HI), he ranks among the top two members of Congress relying the most on bureaucratic information even when looking separately at each Congress.

The proposed measure allows documenting three additional stylized facts about *i*) legislators' attention to bureaucracy, *ii*) partisan differences and *iii*) agency-specific differences in the use of bureaucratic information in Congress. First, as shown in Figure 3, both Democrats and Republicans display similar attention to bureaucracy. The figure shows the share of speeches given in floor and committee sessions by Republicans and Democrats simply mentioning a bureaucratic agency, with the dotted vertical lines marking changing presidencies. Interestingly, Republicans engage with bureaucratic agencies to a similar extent as Democrats, with the average share of both parties' speeches mentioning bureaucracies equal to 18% and 10%, in floor and committee sessions, respectively. Moreover, while the baseline level of attention is larger in the floor than in committees, the trend is markedly growing for both parties.

The second stylized fact is about partisan differences in the use of bureaucratic information. In the aggregate such differences are extremely thin. On average, both Republicans and Democrats use information produced by bureaucracy in approximately 6% of their statements mentioning bureaucracies and – among these sentences – the average *tf-idf* is approximately 2.3. However, differences emerge when looking at the over-time variation.

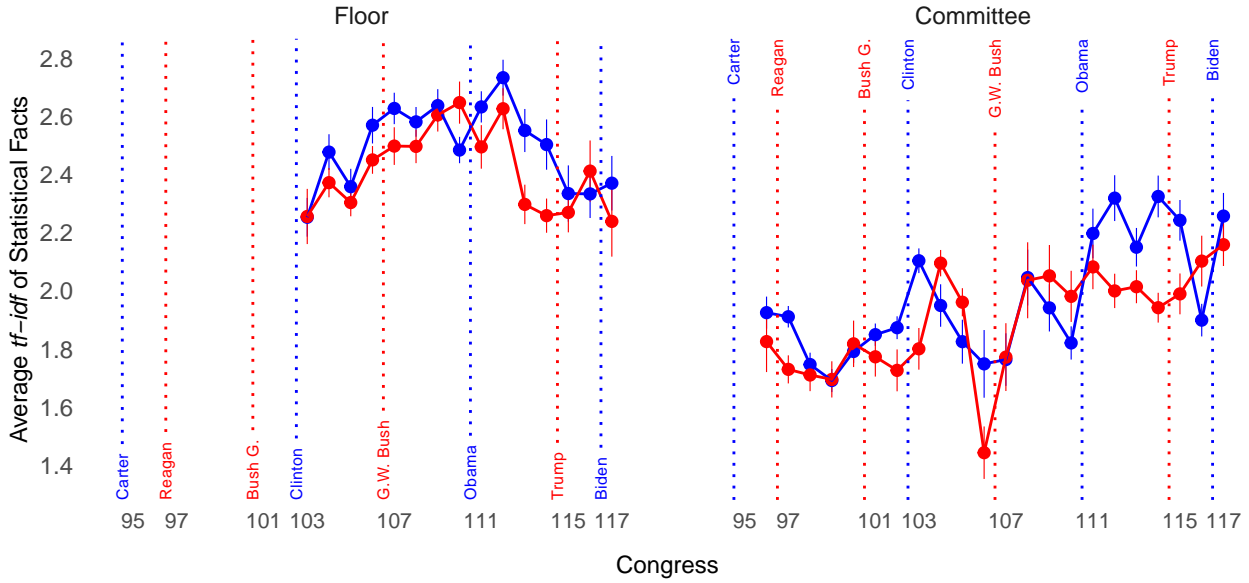
Figure 3: Partisan Attention to Bureaucracy.



Notes: Share of speeches mentioning a bureaucratic agency given by Republican (in red) and Democratic (in blue) members of Congress in floor and committee speeches. Vertical dotted lines mark changing presidencies.

Figure 4 shows the average *tf-idf* of statistical facts in legislators' statements citing bureaucratic agencies. From the 103rd Congress, when data for both floor and committee speeches is available, there is a clear partisan pattern in legislators' use of bureaucratic information. Republicans under Republican presidencies make greater use of bureaucratic information both in floor and in committee speeches (difference between Republican and Democratic Presidents significant at 95% level). The partisan premium affects Democrats too, but in committees alone. This partisan dynamic is particularly marked from the second Obama presidency, when Democrats' use of bureaucratic information is greater than Republicans' by 11% in floor and 20% in committees. The government party, which is responsible for policy-making and for executive action, is associated with greater use of bureaucratic information, especially in committee sessions, when substantive debates around policy occur. The data clearly shows that Republicans and Democrats overtake one another in terms of the average use of information, following a cyclic trend which mirrors the changing presidencies. These patterns are not consistent with accounts which emphasize Republicans' anti-administrative state rhetoric and firmly reject Republicans' alleged disengagement with bureaucratic expertise.

Figure 4: Partisan Reliance on Bureaucratic Information.



Notes: Average use of bureaucratic information across all agencies for every Congress in floor and committee speeches. Blue for Democrats, red for Republicans.

The third stylized fact moves to agency-specific differences. The data shows that the EPA and the Government Accountability Office are the bureaucracies whose information is most often cited by both Republicans and Democrats. However, there are marked partisan differences in the extent to which information produced by specific agencies is used. Perhaps unsurprisingly, Democrats cite the EPA 33% more often than Republicans, and when they cite it, in their quotes the frequency of statistical facts is 9% greater than Republicans'. Similarly, from 2019 Democrats were more likely than Republicans to use the information produced by the Centers for Disease Control and Prevention and the FDA – which were leading actors in the government’s response to the COVID-19 pandemic (difference in means statistically significant at 95% level). Conversely, Republicans’ use of facts produced by drug- and immigration-enforcement agencies is 9% larger compared to Democrats.⁹

Moving from stylized facts to the statistical analysis, Table 5 shows the main regression results. Recall that the unit of analysis is each sentence given by members of Congress mention-

⁹These bureaucracies are: Drug Enforcement Administration, Office of National Drug Control Policy, Department of Homeland Security, Immigration and Customs Enforcement, Customs and Border Protection, and Bureau of Prisons.

Table 5: Ideological Distance and Use of Bureaucratic Information.

	<i>tf-idf</i> of Statistical Facts Agency Type				
	(1)	All Agencies (2)	(3)	Independent (4)	Controlled (5)
Ideological Distance	-0.016* (0.010)	-0.020** (0.010)	-0.021** (0.010)	-0.005 (0.021)	-0.038** (0.017)
MC-level controls:		✓			
Other Controls:		✓	✓	✓	✓
R ²	0.016	0.019	0.027	0.045	0.031
Observations	764,369	764,341	764,369	292,687	471,682
MC FE	✓	✓			
Congress FE	✓	✓			
Agency FE	✓	✓	✓	✓	✓
MC-Congress FE			✓	✓	✓

Notes: OLS estimates. SE clustered by member of Congress. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

ing the name of a bureaucracy, hence the estimates capture the effect of ideological distance on the *tf-idf* of statistical facts at the extensive margin. The third column estimates Equation 1 and represents the preferred specification with MC-by-congress fixed effects. On average, ideological distance has a negative effect on the use of bureaucratic information, even when holding constant everything at the agency and legislator-by-congress level, and hence conditioning the estimated effect on time-changing characteristics of members of Congress (e.g., attention to different issues or co-partisanship with the President).

