Bureaucrats, Politicians, and the Strategic Use of Information

Luca Bellodi

PhD Candidate, UCL Political Science

Research Fellow, Bocconi University

Word Count: 9,129

Abstract

The political control of the bureaucracy remains a classical topic in political science. However, little is known about its reverse: bureaucracies influencing politicians. I conceptualise bureaucratic influence as the extent to which legislators use the information produced by agencies in the legislative process. Building on cheap talk models of strategic communication, I argue that legislators make greater use of bureaucratic information when ideologically closer to agencies and that agency independence – operating as a credibility-enhancing mechanism – mitigates the effect of ideological distance. I introduce a new measurement strategy to estimate legislators' use of bureaucratic information which employs syntactic analysis and apply it to a corpus of 6.8 million speeches given by US congresspersons in floor and committee speeches. I find that ideological distance reduces politicians' use of bureaucratic information, but only in floor speeches. This suggests that in congressional committees, and less partisan arenas, ideology matters less for bureaucratic influence.

1 Introduction

> *Several Senators from the Upper Midwest insisted that the Office of Management and Budget do a study on the effects of the [Dairy] Compact. The OMB report is called "The Economic Effects of the Northeast Interstate Dairy Compact". I will be quoting a lot from that study that those Senators wanted in this floor statement.*
>
> Sen. Patrick Leahy, D–VT

This is one of the opening statements of a speech given by Democratic Senator Leahy, VT on 6th October 1998. In his speech, he cites 13 times what claimed by the OMB in the study. Bureaucratic agencies, due to their expertise, produce a great wealth of information that can be used by politicians to reduce uncertainty over policy outcomes or to strengthen their own argument. Yet we know little about when this occurs. In this paper I build on theories of strategic communication and show that ideological distance prevents politicians from using the information produced by agencies.

Couched within the principal-agent framework, the scholarship on bureaucratic politics made important strides to enhance our understanding of how politicians seek to exert control over bureaucracies in order to prevent agency slack and restrain bureaucratic autonomous policy-making (Epstein and O'Halloran, 1999; Gailmard, 2009; McCubbins et al., 1987; McCubbins and Schwartz, 1984), yet very rarely have scholars looked at the influence that bureaucratic agencies exert on the main political decisions that, in theory, should rest with elected politicians. Indeed, though bureaucrats are legally subordinated in the hierarchy of government, "*they can exert political power over their own superiors.* When this happens […], they can play major roles in determining […] what policies the latter pursue once in office" (Moe, 2012, 37).

Bureaucracies have always been considered rich sources of information for politicians (Niskanen, 1971; Wilson, 1989), but empirical scholarship has struggled to document and study influence behind the closed doors of government organisations. One of the first attempts to interpret agency-political relations as unfolding on a "two-way street" emerges from Krause's work (Krause, 1996, 1999). Krause depicted politicians' decisions to control the agency through budgetary allocation as the result of the interactions between both the agency and the principals. In his study of the US Securities and Exchange Commission (SEC), he shows that the budgetary preferences of the government with respect to the SEC are influenced by the SEC's regulatory performance. Beyond budget preferences, a prominent attempt to theorise the influence of bureaucracy on policy formulation is Carpenter's reputation-based account of bureaucratic autonomy (Carpenter, 2001, 2010). In *The Forging of Bureaucratic Autonomy*, Carpenter argues that bureaucratic reputation – a set of symbolic beliefs about an organisation embedded in a network of multiple audiences – allows agencies to secure their desired policies despite the opposition of elected politicians (Carpenter, 2001, 3–4).

More recent scholarship on bureaucratic politics has started to study role of the bureaucracy in different stages of the legislative process. Nicholson-Crotty and Miller (2012), for instance, find a positive relationship between agency perceived effectiveness and politicians' perceptions of bureaucratic influence on legislative outcomes, and Ingold and Leifeld (2016) find that vertically integrated offices with access to formal decision-making venues are on average perceived as more influential. However, despite few exceptions (e.g., Kroeger, 2020), scholarly work has generally relied on perception measures of influence, easily susceptible of social desirability bias, which could both inflate or deflate the actual influence exerted by the bureaucracy. In the attempt to overcome over- and under-reporting, Blom-Hansen

et al. (2020) implement a series of experiments simulating the decision-making process and find that bureaucrats are willing to use their information to influence politicians' decisions, who are in turn likely to rely on bureaucrats' expertise depending on how the information is framed. While this scholarship made important advancements in the study of the role of bureaucracies in the policy-making process, we know little about the extent to which they capture real world phenomena.

In this paper I propose an information-based conceptualisation of bureaucratic influence, defined as the extent to which legislators' use the information produced by bureaucratic bodies in the legislative process. When legislators pass and discuss policy, they use the information and the expertise of bureaucracy to form their beliefs about policy, strengthen their arguments, or increase the persuasion of their appeals. This information-based definition of influence, although partially different from standard, counterfactual definitions used in the social sciences, is consistent with a long tradition of work in political science that conceives bureaucracies as shaping policies through information (Weber, 1922; Aberbach et al., 1981; Workman, 2015). Importantly, it also allows to capture influence as a political practice, rather than a set of perceptions.

The theoretical gist of an informational definition of bureaucratic influence can be found in Crawford and Sobel's (1982) cheap talk model of strategic communication. Although the model is general in its formulation, it can easily be applied to the communication game between bureaucracy and legislators (Gailmard and Patty, 2012). An expert bureaucracy sends a signal or information to a legislator who will then make a policy decision. Because the legislator cannot verify the quality or veracity of the information, truthful communication is only achieved when both the bureaucracy and the legislator have similar preferences over policy outcomes. The more ideologically apart, the less likely it is that legislators will

use the information produced by agencies. I extend this theory by looking at the role of institutional independence on the politicisation of the information. I argue that, operating as a credibility-enhancing mechanism, agency independence reduces the salience of ideological differences and hence limit the negative effect of ideological divergence on legislators' use of the information sent by the agency.

I test this theory in the US context, with a new measurement strategy that detects when legislators use the evidence and facts produced by bureaucratic bodies in floor and committee speeches. First, I apply syntactic dependency parsing to a large corpus of 6.8 million floor and committee speeches and extract the information produced by agencies and used by legislators. Second, I measure the frequency of words considered as facts and statistical evidence and produce estimates of politicians' use of bureaucratic information for every speech in which the agency is used as a source of information.

I use a series of two-ways fixed effects specifications and find strong support for the theory, but only for floor debates. If we compare the effect size to the average use of facts, a one-unit increase in ideological distance is as large as a reduction in the use of facts and evidence produced by bureaucracy by 20% compared to the average outcome. However, a one-unit increase in the independence of the agency's head and officials reduces the effect of ideological distance by more than half. The effects is not distinguishable from zero for committee speeches. This is surprising, for congressional committees should be the venues where substantive aspects of policies are discussed, and where the strategic communication game should be more pronounced. This difference between floor and committee speeches opens new questions about the role – possibly symbolic – of ideology for political-administrative interactions. A *post-hoc* interpretation of the results might suggest that ideology matters only in highly politicised arenas, where politicians are expected to follow a "partisan script",

whereas its salience weakens in less politicised venues, where politicians can focus more on the merits of the information produced by agencies.

With this paper, I make two major contributions to the literature on bureaucratic politics. First, I present the largest attempt to measure the role of bureaucratic bodies in legislative politics, presenting fine-grained data for about 300 agencies and approximately 40 years of floor and committee speeches. Second, I integrate political and structural explanations of an important example of political-administrative interaction, namely the role of bureaucracies in contributing to the legislative process, showing how institutions can mitigate ideological conflict.

