Legislative Bellwethers: The Role of Committee Membership in Parliamentary Debate

Jorge M. Fernandes[*]   Max Goplerud[†]   Miguel Won[‡]

October 2, 2018

Abstract

Political parties and legislators use legislative debates to establish their reputation, challenge rivals, engage in coalition management, among many other tasks. Yet, existing theories on parliamentary debates have abstracted away from the need for information and expertise, which are costly acquire. Drawing on the "informational" perspective on legislative organization, we address this problem by arguing that party leaders use committees as training arenas for their backbenchers. They task their assigned members with acquiring specific expertise and then rely heavily on those members during the corresponding debates. We turn to the Portuguese legislature, from 2000 to 2015, to discuss how saliency, government dynamics, and party size affect the use of experts. We test this theory using a novel approach to classify speeches that leverages the texts of legislation as training data for a supervised approach.

---

[*]Institute of Social Sciences, University of Lisbon, Av. Prof. Anibal Bettencourt, 9, Lisbon, Portugal

[†]Department of Government, Harvard University, 1737 Cambridge Street, Cambridge, MA, USA

[‡]INESC-ID, Rua Alves Redol, 9, Lisbon, Portugal

1   Introduction

> "[A legislative bellwether] is, in a word,
> one whom his colleagues regard as
> relatively more 'expert' than they, and
> yet sufficiently akin to be trustworthy."
>
> —————————————————
> Dahl (1950, p.60)

Political actors need specialized information and skills to operate effectively in legislatures. These resources, however, are costly to acquire and disseminate. In some contexts, legislators are flooded with information and face difficulties to extract what they need. In others, exogenous actors – for example, interest groups or the bureaucracy – withhold information on purpose, creating difficulties for legislators to map their intents into consequences (Krehbiel, 1991). Parties have an essential role in creating structures to help digest information insofar as they promote hierarchical structures and functional specialization (Saalfeld and Strøm, 2014).

Speechmaking is one of the most prominent activities in legislatures. Existing scholarship on legislative debates focuses on the impact of electoral systems incentives to uphold a cohesive party brand (Proksch and Slapin, 2012), on the ideological distance between legislators and the party leadership (Maltzman and Sigelman, 1996; Diermeier et al., 2012), and how speaker selection has an impact in vote-seeking politics (Herzog and Benoit, 2015). Yet, the analysis of the mechanisms through which parties and legislators obtain information for speechmaking is mainly absent from the literature. Indeed, most speeches are about complex and technical legislation, interacting with the party's brand and objectives. While some speeches can be delivered without much knowledge of the issue, effective speechmaking requires highly informed and trained political actors.

The contribution of this article consists in making a rigorous empirical test of the relationship between committee systems and speechmaking activities. Indeed, we explore the *consequences* of committee assignments, which have been largely overlooked in favor of committee assignment politics. We argue that committee systems serve as a training ground for speechmaking activities.[1]

—————————————————
[1] See previous work by Quinn et al. (2010) and Giannetti and Pedrazzani (2016) on this topic.

Political parties use their prerogatives as gatekeepers of *both* committee assignments *and* floor time to articulate an optimal strategy whose goal is to give the floor to legislators who have gained expertise on a specific topic from the corresponding committee arena.

To account for the pivotal role of political parties in parliamentary democracies (Müller, 2000), we depart from Krehbiel's view of members' expertise being used as a collective good for the *chamber* as a whole. Instead, we suggest that committees are arenas for members to acquire useful information for their *parties*. In return, party leaders rely on these experts as their preferred actors to take the floor and speak on behalf of the party on those topics. By and large, this serves to establish the party's reputation and policy position on a particular jurisdiction and signal the party's positions to actors in the political environment, for example, fellow parties, the electorate, and interest groups. Furthermore, our theoretical argument suggests that the intersection between committee system specialization and floor specialization should be more pronounced the more salient the issue is for the party.

In doing this, we also contribute to the growing literature on text-as-data in political science. Existing research has mostly relied on unsupervised methods—for example, topic models (Blei, Ng and Jordan, 2003; Grimmer and Stewart, 2013)—for uncovering topics in legislative speech. Our contribution introduces a different approach relying on supervised learning methods. We use *bills* to train our supervised model, and then use the estimated model to predict the labels attached to *speeches*. The use of bills in conjunction with parliamentary debates opens up a new frontier to apply text methods to critical questions in parliamentary democracies as it makes supervised learning methods feasible as bills provides a plausible set of training data. Our innovation is particularly useful given that whenever parliamentary debates exist, there are also bills. Many sources of legislative data both historically and comparatively are likely to contain the necessary information to use this supervised procedure.

We test our hypotheses in a causally credible way by using a fixed-effects approach. We include controls for backbencher-committee pairs and thus leverage changes in speaking behavior by comparing backbenchers' behavior *before* and *after* being assigned to a particular committee

jurisdiction. Thus, unobserved time-invariant omitted variables do not undermine our results. We turn to the Portuguese case, over the period of four legislatures (1999-2015). Portugal offers an ideal institutional environment to test our claims because it allows holding constant existing rival explanations. Indeed, Portugal's closed-list proportional electoral system, with associated high levels of party loyalty and cohesion, washes out potential competing explanations from electoral rules and intra-party dissent (Proksch and Slapin, 2012). Furthermore, in Portugal, rules of procedure give parties a monopoly on determining committee assignments and make them gatekeepers of the floor (Fernandes, 2016). Such institutional setting permits us to explore how party leaders make decisions to use their committee delegates to speak on their behalf in the plenary.

2   Legislative Speeches in Parliamentary Democracies

Speeches are one of legislators' most important tools in modern democracies. They serve a variety of functions that, in return, are intimately linked with legislators' goals. First, they help attract media attention, claim credit, and advance their reelection prospects (Mayhew, 1974; Eggers and Spirling, 2014). Second, speeches serve to address information asymmetries and provide cues to other actors (Austen-Smith and Riker, 1987; Austen-Smith, 1990). Third, speeches are signaling mechanisms in coalition governments by partners that want to keep a separate and attractive electoral profile (Martin and Vanberg, 2008). Speeches offer an inexpensive "venting" mechanism to signal dissent without facing the dire consequences of unilaterally changing or reneging on a coalition agreement. Finally, speeches can are mechanisms of substantive representation for minorities, such as women (Bäck and Debus, 2016).

Recently, Proksch and Slapin (2014) presented a framework to analyze speechmaking in legislature, which puts the *party brand* at the core of their argument, an assumption which we share in this work. Political parties need to shore up their brand to be electorally attractive. Proksch and Slapin (2014) model speechmaking as a delegation game in which party leaders (the principal) choose a backbencher (an agent) to deliver the speech on behalf of the party. In an ideal-world,

4

leaders would deliver all speeches to make sure that all speechmaking benefits accrue to collective party brand. Party leaders need to curb the possibility of a legislator going "off message". Yet, the vagaries of real-world politics pose challenges to political parties that make delegation the most efficient solution.

Electoral rules affect the extent to which political parties must be wary of delegation imperiling their brand. The more candidate-centered the electoral system, the less critical the party brand. Consequently, in this context, party leaders can afford to be more lenient with their backbenchers. Party leaders design open-rules of procedure to create incentives for backbenchers to use the floor to cater to their constituencies (Cain, Ferejohn and Fiorina, 1987). Ireland and the United Kingdom are examples of this setting. In party-centered electoral contexts, party leaders need to keep tabs on their backbenchers. Indeed, if a backbencher shirks by delivering a message that does not conform to the party official line, that would hurt the party brand, a collective good that benefits all members running under the same label. According to Proksch and Slapin (2014), party leaders design closed-rules that help control who accesses the floor. Portugal, Spain, Israel, and Norway offer real-world examples of this setting. Ultimately, an equilibrium exists between electoral systems and rules of procedure when party leaders exert an "optimal level of control over their membership, given the political system in which they operate" (Proksch and Slapin, 2014, 78).