For the purpose of this study, the direction of the effect is more relevant than the effect size. While ideological distance has a clear conceptual meaning, it hardly lends itself to substantive interpretations. However, a +1SD increase in ideological distance is associated with a decrease in the *tf-idf* of statistical facts in sentences mentioning a bureaucracy as large as 3.8% compared to the mean in the data. Importantly, I do not find a significant difference between the estimated effects in committee and floor speeches, suggesting that ideology matters both in more public facing venues such as floor debates – where legislators are aware that their constituents have easy access to what they say – and in more specialized meetings such as committee sessions –

where MCs' audience shift from voters to interest groups.¹⁰

To estimate the moderating effect of agency independence (*Hypothesis 2*), I estimate Equation 1 separately on two subsets of the data, one where the agency mentioned in the sentence is an independent agency, and one where the agency mentioned is not an independent agencies. For 84% of the 316 agencies mentioned by members of Congress I have information on their official "status" within the executive branch, as published on the US government website. 65 out of 267 are independent agencies, namely agencies that exist outside executive departments. Columns (4) and (5) of Table 5 shows the estimated effect of ideological distance for independent and departments and executive agencies, respectively. The estimates suggest that the effect observed among the entire sample of agencies is in fact driven by agencies under tighter political control. While the point estimate for independent agencies remain negative, the standard errors become larger and it is not possible to distinguish the effect from 0 at standard confidence levels.¹¹

However, agency independence can be correlated with other features of agencies that can confound the relationship between ideological distance the members' use of bureaucratic information. A stronger test for the moderating effect of independence would require some form of as-if random change to statutory features of agencies determining their level of independence. This form of "natural" experiment is offered by the 2020 *Seila Law v. Consumer Financial Protection Bureau* case, when the Supreme Court ruled that statutory restrictions on the removal of the CFPB director are unconstitutional, hence lifting a key source of independence that insulated the agency from presidential control.¹²

¹⁰In Table H.8 I show that the results are robust to limiting the analysis for committee speeches to the period before 2001, when C-SPAN3 started to televise some important hearings.

¹¹I find similar results when using the independence scores produced by Selin (2015) (see Table H.7 in the Appendix).

¹²I am thankful to Alex Acs for pointing me to this case.

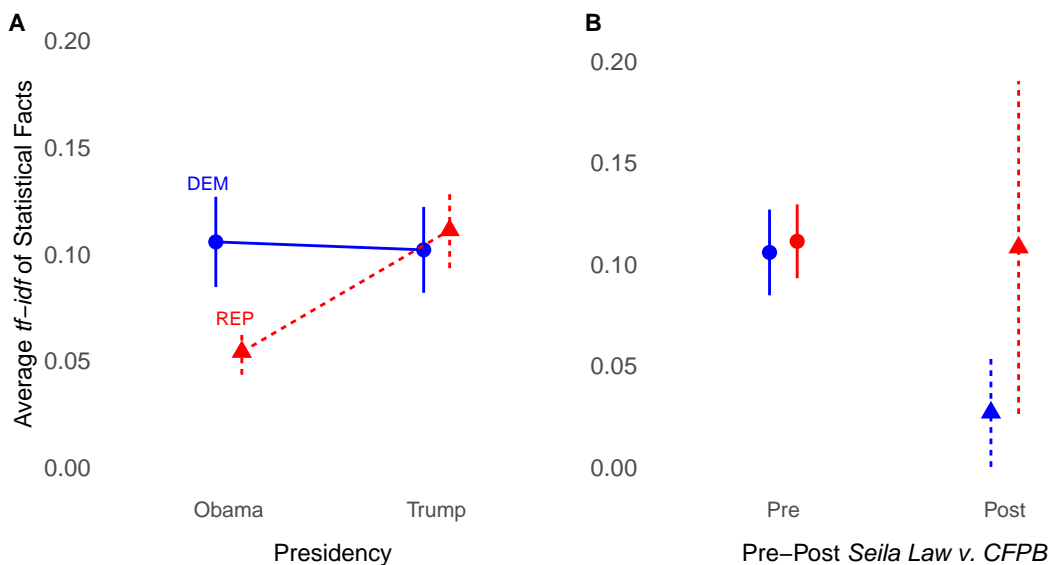
For Cause or at Will?

In response to the 2008 financial crisis, Congress and former President Barack Obama established the Consumer Financial Protection Bureau, an independent federal agency whose goal was to watch over predatory financial services practices. Initially proposed by Senator Elizabeth Warren (D-MA), the agency played an important role in safeguarding consumers' rights by ensuring fair treatment in the financial marketplace, enforcing financial laws, and addressing consumer complaints. Under the Trump administration, the mission (hence ideological leaning) of the agency changed direction. Former President Trump appointed fierce critics of the agency as administrators, accused of using the Bureau "to serve the wishes of the most powerful financial companies in America."¹³ While during the Obama presidencies, Democrats' reliance on the information produced by the CFPB was 82% larger than Republicans', the gap closed under Trump. As evidenced by Plot A in Figure 5, the average *tf-idf* when mentioning the CFPB is larger for Democrats during the second Obama presidency but not under Trump.

As provided by its statute – and unlike other independent agencies – the CFPB was governed by one director which could be removed by the President from her position only for "inefficiency, neglect of duty, or malfeasance in office." This statutory feature was a way of insulating the agency from undue political pressure during the 5-year mandate of the administrator. Political appointees serving as director – while clearly favored by the serving President – were protected from at-will dismissal and enjoyed a significant degree of independence in managing the agency. This guarantee of independence was nonetheless removed with the 2020 *Seila Law v. Consumer Financial Protection Bureau*, where the Supreme Court ruled that the restrictions on the removal of the CFPB director are unconstitutional. The dispute began when Seila Law, a law firm that provides debt-relief services to consumers, was under investigation by the CFPB for possible violations of telemarketing sales rules. Seila Law challenged the CFPB's powers to

¹³See <https://www.forbes.com/sites/advisor/2020/06/30/supreme-court-says-trump-can-fire-consumer-watchdog-director-but-cfpb-here-to-stay/?sh=582cd3218c6a>.

Figure 5: Consumer Financial Protection Bureau and *Seila Law v. CFPB*



Notes: Average use of information produced by the CFPB in the second Obama presidency and during the Trump presidency (Figure A) and before and after the decision of the Supreme Court, during the Trump presidency (Figure B).

obtain documents from the firm, arguing that the Bureau’s organization was unconstitutional due to the one-director structure with substantial power though removable only “for cause”. Instead, *Seila Law* argued, the director should be removable “at will” by the President – that is, for any reason.

The effect of this sudden shock to the CFPB independence on Democrats’ reliance on the information produced by the Bureau is striking. Plot B of Figure 5 shows the average *tf-idf* under Trump – to hold fixed the ideological leaning of the Bureau – before and after the *Seila Law v. CFPB* case.