2 Bureaucrats and Politicians: A Cheap Talk

A vast literature in political science has studied how the ideological leanings of bureaucracies and politicians affect structural characteristics about the agency as well as informal behaviour of agencies and politicians, with topics spanning from delegation of authority (Epstein and O'Halloran, 1999; Huber and Shipan, 2002) to bureaucratic oversight (McCubbins and Schwartz, 1984; Lowande, 2018), and executive policy-making (Bolton et al., 2016; Potter, 2019).

Epstein and O'Halloran (1994), for instance, show that the discretion delegated to bureaucracy decreases as the preferences of Congress and those of executive agencies move apart, and Bolton et al. (2016) find that, as ideological disagreement between the President and the Office of Information and Regulatory Affairs – which is in charge of overseeing every rule passed by administrative agencies – increases, the agency's review times increase, thus inhibiting the entry into force of rules when the agency is ideologically apart from the

President. Similarly, Potter and Lowande (2020) show how legislators who are ideologically distant from the proposals made by the Environmental Protection Agency are more likely to scrutinise the proposal by filing requests for documents, additional hearings, and extended time for public participation.

This strand of the literature shows that political factors such as the ideological leanings of the actors involved in administrative politics affect both formal (e.g., delegation of authority) and informal (ex post oversight) practices. Ideology, I argue, is important for explaining legislators' use of bureaucratic information too.

Politicians rely on bureaucratic offices to acquire information about both the nature and the solutions to the problems they face. However, bureaucratic information can be a channel for bureaucracies to achieve their own goals, and depart from the policy preferences of politicians (Aberbach et al., 1981). Because of the information asymmetries between bureaucracies and politicians, it is hard for politicians to know whether the information provided by bureaucrats is consistent with their own policy preferences (Miller, 2005). As a result, politicians face the moral hazard of trusting bureaucracies who might in fact pursue different policy goals. Politicians will therefore have to decide when to trust bureaucracies and to use the information and the evidence they produce. The canonical model of this strategic form of communication is Crawford and Sobel's (1982) cheap talk communication model. In particular, I follow Gailmard and Patty (2012) and apply this framework to a situation in which an expert bureaucracy produces information in the attempt to shape the decisions taken by legislators. I also extend the original model by looking at the role of the agency's independence on the credibility of the information produced by the agency.

Crawford and Sobel (1982) present a game in which an actor, the sender, tries to influence the decision of another actor, the receiver, who has the power to make authoritative

decisions whose consequences affect the welfare of both actors. Let us consider the following hypothetical scenario. Country C is experiencing a harsh economic crisis. Legislator L needs to pass a law aimed at restarting the economy. Agency A, because of its mission and capacity, is in a strong position to provide L with the necessary information in order to maximise the positive outcome of the law. Both L and A have known policy preferences. Importantly, L cannot verify the quality or veracity of the information produced by A. The key prediction of the model is that the probability of truthful communication increases as A and L's preferences over outcomes become more similar.

When ideological disagreement between a bureaucratic agency and a legislator is high, the probability that the legislator uses the information produced by the agency in debating and passing legislation decreases.

Hypothesis 1: Legislators' use of bureaucratic information decreases with ideological distance.

One important assumption of the model is that legislators cannot rely on an independent system which verifies the information produced by agencies, and hence have to count on ideology as a heuristic when deciding whether the information improves their expected utility or not. Yet, I argue, one such institution exists. When agencies are insulated from political pressures, their ideological leaning is less salient and the legislator-agency ideological distance plays a weaker role in the communication game. Independence, acting as a credibility-enhancing mechanism, tempers the distrust of politicians towards information produced by ideologically distant agencies.

Granting statutory independence to agencies is a powerful signal of credible commitment and can be an effective solution to the policy inconsistency inherent to changing

governments (Miller, 2005). A clear example of this is the independence of central banks for the credibility of monetary policies and for controlling inflationary tendencies (Cukierman et al., 1992; Keeper and Stasavage, 2003). These arguments are widely diffused in studies of central banks and independent regulatory agencies (Jensen, 1997; Gilardi, 2002). They are at the core of the classical work in bureaucratic politics and they are also foundational in the European scholarship on regulatory capitalism and regulatory agencies (Majone, 1997; Gilardi, 2005). Empirically, agency independence has been shown to be important for bureaucratic policy-making, in particular for perceived and objective quality of regulation (Bertelli and Whitford, 2009; Koop and Hanretty, 2018).

Independent agencies, and the statutory provisions which define their relationship with political officials, make the ideological leaning of the agency less salient for legislators. Agency independence counteracts the effect of the ideological position of agencies on legislators' use of bureaucratic information.

Hypothesis 2: The effect of legislator-agency ideological distance on legislators' use of bureaucratic information decreases for more independent agencies.

3 Measuring Bureaucratic Influence

The influence exerted by bureaucratic bodies in the legislative process has generally been measured either qualitatively (Carpenter, 2001; Page, 2012) or through perception measures. If qualitative measures, though benefiting from "deep" observation and multiple sources of data, are limited to few cases, answers to questions like "How influential do you think agency x is?" are easily susceptible of social desirability bias. Both self-reported measures of *received* influence (i.e., legislator attributing an influence score to actors) as well as self-

reported measures of *exerted* influence (i.e., actor self-evaluating their influence) could either deflate or inflate the factual influence exerted by the bureaucracy.

Blom-Hansen et al. (2020) address social desirability bias with a set of experimental designs and find evidence in support of the demand and supply of bureaucratic influence.They first show that a minority of bureaucrats are willing to organise the information they pass on to politicians "in a way that makes it easy for politicians to choose the solution that bureaucrats consider the best". They also show that politicians rely significantly on bureaucrats' expertise and are also susceptible to the way bureaucrats frame the information.

Beyond perception measures, Kroeger (2020) exploits the fact that some US states publish the number of bills sponsored by state departments and is therefore able to measure the success rate of department-sponsored bills. She 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. She also finds that legislators who gain positions of power (i.e., committee chairs and co-partisan with the governor) are more likely to introduce agency-sponsored bills. This is a very accurate way of looking at bureaucratic role in the legislative process. However, as acknowledged by the author, just looking at sponsored bills neglect the possibility that agencies kill, stop, or postpone bills. The sample of bills therefore suffers from selection on the dependent variable, for it just includes a particular group of bills.

I present a new large scale measurement strategy that is more flexible in capturing the extent to which an agency is influential in the legislative process by applying natural language processing methods to a large corpus of floor and committee speeches given by US legislators, detecting when legislators use agencies' information and extracting what type of information they use. This measurement has quantitative and qualitative advantages over

previous methods. First, by looking at floor and committee speeches, I am able to trace how legislators use the evidence produced by bureaucracy on a daily (or debate) basis. Second, I am able to measure how many times legislators use bureaucratic information. Do they briefly mention what said by the Federal Reserve in one speech, or do they use the Federal Reserve's reporting and expertise multiple times in a speech to back their arguments? Qualitatively, by focusing on the frequency of specific evidence-based terms, I can measure the portion of information that taps into the agency's expertise and which is ultimately grounded in facts and statistical evidence attributable to the agency. I can therefore isolate the evidence-based part of information produced by agencies and used by legislators.

Floor and congressional committee debates are appropriate venues where to look for such information. It is in congressional committees where legislators have detailed discussions about policy, advancing their arguments in support of specific bills, or proposing amendments to existing laws. Similarly, in floor debates legislators can focus on the core parts of the law and make more general considerations about policy. In both venues, legislators might find useful to use the information produced by agencies to strengthen their arguments or to persuade members of the opposition. In particular, by looking at both congressional committees and floor debates, I am able to capture both the more political rhetoric of floor speeches as well as the more informal, substantive conversations going on in committees.