The situation facing party leaders when allocating floor time, however, is even more fraught than what Proksch and Slapin (2014) suggest. Beyond merely being constrained in terms of *time*, party leaders face *information* constraints. First, issues under consideration in parliamentary debates are multidimensional and vast in scope – spanning diverse topics, such as foreign affairs, economic policy, environmental regulations, and many others. Second, the fact that many different initiatives are being considered simultaneously means that no single legislator, or group of legislators, can realistically expect to become sufficiently informed on all issues to expect to defend the party on the floor effectively. Third, there are actors operating outside the legislature – for example, interest groups – that have exogenous information that legislators need to acquire for

their speechmaking activities.

As the party leadership does not possess perfect information on all issues, delegation to experts in particular policy areas in the corresponding debates is an efficient institutional solution. The leadership, however, cannot just delegate to *any* backbencher. The leadership cannot risk delegating floor access to a backbencher who does not know enough about the topic and whose performance might jeopardize the party brand in the legislature and beyond. Furthermore, the iterative nature of debates, in which there is often a back-and-forth discussion with other parties, requires highly informed participants. Backbenchers need as much information as possible to anticipate the positions and arguments that are likely to be made by other parties.

In what follows, we explore how and why political parties ensure that certain subsets of their backbenchers become experts who can be effectively deployed in parliamentary democracies. In line with informational theories of legislative organization, we argue that committee systems are the arenas where backbenchers acquire information and are trained for parliamentary debates.[2]

3 Committee Systems as Institutional Solutions for Information Acquisition

Committees systems are arenas that "shun majority rule, seek unanimous agreements via internal deferred payments, and adjust to the outer world, or incorporate its demands, via side payments" (Sartori, 1987, 236). Canonical views of committees see them as serving distributive, information, and partisan functions. The three strands share the assumption that legislators are first and foremost reelection seekers (Mayhew, 1974). They differ, however, in the weight they ascribe to political parties in the incentive structure of legislative actors. Distributive approaches emphasize committees as institutionalized arenas for logrolling and gains from trade. In a weak party envi-

---

[2]Proksch and Slapin (2014) deal only tangentially with committees in their speechmaking model. Their model shows that legislators who have more committee assignments speak more *in aggregate*, although they do not examine whether these legislators speak more on *the committees they serve on* nor do they dwell much on the underlying mechanism of information acquisition as to why committee members may speak more. Giannetti and Pedrazzani (2016) also deal with expertise and speechmaking. However, they focus only on bill debates and whether the legislator sits on the committee that reviewed the legislative proposal.

ronment, legislators self-select to committees and use them to deliver pork to their constituency to heightening their reelection prospects (Shepsle, 1978; Shepsle and Weingast, 1981). The partisan approach sees committees as an extension of the party on the floor (Kiewiet and McCubbins, 1991). Parties and legislators work jointly for the production of a common good: the party brand. Accordingly, Cox and McCubbins (1993) see committee assignments as an electorally valuable reward that is partially given out based on loyalty to the leadership. Legislators who have been more loyal to the party leadership are more likely to be assigned to a committee with tangible policy-influence benefits.

The distributive approach is not particularly suited for our argument on committees as arenas for legislators to acquire information for their speechmaking activities. In this approach, actors are presumed to be high demanders, focused only on their jurisdiction of electoral interest, on which they have full information. Furthermore, this approach presumes the existence of weak parties, self-selection, and committees with fiscal particularism proprieties. Parliamentary democracies do not meet these assumptions (Martin, 2011). The partisan approach models parties as essential actors in legislative organization, which would be more suited to explore strong-parties settings. However, the theory does not make any specific predictions about the role of information in the internal organization of legislatures.

We turn to the informational approach of legislative organization as the most appropriate to develop a model of how parties use committees as institutional solutions for their backbenchers to acquire information for speechmaking. As classically illustrated by Krehbiel, political parties and legislators are risk-averse and highly interested in "policy-making outcomes", i.e.,"the effects of policies upon their enactment and implementation" (Krehbiel, 1991, 66). Labor division and specialization are the building blocks of informational theory. The former consists in the horizontal assignment of legislators to deal with specific jurisdictions. The latter is "the process by which an individual legislator brings to the group's deliberation the product of his personal competence in a subject-matter field" (Buchanan et al., 1960, 649). The informational perspective suggests that legislators are assigned to committees, which are endowed with resources to allow members to

gain specialized information to convey to the chamber how proposed policies map onto outcomes. Krehbiel (1991) conceives committees as creations of the floor majority designed to help the chamber *as a whole* make good policy.

To make the informational approach travel to parliamentary democracies, we need to depart from the weak party and the self-selection assumptions. To be sure, political parties and committees are the two most critical organizational units in legislatures in Europe (Saalfeld and Strøm, 2014). The latter is an endogenous creation of the former insofar as parties control the assignment to committees, as well as their institutional powers via the standing orders of the chamber. In parliamentary democracies, legislators cannot self-select to committees in the expectations of using that path as a mechanism for career advancement. Consequently, we expect committees to be designed to benefit individual *parties* in their legislative and political endeavors.

Four conditions make committees optimal institutional solutions for political parties to train their backbenchers for speechmaking activities. First, the assignment process follows a proportionality rule, in which each party is allowed to designate a number of legislators to each committee commensurate to its size in the legislature (Rasch, 1999). Such arrangement ensures that no party is excluded from the benefits of having its legislators in a committee for informational purposes, regardless of its size.

Second, party leaders considers backbenchers' expertise during the assignment process (Damgaard, 1995). Legislators are more likely to sit in committees on which they have previous knowledge. For example, a legislator with an educational background in public finance is more likely to sit on the finance committee. By and large, this curbs transaction costs that legislators would have to face if they were chosen for committees without previous knowledge. There is extensive research on parliamentary democracies supporting this assumption. For example, Hansen (2011) and Mickler (2017) show for Ireland and Germany, respectively, that educational background and expertise is one of the most important determinants of committee assignments. Also, Yordanova (2009) corroborates this finding in her research on the European Parliament. We contend, however, that being assigned to a committee has an effect above that of pre-existing interest or

knowledge.

Third, committees give access to privileged information, hearings with executive members, access to private documents, and the immersion in a bureaucratic and political network that legislators would not have otherwise. Even those legislators that have previous educational background and experience on topic learn information in committees that is useful for their speechmaking activities (Strøm, 1998).

Finally, beyond the information sources discussed in the previous point, committee members also have access to information from well-informed third parties (Li, 2007). Committees hold regular hearings with interest groups, academics, non-governmental organizations, and so forth that provide them with valuable information from outside of the executive-legislative arena. They acquire and digest information in the committee arena and then act as messengers, transmitting credible exogenous information to their parties (Austen-Smith, 1993; Diermeier and Feddersen, 2000). Information originating from third parties is vital for speechmaking, in that it helps party leaders understand the preferences of exogenous actors and to delineate their policy position accordingly. This discussion leads us to our first key hypothesis.

H1: Party leaders are more likely to select backbenchers to speak in the plenary about topics related to the committee in which they sit.

Political parties do not compete equally across all policy dimensions. Instead, each party has a particular brand made up of a constellation of issues that matter for its electorate. Saliency theories of political competition shed light on how political parties articulate their brand (Robertson, 1976; Dolezal et al., 2014). First, parties deliberately choose not discuss or highlight topics that are not relevant to their brand. Instead, they use their agenda-setting powers to stay on message and focus on a subset of topics that allows them to build a coherent and electorally attractive brand. Second, and relatedly, parties are not expected to address all policy issues.