To strengthen the evidence provided visually, I leverage the shock to the independence of the CFPB to estimate a difference-in-differences model. Democrats are the treated legislators, for they are those ideologically distant from the CFPB under the Trump presidency (as evidenced also by Panel A of Figure 5), and the post-treatment period is the period from June 2020 until the end of 2020 (the case was argued on March 3rd, 2020 and decided on June 29th, 2020). The analysis is performed on the years 2017-2020 during the Trump presidency, to hold fixed the ideological leaning of the Bureau and prevent the transition to the Biden presidency

to confound the effect of the decreased level of statutory independence. Specifically, I estimate the following model

$$tf\text{-idf of Statistical Facts}_{ij\text{at}} = \gamma_{jt} + \delta_a + \tau \mathbb{1}\{j \in |D|, a = \text{CFPB}\} \times Post_{ij\text{at}} + X_{ij\text{at}} \zeta' + u_{ij\text{at}} \quad (2)$$

$\mathbb{1}\{j \in |D|, a = \text{CFPB}\}$ is an indicator equal to one for statements given by Democrats (i.e., when legislator j belongs to the set of Democratic MCs) mentioning the CFPB, and τ is the difference-in-differences estimator, capturing the difference in Democrats' reliance on information produced by the CFPB before and after 29th June 2020, the day when the *Seila Law v. CFPB* case was decided. The key identifying assumption is that, absent the Supreme Court ruling, Democrats' reliance on CFPB information would have experienced parallel trends to Democrats' reliance on information produced by other agencies.¹⁴

Table 6 shows that, as a result of the decreased independence of the agency and holding constant its ideological leaning, after the case, Democrats were less likely to use information and statistical evidence produced by the CFPB in their speeches. As a result of the decreased independence, the *tf-idf* of Democrats quotes from the CFPB drops by -0.07 , which is as large as 83% compared to the sample mean of the *tf-idf* of sentences mentioning the CFPB since its establishment.

Robustness Analysis

In Section H in the Appendix, I include a series of robustness tests. First, because the dependent variable is a frequency variable, I report regression estimates with alternative transformations of the dependent variable in order to reduce the importance of extreme values when estimating β (Table H.4). Second, results are robust to removing from the analysis quotes reported during

¹⁴In Section H10 in the Appendix I indirectly test the parallel trends assumption by estimating Equation (2) with placebo post-treatment indicators.

Table 6: Effect of *Seila v CFPB* Case.

	<i>tf-idf</i> of Statistical Facts		
	(1)	(2)	(3)
Democrat \times CFPB \times Post	-0.072* (0.043)	-0.078* (0.043)	-0.074* (0.044)
MC-level controls:		✓	
Other Controls:		✓	✓
R ²	0.048	0.051	0.054
Observations	181,597	181,574	181,597
Legislator FE	✓	✓	
Congress FE	✓	✓	
Agency FE	✓	✓	✓
Legislator-Congress FE			✓

Notes: OLS estimates. SE clustered by member of Congress. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

oversight hearings, when members of Congress might use the information produced by agencies to question them (Table H.9).¹⁵ Third, I show that alternative measures of agency ideology yield similar results. I use the estimates produced by Bertelli and Grose (2011), Clinton and Lewis (2008) Clinton et al. (2012), Richardson, Clinton, and Lewis (2018), who use a mix of expert surveys and executives’ testimonies about specific bills as input to estimating agency ideology (see Table H.10). Moreover, the results for independent/controlled agencies hold when using as a measure of statutory independence the one produced by Selin (2015) (see Table H.7 in the Appendix). Finally, I show that the effect of ideological distance are overall robust at the intensive margin, namely when restricting the sample from all the sentences *mentioning* a bureaucracy to all the sentences *quoting* a bureaucracy (see Table H.5).

Discussion and Conclusions

The ability to produce expertise and information is one of the main sources of legitimacy of unelected bureaucracies that, despite governed by presidential directives, enjoy large degree of autonomy. By supplying expertise at the disposal of elected representatives, they contribute

¹⁵I adopt a very conservative exclusion criterion, removing every hearing whose title contains the word “oversight”, removing about 8% of the sentences mentioning bureaucratic agencies in committees.

to the policy making process in all its stages, from the agenda setting and problem definition phase to the implementation and administration of programs. However, very little is known about the extent to which this information is used by politicians in Congress. While most studies on the role of bureaucracy in legislative politics resort to perception measures, this paper represents the first attempt at bringing observational evidence to a key question in the scholarship on bureaucracy, Congress, and policy-making.

With this paper, I show that bureaucracies play an important role in the legislative arena and I find that their information is vastly used by members of Congress in floor and committee sessions alike. However, legislators' reliance on bureaucratic information is not immune from political prioritization. Members of Congress favor information coming from agencies inside their ideological camp. Since lawmakers are only imperfectly able to assess the merits of information, they rely on the ideological leaning of bureaucracy as a device to decide whether they can rely on the information when making policy, harboring skepticism towards bureaucracies whose ideological leaning is apart from their own. However, my findings show that institutional features governing bureaucracy – and in particular the extent to which bureaucratic decision-making is not subject to political control – can compensate the ideological differences between agencies and MCs and increase the credibility of information and its compatibility with members' political goals.

These findings have implications for theories of separation of powers, consolidating the role of bureaucracy as a fourth branch of government, as well as for normative considerations about the use of evidence in policy-making. This paper also has implications for the institutional design of bureaucratic bodies, for it suggests that institutional features granting independence to bureaucracy can mitigate the salience of the ideological divide between legislators and bureaucratic agencies.

This paper also opens new research avenues both within the study of politicians and

bureaucracy as well as for other sub-fields in political science. While here I focused on legislators' reliance on information produced by bureaucracies, scholars could study the market for policy-relevant information and the competitive nature that exists between different sources of information. There can be situations in which bureaucracy holds a monopoly in the supply of information compared to other actors, or "environments" in which government agencies are the only credible sources of information. Future work could focus on other influential organizations such as think tanks or interest groups and how changes in congressional staff capacity and composition affect legislators' reliance on bureaucratic expertise. Future work could also look at policy-level outcomes and how they change and possibly improve when bureaucratic expertise enters the legislative arena. For instance, there might be a persuasion effect whereby legislators' deploying bureaucratic information are able to build bipartisan coalitions in support of their favored bills or are more likely to see their proposed amendments passed. Finally, the new flexible measurement strategy proposed in this paper can be used in other sub-fields of political science to observe how information is used by a multitude of actors in different venues, from campaign messages, to social media, and press releases.

References

- Aberbach, Joel, Robert D. Putnam, and Bert Rockman. 1981. *Bureaucrats and Politicians in Western Democracies*. Harvard University Press.
- Atteveldt, Wouter Van, Tamir Sheafer, Shaul R Shenhav, and Yair Fogel-Dror. 2017. “Clause Analysis: Using Syntactic Information to Automatically Extract Source, Subject, and Predicate from Texts with an Application to the 2008-2009 Gaza War.” *Political Analysis* 25: 207–22. <https://doi.org/10.1017/pan.2016.12>.
- Ban, Pamela, Ju Yeon Park, and Hye Young You. 2022. “How Are Politicians Informed? Witnesses and Information Provision in Congress.” *American Political Science Review*, 1–18. <https://doi.org/10.1017/s0003055422000405>.
- . 2023. “Bureaucrats in Congress: Strategic Information Sharing in Policymaking.” <https://www.dropbox.com/s/28idwmhic06bht9/bureaucrats.pdf?dl=0>.
- Bertelli, Anthony M., and Christian R. Grose. 2011. “The Lengthened Shadow of Another Institution? Ideal Point Estimates for the Executive Branch and Congress.” *American Journal of Political Science* 55: 767–81. <https://doi.org/10.1111/j.1540-5907.2011.00527.x>.
- Bertelli, Anthony M., and Andrew B. Whitford. 2009. “Perceiving Credible Commitments: How Independent Regulators Shape Elite Perceptions of Regulatory Quality.” *British Journal of Political Science* 39: 517–37. <https://doi.org/10.1017/S0007123409000623>.
- Blom-Hansen, Jens, Martin Bækgaard, and Søren Serritzlew. 2020. “How Bureaucrats May Shape Political Decisions: The Role of Policy Information.” *Public Administration*. <https://doi.org/10.1111/padm.12709>.
- Bolton, Alexander, John M. de Figueiredo, and David E. Lewis. 2019. “Elections, Ideology, and Turnover in the u.s. Federal Government.” *NBER Working Paper Series*. <https://doi.org/10.1017/CBO9781107415324.004>.
- Bolton, Alexander, Rachel Augustine Potter, and Sharece Thrower. 2016. “Organizational Ca-