4 Legislators' Use of Bureaucratic Information

The key assumption of the proposed measurement strategy is that legislators' use of bureaucratic information can be detected by parsing the syntactic relations of terms in segments of

text (e.g., sentences). Syntactic analysis can in fact identify the action of saying something, the subject carrying out the action, and the object of the action. For instance, let us consider a legislator who wants to use the information produced by the FED to strengthen her argument in a floor speech. She might say something like "The Federal Reserve *[subject]* said *[action]* that higher interest rates will strengthen the economy *[object]*". By creating extraction rules that detect certain syntactic relationships, I can therefore match every instance in which an agency is used as a source of information, while at the same time measuring the frequency of words that pertain to facts and evidence.

Syntactic analysis is a new frontier in political science research, but few promising applications show the benefit of retaining dependency relationships between words when analysing text. Van Atteveldt et al. (2017), for instance, show the differences in how US and English-language Chinese media covered the 2008-9 Gaza war and find how US media underscore Hamas' attacks and Israel's right to defense, whereas Chinese media do not portray Hamas as attacking and focus more on the Israeli military operation and the humanitarian consequences. In a very different context, Vannoni et al. (2020) apply syntactic analysis to a corpus of US state laws to estimate delegation of powers to governors of US states. By extracting syntactic structures encoding obligations, prescriptions, and authorisations, they create a validated measure of delegated authority and test the classical prediction whereby divided government delegate less authority to bureaucratic bodies (Epstein and O'Halloran, 1999; Franchino, 2004). Consistently with theory, they show that the number of provisions delegating powers to the governor is associated with government unity. Similarly, Ash et al. (2020) shows how these methods can efficiently extract workers' rights and duties from labour union contracts.

The measurement strategy I propose consists of three steps. First, I split every speech

mentioning the name of an agency into sentences and process them using a syntactic dependency parser. The parser tags parts of speech (e.g., subject, verb, predicate, etc.) and detect dependency relations. Second, I extract clauses that match pre-defined rules. By pre-defining certain extraction rules (e.g., *subject + say verb + quote*), I can isolate the action of saying something, the source of information, and the content of the quote. I am therefore able to isolate sentences where legislators quote bureaucratic agencies or document and reports produced by agencies. Third, I isolate the quote, namely the actual piece of information used by legislators, and measure the extent to which the quote reports facts and statistical evidence. Eventually, I will obtain a sample of speeches where agencies are used as sources of information, and every speech will receive a continuous measure of the use of facts and statistical evidence. Theory predicts that this measure will be lower when ideological distance between the legislator and the agency used as source is large.

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

First, I tag and parse the text with the spaCy parser (Honnibal et al., 2013; Honnibal and Johnson, 2015).[1] SpaCy operates as a supervised learning algorithm, with the goal of making predictions based on training and labelled data. It achieves state-of-the-art performance on both accuracy and speed (Choi et al., 2015). After splitting speeches into sentences, the parser tags parts of speech and detects dependency relations.

For instance, let us consider the previous example about a congressperson reporting the position of the FED. In her speech, she says: "The FED said that higher interest rates will strengthen the economy". The tokens – namely each single word – within this sentence have syntactic properties and follow specific dependency relations. For instance, "The" refers to

[1] Version 2.1.6 implemented through the R package *spacyr*.

the "FED", which in turns 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 and its lemmatised version, the part-of-speech, the ID of the head token (namely the "parent" token), and the type of dependency relation. For instance, the head token ID of the token "higher" and "interest" is the token ID 7, namely "rates".

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

Table 1: Dependency parsing example.

4.2 Step 2: Extraction Rules

Once the parser has tagged each token of the sentence, I annotate the sentence based on extraction rules that detect when a statement is reported, the source of the statement, and the content of the statement.

I create a comprehensive list of five extraction rules that match who-says-what syntactic structures: one for direct or indirect statements (the FED said; as said by the FED), one for "according-to" structures (according to the FED), one for direct nominal recommendations (the FED's proposal is), and one for indirect nominal recommendations (the FED's proposal to). To match direct and indirect statements, I specify a vector of "say verbs" so that the parser marks the lemmatised version of the verb – therefore capturing verbs declined in

every form (active or passive) or tense – and its respective subject or, in case of indirect statement, the agent. For "according-to" structures, the parser detects the lemmatised token "accord" and the object of the preposition, which will be the source of the information. For direct and indirect nominal recommendations, I specify a vector of recommendation-related tokens for the parser to detect, and their possessive determiner – i.e., the owner of the recommendation – will be labelled as the source of the recommendation. Finally, all the tokens that are dependencies of say verbs, recommendation-related verbs, or according-to structures are labelled as quotes. For instance, for the FED's example, the information about inflation boosting the economy could be used in a speech in five different ways. Table 2 below reports the precise tokens and syntactic structures used to compile the extraction rules, as well as the toy sentences in which a legislator could use the information produced by the FED.

Extraction Rule	**Syntactic Structure**	**Sentence Example**
Direct Statement	subject + say verbs	The FED said that higher interest rates will strengthen the economy.
Indirect Statement	agent + say verbs	As reported by the FED, higher interest rates will strengthen the economy.
According-to Structure	accord + object of preposition	According to the FED, higher interest rates will strengthen the economy.
Direct Nominal Recommendation	recommendation + possession modifier	The FED's recommendation is to increase interest rates.
Indirect Nominal Recommendation	recommendation + possession modifier	I fully endorse the FED's recommendation to increase interest rates.

Table 2: Illustrative examples. Say verbs and recommendation-type words used to match syntactic rules reported in the Appendix.

I then match these parsed sentences with the extraction rules. Figure 1 shows the final output of the syntactic analysis for the indirect nominal recommendation extraction rule, one that might seem particularly challenging to analyse. Dependency trees of other rules are shown in Figure A1 in the Appendix.

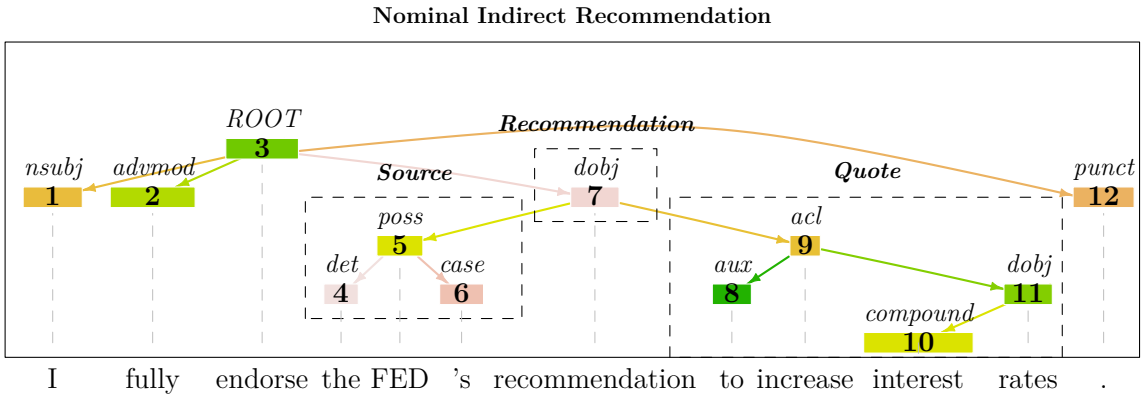

Figure 1: Parsed dependency trees of one illustrative example where the FED is used to support a statement. Implemented through the *rsyntax* package in R.