Saliency theories of political competition matter for our explanation of how and why parties select committee experts to take the floor. We suggest that political parties follow different strategies, depending on the salience of the topic under debate for their brand. If the topic being

debated is highly salient for their brand, party leaders more likely to deploy its best and most competent members in the plenary debate. Thus, it is more likely that committee members are used to deliver speeches on their areas of expertise. Conversely, if the topic under debate is not highly salient for the party brand, the incentives to deploy their most well informed and trained backbenchers will decrease. To be sure, one might think that parties have incentives to deploy experts across the board. Yet, there are trade-offs involved in speaker selection. Experts have time and resources constraints. Their expertise is needed beyond the floor. Indeed, they need to assist party leaders in drafting legislation, in tabling questions to hold the executive accountable, and so forth. Thus, we expect that party leaders will deploy experts to speak about topics that matter for the party. Those who are experts in less critical areas will concentrate their focus on other legislative activities. This leads us to our second hypothesis:

H2: The more salient the policy jurisdiction for the party brand, the more likely party leaders will select a backbencher with committee expertise on the topic.

Political parties face different legislative dynamics depending on whether they are in government or opposition. To be part of the executive offers advantages, but also inflicts costs and risks to the party brand. First, in government parties, party leaders need to devote more time and resources to intra-party maintenance activities to uphold the cohesion of the party group (Kam, 2009). If the government presents a policy proposal that fails to get support from the legislature that hampers the credibility of the party brand. Second, being in government forces parties to make compromises. For example, supranational constraints and agreements may preclude the party from pursuing the most efficient strategy for its brand. Consider the example of many parties that in recent years have accessed power in the European Union and had to make profound changes to their party brand to comply with EU compromises (Mair and Thomassen, 2010). Third, and relatedly, compromises and trade-offs are made at the elite level, the prime beneficiaries of office perks. Party supporters, whose ideological position is more rigid, do not have such tangible benefits and may drift away from the party if its brand qualities weaken (May, 1973).

Opposition parties do not face such competing tensions. First, if they fail to uphold party cohesion, the potentially harmful effects on the party brand are not as strong. Second, the opposition does not have any exogenous force driving it to compromises. By and large, the opposition has the leeway to pursue the most efficient strategy for party brand maximization. Opposition compromises if, and only if, it benefits from such a strategy.

Differences in the legislative dynamics influence the approach that party leaders have to speechmaking. We contend that government parties will have higher incentives to deploy its most qualified backbenchers to speak on behalf of the party. The reason behind differential incentives between the government and the opposition hinges on the need to sell the government position. The government will articulate a strategy with its supporting party to minimize the risks identified above. Crucially, it needs to make sure that backbenchers that are highly specialized and knowledgeable get assigned to speak about the topic. By contrast, opposition parties do not face such challenges. Their decisions on the party brand are mostly endogenous, which facilitates their task of selling it to supporters and electorate at large. This discussion leads us to our third hypothesis:

H3: Government supporting parties are more likely to select backbenchers to speak in the plenary about topics related to the committee in which they sit.

The use of a division of labor system of committee systems should be differentially employed by political parties depending on their size. While all parties would likely benefit from this system, implementing such an arrangement requires a sufficiently large pool of individuals who can specialize. In Krehbiel's original approach, committee backbenchers are agents of the chamber, which is large enough to ensure specialization. However, when we consider backbenchers as agents of their parties in committees, we note that smaller parties are less likely to be able to implement an efficient system of specialization among their members. In smaller parties, it is likely that members will need to specialize in multiple jurisdictions. Consequently, those parties will be less likely to have a compartmentalized assigning of speechmaking depending on committee specialization, with all the efficiency advantages previously exposed. This leads us to our fourth hypothesis:

H4: Major parties should rely more on specialized backbenchers when allocating speaking time.

While some of these core hypotheses are derived fairly straightforwardly from the informational perspective or insights from seasoned scholars of parliamentary politics, we note that we also provide the first large-scale empirical test of whether legislators are differentially active in floor debates on the jurisdictions that correspond to their committee assignments. The rigorous empirical testing of such proposition helps us to understand an overlooked facet of legislative organization and speechmaking.

4   Case Selection: Portugal

Portugal's legislature is a unicameral body whose 230 members are elected in a non-preferential closed-list PR electoral system, with little incentives to cultivate an electoral connection with their home district.[3] Voters have to choose between party lists, with the party brand serving as the important cue guiding vote choice. The country is divided in 22 district with varying magnitudes, which makes party leaders the predominant principal (Carey, 2007). To be sure, districts as such are meaningless in the Portuguese political culture, stemming from the constitutional provisions stating that legislators have a national mandate. (Leston-Bandeira, 2009).

Political parties are the fundamental organizational unit in the legislature.[4] Indeed, there is a Constitutional provision bestowing parties as the principal holders of parliamentary rights. Rules of procedure determine that legislators have few, if any, individual-level prerogatives. According to Leston-Bandeira (2009, 699), "individually, MPs have very little power to intervene in parliamentary work. It is up to the PGs [parliamentary groups] to nominate members to the committees, to set the parliamentary agenda [...] and to manage parliamentary resources". The conjugation of electoral system incentives with rules of procedure gives way to a highly cohesive legislature (Braga da Cruz, 1988). In addition to exogenous reasons, intra-party rules offer further incentives for cohesiveness. Political parties have internal statutes that bind legislators to conform to party

---

[3]Constitutionally, cabinet members are not allowed to hold their parliamentary seats. Thus, we can analyze the allocation of speakers while sidestepping the speaking behavior of ministers that is the subject of most existing analysis.

[4]In the Portuguese legislature there is no formal differentiation between frontbenchers and backbenchers.

discipline.[5]

Rules of procedure determine that floor time should be allocated according to party size. Party leaders actively control the allocation of that time – that is, without their consent, legislators can only take the floor for ten minutes each legislative session to make a personal statement. The most critical debates in Portugal are the annual budget, the fortnightly debates with the prime ministers, and the annual state of the nation debate. In these debates, party leaders carefully select the backbenchers to take the floor and communicate their names to the presidium, whose function in debates is to keep time and to summon selected speakers to take the floor.

The Portuguese legislature has thirteen permanent committees whose jurisdiction shadows cabinet portfolios.[6] Unfortunately, scholarly work has not yet dealt with committee assignments in Portugal. Work on the appointment to committee chairs, however, shows that political parties control the allocation of members to committee positions (Fernandes, 2016). Thus, the ability of legislators to self-select into a particular jurisdiction requires them to persuade the party leadership that such an assignment benefits the party as a whole.

Unlike other European legislatures, for example, the German Bundestag, Portuguese rapporteurs play little political role. With few exceptions, their role is mainly technical insofar as committees tend to operate in a consensual fashion. Similar to most European legislatures, committee chairs are allocated proportionality and exert their functions in a nonpartisan fashion. There is a tacit agreement that the opposition should hold the chair of the finance committee. In line with their European counterparts (Strøm, 1998), they have the powers to review legislation before it reaches the floor and to summon witnesses and document. There are also several *ad-hoc* committees whose nature is mostly investigative.

Portugal offers a compelling case to examine our research question whereby we lay the groundwork for a theory of the role of information in speechmaking in parliamentary democracies. First, we hold constant many of the canonical rival explanations of speechmaking: electoral system

---

[5]PS and PSD do offer some leeway to their legislators in "conscience" issues (i.e., abortion, gay rights, and so forth). Those votes explain the dissent found in recent empirical work by Leston-Bandeira (2009).

[6]All committees have similar powers, including law-making, witness summoning, and document requirement. There are, however, differences in the political importance that each party attaches committees. We account for the party-committee dyad by including a saliency measure in our models.

incentives, party loyalty, and the rules of procedure. Second, Portugal offers the possibility to examine a case where highly cohesive political parties are at the center of legislative organization. Under conditions of well-developed and strong parties, therefore, we can explore the variation in how committee organization affects speechmaking. Third, Portugal allows us to make inferences that can travel to closed-rules party-centered systems, for example, Spain, Norway, and Israel. To be sure, there are limitations in the scope of our argument, a point to which we return in the conclusions. In the following results, we show new empirical confirmation as to how the leaders use of their powers in such an institutional setting to leverage the differential information that their legislators can bring to debates.

