

IS PARTISANSHIP BAD FOR BUREAUCRATIC ACCOUNTABILITY?

Luca Bellodi

Word Count: 9,593

Abstract

Models of bureaucratic accountability rest on the ability of politicians to obtain information about bureaucracies. Yet how politicians process this information is an understudied and central topic. I argue that politicians follow partisan motives when updating their beliefs about bureaucracy. I leverage a new measurement strategy and apply word-embedding techniques to a corpus of legislative speeches and estimate partisan beliefs about 323 agencies over 40 years in the UK and the US. I build panel data models and a generalised difference-in-differences design and find that for the governing party beliefs about bureaucracy are on average more positive by 3-4 percentage points, irrespective of the party-agency ideological distance. I show how bias follows from beliefs to behaviour. When co-partisan with the government, (1) politicians' speeches about bureaucratic agencies are less grounded in facts and evidence, (2) and politicians are less likely to oversee ideologically distant agencies.

1 Introduction

A normative tenet of democratic government is that bureaucracies are accountable to elected politicians. To that end, principals design institutions and procedures to limit agency loss and ensure bureaucracies are held in check (McCubbins et al. 1987; McCubbins and Schwartz 1984). Yet for accountability to be sustained, politicians need to observe the performance of agencies, update their beliefs, and respond accordingly. The literature is nonetheless silent on how politicians form their beliefs about bureaucracy and there is no evidence that they are truly Bayesian actors who see, update, and act. Work in political psychology has demonstrated how individuals – both voters and politicians – are motivated reasoners, and how their opinions follow directional motives aimed at confirming pre-existing ideological or partisan views (Iyengar et al. 2019; Druckman et al. 2013; Baekgaard et al. 2017). In this paper I show that partisanship also biases politicians’ beliefs about bureaucracy and their ability to hold agencies to account.

Models of bureaucratic accountability and political oversight assume that political principals receive information about agencies and update their beliefs accordingly. Within the traditional principal-agent framework, this can happen in three different ways, either relying on politicians’ direct observation, third-party reporting, or bureaucracies’ reporting (Lupia and McCubbins 1994). The main issue at stake is therefore how to build institutions whereby the political principal can access bureaucracies’ private information in an optimal way. Even alternative accounts which move away from the canonical view of accountability aimed at reducing information asymmetries still focus on the information flow between giving-account agencies and holding-account principals (Busuioc and Lodge 2017; Schillemans and Busuioc 2015). What has not been addressed is how this information is processed

by politicians. Even the bureaucratic reputation literature (Carpenter 2001; Carpenter and Krause 2012), that most strongly advocates looking at groups' and audiences' beliefs about bureaucratic agencies, does not deal with possible biases that may affect the process of formation of reputation/beliefs among the agency's audiences.

“Tell me your party, and I'll tell you what you think” is a good summary of empirical and theoretical work on partisan motivated reasoning. Voters and politicians process information using directional motives consistent with pre-existing ideological or partisan preferences (Taber and Lodge 2006; Bartels 2002). Individuals strongly identify with their party-group and therefore make judgements about reality through a partisan lens (Green et al. 2002). In what follows, I present a theory that builds on insights from motivated reasoning and identity theory scholarship to show how partisan bias shapes beliefs and behaviour of politicians vis-à-vis bureaucratic agencies.

Operating as an identity-shaping force, I claim that partisanship drives politicians' beliefs about bureaucratic bodies. Because the executive is responsible for public policies, politicians who are co-partisans with the executive perceive agencies as members of their team, and this co-membership makes beliefs about agencies more positive. This in-group form of bias reduces the importance of substantial policy-oriented aspects such as the ideological distance between politicians and agencies. As a result, how ideologically close politicians and agencies are plays a negligible role in explaining what politicians think about agencies. However, it could be argued that politicians simply pretend to have biased beliefs to protect their party's image. In rebuttal of this statement, I argue that partisan bias hinders politicians' account-holding *behaviour* too, both in public and private forums. When co-partisan with the government, politicians change their rhetorical style when they argue about agencies during legislative debates (public accountability forum), resorting to a

less factual and more politicised style. Furthermore, partisan bias limits legislators' informal oversight activities (private accountability forum), for it weakens the traditional relationship between agency-legislator ideological disagreement and oversight activities. Partisanship hinders politicians' ability to objectively hold agencies to account: all else equal, being co-partisan with the government makes accountability weaker.

Empirically, I introduce new data on partisan beliefs about bureaucratic agencies in the US and the UK. I recover the semantic meaning politicians attach to the agencies they mention in legislative speeches by employing word-embedding techniques. I estimate word vector representations for every agency from local corpora of all the speeches given by Republican and Democratic congresspersons and Labour and Conservative MPs separately and in any given year, producing estimates for each party, agency, and year. I find that beliefs are about 3-4 percentage points more positive when there is partisan alignment between the party and the executive, even after conditioning on agency-party ideological distance, whose effects on beliefs is not distinguishable from zero. The frequency of facts- and evidence-based words in speeches about bureaucracy decreases by about 10% compared to the average value for the US and by 25% compared to the average for the UK. Finally, co-partisans with the executive are less likely to request information from ideologically distant agencies compared to when they are at the opposition, with the difference in probability of oversight between co-partisan and opposition member legislators' resulting from a one-unit increase in ideological distance equal to -0.22 percentage points.

While other studies document accountability deficits (Schillemans 2011; Schillemans and Busuioc 2015), this is the first attempt at unveiling an obstacle to accountability that stems from partisan identity and is not the product of strategic choices or lack of skills of politicians. I present evidence on the key parts of the chain that leads partisan identity

to biased beliefs and weakened accountability. This is the main contribution of this paper. Furthermore, bringing partisan bias inside the study of bureaucracy has important implications for bureaucratic incentives and reputation. While we know how oversight institutions shape agencies incentives to exert effort (Turner 2017) or to make informed policy decisions (Patty and Turner 2020), we do not know how partisan bias distorts such incentive structures. Similarly, while scholars mostly look at how agencies build their unique reputation and its effects on bureaucratic autonomy (Carpenter 2001), or at how agencies respond to reputational threats (Maor and Sulitzeanu-Kenan 2016), more attention should be paid to how reputation (namely beliefs aggregated in some ways) suffers from biases that do not pertain to the political leaning or administrative characteristics of the agency, and how politicians and agencies interact in light of such biases. Finally, understanding how politicians form their beliefs about bureaucracies is central to the effectiveness of accountability mechanisms designed to make sure agencies are kept in check, for partisan bias has the potential to undermine the effective control of bureaucratic policy-making.

2 Partisan Bias in the Administrative State

In democratic government, the executive is responsible for administering public policies through bureaucratic bodies. Despite varying level of autonomy, bureaucratic agencies respond to the political will of the executive, and a bureaucracy doing a poor job has noticeable implications for the consensus of the incumbent party (James 2011; Nielsen and Baekgaard 2015; James and John 2007). This is consistent with the evidence from the retrospective voting literature (e.g. Erikson 1989), both at the national and local level (De Benedictis-Kessner and Warshaw 2020). Because politicians have a strong partisan identity,

they follow partisan reasons when forming beliefs about what the government does or is responsible for, including bureaucratic bodies. Beliefs are more positive when the party is in power, and more negative when the party is at the opposition. This bias, I argue, limits politicians ability to hold agencies to account.

There is a vast literature in political science that characterises partisanship as a political identity which is able to affect opinion and behaviour (Bartels 2002; Mason 2015). While Republicans might view positively what is done by their co-partisan President, they would evaluate the same situation under a Democratic President more negatively just for the fact they are not from the same team (Iyengar and Westwood 2015). In social identity theory, this *a priori* defence of somebody's own team is called in-group bias, for group membership is sufficient for automatically biasing individuals' judgements of the merits of the different groups (Tajfel 1987; Mason 2015). Politicians feel emotionally and socially attached to the party, and tend to view more positively what the party is doing or is responsible for. The process of belief formation about bureaucracy follows this sports-fan dynamic too.

HYPOTHESIS 1: Beliefs about bureaucracy are more positive for the party in government.

In theories of bureaucratic politics, politicians' beliefs about bureaucratic agencies are either neglected or considered equivalent to the ideological agreement between a policy-motivated principal and the agent. As a result, interactions between politicians and agencies are often thought to be driven by the divergence of preferences between the two actors (Epstein and O'Halloran 1999), or by other institutional features of the political environment, such as policy complexity (Huber and Shipan 2002), veto players (Keeper and Stasavage 2003), or accountability mechanisms and agency costs (Bertelli and Sinclair 2018). Yet

beliefs are important for explaining politico-administrative interactions, and can be independent of the ideological alignment between the politician and the agency.

With partisanship being an identity-shaping force, the actual policy preferences of agencies and the degree to which they are controlled by the executive are of secondary importance, for they are not affected by partisan directional motives, which by definition privileges membership over and not members' characteristics (such as ideology or politicisation). Even though ideological disagreement between agencies and politicians is likely to occur, partisan alignment with the executive triggers motivated reasoning, and the resulting bias is independent of the ideological position of each single agency. Partisan identity shapes the way politicians think about everything for which the government is responsible. As a result, the party-agency ideological distance plays a negligible role in shaping politicians' beliefs about bureaucratic agencies.