4.3 Step 3: Analysing Quotes

For each parsed sentence, I extract the quote – namely the information produced by the agency that has been reported by the legislator and estimate the presence of facts and statistical evidence. Sentences which contain the name of an agency but where the agency is not used as a source of information are removed, for I want to compare the use of evidence when the agency is used as a source of information. I follow (Hargrave and Blumenau, 2020) and apply a dictionary-based approach to estimating the use of facts and statistical evidence in speeches. This step is important to ensure legislators' are actually using expertise-based evidence and information produced by the agency.

Every quote is assigned a score which equals the absolute frequency of words belonging to a pre-defined dictionary of facts and statistical evidence.[2] I use the 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

[2]In the Appendix I show the results are robust to using the term-frequency inverse-document-frequency measure of use of facts and evidence (Table A1).

used to express quantities.[3]

Once the dictionary analysis has been applied to all the quotes, I merge all the quotes back at the speech level, and sum the facts-scores of each quote. The key quantity of interest is at speech level and combines all the quotes contained in the speech. More formally, it will be determined by the following metric:

$$\text{Use of Bureaucratic Information}_s = I_s = \sum_{i=1}^{N} facts_i, \quad \text{where} \quad facts_i = f_{Qi \in Dict} \tag{1}$$

where the use of information for each speech I_s is given by the sum of quote-level measures of the use of facts and statistical evidence $facts_i$, which in turn is given by the absolute frequency of the tokens of the quote Q_i which appear in the fact dictionary $Dict$. When more agencies are used as source of information in a speech, I consider the agency with the largest facts-score the one used as source of information.

4.4 Limitations

This method has several advantages. First, it is fully scalable and transparent. Second, once the extraction rules are defined, it does not resort to subjective measures, while it allows for a large time coverage and a large sample of agencies. Finally, it allows to produce micro-level estimates, detecting (1) when the agency is mentioned as a source of information, (2) the extent to which the information is about facts and statistical evidence. However, there are some limitations that are worth being discussed.

[3]The validation of this dictionary has been discussed by Bellodi (2021). Very summarily, the LIWC dictionary has been extensively and successfully validated by Hargrave and Blumenau (2020) in an almost identical setting (legislative speeches in the UK House of Commons). It has also been shown that context-dependence is not particularly concerning for dictionary consisting of facts and statistical evidence, which are more representative of objective attributes and hence less dependent of the context in which are used. Bellodi (2021) shows that the LIWC dictionary performs well at matching a manually labelled corpus of texts from different contexts.

First, the proposed information-based definition of influence is different from the standard counterfactual definition used in the social sciences, which assumes influence to occur if, say, an individual deviates from a predicted path of behaviour (March, 1955). Second, and relatedly, this measurement strategy does not distinguish between politicians genuinely using bureaucratic information to form their opinions about policy and politicians who strategically deploy bureaucratic expertise for strengthening a pre-existing argument. This difference, though subtle, makes the proposed measure silent about the supply and demand of influence, for it is not possible to detect whether influence is exerted by agencies or strategically crafted by politicians.

Another important limitation pertains to the multiple ways legislators can use bureaucratic information. Legislators' can deploy bureaucratic information both explicitly and implicitly. Explicitly, for legislators could use the information produced by bureaucracies while at the same time acknowledging the source. Implicitly, for legislators could be exposed to the information produced by agencies and act without acknowledging the ownership of the information. By anchoring the quote to the name of the agency, the proposed method is suitable for an explicit way of using bureaucratic information.

Nevertheless, as shown in the previous sections, perfect measures of bureaucratic influence are hard to produce, for researchers often face issues related to social desirability, experimental realism, or unobservable behaviours. The proposed measurement strategy, relying on observational data, does not suffer from social desirability bias or simulated reality problems, but it might capture just a partial picture of the process of bureaucratic influence.

5 Sample & Data

5.1 Speeches and Agencies

I apply the proposed method to a corpus of 2,306,300 floor (1980-2016) and 4,545,416 (1990-2019) committee speeches. I downloaded transcripts of all floor speeches from the Stanford University Congressional Repository and I obtained transcripts of congressional committee sessions through ProQuest.[4] After replacing the various ways in which agencies are mentioned with a standardised 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) and Selin (2015), integrated with information directly obtained from the US government website (`usa.gov/federal-agencies`), for a total of 426 agencies. Descriptive statistics about the sample of speeches are reported in Table 3.

Descriptive Statistics	**Floor Speeches**	**Committee Speeches**
Initial Sample of Speeches	2,306,300	4,454,416
N. Speeches mentioning agencies	294,518	498,558
N. Sentences mentioning agencies	782,065	2,933,105
N. Sentences with agency used as source	36,707	95,802
N. Speeches with agency used as source	28,945	42,253
Average use of facts and evidence	1.5	2.2
N. Agencies used as source	262	275

Table 3: Sample of speeches and final sample size.

A total of 294,518 floor speeches and 492,598 committee speeches mention the name of at least one agency, namely 12.7% and 10.8% of the total sample, respectively. I parse these speeches into sentences and remove all sentences which do not contain the name of an agency. I then apply the extraction rules described in Table 2 to each sentence and subset

[4] A note on the quality of congressional committees' data and the speech parsing steps are reported in Section A3 of the Appendix.

the sentences in which the agency is treated as a source of information. I extract the quotes from each sentence and apply the dictionary analysis to the quote, measuring the number of words that belong to the facts and statistical evidence dictionary. The final dataset consists of 28,945 floor and 44,993 committee speeches using the agency as a source of information. The average outcome is 1.5 fact-words (SD = 1.7) in floor speeches, and 2.2 (SD = 4.9) in committee speeches.

5.2 Ideological Distance and Independence

Ideological distance and agency independence are respectively the key independent variable and the moderator. I construct the measure of ideological distance between legislators and the agency as the absolute value between the DW-NOMINATE scores of each legislator and each agency used as source of information in the speech. Data on legislators' ideal point are from Lewis et al. (2020). For agency ideology, I use the dataset assembled by Chen and Johnson (2014), who produce donation-based ideology estimates for 79 federal agencies across five presidencies, from the first Clinton Presidency to the first Obama Presidency (1993-2012). Chen and Johnson (2014) use federal bureaucrats' campaign contributions to individual politicians as input to estimating agency ideology, and produce estimates comparable with the DW-NOMINATE Common Space scores. Because of the limited availability of data on agency ideology – both with respect to time and sample of agencies – the analysis will be limited to speeches given by legislators between 1993 and 2012 for which an agency ideology estimate is available.

I measure independence in two ways. The first one is a dichotomous variable which equals 1 if the agency is listed as an independent agency on the website of the US government. The second one is a continuous estimate of agency statutory independence produced by Selin

(2015). I match the speeches dataset with the name of the agency with Selin's dataset on the structural independence of US federal agencies, which captures agency independence along two dimensions: independence as the ability of an agency to make policy decisions without political interference; and independence as statutory limitations and requirements placed on the officials who manage the agency. The indicators are derived by modelling 50 structural features about the agencies with a Bayesian latent variable model. The estimates range between 0 and 4, with higher values signifying higher independence.

Both these two measures of independence do not capture variation over time. However, while some statutory features could change over time, the structural features with the largest factor loading are likely to remain fixed. In fact, as shown by Selin (2015, 983-4), the estimates about the independence of agency as the ability to take policy decisions without political interference derived from the initial statute establishing the law and from the US Code used by the author in 2013-14 are not different from each other. The second dimension of independence, however, displays some over time change that needs to be taken into account when interpreting the results.