5 Using Bills as Training Data for Speech Analysis

In an ideal world, parliamentary speeches would have clear headings that unambiguously mapped them onto the relevant committee jurisdiction. This would allow us to test our argument of whether committee specialization heightens the likelihood of backbenchers to take the floor in a particular jurisdiction. However, from numerous parliamentary records that we have examined, this does not occur in many cases—especially when considering historical data. When titles do exist on debates, they tend not only to be short but also vague about the topic under consideration.[7] This makes classification challenging to do accurately as the researcher has to decide exactly how to map these different labels onto the categories of interest (committee jurisdictions).

Thus, our approach to mapping speeches to committee jurisdictions is to map *words* to committee jurisdictions and then use that mapping to classify speeches. For that, we need to have some set of reference text where the relevant jurisdiction is *known a priori*. Laver, Benoit and Garry (2003) posit that the reference text should share with the virgin text (1) a similar lexicon, (2) multidimensional constellation of positions, (3) as many different words as possible.

To address these requirements, we turn away from parliamentary debates themselves and

---

[7]Moreover, some debates may span multiple topics, a headings based approach would require us to classify entire groups of speeches as coming from the same topic.

leverage an alternative source of training data: legislative proposals.[8] We see at least three advantages of using bills as the training data for our supervised model. First, bills and speeches have a somewhat similar lexicon. They are instruments of the legislative repertoire meant to convey political parties' policy positions in the legislature. Second, bills make an *explicit* record of the committee to which they are referred.[9] This helps us by linking a particular committee with a particular vocabulary and jargon. Third, bills are introduced on a wide range of topics. Even if they are not passed or considered on the plenary floor, their initial versions still provide highly salient indicators as to the topics that their committees consider. This ensures that our training data consists of multidimensional texts with a diverse vocabulary.

There are major concerns that could be leveled out against the use of bills as the training data for a classifier of parliamentary speeches. First, one might think of bills as consisting *exclusively* of a technical and legalistic lexicon, meant to stand up to judicial scrutiny. This may decrease their usefulness as a reference text for a classifier because speeches are typically non-(or less-) technical, political texts. A careful inspection of bill structure in many continental European democracies shows that most bills often contain a 'motivation' section.[10] Typically, this section is a non-technical, highly political text, with references to the party manifesto or previous positions, providing broad context and explanations on the bill. Parties use this section to defend the need to introduce the bill. As Appendix A shows, such practice is widespread in Europe—with Westminster systems being the major exception. In countries where bills contain a 'motivation' section or similar document, we can leverage both the technical language and this motivation section to classify documents. We examine in detail how classifiers trained on either section alone

---

[8]An an alternative strategy, that we suggest using as a validation exercise rather than the primary procedure, is to hand-annotate speeches into categories by sampling a large and sufficiently representative set of speeches, hand labeling them, and then training the classifier on those speeches. While this procedure would likely work well on a reasonably homogeneous corpus of texts, the fact that we are examining fifteen years of data (and thus must sample representatively to account for the possibility of topical drift) and across many issues means that a sampling approach that we trusted to classify speeches (versus as a validation exercise) correctly would likely be much more costly than relying on the pre-labeled bills.

[9]In this work, out of the 6956 bills introduced in our period of interest, only 11.3% were referred to multiple committees. The overwhelming majority of bills (88.7%) were referred to a single committee. We aim to use bills that have a well-defined text corpus, in which the topic and the labeling is clear, and use it to create a gold-standard "annotated corpus". Thus, we exclude multiple referred bills from our SVM. In systems where multiple referred bills were much more prevalent, researchers could look at other types of classification procedures (e.g., 'multi-label' classifiers) that improve performance.

[10]In the Portuguese case, the motivation section usually has 2-3 pages, and it is called *Exposição de Motivos*.

perform and find that, as expected, models trained only on the formal technical language have weaker performance but models trained on *both* technical and motivations do better than models trained on the motivations section alone.

Second, one might object that bills are often articulate the collective positions of the party whereas speeches articulate individual positions that may diverge from the collective position and thus be challenging to classify correctly. We suggest, following Proksch and Slapin, that backbenchers in parliamentary democracies mostly deliver speeches *on behalf* of the party and thus should articulate positions that are more-or-less consistent with the bill texts drafted by their colleagues. Further, note that since we are seeking to identify *topics* and not *partisan* characteristics in speeches, as long as legislators rely on a shared set of words or concepts to refer to issues in a particular committee's domain, the classifier should do well.

As parliamentary text can be highly complex—encompassing a wide array of technical, complicated, and multidimensional topics—we think that turning to supervised methods using bills as the training data opens up a productive frontier in studying legislative politics. This framework provides an alternative to the traditional unsupervised methods, e.g., topic models, that may prove more tractable for analyzing particular questions in legislative politics that revolve around the critical institutions of committees or government portfolios. For example, existing work on identifying topics in political speech requires extensive work to (1) decide the number of topics and (2) map the topics onto the critical political quantities of interest, e.g., committee jurisdictions.

Our approach avoids this problem by making a trade-off. Supervised method methods fix the categories of interest (committee jurisdictions), and thus avoids the problems of post-estimation interpretability as we are merely trying to find the best mapping of words to pre-defined topics. The cost is the inability to discover categories besides the ones we fixed in advance (e.g., bills on monetary versus bills on fiscal policy) that other unsupervised methods (e.g., topic models) may be able to uncover. However, as our core argument revolves around the fundamental political institutions of committees, a supervised procedure is ideal for tractably classifying speeches.

5.1 Approach and Data

We apply this approach to a fairly long period of legislative history by examining all speeches and bills in Portugal from 2000 until 2015.[11] This constituted about 50,000 speeches and 6,000 bills that we gathered from the Portuguese Parliament Official Website.[12] Appendix B describes the preprocessing procedure, the bills, and the speech corpus. We show that various methods of pre-processing (e.g., including or excluding rare terms or stop words) has little effect on our model's performance.

To analyze our data, we rely on a Support Vector Machine (SVM) that we discuss in more detail in Appendix C. While noting that our procedure allows the researcher to use their preferred classifier, we note that, theoretically, SVMs are well-suited to our purposes for two reasons. First, existing research in political science has used SVMs successfully when analyzing speech in the US Congress (Yu, Kaufmann and Diermeier, 2008a,b; Diermeier et al., 2012). As that is a similar domain with similar challenges (e.g., the use of procedural language), this gives us confidence that this is a reasonable model for our purposes. Second, SVMs are known to deal well with high-dimensional data while maintaining accurate out of sample predictions.[13] Appendix D compares our SVM to other standard methods (logistic regression with (i) LASSO or (ii) Ridge penalties; (iii) a random forest; (iv) a simple neural network) and shows that, while the differences are typically fairly small, the SVM performs better.

We briefly outline how an SVM works, but refer interested readers to more detailed treatments in other sources (Bishop, 2006; Hastie, Tibshirani and Friedman, 2009). To begin, consider a binary case: Is bill j associated with the agriculture committee? Using a set of training data to build our model, we associate bills that known to be agriculture bills as '1' and all other bills as '-1'. The core intuition of an SVM is that we do not simply want to find the coefficients that result in a 'best fit' of the data in the sense of the ordinary least squares (quadratic) loss. Rather,

---

[11]We selected this period because earlier speeches are not available in a similar or easily usable form from the Portuguese Parliament.

[12]www.parlamento.pt

[13]Indeed, some work shows that rather than being problematic, rare words can help classification accuracy in some contexts when using an SVM (Joachims, 1998; Wang and Manning, 2012).

we want to find the set of coefficients that creates a hyper-plane that maximally separates the '1's and '-1's; this is why SVMs can be thought of as a classification procedure that 'maximizes the margin' between the two classes of observations. By finding the line that best separates the data, we are both less sensitive to outliers and thus will likely do better on unseen data.[14] We run each of these binary classifiers for each of the thirteen committee jurisdictions and classify bills based on the one where a bill is given the most positive score; this is the 'one-versus-rest' procedure and is a standard method for multi-class classification with SVMs.