HYPOTHESIS 2: The ideological distance between politicians and agencies does not affect politicians' beliefs about bureaucratic agencies.

The reason why beliefs are important is because they affect behaviour. For instance, it has been shown theoretically that the desensitisation that stems from partisan bias and motivated reasoning – that is, motivated reasoners have a weaker reaction to information about performance – is detrimental to democratic accountability, for voters' beliefs are less likely to change politicians' incentives for effort (Little et al. 2020). If we translate this desensitisation mechanism to the relationship between politicians and bureaucracies, we might expect partisanship to hinder politicians' ability to objectively hold agencies to account: all else equal, being co-partisan with the government makes accountability weaker.

However, distinguishing between politicians' public and private beliefs complicates the picture. It might be possible that, when in power, it is in the interest of politicians to express satisfaction with their own party's achievements, including the performance of the agencies for which the governing party is responsible. Beliefs would therefore seem to be biased, while in fact politicians could privately have an objective view of the agency. Seemingly biased *public* beliefs would have no implications for the effectiveness of bureaucratic accountability, which would be informed by true and objective *private* beliefs. In observational terms, if politicians were not biased, their account-holding behaviour would not change when their party is no longer in power. Although it is hard to present hypotheses about politicians' private beliefs, I can present expectations about observable behavioural implications that would result from beliefs being truly biased. In particular, I argue that partisan bias weakens both public (spoken account-holding) and private (informal oversight) accountability practices.

Legislatures are a primary accountability forum where elected officials can fulfil their account-holding tasks through questions, interrogations, and criticism about the performance of bureaucracy (Miller 2005). When there is partisan alignment between a politician and the government, the former strongly identifies with the latter and evaluates the work of the bureaucracy – their group co-members – in a biased and more positive fashion. Since such more positive beliefs are not the product of anything factual, when majority party politicians debate about bureaucracy they are less likely to use facts and objective statistical evidence. Rather than accountability forums, legislative debates become an extension of partisan conflict. Partisanship weakening accountability is in line with predictions of blame-avoidance accounts of politicians' use of information (Nielsen and Baekgaard 2015) and with accountability deficits identified in the literature (Schillemans 2011; Schillemans

and Busuioc 2015). One key difference, however, the weakened accountability that results from partisan bias is not the result of a strategic choice or the inability of politicians, but it is rather an unintentional consequence of strong partisan identity. In observational terms, partisan motives affect politicians' argumentative style, which follows political rather than objective standards, making less frequent use of facts and statistical evidence. This partisanship-induced politicisation of what politicians say about agencies is detrimental to the ability of legislatures to effectively function as accountability forums.

HYPOTHESIS 3: Partisan alignment between politicians and the government is associated with political speeches about bureaucracy less grounded in facts and evidence.

Another accountability forum consists of political principals directly overseeing agencies, which – according to classical theories of bureaucratic accountability (McCubbins and Schwartz 1984) – should be more likely to happen when legislators and agencies are far apart on the ideological space. This proposition is the hallmark of bureaucratic accountability, for it entails that, faced with limited resources, elected legislators pay more attention to the activity of agencies that are ideologically far from their policy preferences, and hence deserve tighter control. Yet this relationship is not immune to partisan bias. Being co-partisan with the executive makes politicians less attentive to what agencies do and therefore inhibit their ability to hold bureaucracies to account by informally requesting information from agencies. Although recent work has found that, against common wisdom, legislators' individual oversight is mostly driven by policy valence rather than ideological disagreement (Lowande 2018), I claim that whatever role ideology plays in explaining oversight, it will be weaker for co-partisans with the government.

HYPOTHESIS 4: The positive relationship between agency-politician ideological distance and oversight is weaker for co-partisans with the government.

Moving from beliefs to public and private accountability forums, namely legislative debates and informal oversight, this theory links partisan motivated reasoning with theories of bureaucratic accountability. When there is partisan alignment between the government and politicians, beliefs about agencies are more positive. This has implications for bureaucratic accountability: argumentative style is less grounded in facts and evidence, and informal oversight is weaker.

3 Outcome Variables

3.1 Beliefs about Bureaucracy

One obvious way to measure politicians' beliefs about bureaucracy would be to ask them directly what they think about agencies. Yet this approach would result in a snapshot of what *current* politicians think about agencies, and it would not allow to compare beliefs over time. To address these limitations, I measure beliefs about bureaucratic agencies at the party level from what politicians belonging to the same party say in parliament. I ultimately produce party-agency-year estimates of beliefs about 323 agencies (191 for the US and 132 for the UK) over a time-frame of approximately 40 years. Legislative speeches can in fact aptly capture what politicians think about bureaucracy while allowing for a large time-coverage.

I measure partisan beliefs about bureaucratic agencies on a unidimensional scale, with larger values signifying more positive beliefs. The measurement strategy follows Bellodi

(2020), who uses word-embedding models to produce reputation estimates for bureaucratic bodies in the US and the UK. Because reputation is defined as a “set of symbolic beliefs” (Carpenter 2001, 3-4), this measurement strategy is particularly suitable for estimating partisan beliefs about bureaucracy. Here I use a similar approach, but estimate embeddings separately from local corpora of speeches given by the main parties in each year and country. In particular, I estimate party p ’s beliefs about agency a by pooling all speeches given by legislators of party p in year t . The output of this strategy is a set of estimates of party p ’s beliefs about all agencies that are mentioned in the speeches given by the same party’s legislators in any given year. While this measurement strategy does not allow me to estimate the beliefs of each single legislator, it leaves me with sufficiently large corpora to estimate word embeddings for each year and each party separately, capturing variation along three key dimensions: party, time, and agency.¹

The intuition behind word-embedding models is that we can learn about the relationship between words and discriminate between words related to one word but not another by looking at the ratio of co-occurrence probabilities. Suppose we want to learn the relationship between the words *FED* and *EPA* in a year when the EPA is highly criticised. To do so, we compare the probabilities of these two words co-occurring within a pre-defined segment of text with various context words k . We might expect word $k = independence$ to co-occur with the word *FED* more often than with the word *EPA*, and word $k = critic$ to co-occur more often with the word *EPA* than with the word *FED*. Similarly, we expect the word $k = dog$ to be very unlikely to co-occur with any of the words, and the word $k = policy$ with both. Table 1 represents these expectations in terms of hypothetical probabilities.

¹Estimating word embeddings for each legislator in each year would result in very small corpora and hence highly unstable embeddings. Different estimations would yield very different results. For instance, the median number of speeches per year given by US legislators is 60, and the median length is just 60 words.

Probability	$k = \text{independence}$	$k = \text{critic}$	$k = \text{dog}$	$k = \text{policy}$
$Pr(k FED)$.1	.01	.001	.15
$Pr(k EPA)$.01	.1	.001	.15
$Pr(k FED)/Pr(k EPA)$.1/.01 = 10	.01/.1 = 0.1	.001/.001 = 1	.15/.15 = 1

Table 1: Example of co-occurrence probabilities for target words FED and EPA with related and unrelated context words. Only in the ratio does noise from non-discriminative words like dog and policy cancel out, so that large values correlate well with words associated with FED, and small values correlate well with words associated with EPA.

The ratio of co-occurrence probabilities for words related to *FED* is large, whereas for words related to *EPA* is low. Words related to both or neither have a ratio that approximates 1, because they do not help discriminate between which word is related to which. Co-occurrence ratios are therefore very powerful at encoding relevant semantic relations and at estimating the semantic similarity between terms. These probabilities can be calculated for each unique word that appears in a corpus of texts. Every word can therefore be represented as a real-value vector that encodes the semantic characteristic of the word. Slightly more formally, word embeddings are the coefficients of statistical models (i.e., neural networks) that capture the relationship between the ratio of co-occurrence probabilities. Like the genetic information encoded in a strand of DNA, the elements of such vectors carry semantic information about the word. Words that co-occur with similar words will therefore have a similar “structure” and hence more similar vector representations. Distances between these vectors are informative about the semantic similarity of the words as used in the corpus from which they have been estimated (Spirling and Rodriguez 2019; Pennington et al. 2014). Therefore, by comparing the word embeddings of *EPA* and *FED* with a vector that combines several clearly positive terms such as good, excellent, great, and so on, we can learn about the similarity between the *EPA* vector and a positivity vector. Similarly, if we do this separately for Democrats and Republicans, we can see how different these agency

embeddings are from such positivity vectors.

More generally, let us consider a very large corpus of, say, parliamentary speeches given by Republican congresspersons in 2015. Let V be a vocabulary consisting of all the unique words that appear in the corpus. The absolute frequency with which words in V co-occur with each other can be reported in a matrix X of dimension $V \times V$, whose elements X_{ij} represent how many times word i co-occurs in the context of word j , with the context simply being a pre-defined window of words. Let X_j be the sum of the co-occurrences of any word i with the context word j (i.e., the sum of the j^{th} column), and $P(i|j) = X_{ij}/X_j$ be the probability that word i appears in the context of word j . Word vectors are then estimated with a neural network, namely a statistical model containing one layer of latent variables (the dimension of the word vectors) between the textual input (term co-occurrence matrix) and the output data (the word vectors). Technical details about the model parameters and estimation steps as well as information about the speeches and sample of agencies are reported in the Online Appendix.