6 Methods

I am interested in two relationships: the effect of ideological distance on the use of bureaucratic information (H1), and how agency independence moderates such effect (H2).

There are three methodological concerns for identifying these effects. 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. Education, socio-economic background, but also their level of inter-

est towards bureaucratic policy-making. Second, on the agency side, some policy domains – e.g., financial or environmental regulation – might be more salient and therefore legislators might mention and use the information of the Federal Reserve or the Environmental Protection Agency more often than that produced by other agencies. Third, the salience of agencies/sectors can also change over time, and the fact that legislators use frequently the information produced by one agency might be the result of the agency being highly salient in that particular period of time, rather than being the result of more similar ideological positions.

I address these sources of omitted variable bias with a two-way fixed effects estimator. I leverage within-legislator variation in the use of bureaucratic information holding constant all time-invariant characteristics of legislators as well as yearly shocks that could affect the use of the information produced by agencies. I also include agency fixed effects to account for all time-constant agency differences. In particular, I estimate the following model:

$$I_{s[l,t,a]} = \eta_l + \phi_t + \alpha_a + \beta Distance_{l[a,t]} + u_s \tag{2}$$

where $I_{s[l,t,a]}$ is the use of bureaucratic information in speech s, year t, and given by legislator l, quoting agency a. η_l and ϕ_t are legislator and year dummies, α_a agency fixed effects, and $Distance_{l[a,t]}$ is the ideological distance between legislator l and agency a in time t. In particular, β is the marginal effect of ideological distance on the use of fact-and-evidence when quoting what said by a bureaucratic agency in the speech.

Despite Model 2 accounting for all the agency, year, and legislator-level characteristics that remain fixed, legislators might use more or less information produced by agencies as a result of the changing salience of the agency. In times of financial instability, politicians

will be more likely to use the information of the Federal Reserve, whereas in times of environmental disasters, they might count more often on what said by the Environmental Protection Agency. To account for this source unobserved heterogeneity, I include in all the specifications the number of mentions the agency receives in floor debates every year, which is a good proxy of the time-changing nature of agency salience. Furthermore, I estimate agency-year fixed effects to account for time-changing agency-level confounders which might affect the credibility of the agency's expertise and hence legislators' use of agencies' information. Standard errors are clustered by agency.

This specification has several advantages. First, legislator dummies sweep out all the variation at legislator level. Second, agency dummies account for all time-invariant features of agencies such as the history, culture, mission, and policy sector. Third, the changing saliency and agency characteristics are captured by agency-year fixed effects. Fourth, common shocks or reforms affecting the bureaucracy as a whole are captured by year fixed effects. However, I cannot condition on time-varying characteristics of legislators, which might bias the estimation.

To estimate the moderating effect of agency independence (H2), I build the following multiplicative interaction model:

$$I_{s[l,t,a]} = \eta_l + \phi_t + \beta Distance_{l[a,t]} \times Ind_a + u_s \tag{3}$$

where Ind_a is a measure of agency independence. Because Ind_a does not change over time, I cannot estimate agency fixed effects. I nonetheless include the number of mentions of agencies in floor debates every year as a proxy of the changing salience of the agency.

7 Results

Democratic Senator Leahy's speech reported in the introduction makes the largest use of bureaucratic information. The extraction strategy outlined above matched 13 instances in which the Office of Management and Budget (OMB) was used to strengthen the Senator's argument about dairy industry in New England. His statement is dense of evidence produced by the OBM. For instance, the Senator claims that "during the first 6 months of operation, the OMB reported that New England's dairy farm income rose by an estimated $2,227 million", or again that "the OMB reports that New England suffered a 20% decline in the number of farms with milk cows from 1990 to 1996.", and that "Evidence reported by the OMB shows that neighbouring farmers get the benefit of the higher Compact price", or that "The OMB's report states that the compact could support a small increase in participation during the demonstration period." Overall, 21 terms belonging to the "facts" dictionary appears in the 13 quotes of the OBM. This example shows the precision of the extraction rules, which are able to capture both the OBM's action of saying something, as well as the contents of the OBM's report.

The general pattern of legislators using bureaucratic information is far from being uniform over time, across agency types, and in floor and committee speeches. Figure 2 shows the time trend of the use of bureaucratic information for every year and across three types of agencies: government departments, executive sub-agencies, and independent agencies, for both floor and committee speeches. Perhaps not surprisingly, in committee speeches politicians use bureaucratic information more frequently. The y-axis shows the sum of the number of facts-words used in politicians speeches using agencies as a source of information. Interestingly, while Republicans' average use of bureaucratic information is on average 8.5%

larger than Democrats' in committee speeches, Democrats use more bureaucratic information than Republicans in floor speeches (+22.1%). The EPA, followed by the Treasury and Office of Management and Budget are the three bureaucracies whose information is most often used by legislators in floor speeches, whereas the EPA, the FDA, and the Internal Revenue Service are those with the average highest score in committee speeches.

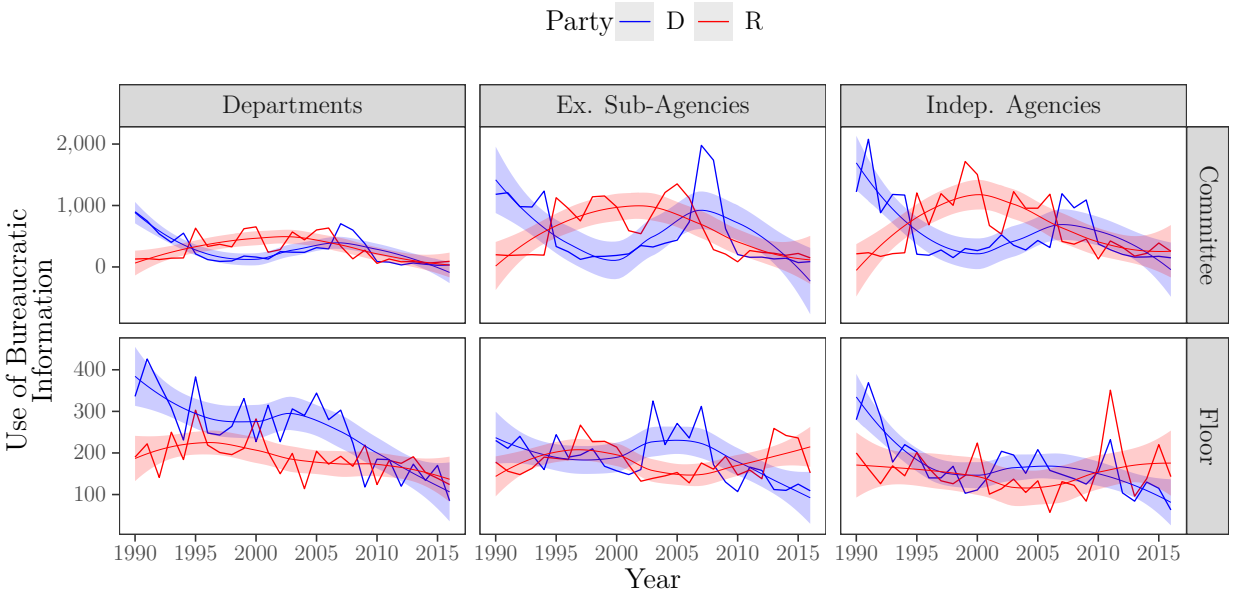

Figure 2: Sum of use of bureaucratic information across all agencies for every year and across three types of bureaucracy, with superimposed loess approximation. Blue line for Democrats, red line for Republicans.