6   Validation of the Model

Validation is crucial to test our argument that bills are a good predictor of speeches. We do this in two stages. First, we confirm that the SVM can do a good job of predicting the committee labels of unseen bills that are not included when training the model. This gives us confidence that our procedure is good at capturing the underlying words that differentiate bills. This implies that the committee jurisdictions are coherent enough that we can accurately predict, say, unseen agriculture bills using the agriculture bills in our original training data. Second, we confirm that the SVM can work on *speeches*. To do this, we hand classified a moderate number of speeches and then examine the performance of our model. If this shows good results, it confirms that there is enough similarity across the political texts to justify using this procedure.

6.1   Validating the Model on Bills

Given the above set up, it is important to tune the parameters of the SVM, as well as to verify that it performs acceptably well on our training data. Cross-validation is the canonical procedure to do this (Hastie, Tibshirani and Friedman, 2009). To do this, we split our dataset of bills randomly into two parts—one constitutes 80% of the data (the training set) and the other 20% we temporarily

---

[14]There is a slight wrinkle in that for our data, and, indeed most data, there are some points that are on the 'wrong side' of the margin boundary—i.e., the data cannot be linearly separated by the covariates. To address this problem, we use the 'soft' formulation of the SVM where we introduce a penalty C tuned by cross-validation that allows there to be some 'mistakes' that are penalized. This tends to increase the robustness of the SVM further. See Bishop (2006) for a thorough discussion.

set aside as 'validation' data. We train the SVM on the 80% of the data and then see how it performances on the remaining 20% where the labels are known. This gives us an approximation of how the SVM will perform on new documents. To examine how the model performs, we use the standard F-measure. This combines two common metrics: Precision (P) asks, of documents *classified* as '1' (in a particular jurisdiction), what proportion are *correctly* classified? Recall (R) asks, of all documents that are *truly* '1', what proportion are classified correctly? Classifiers could do well at one at the expense of the other; for example, classifying everything as '1' gives perfect recall but is a flawed model. Thus, the F-measure takes the harmonic average of the two to ensure that we score models based on how they do at both:

$$F = 2 \cdot \frac{1}{1/R + 1/P} \tag{1}$$

Note this has the desirable property that if one measure is bad, i.e., R or P is close to zero, this drives the entire measure towards zero. Thus, in the example above, the poor recall of the model would suggest that the 'good' precision does not lead us to assume our SVM is working well. Table 1 reports our classifier's F-measures for each of the binary classification algorithms employed. We then calculate a 'global' F-measure by taking the average of the individual F-measures, weighted by the size of each category. We note that the F-measure is quite good overall. The one category where the measure is quite poor (Ethics) has only a minimal number of bills (6 in total) and thus should not unduly bias our results.

These results give us reasonable confidence that our SVM trained on bills performs well and avoids key problems of over-fitting. To project to speeches, we then re-train the SVM on the entire corpus of bills and predict the categories of all speeches using the procedure outlined above. Appendix D discusses our validation of the particular parameters of the SVM selected and compares its performance against common other models.

Table 1: Validating the SVM on Held-Out Bills

Class	Precision	Recall	*F*-score	Number
Agriculture and Fisheries	0.77	0.73	0.75	41
Constitution, Civil Rights and Liberties	0.89	0.91	0.90	202
Culture	0.69	0.81	0.75	42
National Defense	0.89	0.80	0.84	20
Economy and Public Works	0.80	0.73	0.77	143
Education and Science	0.91	0.91	0.91	116
Foreign Affairs	0.89	0.96	0.93	84
Ethics, Society and Assembly House Keeping	0.50	0.33	0.40	6
European Affairs	0.89	0.67	0.76	24
Public Administration and Budget	0.85	0.85	0.85	116
Environment, Territory and Local Government	0.91	0.97	0.94	203
Health	0.91	0.88	0.89	81
Labor and Social Security	0.84	0.83	0.83	156
Overall	0.86	0.87	0.86	1234

6.2 Validating the Model on Speeches

The above validation exercise only confirms, however, that a model trained on the text of bills can predict the text of other unseen bills accurate. Next, we need to show that this model can also correctly classify speeches. To do this, we examine a subset of speeches that we classified using human coders. We have manually annotated a randomly selected sample of 250 speeches. Two independent coders have been asked to assign speeches to one of the thirteen categories of our classification schemes, resulting in a Cohen's kappa inter-rater agreement of 0.59. While this is only a moderate ranking (Landis and Koch, 1977), we note that speeches — given their complexity and multidimensionality — are sometimes hard to classify and thus we should expect reasonably high levels of inter-coder disagreement. We show qualitative evidence of 'difficult' speeches that both our coders and the SVM found more 'difficult' to classify.

From those hand-coded speeches, we selected the 150 on which both coders agreed. These are the clear cases that our SVM should also correctly identify if extrapolating from bills to speeches works successfully. This can be thought of as a 'ground truth' test for the procedure. The results, by the human coded jurisdiction as 'truth', are shown in Table 2.

We see that the SVM performs well and has *F*-scores that are roughly comparable to those seen in the cross-validation on bills shown earlier. This gives us confidence that extrapolating

Table 2: Validating the SVM on Speeches

Class	Precision	Recall	F-score	Number
Agriculture and Fisheries	0.75	0.75	0.75	4
Constitution, Civil Rights and Liberties	0.64	0.96	0.77	24
Culture	1.00	0.71	0.83	7
National Defense	1.00	0.50	0.67	2
Economy and Public Works	0.73	0.55	0.63	20
Education and Science	1.00	0.94	0.97	18
Foreign Affairs	0.71	0.71	0.71	7
Ethics, Society and Assembly House Keeping	0.00	0.00	0.00	1
European Affairs	0.67	0.67	0.67	3
Public Administration and Budget	0.80	0.76	0.78	21
Environment, Territory and Local Government	0.93	0.87	0.90	15
Health	1.00	0.92	0.96	13
Labor and Social Security	0.81	0.87	0.84	15
Overall	0.82	0.81	0.80	150

from bills to speeches using our supervised procedure works reasonably well, especially given the fact that speeches are 'messier' than bills regarding their content.

To further validate our procedure, we provide two examples of where our approach works well and where it has difficulties in classifying speeches. On the one hand, some speeches are easy to classify. The author stays on topic and makes a well-structured intervention. For example, Bernardino Soares, from the PCP (Portuguese Communist Party) said on 22 July 2009:

> "[...] there is a clear need to have hospitals solely dedicated to infant health care [...]
> in order to make for more specialized and qualified health policy for children. [...] We
> need to consider a management plan for such kind of hospitals, not just in Lisbon, but
> also in Coimbra and Oporto."

For this particular example, the SVM classifier gave a positive score (0.94) on Health, in line with the two human coders, and negative scores for all other categories.

On the other hand, some speeches are complicated to classify, in that the speaker refers to a topic, but also touches upon other topics. For example, Agostinho Lopes from the PCP (Portuguese Communist Party) said on 1 September 2011:

> "[...] witness the trouble of PSD to recognize its responsibilities in the destruction of
> the Portuguese agriculture over the past 30 years, alongside with the PS. Specifically,

regarding the two PAC reforms and the policy towards small and medium producers. [...] Additionally, I would like to inquire about why the Agriculture Ministry has not yet dealt with the benefits to the Douro wine producers, after a strong increase in Port Wine exports."

First, the legislator talks about several, and distinct, topics in the same speech. This is often the case with small parties. Because they are allocated floor time in proportion to their size (and thus have less of it), they tend to condense their points on multiple issues into a single speech. Larger parties, which are allocated more time, could deliver different speeches on different aspects of the same debate—which fits into our earlier hypothesis about them being more able to use a division-of-labor system.