- capacity, Regulatory Review, and the Limits of Political Control.” *Journal of Law, Economics, and Organization* 32: 242–71. <https://doi.org/10.1093/jleo/ewv025>.
- Bonica, Adam. 2019. “Are Donation-Based Measures of Ideology Valid Predictors of Individual-Level Policy Preferences?” *Journal of Politics* 81: 327–33. <https://doi.org/10.1086/700722>.
- Carpenter, Daniel P. 2001. *The Forging of Bureaucratic Autonomy*. Princeton University Press.
- Chen, Jowei, and Tim Johnson. 2014. “Federal Employee Unionization and Presidential Control of the Bureaucracy: Estimating and Explaining Ideological Change in Executive Agencies.” *Journal of Theoretical Politics* 27: 151–74. <https://doi.org/10.1177/0951629813518126>.
- Choi, Jinho D., Joel Tetreault, and Amanda Stent. 2015. “It Depends: Dependency Parser Comparison Using a Web-Based Evaluation Tool.” *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference* 1: 387–96. <https://doi.org/10.3115/v1/p15-1038>.
- Clinton, Joshua D., Anthony Bertelli, Christian R. Grose, David E. Lewis, and David C. Nixon. 2012. “Separated Powers in the United States: The Ideology of Agencies, Presidents, and Congress.” *American Journal of Political Science* 56: 341–54. <https://doi.org/10.1111/j.1540-5907.2011.00559.x>.
- Clinton, Joshua D., and David E. Lewis. 2008. “Expert Opinion, Agency Characteristics, and Agency Preferences.” *Political Analysis* 16: 3–20. <https://doi.org/10.1093/pan/mpm009>.
- Crawford, VP, and J Sobel. 1982. “Strategic Information Transformation.” *Econometrica* 50: 1431–51.
- Cukierman, Alex, Steven B. Webb, and Bilin Neyapti. 1992. “Measuring the Independence of Central Banks and Its Effects on Policy Outcomes.” *The World Bank Economic Review* 6: 353–98.

- Ellig, Jerry, and Christopher J. Conover. 2014. "Presidential Priorities, Congressional Control, and the Quality of Regulatory Analysis: An Application to Healthcare and Homeland Security." *Public Choice* 161: 305–20. <https://doi.org/10.1007/s11127-014-0201-3>.
- Fiorina, Morris P. 1981. "Congressional Control of the Bureaucracy: A Mismatch of Incentives and Capabilities." In *Congress Reconsidered*, edited by Dodd C. Lawrence and Bruce I. Oppenheimer, 2nd ed. Congressional Quarterly Press.
- Gailmard, Sean, and John Patty. 2012. "Formal Models of Bureaucracy." *Annual Review of Political Science* 15: 353–77. <https://doi.org/10.1146/annurev-polisci-031710-103314>.
- . 2013. *Learning While Governing: Expertise and Accountability in the Executive Branch*. University of Chicago Press.
- Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. 2023. "ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks." *Proceedings of the National Academy of Sciences* 120. <https://doi.org/10.1073/pnas.2305016120>.
- Gilligan, Thomas W., and Keith Krehbiel. 1987. "Collective Decisionmaking and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures." *Journal of Law, Economics, and Organization* 3: 287–335. <https://doi.org/10.1093/oxfordjournals.jleo.a036932>.
- Grimmer, Justin. 2013. *Representational Style in Congress : What Legislators Say and Why It Matters*. Cambridge: Cambridge University Press.
- Grimmer, Justin, Sean Westwood, and Solomon Messing. 2014. *The Impression of Influence: Legislator Communication, Representation, and Democratic Accountability*. Princeton University Press.
- Honnibal, Matthew, and Mark Johnson. 2015. "An Improved Non-Monotonic Transition System for Dependency Parsing." *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing*, 1373–78. <https://doi.org/10.18653/v1/d15->

- Ingold, Karin, and Philip Leifeld. 2016. "Structural and Institutional Determinants of Influence Reputation: A Comparison of Collaborative and Adversarial Policy Networks in Decision Making and Implementation." *Journal of Public Administration Research and Theory* 26: 1–18. <https://doi.org/10.1093/jopart/muu043>.
- Keeper, Philip, and David Stasavage. 2003. "The Limits of Delegation: Veto Players, Central Bank Independence, and the Credibility of Monetary Policy." *American Political Science Review* 97: 407–23. <https://doi.org/10.1017/S0003055403000777>.
- Koop, Christel, and Chris Hanretty. 2018. "Political Independence, Accountability, and the Quality of Regulatory Decision-Making." *Comparative Political Studies* 51: 38–75. <https://doi.org/10.1177/0010414017695329>.
- Krause, George A. 1996. "The Institutional Dynamics of Policy Administration : Bureaucratic Influence over Securities Regulation." *American Journal of Political Science* 40: 1083–1121.
- Kroeger, Mary A. 2022. "Bureaucrats as Lawmakers." *Legislative Studies Quarterly* 47: 257–89. <https://doi.org/10.1111/lsq.12320>.
- Lee, Frances E. 2016. *Insecure Majorities : Congress and the Perpetual Campaign*. Chicago; The University of Chicago Press.
- Lewis, Jeffrey, Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet. 2020. "Voteview: Congressional Roll-Call Votes Database." voteview.com.
- Lowande, Kenneth. 2018. "Who Polices the Administrative State?" *American Political Science Review* 112: 874–90. <https://doi.org/10.1017/S0003055418000497>.
- Maltzman, Forrest, and Lee Sigelman. 1996. "The Politics of Talk: Unconstrained Floor Time in the u.s. House of Representatives." *The Journal of Politics* 58 (3): 819–30. <https://doi.org/10.2307/2960448>.
- Mayhew, David. 1974. *Congress: The Electoral Connection*. Yale University Press.