Once I have word embeddings for every word in V , I can exploit the arithmetic properties of vector representations of words and build a vector that combines some unambiguously positive and negative words that will act as benchmark to measure the partisan belief about the agency. By deducting clearly negative embeddings from the sum of clearly positive embeddings, I obtain a word vector that uncontroversially captures positivity. The specific word vectors I used are:²

$$\begin{aligned} \vec{positivity} = & \vec{successful} + \vec{effective} + \vec{great} + \vec{excellent} \\ & - \vec{poor} - \vec{negative} - \vec{terrible} - \vec{bad} \end{aligned} \tag{1}$$

²The words have been chosen arbitrarily among clearly positive and negative words whose meaning is the same in both countries and did not change over time. This is similar to the seed words chosen by Rice and Zorn (2019) to set the benchmark for positivity and negativity dictionaries.

where the arrows signify the words are vectors. I finally measure the cosine similarity between the word embeddings of each agency and the *positivity* vector. The beliefs score will thus be the angular distance between the two embeddings. Formally,

$$\theta_j = \theta_{(\vec{a}, \vec{p})} = \frac{\vec{a} \times \vec{p}}{\|\vec{a}\| \times \|\vec{p}\|} = \frac{\sum_1^n a_i \times p_i}{\sqrt{\sum_1^n a_i^2} \times \sqrt{\sum_1^n p_i^2}} \quad (2)$$

where $\theta_{(\vec{a}, \vec{p})}$ is the cosine similarity between the agency vector \vec{a} and the positivity vector \vec{p} , namely the ratio between the sum of the products of the i^{th} elements of the two vectors (the nominator) and the product of the square root of the vectors to the power of two (the denominator). For instance, if the embedding $F\vec{E}D$ is semantically very similar to \vec{p} , beliefs about the FED will be more positive, whereas if the $E\vec{P}A$ embedding is semantically distant, beliefs will be more negative. The resulting metric is normalised to take up values between 0 and 1, where greater values signify more positive beliefs.

I apply this method to a corpus of all floor speeches given by the two main parties between 1982-2016 for the US and 1980-2018 for the UK. Detailed information about speeches and the sample of agencies is reported in the Online Appendix. First, I merge all the speeches at party level. Then I split them by year and obtain 86 local party-year corpora for the US and 78 for the UK. I then estimate word embeddings from each single local corpus. By doing so, I allow the semantic meaning of words to vary across country, party, and over time, for every estimation is performed on a different country-party-year corpora. The key advantage of this flexible estimation is that, for instance, the positivity vectors are allowed to vary based on the language used by each party in any given year. The final dataset consists of 8,824 party-agency-year observations, 6,464 for the US and 2,360 for the UK.

3.2 Weaker Accountability: Argumentative Style and Informal Oversight

I am interested in two types of accountability behaviour: politicians arguing about bureaucracies (public accountability forum), and politicians overseeing bureaucracies (private accountability forum).

I measure politicians' use of facts and statistical evidence when arguing about bureaucracy through a dictionary-based analysis of legislative speeches. Legislative speeches are assigned a score capturing the frequency with which words contained in a pre-defined list appear in the text. I use the 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.

Dictionary-based approaches are deemed to be highly context-dependent and therefore need careful validation (Grimmer and Stewart 2013). To this end, the “fact-dictionary” derived from the LIWC lists of words has been extensively and successfully validated by Hargrave and Blumenau (2020) in an almost identical setting as the one I study here: legislative speeches in the UK House of Commons. Furthermore, contrary to sentiment analysis tasks – where the meaning of words is likely to change across domains and over time (Rice and Zorn 2019) – words pertaining to facts and statistical evidence should be more representative of objective attributes and hence less dependent of the context in which are used. To support this claim, in the Online Appendix I report the results of an additional validation test which shows that the LIWC fact-dictionary performs well at matching a manually labelled corpus of texts from different contexts. I find a positive and significant correlation

between manual labels and the estimates of the dictionary method, thus strengthening our confidence in the low context-dependence of the dictionary.

I estimate the use of facts and evidence for more than half a million speeches (177,394 for the UK and 337,884 for the US) mentioning 268 agencies in the US and 308 in the UK, given by a total of 3,836 different legislators. Text pre-processing steps are reported in the Online Appendix. To ensure the analysis is performed over segments of text which are about the agencies, I limit the analysis to various symmetric windows of words centred around the names of the agencies, namely to segments of text that are 20, 50, and 100 words before and after any name of agencies. Speeches can be long and about several topics. By looking at word usage within small segments of text around the names of agencies, I increase the likelihood that what legislators are saying is in fact about bureaucracy.

I then compute for every term in each speech the term frequency-inverse document frequency (tf-idf), in order to reduce the importance of words that appear very often and in many speeches (Welbers et al. 2017). More precisely, tokens are assigned a weight which is equal to the logarithm of the inverse fraction of the speeches that contain the word. For instance, let us consider the following words which belong to the “fact-dictionary”: “approximately” and “less”. If “less” appears in more speeches than “approximately”, then “less” will receive a lower weight, for it is less helpful in discriminating between which word is more strongly representing the use of facts and evidence. For each speech, 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 speeches, and each speech a set of words, whose cardinality represents the number of unique words in the speech. For each speech mentioning a bureaucratic agency, the use of facts and evidence is given by the following

formula

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

where t is each token within the pre-defined windows of words for speech s , $Dict$ the list of words capturing the use of facts and statistical evidence, and $tf - idf$ is the term frequency-inverse document frequency of token t in speech s . The $Fact$ score is ultimately a function of the absolute frequency of the token t ($f_{t,s}$), the number of words in speech s ($|s|$), the number of speeches of the total corpus S , and the number of documents in the corpus that contain the token t ($|\{s \in S : t \in Dict\}|$).³

As far as bureaucratic oversight is concerned, I use data on informal oversight collected by Lowande (2018), who obtained records of correspondence between 16 bureaucratic agencies and members of Congress during the 110th and 111th Congress. The measure of oversight is dichotomous, and equals one if legislator i during Congress t sent a request or an inquiry to the agency. He classifies legislators' requests into two different categories, depending on the subject of the request. *Casework* requests are those sent on behalf of constituents with particular grievances. *Policy* request are concerns about the substance of policy or programme implementation. Informal oversight is particularly suitable for the purpose of the test I present below, for it is the most private and behind-the-scene form of oversight legislators could make, and should therefore be insulated from the strategic logic of the public sphere.

³Results are robust to using absolute frequency of fact-words appearing in the window of words. See Table A2 in the Online Appendix.

4 Methods

I am interested in four relationships, between partisan alignment and beliefs (Hypothesis 1), between ideology and beliefs (Hypothesis 2), and between partisan alignment and politicians' argumentative style (Hypothesis 3) and oversight activities (Hypothesis 4). Since the measure of beliefs is at party level (i.e., I estimate party p beliefs about agency a in year t), the unit of analysis for Hypotheses 1 and 2 is party-agency-year observations, whereas tests of Hypotheses 3 and 4 model individual speeches given by legislators in any given day, and legislators' oversight activities in any given Congress.

To investigate these relationships, I present four tests. First, I build panel data models and estimate the effect of partisan alignment on beliefs in both the US and the UK. Second, I show how this effect holds also when conditioning on the ideological distance between the party and the agency. Third, I estimate two-way fixed effects regressions with legislator fixed effects to compare the use of facts and evidence when arguing about bureaucracy between co-partisans with the executive and opposition members. Finally, I employ a generalised difference-in-differences estimator and estimate the moderating effect of partisan alignment on the relationship between ideological disagreement and oversight.

First, I estimate the following model for both the US and the UK.

$$Beliefs_{p[a,t]} = \delta_p + \phi_a + \alpha_t + \beta Alignment_{p[t]} + u_p \quad (4)$$

where $Beliefs_{p[a,t]}$ represents beliefs among party p , about agency a , and in year t . $Alignment_{p[t]}$ is a dummy variable indicating whether there is party-executive alignment in year t , ϕ_a are agency fixed effects to account for all time-invariant agency characteristics, δ_p

are party fixed effects, and α_t are year fixed effects to account for common shocks.

In the US, because of high levels of turnover in agency staff as a result of a new presidency, the effect of partisan alignment on beliefs might be confounded by a change in agency ideology. Republican presidents might appoint conservative bureaucrats and Democrats' beliefs about the agency might decrease for ideological reasons rather than for partisan misalignment. If this were the case, my main hypothesis for which politicians follow partisan motives when it comes to bureaucratic agencies would be rejected. To test this, I build a measure of the distance between the ideology of the agency and the average of politicians' ideology within each party and estimate the following model:

$$Beliefs_{p[a,t]} = \delta_p + \phi_a + \alpha_t + \beta Alignment_{p[t]} + \omega Distance_{p[a,t]} + \rho X_{a[t]} + u_p \quad (5)$$

The notation is the same as the one in Equation 4. $Distance_{p[a,t]}$ captures ideological distance between party p and agency a in year t , measured as the absolute value of the difference between the agency ideology score and the average DW-NOMINATE score of the members of Congress belonging to party p (Lewis, Jeffrey B. et al. 2020). I use Chen and Johnson (2014) donation-based estimates of agency ideology, which covers the largest sample of agencies over the longest time-frame (79 federal agencies across five presidencies, from the first Clinton Presidency to the first Obama Presidency).⁴ Contrary to the US context, there is no available data on the political leaning of UK agencies and as a result, this second test will be limited to the US. More details on the ideology estimates are reported in the Online Appendix.