The results of the first test are reported in Table 4. Across various specifications, ideological distance has a negative effect on the use of bureaucratic information, but only for floor speeches. A one-unit increase in ideological distance leads to a decrease in the frequency of facts and statistical evidence by .27-.29. If we compare the change in terms of the average use of information (1.47), the marginal effect of ideological distance is associated with a decrease equal to 19% of the average. Surprisingly, the effect is not distinguishable from zero and in the opposite direction for committee speeches.

	DV: Use of Bureaucratic Information					
	Floor Speeches			Committee Speeches		
	(1)	(2)	(3)	(4)	(5)	(6)
Ideol. Distance	-0.25^{***}	-0.26^{***}	-0.28^{**}	0.34	0.34	0.05
	(0.08)	(0.08)	(0.11)	(0.29)	(0.31)	(0.31)
Log N. Mentions	0.03			-0.27^{**}		
	(0.04)			(0.10)		
Legislator FE			✓			✓
Agency-Year FE		✓	✓		✓	✓
Committee FE				✓	✓	✓
Mean DV	1.47	1.47	1.47	2.38	2.38	2.38
Num. obs.	10,864	10,864	10,864	17,391	17,391	17,391
R^2	0.02	0.09	0.35	0.06	0.09	0.13
N. Years	20	20	20	20	20	20
N. Agencies	53	53	53	57	57	57
N. Legislators			3,010			983

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 4: OLS estimates. SE clustered by agency in parenthesis. DV is frequency of of facts and statistical evidence in quotes of agencies mentioned in legislators' speeches. Agency and year FE always estimated. Other fixed-effects included in the specifications as reported in the table.

Table 5 reports the results of the multiplicative interactive models for floor speeches. Consistently wit the findings for H1, ideological distance has a negative effect on legislators' use of bureaucratic information across all the four specifications. Model (1) reports the results of a regression where the simple dummy indicator of independence (i.e., whether the agency is an independent body) is used. Models (2) and (3) employs Selin's (2015) independence estimates separately. Model (4) includes both dimensions of independence, the statutory requirements and limitations imposed on agency officials and the degree to which the decision-making process of agencies is subject to political review.[5]

The positive interaction coefficients lend support to the counteracting effect of agency independence. The effect seems to be mostly driven by the independence of the agency officials rather than the political review process. Importantly, these effects are distinguishable

[5]In Table A2 in the Appendix I report the results for committee speeches.

from zero even when conditioning on the number of mentions of agencies, and hence are not the driven by the changing salience of the agency. As shown in Model (4), a one-unit increase in the independence of agency officials reduces the effect of ideological distance on the use of bureaucratic information by almost 75% ($.252/-.340=-.741$).

	DV: Use of Bureaucratic Information Floor Speeches			
	(1)	(2)	(3)	(4)
Ideological Distance	-0.424^{***}	-0.313^{**}	-0.337^{**}	-0.340^{**}
	(0.161)	(0.134)	(0.148)	(0.138)
Independent Agency (dummy)	-0.067			
	(0.134)			
Ideol. Dist. × Ind. Agency (dummy)	0.516^{**}			
	(0.221)			
Independence: Decision Makers		-0.116^{*}		-0.025
		(0.060)		(0.081)
Ideol. Dist. × Ind: Dec. Makers		0.319^{**}		0.252^{*}
		(0.126)		(0.147)
Independence: Political Review			-0.101^{***}	-0.090
			(0.039)	(0.056)
Ideol. Dist. × Ind: Pol. Review			0.179^{**}	0.064
			(0.089)	(0.102)
Log Mentions	0.014	0.002	-0.010	0.002
	(0.036)	(0.039)	(0.036)	(0.037)
Legislator FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Num. obs.	10,043	10,038	10,038	10,038
R^2	0.292	0.287	0.287	0.288
N. Legislators	2,926	2,926	2,926	2,926
N. Years	20	20	20	20

$^{***}p<0.01$; $^{**}p<0.05$; $^{*}p<0.1$

Table 5: OLS estmates of multiplicative interaction models. SE clustered by agency in parenthesis. Ideological distance is interacted with three different independence indicators: whether the agency is an independent body (dummy) and with Selin's (2015) estimates of independence as limitations and requirements for officials who manage the agency (decision makers) and independence as absence of political interference (political review).

8 Discussion and Conclusions

The role of unelected bureaucracies in democratic government has received increasing attention in the last decades (Tucker, 2018). Empirical knowledge on the conditions under which this happens has two important normative implications. The first one concerns the role unelected officials exert on democratic functions. In the quest for bureaucratic legitimacy, it is important to understand the extent to which the information and evidence produced by bureaucratic bodies penetrate the legislative process. The second one concerns the use of evidence in decision-making aimed at increasing the quality of policies. Bureaucratic expertise can in fact enhance the quality of legislation and, as a result, improve policy outcomes.

In this paper I presented a strategic communication game, which predicts that actors are less likely to undertake constructive communication when they have divergent preferences over policy outcomes. I extended this model and proposed that the institutional independence of agencies can counteract the negative effect of ideological distance on the quality of communication. Operating as a credibility enhancing mechanisms, agency independence represents a pledge of non-political information, which is more likely to be employed by legislators despite being ideologically far from the agency.

I introduced a new large scale measurement strategy that employs natural language processing techniques and syntactic analysis to detect when legislators use the information produced by agencies in floor and committee debates and measure the extent to which the information consists of facts and statistical evidence. I presented fine-grained data at speech level for over 40 years and more than 250 agencies. The findings support the theoretical expectations, but only for floor speeches. When legislators use bureaucratic information in floor debates, the legislator-agency ideological distance reduces the use of

facts and statistical evidence. However, this effect is less than half when the agency is insulated from political pressures, especially when the agency head or officials are required high standards of independence and neutrality. Surprisingly, the effect disappears when we look at committee speeches. The use of bureaucratic information is large in committees, but ideological distance seems not to be its driver.

These findings are important for theories of bureaucratic influence and models of strategic communication. The effect of ideology might be highly heterogeneous based on the "venue" of communication, which might shape politicians' incentives to close ranks along partisan or ideological line, whereas in less politicised venues, other dynamics prevail. Future research could build on this measurement strategy and advance our understanding of the role – possibly symbolic – of ideology in political-administrative interaction. In the same vein as Lowande (2018), who finds that ideological distance between legislators and agencies has a negligible effects on oversight, it might be the case the the effect of ideological differences becomes immaterial when politicians operate in more protected spaces where they can focus on the substantive issues of policies and hence consider the information produced by agencies more objectively.

This paper makes three contributions to the literature on bureaucratic politics. Theoretically, it combines rational choice models of inter-institutional communication with structural aspects of politician-bureaucracy interactions. Methodologically, it introduces a new strategy to measure the role of bureaucracy in legislative politics. Finally, it opens new path for future research to investigate the heterogeneous effects of ideology on political-administrative interactions.

References

Aberbach, J., Putnam, R. D. and Rockman, B. (1981), *Bureaucrats and politicians in Western democracies.*, Harvard University Press, Cambridge, MA.

Ash, E., Jacobs, J., Macleod, B. and Stammbach, D. (2020), 'Unsupervised Extraction of Workplace Rights and Duties from Collective Bargaining Agreements', *International Workshop on Mining and Learning in the Legal Domain* .

Bellodi, L. (2021), Is Partisanship Bad for Bureaucratic Accountability?

Bertelli, A. M. and Grose, C. R. (2011), 'The Lengthened Shadow of Another Institution? Ideal Point Estimates for the Executive Branch and Congress', *American Journal of Political Science* **55**(4), 767–781.

Bertelli, A. M. and Whitford, A. B. (2009), 'Perceiving credible commitments: How independent regulators shape elite perceptions of regulatory quality', *British Journal of Political Science* **39**(3), 517–537.