- McCubbins, M, Roger G Noll, and Barry R Weingast. 1987. "Administrative Procedures as Instruments of Political Control." *Journal of Law, Economics, & Organization* 3: 243–77.
- Miller, Gary J. 2005. "The Political Evolution of Principal-Agent Models." *Annual Review of Political Science* 8: 203–25. <https://doi.org/10.1146/annurev.polisci.8.082103.104840>.
- Moe, Terry M. 2006. "Political Control and the Power of the Agent." *Journal of Law, Economics, and Organization* 22: 1–29. <https://doi.org/10.1093/jleo/ewj011>.
- . 2012. "Delegation, Control, and the Study of Public Bureaucracy." *The Forum* 10. <https://doi.org/10.1515/1540-8884.1508>.
- Moore, Emily H. 2018. "Polarization, Excepted Appointments, and the Administrative Presidency." *Presidential Studies Quarterly* 48: 72–92. <https://doi.org/10.1111/psq.12417>.
- Nicholson-Crotty, Jill, and Susan M. Miller. 2012. "Bureaucratic Effectiveness and Influence in the Legislature." *Journal of Public Administration Research and Theory* 22: 347–71. <https://doi.org/10.1093/jopart/mur054>.
- Niskanen, WA. 1971. *Bureaucracy and Public Economics*. Aldine-Atherton.
- Pennebaker, James W., Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The Development and Psychometric Properties of LIWC2015*. University of Texas at Austin. <https://doi.org/10.1068/d010163>.
- Potter, Rachel Augustine. 2019. *Bending the Rules. Pocedural Politicking in the Bureaucracy*. University of Chicago Press.
- Przeworski, A., S. Stokes, and B. Manin. 1999. *Democracy, Accountability, and Representation*. Edited by A. Przeworski, S. Stokes, and B. Manin. Cambridge Studies in the Theory of Democracy. Cambridge University Press.
- Richardson, Mark D., Joshua D. Clinton, and David E. Lewis. 2018. "Elite Perceptions of Agency Ideology and Workforce Skill." *Journal of Politics* 80: 303–8. <https://doi.org/10.1086/694846>.

- Selin, Jennifer L. 2015. "What Makes an Agency Independent?" *American Journal of Political Science* 59: 971–87. <https://doi.org/10.1111/ajps.12161>.
- Shobe, Jarrod. 2017. "Agencies as Legislators: An Empirical Study of the Role of Agencies in the Legislative Process." *The George Washington Law Review* 85: 451–535.
- Vannoni, Matia, Elliott Ash, and Massimo Morelli. 2019. "Measuring Discretion and Delegation in Legislative Texts: Methods and Application to u. S. States."
- Volden, Craig, and Alan Wiseman. 2020. "Centre for Effective Lawmaking." <https://thelawmakers.org/data-download>.
- Weber, Max. 1922. *Economy and Society*. Edited by Roth Guenther and Wittich Claus. 1968th ed. University of California Press.
- Wilson, J. Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. Basic Books.
- Workman, Samuel. 2015. *The Dynamics of Bureaucracy in the US Government: How Congress and Federal Agencies Process Information and Solve Problems*. Cambridge University Press.

Appendix

A	Extraction Rules	1
B	Dependency Parsing: Examples	2
C	LIWC Dictionary	3
D	Building the <i>tf-idf</i>	4
E	Validation with GPT	5
F	Sentiment Analysis of Speeches Mentioning Bureaucracies	6
G	Speeches: Data Quality	7
H	Robustness Checks	8
H1	Transformation of DV	8
H2	Intensive Margin. Sample is All Sentences Quoting a Bureaucracy	9
H3	Ideological distance and MC Mentioning Bureaucracies	10
H4	Alternative Measures of Independence	11
H5	Committee Results Pre-Post C-SPAN3	12
H6	Committee Results Without Oversight Hearings	13
H7	Alternative Measures of Agency Ideology	14
H8	Committee and Chamber FE	15
H9	Differences between House and Senate	16
H10	Parallel Trends	17

A Extraction Rules

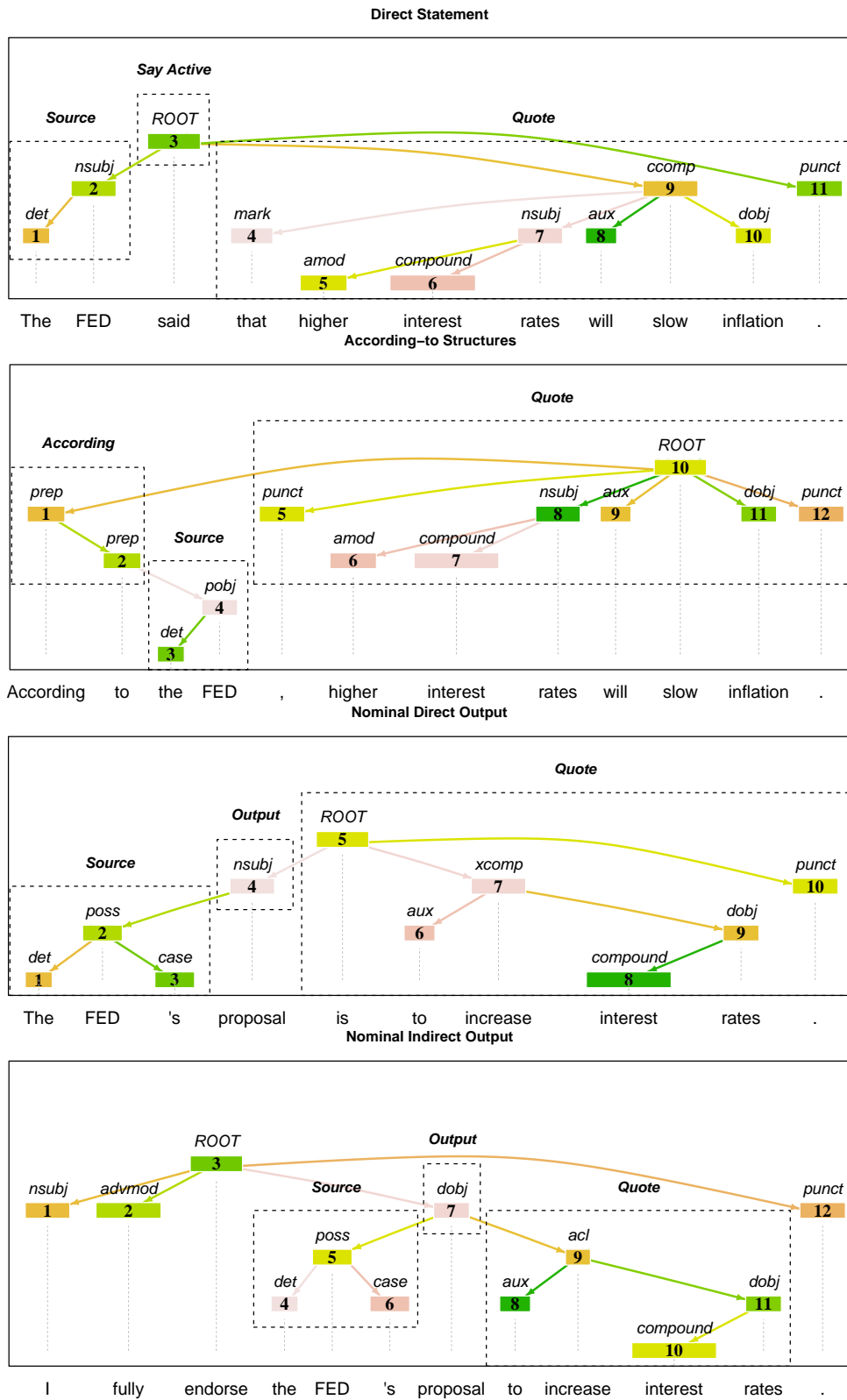
Below I report the lists of say-verbs and output-type words used to extract legislators' quotes of bureaucratic information.

Say verbs used to match syntactic rules: acknowledge, admit, advance, advise, advise, affirm, agree, argue, assert, assume, assume, assure, believe, claim, clarify, complain, concede, conclude, confirm, consider, contend, convince, decide, decide, define, demonstrate, document, encourage, estimate, evaluate, explain, find, identify, indicate, inform, predict, present, presume, project, propose, propose, prove, realise, realize, recommend, refer, remind, report, respond, reveal, say, see, set out, show, state, suggest, tell, testify, think, warn.