X is a matrix of time-varying agency-level covariates, including agency politicisation,

⁴This 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 et al. 2019), and rule-making (Ellig and Conover 2014; Potter 2019).

budget, and number of employees, obtained from Lewis (2008). Agency politicisation is measured as the ratio of managers who are presidential appointees, whereas (authorised) budget and employees are measured in dollars and units. All these covariates are log transformed.⁵

These specifications have a series of advantages. First and most importantly, agency dummies account for all time-invariant agency characteristics such as the history, culture, mission, statutory features, and policy sector of the agency, which might well confound the relationship between partisan alignment and beliefs. Some agencies might in fact perform functions or be in charge of policies more likely to be associated with more positive beliefs. Second, common shocks or reforms affecting the bureaucracy as a whole which might impact agencies' reputation or performance are captured by year fixed effects. Third, because different party might have systematically different beliefs, party dummies account for different party characteristics, organisation, history, and time-constant views about the state and the bureaucracy. Fourth, by conditioning on budget size and number of employees, I can hold constant the economic relevance of the agency and its size/capacity, which can be important predictors of what politicians think about federal bureaucracy. Fifth, by including a variable capturing the degree of politicisation of the agency, I can sweep out the confounding effect of the degree to which the agency is "controlled" by the executive. In fact, it could be argued that beliefs are more biased for agencies that are particularly close to the President.

The third test moves from party-level to legislator-level analysis, and looks at the within-legislator change in argumentative style when arguing about bureaucracy as a result of being co-partisan with the executive. I model the measure of use of facts and evidence as

⁵Data available only from 1988 to 2005.

a function of partisan alignment using a two-way fixed effects estimator, in order to look at change in the use of facts within legislators and legislative debate (i.e., date) and therefore holding constant all unobserved sources of heterogeneity at the legislator and debate level. Since language can be correlated with specific agencies, I also include agency dummies to account for time-invariant agency-specific characteristics. Formally, I estimate the following model:

$$Facts_{i[l,a,d]} = \eta_l + \phi_a + \alpha_d + \beta Alignment_{l[d]} + \rho X + u_i \quad (6)$$

where $Facts_{i[l,a,d]}$ is the frequency of tokens considered facts and evidence in speech i given by legislator l about agency a in day d . η_l and ϕ_a are legislator and agency dummies, α_d date fixed effects, and β the effect of being a co-partisan with the executive.

Finally, I estimate the effect of partisan alignment on bureaucratic oversight activities with a generalised difference-in-differences model with a multiplicative interaction term between ideological distance and partisan alignment. I follow Lowande (2018) and exploit the transition from the second Bush presidency to the first Obama presidency as source of variation in the ideology of the agency and estimate the causal effect of ideological distance with data consisting of only the 110th and 111th Congress.

The appealing feature of this design is that it allows to isolate the variation in ideological distance resulting from a change in the presidency while committees' control and the internal organisation of the House and Senate remained fixed because Democrats maintained the majority in both chambers. In fact, the changing roles of legislators as a result of a new Congress might affect legislators' incentives to oversee agencies. I then build the following linear probability model:

$$Oversight_{l[a,c]} = \eta_l + \phi_a + \alpha_c + \beta_1 Dist_{l[a,t]} + \beta_2 Align_{l[t]} + \beta_3 Dist \times Align + \rho X + u_l \quad (7)$$

where $Oversight_{l[a,c]}$ is equal to 1 if legislator l in Congress c sent a request to agency a . The model includes agency, legislator, and congress dummies. I follow the original specification of Lowande (2018) and include both legislator- and agency-level time-varying covariates: the seniority of the legislator, whether the legislator is a member, chair, or ranking minority member of an oversight committee, agency budget, politicisation and logged number of employees. The continuous treatment $Dist$, which represents the ideological distance between legislator l and agency a , is the classical predictor of bureaucratic oversight. Here I am interested in estimating whether partisan alignment biases the oversight activities of legislators, which should be a function of ideological distance. The causal quantity of interest is therefore β_3 , which accounts for the moderating effect of partisan alignment on the relationship between ideological distance and oversight.

5 Results

By just looking at Figures 1 and 2, it is clear how partisanship matters for beliefs. The figures plot the political parties' beliefs averaged across all agencies for the Democratic and Republican parties for the US, and for the Conservative and Labour parties for the UK, together with the party in government (horizontal bar at the bottom of the plots). Horizontal dashed lines represent the average beliefs across each government.

On average, when there is a Democratic President, the beliefs of the Democratic party are more positive compared to when there is a Republican President, and *vice versa* for the Republican party's beliefs. The UK shows a similar trend. When the government changes color, partisan beliefs change too. The Conservative party's beliefs about bureaucracy are more positive during the Thatcher governments than during the Blair premiership, and they

increase again during the Cameron and May governments. The Labour party's beliefs too, despite being always more positive than the Conservative's, follow government cycles, more positive under Labour governments, more negative under Conservative governments.

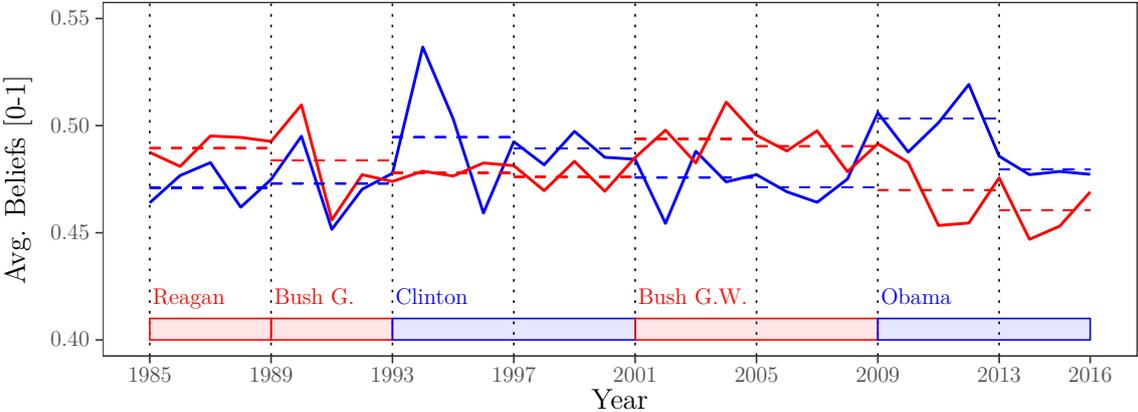


Figure 1: Parties' beliefs averaged across all agencies over time and across presidencies. Red for the Republican and blue for the Democratic party.

Even though I am interested in within-party change in beliefs as a result of being co-partisan with the executive, the figures also show a striking difference between the UK and the US in terms of level of partisan polarisation with respect to bureaucracy (McCarty et al. 2006). Democrats and Republicans hold more similar beliefs compared to Labour and Conservative MPs. In particular, Republicans' and Democrats' beliefs overtake each other as government changes, whereas the Labour and Conservative parties' beliefs remain far apart throughout the entire time period. Except for the golden era of the New Public Management reforms of the mid-80s, when bureaucracy was considered a neutral body of experts (Hood 1991), partisan beliefs remain highly polarised, particularly during the Blair and Brown governments.

Who is in charge of government matters for the process of belief formation. For both the US and the UK, partisan alignment has a significant effect on beliefs about bureaucratic bodies. Even when conditioning on agency time-invariant characteristics, common shocks,

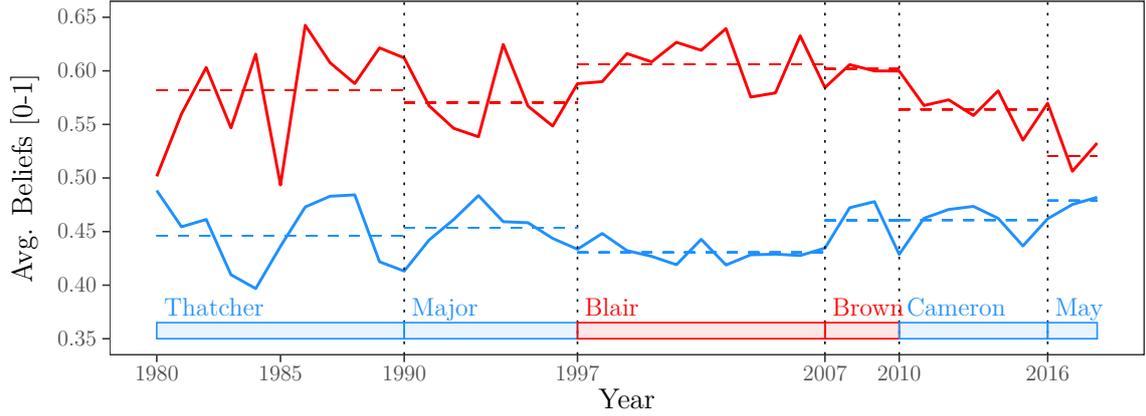


Figure 2: Parties' beliefs averaged across all agencies over time and across governments. Red for the Labour and light blue for the Conservative party.

and party fixed effects, the results are unambiguously positive and different from 0. In Table 2 I report the results of the first test. Sample size is large and estimates are highly precise. Because beliefs are measured from 0 to 1, the effect of partisan alignment can be interpreted as changes in percentage points. All else being equal, partisan alignment is associated with beliefs on average 1.7 and 3 percentage points more positive in the US and the UK, respectively. To ease the interpretation of the effect size, let σ_{US} and σ_{UK} be the standard deviations of beliefs in the US and the UK. The effect size of partisan alignment is $.12 \times \sigma_{US}$ in the US and $.18 \times \sigma_{UK}$ in the UK.