In these ambiguous cases, where the speaker mixes phrases and words from multiple categories, multiple of the 'one-versus-the-rest' classifiers that constitute the multi-category SVM receive high scores. The standard usage of an SVM is to return a single deterministic prediction based on the one with the highest score. As some of these ambiguous cases will be narrowly misclassified, the F-scores reported earlier are a conservative test of our classifier's effectiveness. For example, consider cases where both human coders were conflicted but narrowly decided on the same label. That case may be misclassified by the SVM when faced with a narrow choice.

We conduct extensive further validation in Appendix E where we examine which types of words drive performance; in short, we find that using only the 'technical' language has relatively poor performance, but the joint usage of the 'motivations' section (analogous to a political speech) and 'technical' section outperforms 'motivations' alone. We further show that the top words selected by the SVM for each category are nouns that relate to the core jurisdiction of each committee and thus we are classifying based on substantive content rather than idiosyncratic words.

7   Methods

To format the data for analysis, we created a panel dataset from the corpus of speeches delivered. For each legislator i who served during some legislature l, we created thirteen observations corresponding to the number of speeches delivered in each of the thirteen committees indexed by c. We create two dependent variables. First, we measure the proportion of time a backbencher allocates to a jurisdiction c during a given legislature. Formally, if n_{ilc} is the number of speeches by backbencher i in legislature l and committee c, the dependent variable:[15]

$$p_{ilc} = \frac{n_{ilc}}{\sum_j n_{ilj}} \qquad (2)$$

Second, we create a dummy variable that takes a value of 1 if a backbencher delivered *any* speech in a jurisdiction c, and 0 otherwise.[16]

Our independent variables are measured as follows. First, we create a dummy variable to identify the legislators who serve on the relevant committee ('Members') versus those who do not ('Non-Members').[17] Second, we measure *Saliency* using data from Comparative Manifesto Project (Volkens et al., 2016). We extract a measure of salience for each party across the committee jurisdictions.[18] To facilitate comparability across parties, we rank the salience of jurisdictions for each party and then order them at equal intervals from '-1' to '1', where '1' indicates the most salient portfolio.[19] We construct our measure by assigning each CMP indicators according to committee jurisdictions. Our third independent variable of interest is *Government*, a dummy that takes a value of 1 if the legislator's party belongs to the executive, and 0 otherwise. Fourth, we include a dummy variable that accounts for *Major Parties*, whereby we define as major those parties that have more than ten percent of seats in the chamber. In Portugal, the Socialist

---

[15] Appendix G shows the specification where we look at the proportion of *words* allocated to each jurisdiction. These two outcome variables are nearly perfectly correlated, $\rho = 0.98$, and the results are nearly identical.

[16] Both variables are normalized inside each legislature. If there are idiosyncratic reasons why some backbencher speaks more (in aggregate) than others, looking only at their percentages or any participation means they are unlikely to actively drive the results versus, say, using the count of speeches in a given jurisdiction.

[17] The small number of legislators who switch committees during a government are coded as 'Members' for any committee that they served on during a government.

[18] We match committee jurisdictions with the CMP coding scheme following work previously developed by Bäck, Debus and Dumont (2011).

[19] The most salient issues by each party are shown in Appendix F.

Party (PS) and the Social Democratic Party (PSD) are much larger than all other parties in the legislature.

In all of our analyses, we use ordinary least squares (OLS) in a fixed effects framework to address the major threat to testing our hypotheses: unobserved factors that affect speaking behavior.[20] We include fixed effects for backbencher-committee (α_{ic}) in our main analyses in the text to perform a tough and causally credible test of our hypotheses. The inclusion of these fixed effects means that our results are only driven by within-backbencher-committee changes – that is, how does being assigned to a committee change a backbencher's behavior relative to their behavior when they were not assigned to the committee. This is a stringent test as it means that only backbenchers who move-on-and-off of a committee during the term we examine can be leveraged in estimating our causal effect.[21] Fortunately, this does occur with some frequency. Of the 9867 backbenchers-committee pairs (756 backbenchers; 13 committees) in our sample, approximately 9% (896) have variation in committee assignment.[22] Appendix G shows less robust specifications that include separate committee and backbencher fixed effects ($\alpha_i + \gamma_c$). The results are mostly in agreement regarding sign and statistical significance, although we will comment on differences that occur.[23] In all models, we include fixed effects for the legislative term to deal with other possible omitted confounders such as differential legislative agendas across parliamentary terms.

Our approach allows us to control for various unobservable underlying variables. For example, imagine a backbencher who spoke more on financial issues given her background as a banker (or some other arbitrary unobservable like 'competence' or 'knowledge' on a particular issue).

---

[20]Our first dependent variable is binary, and thus we estimate a linear probability model as is common in the fixed effects literature. We show in Appendix G that using a fixed effects logistic regression returns identical results regarding which coefficients are statistically significant. Our second dependent variable is bounded between zero and one; while other forms of modeling this relationship exist (e.g., fractional logistic regression), performing OLS has a 'linear probability model' type interpretation and allows us to easily include fixed effects to control for unobserved confounding. Thus, following the majority of the literature on estimating fixed effects in political science, we use a linear model.

[21]This also means that backbenchers who serve only one term are unable to contribute to the estimation of the causal effect

[22]4316 cannot contribute because of serving only a single term; 4655 see no variation in the assignment.

[23]We cluster our standard errors on legislators in this analysis and on legislator-committee (ic) when legislator-committee fixed effects are included. Results available upon request show models estimated in STATA (lmer in R proved intractable to get convergence in a sufficient time-line) where we estimated an intermediate model that has fixed effects for legislators and committee ($\alpha_i + \gamma_c$) and *random* effects for legislator-committees, i.e. $\alpha_i + \gamma_c + \zeta_{ic}$; $\zeta_{ic} \sim N(0, \sigma^2)$. The results are, as expected, somewhere between the two fixed effects specifications we show in the main paper.

Rather than engaging in the (heroic) task of claiming we have measured and included variables related to every possible individual confounder, we can leverage the fact that we can observe said backbencher's behavior *before* and *after* she was assigned to the, say, Budget Committee. By including backbencher-committee fixed effects, we can conclude that the change in speaking behavior after an assignment is not due to some underlying time-invariant propensity to speak, but rather because of the change in their behavior after the assignment took place. A downside of this approach is that we 'black box' the analysis of which individual characteristics matter, although we gain the benefit of being able to correctly test our hypotheses in the presence of certain types of unknown and unobserved heterogeneity. Examining the details of how the background of legislators relates to both committee assignment and plenary time is a fruitful area for further research.

We address possible time-varying confounding by including two control variables (logged seniority and logged electoral vulnerability) that existing literature suspects are also related to speaking activity.[24] As a threat to inference when using fixed effects is the existence of omitted time-varying predictors, Appendix G provides visual evidence that, in the two years before appointment, there is no significant trending in speaking behavior.

8   Results

Before proceeding to our analysis, we first give a descriptive sense of how often each category occurs, Figure 1 shows the number of speeches across committees by legislature. Next, in Figure 2, we turn to show the proportion of legislators who spoke in each jurisdiction (at all) during a legislature. We see that most legislators do not speak on most jurisdictions, which is initial evidence favor of a division-of-labor inside of parliamentary parties.

Descriptive data show that "Constitution" is, by and large, the predominant category in all legislative terms, except 2009 and 2011. This makes sense given the nature of that committee as

---

[24]We measure Electoral Vulnerability following André, Depauw and Martin (2015), in which the position on the list of legislator i is divided by the total number of elected number of legislators elected in the district d by party p. Seniority is defined as the length of time (in years) that a legislator had been in parliament when they delivered their first speech during government g.