Output-type words are: advice, advise, analysis, argument, article, assessment, brief, comment, complaint, conclusion, copy, data, datum, decision, directive, document, estimate, evidence, figure, forecast, guidance, guideline, idea, indication, information, input, inquiry, instruction, memo, observation, opinion, paper, plan, position, prediction, prevision, program, programme, project, projection, proposal, rationale, reasoning, recommendation, report, statement, statistic, statistics, strategy, study, suggestion, survey, testimony, thesis, view.

B Dependency Parsing: Examples

Figure B.1: Dependency Tree.



Notes: Parsed dependency trees of the four remaining illustrative examples where the FED is used in speech. Implemented through the *rsyntax* package in R.

C LIWC Dictionary

In the table below I report the LIWC dictionary on numbers and quantifiers as well as a list of verbs capturing quantitative description.

Table C.1: Dictionary of Statistical Facts and Quantitative Evidence.

Source	Words
LIWC Dictionary	billion*, doubl*, dozen*, eight*, eleven, fift*, first, five, four*, half, hundred*, infinit*, million*, nine*, onc, one, quarter*, second, seven*, singl, six*, ten, tenth, third, thirt*, thousand*, three, trillion*, twel*, twent*, twice, two, zero, zillion*, add, ad, all, allot, alot, amount, anoth, ani, approxim*, averag, bit, both, bunch, chapter, coupl, each, either, entire*, equal*, everi, extra, few, fewer, fewest, group*, unequal*, least, less, lot, lotof, lotsa, lotta, major, mani, mo, more, most, much, mucho, multipl, nada, none, part, percent*, piec, plenti, remain, sampl*, scarc, scarcer, scarcest, section, segment*, seri, several*, some, somewhat, ton, total, triple*, tripl, varieti, various, whole, [All digits that are not dates]
Verbs (Quantitative Description)	increase, decrease, reduce, boost, lower, decline, skyrocket, eliminate, enhance, rise, limit, accelerate, significantly, plummet, spike, overall, large, face, hurt, harm, end, nurture, criticize, cause, induce, suffer, exacerbate, result, inflict, prevent, worsen, consequence, impact, affect, effect, combat, minimize, maximize, ensure, allow, curb, avoid, curtail, save, mitigate, promote, cultivate, facilitate, create, adopt, sustain, develop, bolster, improve, expand, maintain, restore, intensify, decay, crumble, erode, collapse, evolve, neglect, stop, budget, lose, fund, regulate, provide, discourage, encourage, go, plan

Notes: Stemmed tokens of LIWC dictionary of numbers and quantifiers and verbs capturing quantitative description. Numbers in digit format, except for dates and years, are included in the dictionary.

D Building the *tf-idf*

To build the *tf-idf*, I first build a document-token matrix, with one row for every quote, and one column for every unique token used in the corpus as a whole. Tokens are assigned a weight which is equal to the logarithm of the inverse fraction of the quotes that contain the word.

For each quote mentioning the information produced by a bureaucratic body, the final score is the sum of the *tf-idf* frequencies of tokens that appear in the dictionary. More formally, consider the full corpus a set of quotes, and each quote a set of words, whose cardinality represents the number of unique words in the quote. For each quote, the use of facts and evidence is given by the following formula:

$$Facts_q = \sum_{t \in Dict} tf - idf_{t,q} \quad \text{with} \quad tf - idf_{t,q} = \frac{f_{t,q}}{|q|} \times \log \frac{|Q|}{|\{q \in Q : t \in Dict\}|} \quad (3)$$

where t is each token of quote q , $Dict$ the list of words capturing the use of statistical facts, and $tf - idf$ is the term frequency-inverse document frequency of token t in quote q . The *Fact* score is ultimately a function of the absolute frequency of the token t ($f_{t,q}$), the number of words in quote q ($|q|$), the number of quotes of the total corpus Q , and the number of documents in the corpus that contain the token t ($|\{q \in Q : t \in Dict\}|$).

E Validation with GPT

In the table below, I report an example of the prompt/reply of the validation task performed with ChatGPT.

Table E.2: GPT Prompt for Validation Task.

Prompt Input: I will show you a statement given by a US member of Congress mentioning the Bureau of the Census. Please tell me whether the member of Congress is explicitly citing policy information or evidence produced by the Bureau of the Census. Please reply YES or NO and justify your response. This is the sentence: "The Bureau of the Census projects that by the year 2030, Hispanic children will represent 25 percent of the total student population, and even the most recent Census figures show that Hispanics are now on pace to become the nation 's largest minority sooner than expected."

GPT Output: YES, the member of Congress is explicitly citing policy information or evidence produced by the Bureau of the Census. The statement refers to projections made by the Bureau of the Census about the future representation of Hispanic children in the total student population, and also cites recent Census figures about the pace of growth of the Hispanic population.

Notes: Input and output of zero-shot text classification with GPT. Model used gpt-4 with temperature set to 0 to minimize idiosyncrasy in responses.

F Sentiment Analysis of Speeches Mentioning Bureaucracies

In Table F.3 below I report OLS estimates of the effect of whether the speech is using one or more bureaucracies as as source of information (dummy) on the probability of the speech being assigned a negative sentiment. I perform the dictionary analysis with the `nltk` library in Python, which automatically assign a continuous score to each text ranging between -1 and 1, where larger value signify more positive sentiment. To create a measure of negative sentiment, I then dichotomize the score and convert it to 1 if the score is below its median value and 0 otherwise.

Table F.3: Sentiment Analysis.

	Negative Sentiment			
	(1)	(2)	(3)	(4)
Agency Used as Source	-0.097*** (0.004)	-0.061*** (0.004)	-0.063*** (0.004)	-0.061*** (0.004)
MC-level controls:			✓	
Other Controls:		✓	✓	✓
R ²	0.002	0.123	0.123	0.144
Observations	924,340	924,243	897,250	924,243
Legislator FE		✓	✓	
Congress FE		✓	✓	
Agency FE		✓	✓	✓
Legislator-Congress FE				✓

Notes: OLS estimates. DV is the probability of the speech's sentiment being classified as negative (below median of sentiment score). Independent variable equal to 1 if speech uses agency as source and 0 if it just mentions the agency. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

G Speeches: Data Quality

I access transcripts of speeches from two sources. For floor speeches (1994-2022) and for committee speeches (2010-2020), I scraped the digitalized version of the Congressional Record. For committee speeches (1980-2009) I relied on transcripts available through ProQuest.

Online Version of Congressional Record

To access speeches on the online version of the Congressional Record I obtain the universe of available links to congressional hearings material and to floor speeches via the website <https://www.govinfo.gov/sitemaps>. The sitemaps contain one main url for each year, and each url contains as many urls as the number of packages in which the record has been grouped. For instance, the material for the Senate Hearing with ID 76804 can be accessed via the following link: <https://www.govinfo.gov/content/pkg/CHRG-107shrg76804/html/CHRG-107shrg76804.htm>, which re-directs to the text of the entire committee session.

Thanks to metadata listing the members of Congress who gave speeches in the session, it is possible to extract the speech with a set of flexible regular expressions that capture the structure “title + surname + period + white space + start of sentence”.