	DV: Beliefs [0,1]							
	US				UK			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Partisan Alignment	0.017*** (0.003)	0.017*** (0.003)	0.017*** (0.003)	0.017*** (0.003)	-0.013 (0.012)	0.030** (0.010)	0.030** (0.010)	0.030** (0.010)
Party FE		✓	✓	✓		✓	✓	✓
Agency FE			✓	✓			✓	✓
Year FE				✓				✓
R ²	0.004	0.004	0.250	0.259	0.002	0.151	0.341	0.357
Num. obs.	6,464	6,464	6,464	6,464	2,360	2,360	2,360	2,360

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 2: OLS estimates. Robust SE clustered by agency. Units are party-agency-year observations.

Interestingly, the effect of partisan alignment is larger in the UK than in the US. This is surprising given the lack of strategic appointments and partisan allegiance of the British civil service, where governments cannot refresh agencies to the same extent that US Presidents can do and bureaucrats cannot be asked about their political views.

As hypothesised, politicians manifest partisan bias when they build their beliefs about bureaucracy, more positive when their party is in government and therefore share the responsibility of the performance of government as a whole. To strengthen these results and to show that beliefs are not driven by substantive ideological differences between the party and the agency, I now turn to the second test and estimate the effect of partisan alignment on beliefs, conditioning on a set of time-varying agency characteristics, including the ideological distance between the agency and the party. Because of data availability, this test is carried out on the US, over a time-frame of almost 20 years, from 1993 to 2012, across five different presidencies. Table 3 shows the results. When conditioning on ideological distance, the effect of partisan alignment remains positive and much larger than the estimate over the entire sample in Table 2. Depending on the specification, partisan alignment makes beliefs more positive by 2.6-3.6 percentage points.

The first model has the same specification of Model (1) in Table 2, but now the effect size is larger (from 1.7 to 2.6 percentage points). This is just the result of sub-setting the data from 1993 to 2012, the only time-frame for which agency ideology data are available. Being co-partisan with the President increases beliefs about the bureaucracy by 2.6-3.6 percentage points. Despite the smaller sample size of specifications (3) and (6) (for time-varying agency covariates are only available from 1988 to 2005), the estimates remain positive and highly significant. Importantly, the effect of ideological distance is highly uncertain, with different direction across the different specifications. This lends support to Hypothesis 1

	DV: Beliefs [0,1]					
	(1)	(2)	(3)	(4)	(5)	(6)
Partisan Alignment	0.026*** (0.006)	0.026*** (0.006)	0.038** (0.011)	0.026*** (0.006)	0.027*** (0.006)	0.036** (0.010)
Ideological Distance		-0.000 (0.015)	0.013 (0.025)	-0.015 (0.019)	-0.011 (0.019)	0.014 (0.023)
Covariates			✓			✓
Party FE				✓	✓	✓
Agency FE				✓	✓	✓
Year FE					✓	✓
R ²	0.009	0.009	0.086	0.341	0.350	0.363
Num. obs.	1,672	1,672	474	1,672	1,672	474

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 3: OLS estimates. Robust SE clustered by agency. Units are party-agency-year observations. Covariates include agency politicisation, budget, and number of employees (all log transformed).

and Hypothesis 2. Partisan alignment with the government changes beliefs, and ideological differences have a negligible effect.

In the Online Appendix (Table A4), I address the concerns of units differentially contributing to the regression estimates by calculating the weights that are used in multiple regression and find no marked differences between the nominal and effective samples (Aronow and Samii 2016).

5.1 From Beliefs to Behaviour

What I have shown so far is that beliefs are biased. To show partisan bias affects accountability practices too, I now turn to the third and fourth tests, where the analysis is performed at the legislator rather than at party level. In support of Hypothesis 3, the two-way fixed effects results reported in Table 4 show that, compared to when they are at the opposition, when politicians argue about bureaucracy and are aligned with the government, they are

less likely to use analytical language grounded in facts and statistical evidence. The effect of partisan alignment is statistically significant across the various windows of words used and for both the UK and the US. In particular, focusing of segments of speeches 20 words before and after the name of the agency, the frequency of “fact-words” when there is partisan alignment decreases by 0.025 points for the US (average frequency among all speeches is .37) and by 0.047 points for the UK (average frequency among all speeches is .28). If we consider the average use of facts as a baseline, being co-partisan reduces the use of facts by approximately 7% for the US and 17% for the UK.

	US			UK		
	(1)	(2)	(3)	(1)	(2)	(3)
<i>Window Size</i>	20 tokens	50 tokens	100 tokens	20 tokens	50 tokens	100 tokens
Partisan Alignment	-0.025*** (0.007)	-0.023*** (0.006)	-0.019** (0.006)	-0.047*** (0.004)	-0.059*** (0.004)	-0.066*** (0.004)
Log. Days in House				-0.002 (0.003)	-0.007* (0.003)	-0.010*** (0.003)
Legislator Age	-0.002 (0.001)	-0.001 (0.001)	-0.002** (0.001)	-0.003 (0.002)	-0.003 (0.002)	-0.002 (0.002)
Ideological Distance	0.021 (0.017)	0.019 (0.016)	0.017 (0.014)			
Legislator FE	✓	✓	✓	✓	✓	✓
Agency FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
R ²	0.103	0.113	0.118	0.114	0.106	0.124
Num. obs.	95,987	95,987	95,987	170,391	170,391	170,391
Num. Legislators	1,066	1,066	1,066	2,075	2,075	2,075
Num. Agencies	53	53	53	287	287	287

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4: OLS estimates. SE clustered by legislator. Dependent variable is the tf-idf of fact-words in speeches. The fewer observations for the US compared to the UK depends on the limited number of agencies for which there is data on ideological leaning. Results are robust when omitting the *Ideological Distance* variable and therefore recovering the full dataset and the entire time-frame (see Table A3 in the Online Appendix).

Finally, the classical prediction for which ideologically divergent agencies are more

likely to be subject to legislators’ scrutiny is confirmed but only for politicians who are at the opposition. Partisan alignment biases the accountability practices of legislators, who are less likely to oversee ideologically distant agencies when they are aligned with the executive.

While ideological distance seems to have an insignificant effect when excluding partisan alignment, its effect is much clearer once partisan alignment enters the specification. Most importantly, there is a large slope change between co-partisan legislators and members of the opposition, even when holding constant all time-invariant factors at legislator, agency, and congress level. The results are robust across each category of oversight but the moderating effect of co-partisanship is particularly strong for policy-relevant request. A one-unit increase in ideological distance increases the probability of oversight by 11.3 percentage points for members of the opposition, whereas it decreases by 11.1 percentage points for co-partisans. Figure 3 plots the predicted probability against ideological distance for both co-partisans and opposition members. The slope change is large and highly significant. Regression results are reported in Table A5 in the Online Appendix.

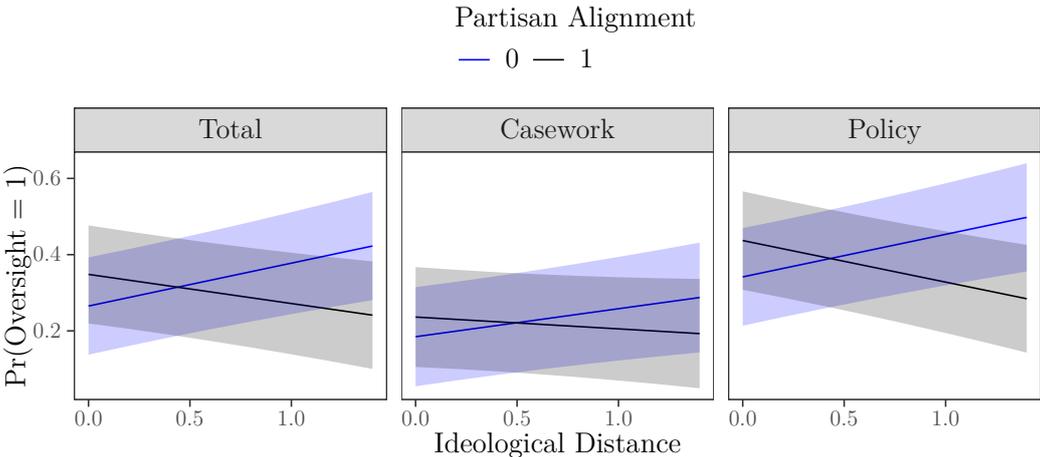


Figure 3: Predicted probability with 95% confidence intervals of legislator oversight based on ideological distance and partisan alignment with the executive. Labels indicate the type of legislators’ request to agencies.