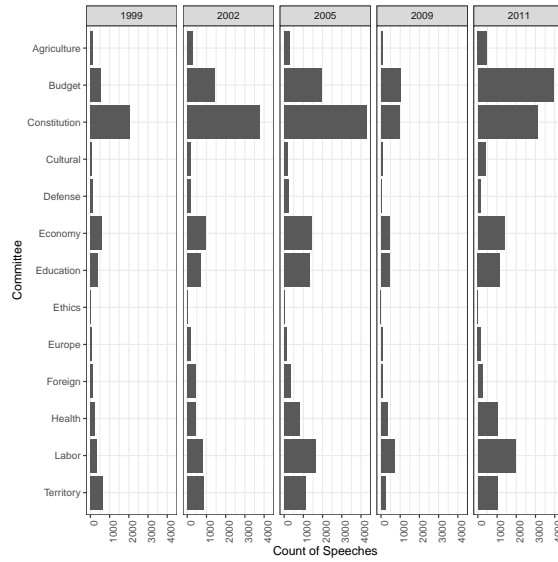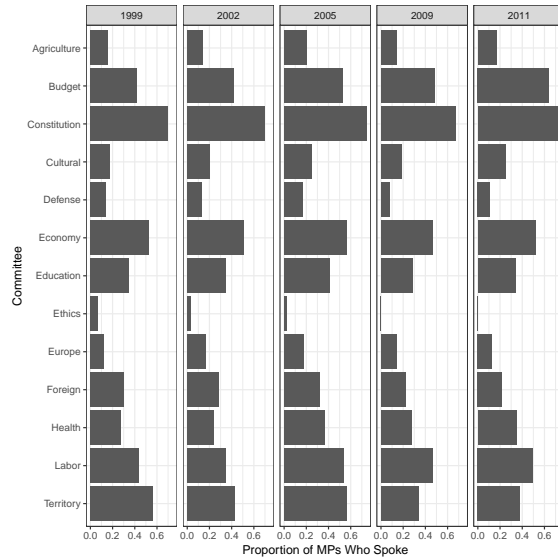Figure 1: Number of Speeches by Committee Jurisdiction

Figure 2: Proportion of Legislators Giving Any Speech by Jurisdiction Over Time

Note: This figure shows the proportion of legislators who participated in each jurisdiction c by legislature. The title of each panel corresponds to the election year for the 8th through 12th Parliaments.

it deals with all procedural issues as well as a variety of legal issues as well as house-keeping for the legislature itself. As many speeches by MPs are making procedural points, it makes sense that these speeches fall in the "Constitution" category .

To test the first core hypothesis, we look for evidence that members who are assigned to committees participate more in floor debates on their jurisdictions. All models in our analysis are

shown in Tables 3 and 4. Model (1) corresponds to a simple hypothesis with a variable for committee membership, the relevant backbencher-committee fixed effects and other controls outlined above. We show that there is a large and highly-significant effect of committee membership on speaking behavior: backbenchers who are on the relevant committee allocate about 10 percentage points more of their time (versus the counter-factual legislator who is not on the committee) and are 20 percentage points more likely to speak on that jurisdiction. This suggests strong support for H1 that political parties use their members who have been assigned to committees as experts to take the floor in speechmaking.

Table 3: Modeling the Proportion of Speeches Allocated To Each Jurisdiction

	(1)	(2)
Member	0.116***	0.104***
	(0.008)	(0.012)
Salience		-0.018***
		(0.004)
Member X Salience		0.043***
		(0.012)
Government		-0.002
		(0.003)
Member X Government		0.021
		(0.013)
Member X Major Party		0.007
		(0.016)
Number of Observations	21411	21372
Number of Legislators	759	759

Note: *** : $p < 0.01$; ** : $p < 0.05$; * : $p < 0.1$ Standard errors in parentheses. The dependent variable is the proportion of time a backbencher allocates to a jurisdiction (i.e. what proportion of the total speeches that a legislator gives in legislature l are in committee c?). Standard errors are clustered on the legislator-committee. All models include the following controls (not shown): legislator-committee fixed effects, legislature fixed effects, and other controls noted in the main text. Note that for Model (2), the lower order effect 'Major Party' cannot be included as no individuals switch party in the period under observation.

Our second hypothesis stated that parties rely on experts more on the issues that are most salient to their party. Model (2) in Tables 3 and 4 shows the results of adding salience to the baseline specification and interacting it with a member's committee status.[25]

As this is a binary-by-continuous interaction, we can assess these hypotheses by examining

___

[25]As a different test of the importance of salience, we re-aggregate our data to the party-jurisdiction level, e.g., how many speeches does some party p give on jurisdiction c? We see that the simple bivariate correlation between the proportion of speeches and salience is moderately high ($\rho = 0.43$; t-statistic of 9.213). This also accords with our expectation and existing work that parties spend more time, in aggregate, on highly salient issues.

Table 4: Modeling Whether a Legislator Participates At All

	(1)	(2)
Member	0.211***	0.166***
	(0.015)	(0.025)
Salience		-0.002
		(0.013)
Member X Salience		-0.011
		(0.022)
Government		-0.033***
		(0.008)
Member X Government		0.083***
		(0.022)
Member X Major Party		0.016
		(0.032)
Number of Observations	21411	21372
Number of Legislators	759	759

Note: *** : $p < 0.01$; ** : $p < 0.05$; * : $p < 0.1$ Standard errors in parentheses. The dependent variable is the extensive margin (i.e. does a legislator participate at all in jurisdiction c?). Standard errors are clustered on the legislator-committee pair. All models include the following controls (not shown): legislator-committee fixed effects, legislature fixed effects, and other controls noted in the main text. Note that for Model (2), the lower order effect 'Major Party' cannot be included as no individuals switch party in the period under observation.

the regression terms directly. First, we see that for MPs who are not on a committee, there is a statistically significant negative effect: If the committee jurisdiction changed from moderate (0) to highest (1) salience, this would lead to an expected decrease of about 2 percentage points for non-members. However, for members, we see the opposite effect; for them, their effect of speaking is significantly larger than non-members as the significant interaction term notes (i.e., the effect of salience for members is 4 percentage points larger than for non-members). Thus, a one-unit increase in salience for *members* increases the amount of time they allocate to jurisdiction by about two percentage points. The significance of these coefficients is replicated when we only include member and committee fixed effects, although the magnitude of the interaction term increases noticeably (0.10).[26] When we look at Table 4 (i.e., modeling whether a member participates at all), we do not see any statistically significant effect of salience for either non-members or members. This perhaps suggests that the effects of salience are concentrated more in the extensive margin: Of those committee members who are chosen to participate on salient issues, they allocate more

---

[26]The smaller effect here could be that we observe only a moderate amount of variation in the salience of jurisdictions over time and thus the limited amount of variation may limit our ability large effects if we are constrained to leveraging only within-committee variation in salience as the member-committee fixed effects approach implies.

of their scare time to those issues to defend the party's reputation.

To provide further evidence for a division of labor, we turn to test H3 on whether government status matters for the reliance on committee member as the stakes are higher. Government parties have a stronger incentive to facilitate a division of labor as they are trying to sell their policies to the public, rather than establishing a coherent alternative reputation. Model (2) tests this and finds mixed effects. As these are binary-by-binary interactions, we can interpret the coefficients directly. We can safely reject the null of the interaction term being zero in the case of Table 4 and note a non-member government legislator is less likely to participate by about 3 percentage points, a member from the government is about 4 percentage points *more* likely to participate. We cannot, however, reject the null at the 0.05 significance level for Table 3 ($p = 0.106$).[27] This suggests, perhaps, that while government membership interacts with committee membership, the effects are more pronounced for whether backbenchers get involved (governments tend to rely more on their committee members than opposition parties) versus the extent to which backbenchers participate.