Pro-Quest Data

For older committee sessions, I accessed transcripts directly from ProQuest. I obtained 42,277 transcripts of congressional committee sessions, each one consisting of one large html file, and no metadata exists to facilitate the extraction of single speeches. Speeches are nonetheless identifiable thanks to the way they appear in the text. The title and SURNAME of the speaker precedes the speech and is reported in capital cases. “Mr. FORD”, for instance, marks a new speech. Many individuals give speeches or statements and to extract speeches given by politicians alone, I exploit the fact that at the beginning of each transcript, the names of all members of Congress are reported followed by their home state. From every transcript I therefore extract all the name of politicians with a regular expression that matches the name and surname of individuals followed by the name of their respective state. Only speeches given by any of the extracted names are parsed from the transcript.

Despite some typos in the full text, a careful look at a random sample of parsed speeches suggests the quality of the parsing procedure is sufficiently high to confidently attribute speeches to legislators. By merging surname, date of congressional session, and state of the legislators I am then able to match data on committee speeches with the DW-NOMINATE score of each legislator.

H Robustness Checks

H1 Transformation of DV

Table H.4: Alternative Measures of Outcome Variable.

	Abs. Frequency (1)	Log Abs. Frequency (2)	Log <i>tf-idf</i> (3)	Dummy Measure (4)	Abs. Frequency / Length Quote (5)
Ideological Distance	-0.018*** (0.007)	-0.007** (0.003)	-0.007** (0.003)	-0.005* (0.002)	-0.007** (0.003)
MC-level controls:					
Other Controls:	✓	✓	✓	✓	✓
R ²	0.028	0.031	0.030	0.030	0.185
Observations	764,369	764,369	764,369	764,369	40,342
Legislator-Congress FE	✓	✓	✓	✓	✓
Agency FE	✓	✓	✓	✓	✓

Notes: OLS estimates. SE clustered by member of Congress. Alternative measures of use of information from left to right: absolute frequency (1), log of absolute frequency (2), log of extittf-idf (3), dummy measure (equal to 1 if the absolute frequency is greater than 0 and 0 otherwise) (4), absolute frequency divided by the length of the sentence (5). Frequency refers to statistical facts and evidence in quotes of agencies mentioned in legislators' speeches. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

H2 Intensive Margin. Sample is All Sentences Quoting a Bureaucracy

In the table below I report the results when looking at the intensive margin of the effect of ideological distance. Here, the sample consists of all sentences quoting a bureaucracy using several measures of quantitative evidence and statistical facts in the quote as outcome.

Table H.5: Intensive Margin Analysis.

	<i>tf-idf</i> (1)	Absolute Frequency (2)	Log Abs. Frequency (3)	Log <i>tf-idf</i> (4)	Dummy Measure (5)	Abs. Freq./ Length Quote (6)
Ideological Distance	-0.230 (0.141)	-0.184* (0.100)	-0.072* (0.037)	-0.081* (0.044)	-0.049* (0.026)	-0.007** (0.003)
MC-level Controls:						
Other Controls:	✓	✓	✓	✓	✓	✓
R ²	0.186	0.188	0.184	0.179	0.148	0.185
Observations	40,345	40,345	40,345	40,345	40,345	40,345
Legislator-Congress FE	✓	✓	✓	✓	✓	✓
Agency FE	✓	✓	✓	✓	✓	✓

Notes: OLS estimates. DV is *tf-idf* of statistical facts in sentences quoting bureaucracy and 0 otherwise. Sample of all sentences mentioning at least one bureaucratic agency. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

H3 Ideological distance and MC Mentioning Bureaucracies

To test whether members of Congress are less likely to mention ideologically distant agencies, I assemble a dyadic dataset at the MC-agency-congress level, combining every MC giving at least one speech for every Congress with agencies mentioned in any given Congress, to make sure that agencies not yet established are not treated as never mentioned. The dataset includes 2,099 unique MCs and 48 unique bureaucracies for which I also have data on ideology.

I then estimate the following dyadic regression model via OLS:

$$y_{ijat} = \beta \text{Ideological Distance}_{jat} + \zeta_{ja} + \delta_t + \epsilon_{ijat}$$

where y_{ijat} is either a dummy variable equal to 1 if legislator j in Congress t mentioned agency a in her speeches and 0 otherwise, or alternatively a continuous number of mentions of agency a . Ideological distance is the absolute value between the agency ideal point and the first dimension of the MC DW-NOMINATE score. ζ_{ja} are dyad fixed effects and δ_t are Congress fixed effects. SE are clustered by MC-agency dyad. Table H.6 reports the results. There is no effect of MC-agency ideological distance on its probability or the number of mentions of bureaucracies in speeches.

Table H.6: Ideology and Mentioning Bureaucracies.

	Agency Mentions	
	Count (1)	Dummy (2)
MC-Agency Ideological Distance	-0.122 (0.127)	0.005 (0.005)
R ²	0.526	0.522
Count	417,023	417,023
Legislator-Agency FE	✓	✓
Congress FE	✓	✓

Notes: Dyadic regression, OLS estimates. Outcomes are number of agency mentions and dichotomous measure equal to one if the number of agency mentions is greater than 0 and 0 otherwise. SE clustered by MC-agency dyad. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

H4 Alternative Measures of Independence

In Table H.7 I show that the effect of ideological distance is larger for less independent agencies using an alternative measure of agency independence. I use the two different indicators produced by Selin (2015), capturing independence along two dimensions: independence as the ability of an agency to make policy decisions without political interference; and independence as requirements imposed on agency officials and limitations to presidential power of appointment/removal of agency heads. The indicators are derived by modelling 50 structural features about agencies with a Bayesian latent variable model. The estimates range between 0 and 4, with higher values signifying higher independence. These measures do not capture variation over time, therefore I estimate Equation 1 on two subsets of agencies above and below the average of each of the two indicators separately.

Table H.7: Alternative Measures of Independence.

	Above Average Independence		Below Average Independence	
	Decision Makers (1)	Political Review (2)	Decision Makers (3)	Political Review (4)
Ideological Distance	-0.006 (0.020)	0.030 (0.024)	-0.041** (0.018)	-0.041*** (0.016)
MC-level controls:				
Other Controls:	✓	✓	✓	✓
R ²	0.044	0.068	0.031	0.028
Observations	308,075	192,433	456,294	571,936
Legislator-Congress FE	✓	✓	✓	✓
Agency FE	✓	✓	✓	✓

Notes: OLS estimates. SE clustered by member of Congress. .. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

H5 Committee Results Pre-Post C-SPAN3

To ensure the results for committee speeches are not driven by hearings televised on C-SPAN3, I split the analysis into a pre and post C-SPAN3 period. In fact, only from 2001 have some salient hearings started being televised. I find similar results across the two periods hence there is no evidence that the results in committee sessions are driven by a C-SPAN3 effect.

Table H.8: Committee Results and C-SPAN 3.

	<i>tf-idf</i> of Statistical Facts		
	Full Sample	Pre C-SPAN3	Post C-SPAN3
	(1)	(2)	(3)
Ideological Distance	-0.029** (0.013)	-0.016 (0.015)	-0.034 (0.022)
Legislator-level Controls:			
Other Controls:	✓	✓	✓
R ²	0.023	0.022	0.024
Observations	273,580	145,876	127,704
Legislator-Congress FE	✓	✓	✓
Agency FE	✓	✓	✓
Committee FE	✓	✓	✓

Notes: OLS estimates. DV is the *tf-idf* of statistical facts and evidence in quotes of agencies mentioned in legislators' speeches. Separate regressions on three periods: entire period, before introduction of televised hearings (pre-2001) and after introduction of televised hearings (post-2001).