The effect of partisan alignment on the relationship between ideological distance and

oversight is very different based on the majority or opposition status of legislators. In fact, for co-partisans with the government the probability of oversight *decreases* with ideological distance. These findings are surprising and deserve further research, but a *post-hoc* explanation could be that legislators who are co-partisans with the president avoid the uncomfortable oversight of agencies with very different ideological leaning, while they intensify the oversight of agencies with similar policy preferences to show they comply with their accountability duties. Overseeing agencies for which they are still overall responsible but whose policies are likely to clash with the government policy goals is an uncomfortable activity for politicians of the same party of the executive.

6 Discussion

Holding bureaucracies to account is a central topic in political science. While scholars have shown how politicians design institutions to hold agencies to account, no attention has been paid to the possible biases that systematically affect the way politicians form their beliefs about bureaucracy, and how such biases affect the practice of accountability. This paper represents the first attempt at solving this puzzle.

Politicians follow partisan motives when forming their beliefs about bureaucracy. When their party is in power, beliefs are more positive, when they are at the opposition, more negative. Bias quickly moves from beliefs to argumentative style and oversight activities. Since the reasons which make beliefs more positive belong to partisan identity, when politicians talk about bureaucracies and they are of the same team as the government, they use a more politicised style and resort to facts and evidence less frequently. Similarly, the commonly held proposition whereby ideological distance triggers oversight is much weaker when legis-

lators are co-partisan with the President. Doubts about the efficacy of legislators' oversight when aligned with the government might be legitimate.

This paper represents the first attempt at studying the effect of partisan motivated reasoning on belief formation and accountability behaviour in the administrative state. However, there can be many reasons why these findings could not generalise to other settings. Most importantly, the US and the UK are advance democracies with increasingly polarised political systems and majoritarian institutions. The relationship between elected politicians, partisanship, and bureaucracy in more proportional types of democracies or in countries with weaker administrative capacity might display different patterns from those observed in the cases studied here. Future research could look at such different contexts and at how bureaucracies – knowing politicians are motivated reasoners – make policies and interact with politicians.

References

- Peter M. Aronow and Cyrus Samii. Does Regression Produce Representative Estimates of Causal Effects? *American Journal of Political Science*, 60(1):250–267, 2016. ISSN 15405907. doi: 10.1111/ajps.12185.
- Martin Baekgaard, Julian Christensen, Casper Mondrup Dahlmann, Asbjørn Mathiasen, and Niels Bjørn Grund Petersen. The Role of Evidence in Politics: Motivated Reasoning and Persuasion among Politicians. *British Journal of Political Science*, 49(3):1117–1140, 2017. ISSN 14692112. doi: 10.1017/S0007123417000084.
- M Bartels. Beyond the Running Tally: Partisan Bias in Political Perceptions. *Political Behavior*, 24(2):117–150, 2002. ISSN 1572-0241. doi: 10.1038/ajg.2011.482.
- Luca Bellodi. A Dynamic Measure of Bureaucratic Reputation: New Data for New Theory. 2020.
- Anthony M. Bertelli and Christian R. Grose. The Lengthened Shadow of Another Institution? Ideal Point Estimates for the Executive Branch and Congress. *American Journal of Political Science*, 55(4):767–781, 2011. ISSN 00925853. doi: 10.1111/j.1540-5907.2011.00527.x.
- Anthony M. Bertelli and J. Andrew Sinclair. Democratic Accountability and the Politics of Mass Administrative Reorganization. *British Journal of Political Science*, 48(3):691–711, 2018. ISSN 14692112. doi: 10.1017/S0007123416000077.
- Alexander Bolton, John M. de Figueiredo, and David E. Lewis. Elections, Ideology, and Turnover in the U.S. Federal Government. 2019.
- Adam Bonica. Are donation-based measures of ideology valid predictors of individual-level policy preferences? *Journal of Politics*, 81(1):327–333, 2019. ISSN 14682508. doi: 10.1086/700722.
- E. Madalina Busuioc and Martin Lodge. Reputation and Accountability Relationships: Managing Accountability Expectations through Reputation. *Public Administration Review*, 77(1):91–100, 2017. ISSN 15406210. doi: 10.1111/puar.12612.
- Daniel P. Carpenter. *The Forging of Bureaucratic Autonomy*. Princeton University Press, Princeton, 2001.
- Daniel P. Carpenter and George A. Krause. Reputation and public administration. *Public Administration Review*, 72(1):26–33, 2012.
- Jorge Carrillo-de Albornoz, Javier Rodríguez Vidal, and Laura Plaza. Feature engineering for sentiment analysis in e-health forums. *PLoS ONE*, 13(11):1–25, 2018. ISSN 19326203. doi: 10.1371/journal.pone.0207996.
- Jowei Chen and Tim Johnson. 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, 2014. ISSN 14603667. doi: 10.1177/0951629813518126.

- Joshua D. Clinton and David E. Lewis. Expert opinion, agency characteristics, and agency preferences. *Political Analysis*, 16(1):3–20, 2008. ISSN 10471987. doi: 10.1093/pan/mpm009.
- Justin De Benedictis-Kessner and Christopher Warshaw. Accountability for the Local Economy at All Levels of Government in United States Elections. *American Political Science Review*, pages 1–17, 2020. ISSN 15375943. doi: 10.1017/S0003055420000027.
- James N. Druckman, Erik Peterson, and Rune Slothuus. How elite partisan polarization affects public opinion formation. *American Political Science Review*, 107(1):57–79, 2013. ISSN 15375943. doi: 10.1017/S0003055412000500.
- Jerry Ellig and Christopher J. Conover. Presidential priorities, congressional control, and the quality of regulatory analysis: an application to healthcare and homeland security. *Public Choice*, 161(3-4):305–320, 2014. ISSN 00485829. doi: 10.1007/s11127-014-0201-3.
- David Epstein and Sharyn O’Halloran. *Delegating Powers: A Transaction Cost Politics Approach to Policy Making under Separate Powers*. Cambridge University Press, Cambridge, UK, 1999.
- Robert S Erikson. Economic Conditions and The Presidential Vote. *American Political Science Review*, 83(2):567–573, 1989.
- Donald Green, Bradley Palmquist, and Eric Schickler. *Partisan Hearts and Minds*. Yale University Press, New Haven, CT, 2002.
- Justin Grimmer and Brandon M. Stewart. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3):267–297, 2013. ISSN 14764989. doi: 10.1093/pan/mps028.
- Lotte Hargrave and Jack Blumenau. The Declining Importance of Gender Stereotypes for Politicians’ Style in the UK. 2020.
- Christopher Hood. A Public Management For All Seasons? *Public Administration*, 69:3–19, 1991. ISSN 00093696. doi: 10.2307/25305228.
- John D. Huber and Charles R. Shipan. *Deliberate Discretion? The institutional foundations of bureaucratic autonomy*. Cambridge University Press, New York, 2002.
- Shanto Iyengar and Sean J. Westwood. Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, 59(3):690–707, 2015. ISSN 15405907. doi: 10.1111/ajps.12152.
- Shanto Iyengar, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22:129–146, 2019. ISSN 10942939. doi: 10.1146/annurev-polisci-051117-073034.
- Oliver James. Performance measures and democracy: Information effects on citizens in field and laboratory experiments. *Journal of Public Administration Research and Theory*, 21(3):399–418, 2011. ISSN 10531858. doi: 10.1093/jopart/muq057.

- Oliver James and Peter John. Public management at the ballot box: Performance information and electoral support for incumbent english local governments. *Journal of Public Administration Research and Theory*, 17(4):567–580, 2007. ISSN 10531858. doi: 10.1093/jopart/mul020.
- Philip Keeper and David Stasavage. The Limits of Delegation: Veto Players, Central Bank Independence, and the Credibility of Monetary Policy. *American Political Science Review*, 97(3):407–423, 2003. ISSN 00030554. doi: 10.1017/S0003055403000777.
- David E. Lewis. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance*. Princeton University Press, 2008.
- Lewis, Jeffrey B., K. Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet. Voteview: Congressional Roll-Call Votes Database. 2020. URL voteview.com.
- Andrew Little, Keith Schnakenberg, and Ian Turner. Motivated reasoning and democratic accountability. 2020.
- Kenneth Lowande. Who Polices the Administrative State? *American Political Science Review*, 112(4):874–890, 2018. ISSN 15375943. doi: 10.1017/S0003055418000497.
- Arthur Lupia and Mathew D. McCubbins. Designing bureaucratic accountability. *Law and Contemporary Problems*, 57(1):91–126, 1994. ISSN 0023-9186. doi: 10.2307/1191988.
- Moshe Maor and Raanan Sulitzeanu-Kenan. Responsive Change: Agency Output Response to Reputational Threats. *Journal of Public Administration Research and Theory*, 26(1): 31–44, 2016. ISSN 14779803. doi: 10.1093/jopart/muv001.
- Lilliana Mason. "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, 59(1):128–145, 2015. ISSN 15405907. doi: 10.1111/ajps.12089.
- Nolan McCarty, Keith T. Poole, and Howard Rosenthal. *Polarized America: The Dance of Ideology and Unequal Riches*. The MIT Press, Cambridge, MA, 2006.
- M McCubbins and T Schwartz. Congressional Oversight Overlooked: Police Patrols versus Fire Alarms. *American Journal of Political Science*, 28(1):165–179, 1984.
- M McCubbins, Roger G Noll, and Barry R Weingast. Administrative Procedures as Instruments of Political Control. *Journal of Law, Economics, & Organization*, 3(2):243–277, 1987.
- Gary J. Miller. the Political Evolution of Principal-Agent Models. *Annual Review of Political Science*, 8(1):203–225, 2005. ISSN 1094-2939. doi: 10.1146/annurev.polisci.8.082103.104840.
- Emily H. Moore. Polarization, Excepted Appointments, and the Administrative Presidency. *Presidential Studies Quarterly*, 48(1):72–92, 2018. ISSN 17415705. doi: 10.1111/psq.12417.