Blom-Hansen, J., Bækgaard, M. and Serritzlew, S. (2020), 'How Bureaucrats May Shape Political Decisions: The Role of Policy Information', *Public Administration* .

Bolton, A., Potter, R. A. and Thrower, S. (2016), 'Organizational Capacity, Regulatory Review, and the Limits of Political Control', *Journal of Law, Economics, and Organization* **32**(2), 242–271.

Carpenter, D. P. (2001), *The Forging of Bureaucratic Autonomy*, Princeton University Press, Princeton.

Carpenter, D. P. (2010), *Reputation and Power. Organizational Image and Pharmaceutical Regulation at the FDA*, Princeton University Press, Princeton.

Chen, J. and Johnson, T. (2014), 'Federal employee unionization and presidential control of the bureaucracy: Estimating and explaining ideological change in executive agencies', *Journal of Theoretical Politics* **27**(1), 151–174.

Choi, J. D., Tetreault, J. and Stent, A. (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**, 387–396.

Crawford, V. and Sobel, J. (1982), 'Strategic Information Transformation', *Econometrica* **50**(6), 1431–1451.

Cukierman, A., Webb, S. B. and Neyapti, B. (1992), 'Measuring the independence of Central Banks and Its Effects on Policy Outcomes', *The World Bank Economic Review* **6**(3), 353–398.

Epstein, D. and O'Halloran, S. (1994), 'Administrative Procedures, Information, and Agency Discretion', *American Journal of Political Science* **38**(3), 697–722.

Epstein, D. and O'Halloran, S. (1999), *Delegating Powers: A Transaction Cost Politics Approach to Policy Making under Separate Powers*, Cambridge University Press, Cambridge, UK.

Franchino, F. (2004), 'Delegating Powers in the European Community', *British Journal of Political Science* **34**(2), 269–293.

Gailmard, S. (2009), 'Multiple principals and oversight of bureaucratic policy-making', *Journal of Theoretical Politics* **21**(2), 161–186.

Gailmard, S. and Patty, J. (2012), 'Formal Models of Bureaucracy', *Annual Review of Political Science* **15**, 353–377.

Gilardi, F. (2002), 'Policy credibility and delegation to independent regulatory agencies: A comparative empirical analysis', *Journal of European Public Policy* **9**(6), 873–893.

Gilardi, F. (2005), 'The institutional foundations of regulatory capitalism: The diffusion of independent regulatory agencies in Western Europe', *Annals of the American Academy of Political and Social Science* **598**, 84–101.

Hargrave, L. and Blumenau, J. (2020), The Declining Importance of Gender Stereotypes for Politicians' Style in the UK.

Honnibal, M., Goldberg, Y. and Johnson, M. (2013), 'A non-monotonic Arc-Eager transition system for dependency parsing', *CoNLL 2013 - 17th Conference on Computational Natural Language Learning, Proceedings* pp. 163–172.

Honnibal, M. and Johnson, M. (2015), 'An improved non-monotonic transition system for dependency parsing', *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing* (September), 1373–1378.

Huber, J. D. and Shipan, C. R. (2002), *Deliberate Discretion? The institutional foundations of bureaucratic autonomy*, Cambridge University Press, New York.

Ingold, K. and Leifeld, P. (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), 1–18.

Jensen, H. (1997), 'Credibility of Optimal Monetary Delegation', *American Economic Review* **87**(5), 911–920.

Keeper, P. and Stasavage, D. (2003), 'The Limits of Delegation: Veto Players, Central Bank Independence, and the Credibility of Monetary Policy', *American Political Science Review* **97**(3), 407–423.

Koop, C. and Hanretty, C. (2018), 'Political Independence, Accountability, and the Quality of Regulatory Decision-Making', *Comparative Political Studies* **51**(1), 38–75.

Krause, G. A. (1996), 'The Institutional Dynamics of Policy Administration : Bureaucratic Influence over Securities Regulation', *American Journal of Political Science* **40**(4), 1083–1121.

Krause, G. A. (1999), *A Two-Way Street: The Institutional Dynamics of the Modern Administrative State*, University of Pittsburgh Press, Pittsburgh, PA.

Kroeger, M. A. (2020), Bureaucrats as Lobbyists.

Lewis, J., Poole, K., Rosenthal, H., Boche, A., Rudkin, A. and Sonnet, L. (2020), 'Voteview: Congressional Roll-Call Votes Database'.
URL: *voteview.com*

Lowande, K. (2018), 'Who Polices the Administrative State?', *American Political Science Review* **112**(4), 874–890.

Majone, G. (1997), 'The new European agencies: Regulation by information', *Journal of European Public Policy* **4**(2), 262–275.

March, J. G. (1955), 'An Introduction to the Theory and Measurement of Influence', *American Political Science Review* **49**(2), 431–451.

McCubbins, M., Noll, R. G. and Weingast, B. R. (1987), 'Administrative Procedures as Instruments of Political Control', *Journal of Law, Economics, & Organization* **3**(2), 243–277.

McCubbins, M. and Schwartz, T. (1984), 'Congressional Oversight Overlooked: Police Patrols versus Fire Alarms', *American Journal of Political Science* **28**(1), 165–179.

Miller, G. J. (2005), 'the Political Evolution of Principal-Agent Models', *Annual Review of Political Science* **8**(1), 203–225.
URL: *http://www.annualreviews.org/doi/10.1146/annurev.polisci.8.082103.104840*

Moe, T. M. (2012), 'Delegation, Control, and the Study of Public Bureaucracy', *The Forum* **10**(2).

Nicholson-Crotty, J. and Miller, S. M. (2012), 'Bureaucratic effectiveness and influence in the legislature', *Journal of Public Administration Research and Theory* **22**(2), 347–371.

Niskanen, W. (1971), *Bureaucracy and Public Economics*, Aldine-Atherton, Chicago.

Page, E. (2012), *Policy Without Politicians. Bureaucratic Influence in Comparative Perspective*, Oxford University Press, Oxford.

Pennebaker, J. W., Boyd, R. L., Jordan, K. and Blackburn, K. (2015), *The Development and Psychometric Properties of LIWC2015*, University of Texas at Austin, Austin.

Potter, R. A. (2019), *Bending the Rules. Pocedural Politicking in the Bureaucracy*, University of Chicago Press, Chicago.

Potter, R. A. and Lowande, K. (2020), 'Congressional Oversight Revisited: Politics and Procedure in Agency Rulemaking', *The Journal of Politics* .

Selin, J. L. (2015), 'What Makes an Agency Independent?', *American Journal of Political Science* **59**(4), 971–987.

Tucker, P. (2018), *Unelected power: the quest for legitimacy in central banking and the regulatory state*, Princeton University Press, Oxford.

Van Atteveldt, W., Sheafer, T., Shenhav, S. R. and Fogel-Dror, Y. (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**, 207–222.

Vannoni, M., Ash, E. and Morelli, M. (2020), 'Measuring Discretion and Delegation in Legislative Texts: Methods and Application to US States', *Political Analysis* .

Weber, M. (1922), *Economy and Society*, 1968 edn, University of California Press, Berkeley.

Wilson, J. (1989), *Bureaucracy: What government agencies do and why they do it.*, Basic Books, New York.

Workman, S. (2015), *The dynamics of bureaucracy in the US government: How Congress and federal agencies process information and solve problems.*, Cambridge University Press, Cambridge.