H6 Committee Results Without Oversight Hearings

Because in committees legislators might report what said by bureaucracies to hold them to account, in the table below I report the results of the statistical tests performed after removing all oversight hearings. I adopt a very conservative exclusion criterion, removing every speech which contains the word “oversight” either in the short description of the hearing, in the list of topics, title, and name of committee, removing about 7,000 quotes. The results are robust to omitting such speeches. For instance, “Automobile Fuel Economy: EPA Oversight. Congressional Hearing, Jan. 29, Feb. 1, 1980”.

Table H.9: Analysis Removing Oversight Hearings.

	<i>tf-idf</i> of Statistical Facts		
	(1)	(2)	(3)
Ideological Distance	-0.029** (0.013)	-0.029** (0.013)	-0.035*** (0.013)
Oversight Hearing		-0.005 (0.007)	
MC-level Controls:			
Other Controls:	✓	✓	✓
R ²	0.023	0.023	0.024
Observations	273,580	273,580	254,000
Legislator-Congress FE	✓	✓	✓
Agency FE	✓	✓	✓
Committee FE	✓	✓	✓

Notes: OLS estimates. DV is the *tf-idf* of statistical facts and evidence in legislators’ sentences mentioning agencies. Col (1) all sentences, Col (2) specification conditioning on oversight hearing, Col (3) removing oversight hearings. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

H7 Alternative Measures of Agency Ideology

Results using alternative measures of agency ideology. When the measure is fixed over time, agency fixed effects are replaced by agency-type fixed effects (i.e., independent agencies, executive departments, executive sub-agencies, etc.)

Table H.10: Ideological Distance and Use of Bureaucratic Information.

	<i>tf-idf</i> of Statistical Facts				
	(1)	(2)	(3)	(4)	(5)
Ideological Distance (Chen and Johnson 2014)	-0.021** (0.010)				
Ideological Distance (Bertelli and Grose 2011)		-0.051*** (0.019)			
Ideological Distance (Richardson et al. 2018)			-0.014*** (0.002)		
Ideological Distance (Clinton and Lewis 2008)				-0.010*** (0.002)	
Ideological Distance (Clinton et al. 2012)					-0.022*** (0.007)
MC-level controls:					
Other Controls:	✓	✓	✓	✓	✓
R ²	0.027	0.034	0.016	0.019	0.022
Observations	764,369	222,264	1,977,499	1,404,339	1,009,058
Legislator-Congress FE	✓	✓	✓	✓	✓
Agency FE	✓	✓			
Agency Type FE			✓	✓	✓

Notes: OLS estimates. SE clustered by member of Congress. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

H8 Committee and Chamber FE

In the table below I show the results are robust to including committee and chamber fixed effects. In particular, committee fixed effects can be interpreted as topic fixed effects, for committees to conditioning on the topic of the speech, and hence limiting the possible sources of information to agencies (and other competing actors) that operate within one specific policy sector. There are 204 unique committees in the data, hence it is possible to control for very subtle topics/policy sectors such as the Committee on Agriculture, Nutrition, and Forestry or the Committee on Banking, Housing, and Urban Affairs.

Table H.11: Analysis at Legislator-Congress Level.

	<i>tf-idf</i> of Statistical Facts		
	(1)	(2)	(3)
Ideological Distance	-0.028** (0.012)	-0.029** (0.013)	-0.029** (0.013)
MC-level Controls:			
Other Controls:	✓	✓	✓
R ²	0.022	0.023	0.023
Observations	273,639	273,580	273,580
Legislator-Congress FE	✓	✓	✓
Agency FE	✓	✓	✓
Committee FE		✓	✓
Chamber FE			✓

Notes: OLS estimates. DV is the average *tf-idf* of statistical facts and evidence in quotes of agencies mentioned in legislators' speeches. SE clustered by legislator. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

H9 Differences between House and Senate

I do not detect differences in the estimated effect of ideological distance on the *tf-idf* of statistical facts between speeches given in the House and Senate. This test, by keeping MC fixed effects, exploits variation at the individual level in speeches given in the floor and the house, and hence effectively relies on legislators who held office in both the House and Senate throughout their career in Congress. In my data, these account to 5% of the legislators.

Table H.12: Ideological Distance and Use of Bureaucratic Information.

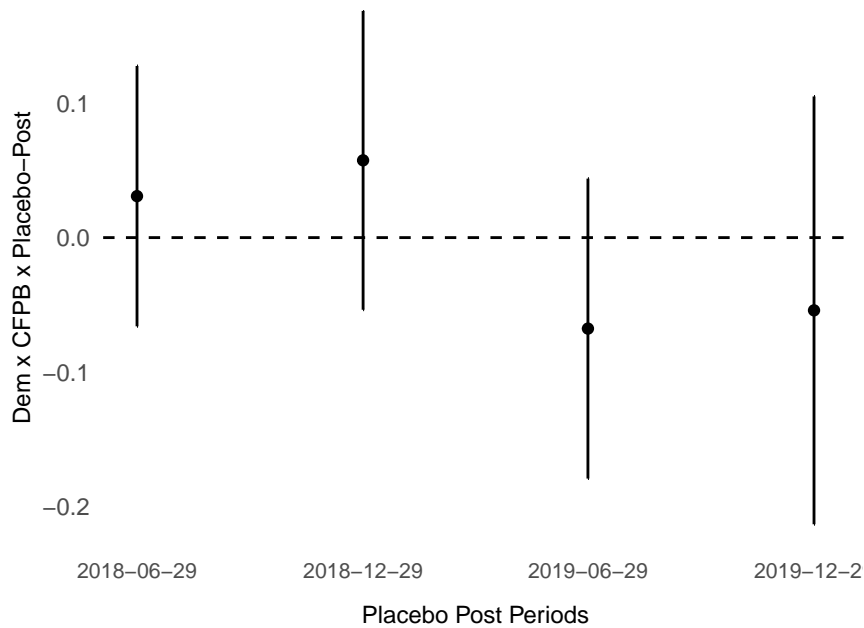
	<i>tf-idf</i> of Statistical Facts	
	(1)	(2)
Ideological Distance	-0.027** (0.013)	-0.031** (0.012)
Senate	0.016 (0.021)	0.024 (0.016)
Ideological Distance \times Senate	0.024 (0.019)	0.024 (0.019)
MC-level controls:		✓
Other Controls:		✓
R ²	0.016	0.019
Observations	764,369	764,341
Legislator FE	✓	✓
Congress FE	✓	✓
Agency FE	✓	✓

Notes: OLS estimates. SE clustered by member of Congress. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

H10 Parallel Trends

In the figure below I report the results of a robustness test showing that Democrats' reliance on information from the CFPB was on parallel trends compared to Democrats' reliance on other agencies during the Trump presidencies. I do so by estimating Equation (2) on a sample of sentences mentioning bureaucracies in the pre-treatment period (i.e., sentences given before 29th June 2020) and use a series of placebo post-treatment indicators (reported on the horizontal axis). On the vertical axis I plot the estimated effect of the triple interaction Democrat \times CFPB \times Placebo Post. The 95% confidence intervals largely include 0.

Figure H.2: Placebo Post-Treatment Periods.



Notes: Estimated effect of the triple interaction of Democrat \times CFPB \times Placebo Post-treatment period on the *tf-idf* of statistical facts in sentences mentioning bureaucratic agencies. SE clustered by member of Congress. The period of analysis is the Trump presidency up until the day before the Supreme Court ruling (i.e., 25th June 2016).