- Poul A. Nielsen and Martin Baekgaard. Performance information, blame avoidance, and politicians' attitudes to spending and reform: Evidence from an experiment. *Journal of Public Administration Research and Theory*, 25(2):545–569, 2015. ISSN 14779803. doi: 10.1093/jopart/mut051.
- John W. Patty and Ian R. Turner. Ex Post Review and Expert Policymaking: When Does Oversight Reduce Accountability? *Journal of Politics*, 2020. doi: <https://doi.org/10.1086/708913>.
- James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. *The Development and Psychometric Properties of LIWC2015*. University of Texas at Austin, Austin, 2015. doi: 10.1068/d010163.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. GloVe: Global Vectors for Word Representation. *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.
- Keith T Poole and Howard Rosenthal. A Spatial Model for Legislative Roll Call Analysis. *American Journal of Political Science*, 29(2):357–384, 1985.
- Rachel A. Potter. *Bending the Rules. Pocedural Politicking in the Bureaucracy*. University of Chicago Press, Chicago, 2019.
- Douglas R. Rice and Christopher Zorn. Corpus-based dictionaries for sentiment analysis of specialized vocabularies. *Political Science Research and Methods*, pages 1–17, 2019. doi: 10.1017/psrm.2019.10.
- Mark D. Richardson, Joshua D. Clinton, and David E. Lewis. Elite perceptions of agency ideology and workforce skill. *Journal of Politics*, 80(1):303–308, 2018. ISSN 14682508. doi: 10.1086/694846.
- Thomas Schillemans. Does horizontal accountability work? Evaluating potential remedies for the accountability deficit of agencies. *Administration and Society*, 43(4):387–416, 2011. ISSN 00953997. doi: 10.1177/0095399711412931.
- Thomas Schillemans and E. Madalina Busuioc. Predicting public sector accountability: From agency drift to forum drift. *Journal of Public Administration Research and Theory*, 25(1):191–215, 2015. ISSN 14779803. doi: 10.1093/jopart/muu024.
- Arthur Spirling and Pedro L Rodriguez. Word Embeddings What works, what doesn't, and how to tell the difference for applied research. 2019.
- Charles S Taber and Milton Lodge. Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, 50(3):755–769, 2006. ISSN 0092-5853.
- Henri Tajfel. The achievement of inter-group differentiation. In Henri Tajfel, editor, *Differentiation between social groups*, pages 77–100. Academic Press, London, 1987.
- Ian R. Turner. Working smart and hard? Agency effort, judicial review, and policy precision. *Journal of Theoretical Politics*, 29(1):69–96, 2017. ISSN 14603667. doi: 10.1177/0951629816630430.

Kasper Welbers, Wouter Van Atteveldt, and Kenneth Benoit. Text Analysis in R. *Communication Methods and Measures*, 11(4):245–265, 2017. ISSN 19312466. doi: 10.1080/19312458.2017.1387238.

Online Appendix

A1 Measuring Partisan Beliefs About Bureaucracy

The core assumption of the measurement strategy is that we can capture politicians beliefs by studying what politicians say in legislatures.

The corpus from which I estimate the word embeddings are all the legislative speeches from 1980 to the most recent available data in the two chambers of the US Congress and the UK House of Commons. I downloaded the US congressional debates from the Congressional Repository of Stanford University (data.stanford.edu/congress_text), while the UK parliamentary speeches were accessed through the *TheyWorkForYou* API, for a total of more than 4,6 million speeches. US speeches for the years 1983-84 are not available because the Stanford University Congressional Repository does not have speeches for those years. The sample of agencies consists of the 432 unique entries in Bellodi’s (2020) dataset.

I employ the unsupervised learning algorithm GloVe (Pennington et al. 2014), a count-based model that produces vector representations of words by doing dimensionality reduction on a co-occurrence matrix. The first step is to create a term co-occurrence matrix X of dimension $V \times V$, where V is a vocabulary consisting of all the unique tokens that appear in the corpus, where each corpus consists of all speeches of legislators belonging to each party in any given year. Each element X_{ij} is a number representing how many times word i co-occurs in the context of word j , with the context simply being a pre-defined window of words whose size depends on the particular task at hand. Let X_j be the sum of the co-occurrences of any word i with the context word j (i.e., the sum of the j^{th} column), and $P(i|j) = X_{ij}/X_j$ be the probability that word i appears in the context of word j .

Word vectors are then estimated with a neural network, namely a statistical model containing one layer of latent variables (the dimensions of the word vectors) between the textual input (term co-occurrence matrix) and the output data (the word vectors). To avoid the model from weighting all the co-occurrences equally, word vectors are estimated for every word in V by training a log-bilinear model with a weighted least-squares objective that tries to predict the context word j in which word i is used. Very summarily, the model minimises the following equation J ,

$$J = \sum_{i=1}^V \sum_{j=1}^V f(X_{ij})(w_i^T w_j + b_i + b_j - \log(X_{ij}))^2 \quad (8)$$

where $V = \{v_1, v_2, \dots, v_V\}$ is the vocabulary, w_i is the vector of the target word, w_j is the vector of the context word, and b_i and b_j are scalar bias terms. $f(X_{ij})$ is a function that determines the weight to each pair of words based on how often they co-occur; pairs of words that co-occur more often will have greater weight. The final output is a word embedding for every word in the vocabulary.

I train the GloVe algorithm on a local corpus of parliamentary speeches for every year, every country, and for the two main parties in the US and the UK (Republican, Democratic, Labour, Conservative). I follow standard practice in text-analysis and I lemmatise the

tokens, remove punctuations, digits, capitalisation, and stop-words to increase the precision of the estimation. I also remove two-character words for they are deemed to carry no semantic information. Agencies referred to in more than one way (e.g., CIA and Central Intelligence Agency) were replaced in the text with standardised tokens. I then create a vocabulary with all the tokens appearing at least five times in all the corpus, because words appearing very few times do not convey semantic information. I create a term co-occurrence matrix specifying a window size of 12 tokens and estimate 300-dimensional word vectors with a weighting function $X_{max} = 10$. This means that any pair of words for which the co-occurrence count is greater than 10 will receive a weight of 1, whilst the other weights $w_i \in [0, 1)$. I estimate the model through 100 iterations, with a convergence threshold of 0.001, and a learning rate appropriate to the size of the corpus, equal to 0.1. I use these parameters because they are deemed to be the most appropriate for semantic tasks (Spirling and Rodriguez 2019).

A2 Text Pre-Processing

I implement the dictionary-based measurement through the following steps: the speeches are the same used to estimate beliefs (see Section A1 for more details on the corpus of speeches).⁶ To compare speeches about bureaucracy, I keep only the speeches which mention at least one agency. I removed punctuation and converted all the tokens to lower case. Agencies referred to in more than one way (e.g., CIA and Central Intelligence Agency) were replaced in the text with standardised token.

A3 Agency Ideology

Various methods have been proposed to measure agency ideology, either relying on the political party establishing the agency or appointing the agency head (Epstein and O’Halloran 1999), or using bureaucrats’ characteristics aggregated at agency level. These methods produce agency ideal points modelling expert surveys (Clinton and Lewis 2008), bureaucrats’ voting preferences (Bertelli and Grose 2011), bureaucrats’ campaign contributions (Chen and Johnson 2014), or perceptions of federal executives (Richardson et al. 2018). Given the theory I present in this paper, I use the dataset that covers the largest sample of agencies over the longest time-frame. I therefore 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). This 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 et al. 2019), and rule-making (Ellig and Conover 2014; Potter 2019).

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 developed by Poole and Rosenthal (1985). The key assumption – recently validated by Bonica (2019) – is that campaign donations capture and are strong predictors of policy preferences. The resulting estimates are weighted averages the Common Space scores of legislators who receive donations from bureaucrats, with weights accounting for the amount of the donation, so that larger contributions, which

⁶Since I do not need a minimum number of speeches to measure politicians’ use of facts and evidence, I keep all the speeches given by every political party.

are more likely to come from better-paid, upper-level bureaucrats, receive a larger weight. 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.