Appendix

A1 Extraction Rules

Say verbs used to match syntactic rules : say, tell, show, claim, report, admit, acknowledge, present, explain, state, indicate, recommend, propose, advance, believe, think, affirm, conclude, propose, advise, encourage, argue, contend, set out, inform. Recommendation-type words are: recommendation, advise, suggestion, indication, proposal, attempt, document, idea, project, programme, conclusion, report, program, brief, paper, argument, thesis, statement, survey, study.

A2 Dependency Parsing: Examples

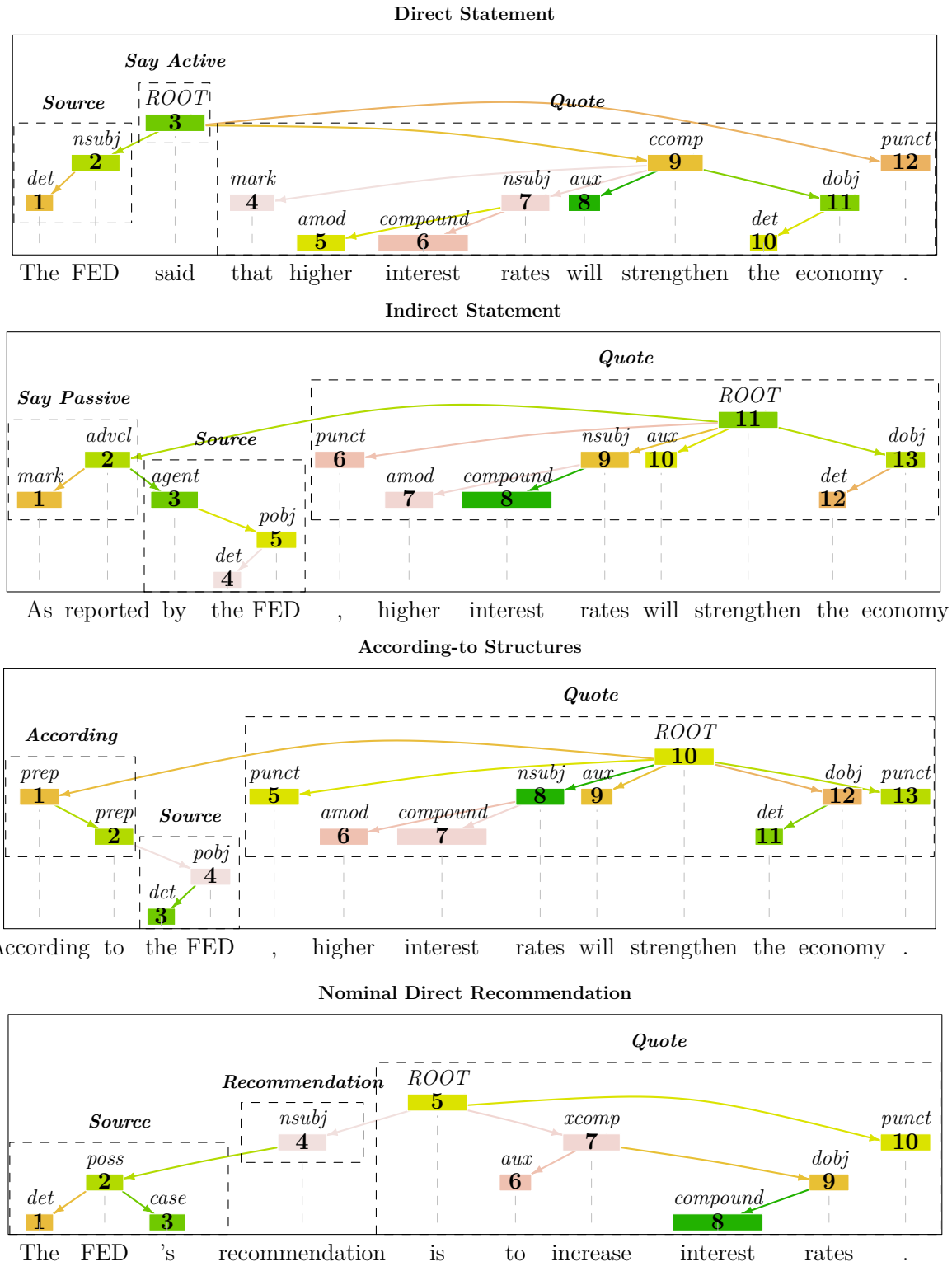

Figure A1: Parsed dependency trees of the three remaining illustrative examples where the FED is used to support a statement. Implemented through the *rsyntax* package in R.

A3 Committees' Speeches: Data Quality

I accessed transcripts of 42,277 congressional committee sessions from ProQuest. Each transcript consists of one text 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 in fact precedes the speech and is reported in capital cases. "Mr. FORD", for instance, marks a new speech. Many individuals are heard in congressional committees. To extract speeches given by politicians, I exploit the fact that at the beginning of each transcript, the names of all congresspersons 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.

A4 Robustness Checks

	DV: Use of Bureaucratic Information (tf-idf)					
	Floor Speeches			Committee Speeches		
	(1)	(2)	(3)	(4)	(5)	(6)
Ideol. Distance	-0.05^{***}	-0.05^{***}	-0.06^{**}	0.08	0.08	0.01
	(0.02)	(0.02)	(0.02)	(0.07)	(0.07)	(0.07)
Log N. Mentions	0.01			-0.06^{**}		
	(0.01)			(0.02)		
Legislator FE			✓			✓
Agency-Year FE		✓	✓		✓	✓
Committee FE				✓	✓	✓
Mean DV	0.31	0.31	0.31	0.55	0.55	0.55
Num. obs.	10,864	10,864	10,864	17,391	17,391	17,391
R^2	0.02	0.09	0.35	0.06	0.09	0.13
N. Years	20	20	20	20	20	20
N. Agencies	53	53	53	57	57	57
N. Agency-Years		758	758		821	821
N. Legislators			3,010			983

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A1: OLS estimates. Dependent variable is sum of tf-idf of fact-words. SE clustered by agency in parenthesis. DV is frequency of of facts and statistical evidence in quotes of agencies mentioned in legislators' speeches. Agency and year FE always estimated. Other fixed-effects included in the specifications as reported in the table.

A5 Multiplicative Interactive Models - Committee Speeches

	DV: Use of Bureaucratic Information Committee Speeches			
	(1)	(2)	(3)	(4)
Ideological Distance	−0.128	0.063	0.027	−0.019
	(0.265)	(0.264)	(0.243)	(0.244)
Independent Agency	0.992***			
	(0.185)			
Ideol. Dist. x Ind. Agency	0.197			
	(0.363)			
Independence: Decision Makers		0.392***		0.555***
		(0.139)		(0.184)
Ideol. Dist. x Ind: Dec. Makers		−0.283		−0.549**
		(0.260)		(0.265)
Independence: Political Review			0.087	−0.159
			(0.091)	(0.131)
Ideol. Dist. x Ind: Pol. Review			−0.019	0.249
			(0.249)	(0.303)
Log Mentions	0.029	−0.011	−0.089	−0.009
	(0.059)	(0.095)	(0.094)	(0.095)
Legislator FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Committee FE	✓	✓	✓	✓
Num. obs.	17,269	17,357	17,357	17,357
R^2	0.093	0.087	0.086	0.087
N. Legislators	981	983	983	983
N. Years	20	20	20	20
N. Committees	174	174	174	174

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A2: OLS estmates of multiplicative interaction models for committee speeches. SE clustered by agency in parenthesis. Ideological distance is interacted with three different independence indicators: whether the agency is an independent body (dummy) and with Selin's (2015) estimates of independence as limitations and requirements for officials who manage the agency (decision makers) and independence as absence of political interference (political review).