Figure 4 shows the distribution of agency ideology across presidencies. Variation in ideological composition is the result of two factors: voluntary departure of senior employees with divergent views to that of the new presidency, and political appointments (Bolton et al. 2019).

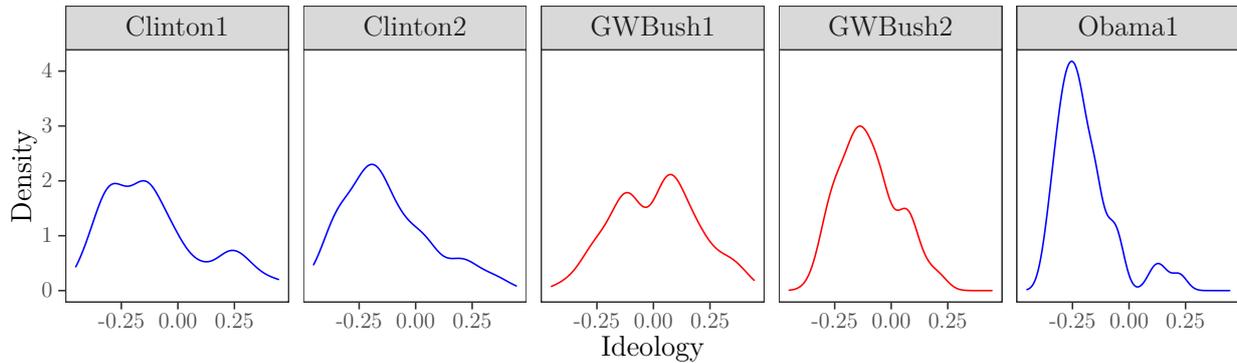


Figure 4: Density plots of agency ideology across presidencies. On the x axis, DW-NOMINATE common space scores.

Contrary to the US context, there is no available data on the political leaning of UK agencies. This is possibly the result of the norms of administrative neutrality and impartiality that are rooted in the UK civil service (Hood 1991). Both practice and research generally assume that the UK administrative system is immune to bureaucrats’ political preferences. Moreover, stringent regulations prevent bureaucrats from identifying with a political party and therefore make the use of proxies very hard. The Civil Service Management Conduct, for instance, prescribes that top-level bureaucrats “must retain at all times a proper reticence in matters of political controversy so that their impartiality is beyond question”.⁷ These are the reasons why researchers have little material to input in models of agency ideology. As a result, all the tests which I will present including ideological positions of the agencies will be limited to the US.

A4 Dictionary Approach: Validation

Dictionary-based approaches to analyse text are deemed to be highly context-dependent and therefore need careful validation (Grimmer and Stewart 2013). Words’ semantics can in fact change from one context to another. This issue is particularly concerning for sentiment analysis tasks, for the valence of words is likely to change over time and across domains.

The “fact-dictionary” derived from the LIWC lists of words I use to measure legislators’ use of facts and evidence when arguing about bureaucracy has been extensively validated by Hargrave and Blumenau (2020) in an almost identical setting as the one I study here: legislative speeches in the UK House of Commons. Moreover, context-dependence seems less problematic for facts and evidence-related dictionaries, whose words are more representative of quantities and objective attributes and less reflective of emotions. To back this claim with

⁷Document accessible at gov.uk/government/publications/civil-servants-terms-and-conditions.

data, I compare the estimates of the dictionary approach with a manually labelled corpus of text from a very different context: posts and comments of medical online forums on breast cancer, crohn, and various allergies.

The corpus is assembled by Carrillo-de Albornoz et al. (2018), who train a classification model to estimate patients’ opinion about health services. Coders classified each sentence of each post as communicating “experience”, “fact”, or “opinion”. The benchmark I use to assess the validity of the dictionary is thus the number of sentences classified as “fact” in each post. I then apply the dictionary-approach to the corpus of posts ($N = 480$) and I model the relationship between the dictionary and manual estimates. Table A1 below reports regression estimates of OLS and various count models where the number of fact sentences is regressed on the *fact* estimates consisting of the sum of the tf-idf of each term in the fact dictionary that appears in the post, as per Equation 3. The coefficients suggest that the *Dictionary Measure* is a strong predictor of the number of fact sentences in forum posts. This strengthens our confidence of the validity of the dictionary for capturing the use of facts and evidence in texts and its weak dependence of context.

	OLS	Poisson	Zero-Infl.	Neg.-Bin.	Logit
Dictionary Measure	2.76** (0.96)	0.75*** (0.14)	0.65*** (0.14)	0.97*** (0.15)	0.72** (0.23)
R ²	0.16				
Num. obs.	480	480	480	480	480
AIC		1916.46	1799.65	1567.39	640.43
Log Likelihood		-956.23	-895.83	-780.70	-318.22

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table A1: Regression estimates. Robust SE in parentheses with *sandwich* variance estimator. Logistic regression with dichotomised outcome, 1 if number of fact sentences in post > 1 , 0 otherwise.

	US			UK		
	(1)	(2)	(3)	(1)	(2)	(3)
<i>Window Size</i>	20 tokens	50 tokens	100 tokens	20 tokens	50 tokens	100 tokens
Partisan Alignment	-0.099*** (0.029)	-0.191*** (0.053)	-0.260** (0.081)	-0.113*** (0.011)	-0.326*** (0.023)	-0.584*** (0.039)
Log. Days in House				-0.005 (0.008)	-0.039* (0.017)	-0.090*** (0.027)
Legislator Age	-0.009 (0.005)	-0.012 (0.007)	-0.028** (0.011)	-0.006 (0.005)	-0.018 (0.010)	-0.022 (0.014)
Ideological Distance	0.082 (0.068)	0.157 (0.131)	0.234 (0.188)			
Legislator FE	✓	✓	✓	✓	✓	✓
Agency FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
R ²	0.103	0.113	0.118	0.114	0.106	0.124
Num. obs.	95,987	95,987	95,987	170,391	170,391	170,391
Num. Legislators	1,066	1,066	1,066	2,075	2,075	2,075
Num. Agencies	53	53	53	287	287	287

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table A2: OLS estimates. SE clustered by legislator. Dependent variable is the absolute frequency of facts-words in speeches.

A5 Robustness Checks

<i>Window Size</i>	(1)	(2)	(3)
	20 tokens	50 tokens	100 tokens
Partisan Alignment	-0.019*** (0.005)	-0.014** (0.004)	-0.011** (0.004)
Legislator Ideology	-0.052 (0.054)	-0.049 (0.044)	-0.058 (0.042)
Legislator Age	0.000 (0.002)	0.001 (0.001)	0.000 (0.001)
Legislator FE	✓	✓	✓
Agency FE	✓	✓	✓
Date FE	✓	✓	✓
R ²	0.084	0.089	0.095
Num. obs.	256,564	256,564	256,564
Num. Legislators	1,590	1,590	1,590
Num. Agencies	267	267	267

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table A3: OLS estimates. Dependent variable is the tf-idf of facts-words in US speeches omitting the ideological distance between agency and legislator and hence recovering the full sample of speeches. SE clustered by legislator.

Variable	Nominal Sample		Effective Sample	
	Mean	Std. D.	Mean	Std. D.
Beliefs	0.49	0.14	0.50	0.22
Ideological Distance	0.36	0.24	0.40	0.23
Agency Ideology	-0.10	0.22	-0.10	0.20
Agency Types				
Independent Agency	86.0 %		86.0 %	
Executive Department	10.2 %		12.7 %	
Exec. Sub-Agencies	0.0 %		0.0 %	
Office President	3.8 %		1.4 %	

Table A4: Nominal and Effective Samples. Values for the effective sample are weighted averages and weighted standard deviations of each variable. Weights computed following Aronow and Samii (2016) as the squared residual of a regression model of the outcome on all the covariates and fixed effects but with no treatment. I used the same specification as the one reported in Equation 5.

A6 Test 4: Regression Table

	No Interaction			Multiplicative Interaction		
	Total	Casework	Policy	Total	Casework	Policy
Ideological Distance	0.015 (0.020)	0.014 (0.019)	-0.001 (0.018)	0.114*** (0.027)	0.074** (0.025)	0.113*** (0.025)
Partisan Alignment				0.084*** (0.014)	0.052*** (0.014)	0.096*** (0.015)
Part. Alignment \times Id. Distance				-0.191*** (0.033)	-0.105*** (0.031)	-0.224*** (0.035)
Congress FE	✓	✓	✓	✓	✓	✓
Legislator FE	✓	✓	✓	✓	✓	✓
Covariates	✓	✓	✓	✓	✓	✓
Num. obs.	17,552	16,455	16,455	17,552	16,455	16,455
R ²	0.503	0.500	0.430	0.504	0.500	0.431
Adj. R ²	0.484	0.480	0.407	0.484	0.480	0.408

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table A5: Linear probability models. SE clustered by legislator in parenthesis. Binary variable, 1 for informal legislator oversight of agency (for each type of request: casework, policy, and both). Covariates include committee membership, committee chair, ranking, budget in millions of dollars, agency politicisation, log number of staff.