Finally, we examine whether it is only larger parties who can maintain a specialized division of labor. We interact a dummy variable for whether a legislator is from a major party against the dummies for committee membership. The large parties in our data are the PS and the PSD. However, given that we include individual fixed effects and no legislator switch parties in the period we observe, we can only estimate the interaction term which corresponds to the quantity of interest: Are committee members used more heavily by large parties? We find no significant effects for this hypothesis, although note that a model estimated with backbencher and committee (vs. backbencher-committee) fixed effects does find a highly significant interaction term.

Overall, our results provide explicit support for H1 and H2 and moderate support for H3. We see that Portuguese parties rely on a division of labor system when deciding who participates in plenary debates. Even controlling for individual proclivities to participate in specific topics based on unobserved background characteristics, we see that committee members are markedly more likely to participate and to spend a large amount of their speaking time on their assigned

_____

[27]Note that in Appendix G, we show that with a less causally credible but less underpowered specification, the effect is highly statistically significant.

jurisdictions. We find that when the stakes of speeches are higher—they are on a highly salient issue to the party or the party is in government, parties rely more heavily on their committee members.

9 Conclusion

Our investigation on speechmaking in parliamentary democracies has highlighted the importance of information as a critical factor in shaping effective participation in legislative debate. We have drawn on the 'informational' perspective of legislative organization (Krehbiel, 1991) to explore how party leaders use committee systems to acquire information and train backbenchers for parliamentary debate. From our results, we draw several conclusions.

First, political parties use the division of labor in the committee system to guide how they allocate parliamentary speaking time. Members who sit in specific jurisdictions have a higher likelihood of spending their speaking time talking about that jurisdiction. While this may seem an intuitive proposition to seasoned legislative scholars, our paper makes the first empirically rigorous analysis on this facet of speechmaking by committee membership and speechmaking.

Second, we demonstrated that specialists are relied upon more heavily in high stakes situations, further corroborating the canonical argument that party brand is vital in speechmaking. When issues under consideration matter more to the party's electoral prospects, party leaders rely more heavily on their most reliable and experienced members. Our results suggest that the more salient the policy area, the more likely that party leaders deploy experts to shore up the party's brand in the plenary.

Third, our evidence shows that, for governing parties, whose need for cohesion is higher, there was a greater reliance on committee members to take the floor. Against our expectations, we did not find any differences suggesting that party size matters in the incentives to mimic committee labor division in speechmaking. Future research should explore how small parties' difficulties in having a more complex labor division affect their speechmaking activities.

These results are only being driven by variation in behavior of those backbenchers who switched on to (or off of) a committee during the time window of observation. Our fixed-effects strategy, as well as numerous additional controls, help us to trust the robustness of our results to rival explanations. To be sure, we do not rule out, and indeed can see evidence for, the fact that people tend to speak more on issues that they have background connections to and experience on. We show, however, that, even controlling for those effects, being assigned to a specific committee—and gaining access to privileged information that non-members cannot easily access—has an impact on the likelihood of being selected to speak on behalf of the party on that topic.

This paper turns to Portugal as a case-study. Our results, however, are useful to learn about information and speechmaking in legislatures across Europe. Indeed, Portugal offers us the possibility to test the role of committees in speechmaking while holding constant rival canonical explanations: rules of procedure, electoral systems, and party loyalty. To be sure, Portugal offers a particular setting, in that its institutional environment is highly party-centered. Rules of procedure make parties the *only* gatekeepers of committee system and floor access.

Furthermore, the electoral system gives parties full control of candidate selection and ranking. All these factors push towards high party cohesion. Proksch and Slapin (2014, 83) offer a classification of legislatures according to parliamentary rules and electoral incentives that allow us to map our results onto other legislatures. Our results are most generalizable to systems where party leaders control floor access, and there is weak to moderate incentives to cultivate a personal vote (i.e., Spain, Israel, Norway, Denmark, Slovenia).

Besides evidence on the relationship between committee membership and speechmaking, our paper makes a methods contribution by using bills as training data for speech analysis. Rather than relying on unsupervised methods that attempt to discover topics in speeches without human intervention beforehand, we noted that committee jurisdictions provide an *ex ante* set of topics that map onto critical political issues. Thus, by trying to attach speeches to these existing and salient categories, we can map legislative speech onto relevant political debates and avoid the challenges of trying to map unsupervised topics (e.g., from a topic model) onto the relevant ministries or

committees. To do this, we leveraged a widely applicable source of training data—legislative bills consisting of technical language to enact a statue as well as a 'motivations' section that resembles a political speech. We showed that while technical language alone performs somewhat poorly, using the technical language in conjunction with the motivations gives the best performance. Such an approach can be widely applicable to other contexts far beyond the case of Portugal.

Future research on the role of information for speechmaking must turn to legislatures where parties have less control of the floor and where incentives for a personal vote are stronger. In such settings, party leaders face more complicated decisions. While there is still a need to specialize and obtain information, party leaders also need to consider a trade-off with loyalty. Trying to characterize and understand that dynamic and how those challenges are met seems to be an exciting avenue for future research.

References

André, Audrey, Sam Depauw and Shane Martin. 2015. "Electoral Systems and Legislators' Constituency Effort: The Mediating Effect of Electoral Vulnerability." *Comparative Political Studies* 48(4):464–496.

Austen-Smith, David. 1990. "Information transmission in debate." *American Journal of Political Science* 34(1):124–152.

Austen-Smith, David. 1993. "Information and influence: Lobbying for agendas and votes." *American Journal of Political Science* 37(3):799–833.

Austen-Smith, David and William H Riker. 1987. "Asymmetric information and the coherence of legislation." *American Political Science Review* 81(3):897–918.

Bäck, Hanna and Marc Debus. 2016. *Political parties, parliaments and legislative speechmaking.* New York: Palgrave.

Bäck, Hanna, Marc Debus and Patrick Dumont. 2011. "Who gets what in coalition governments? Predictors of portfolio allocation in parliamentary democracies." *European Journal of Political Research* 50(4):441–478.

Bishop, Christopher M. 2006. *Pattern recognition and machine learning.* New York: Springer-Verlang.

Blei, David M., Andrew Y. Ng and Michael I. Jordan. 2003. "Latent dirichlet allocation." *Journal of Machine Learning Research* 3:993–1022.

Braga da Cruz, Manuel. 1988. "Sobre o Parlamento português: partidarização parlamentar e parlamentarização partidária." *Análise Social* pp. 97–125.

Buchanan, William, Heinz Eulau, LeRoy C Ferguson and John C Wahlke. 1960. "The legislator as specialist." *Western Political Quarterly* 13(3):636–651.

Cain, Bruce, John Ferejohn and Morris Fiorina. 1987. *The personal vote: Constituency service and electoral independence.* Cambridge, MA: Harvard University Press.

Carey, John M. 2007. "Competing principals, political institutions, and party unity in legislative voting." *American Journal of Political Science* 51(1):92–107.

Cox, Gary W and Mathew D McCubbins. 1993. *Legislative Leviathan: Party government in the House.* Berkeley: University of California Press.

Dahl, Robert A. 1950. *Congress and foreign policy.* New York: W. W. Norton.

Damgaard, Erik. 1995. How parties control committee members. New York: St. Martin's Press pp. 308–325.

Diermeier, Daniel, Jean-François Godbout, Bei Yu and Stefan Kaufmann. 2012. "Language and ideology in Congress." *British Journal of Political Science* 42(1):31–55.

Diermeier, Daniel and Timothy J Feddersen. 2000. "Information and congressional hearings." *American Journal of Political Science* 44(1):51–65.

Dolezal, Martin, Laurenz Ennser-Jedenastik, Wolfgang C. Müller and Anna Katharina Winkler. 2014. "How parties compete for votes: A test of saliency theory." *European Journal of Political Research* 53(1):57–76.

Eggers, Andrew C and Arthur Spirling. 2014. "Ministerial Responsiveness in Westminster Systems: Institutional Choices and House of Commons Debate, 1832–1915." *American Journal of Political Science* 58(4):873–887.

Fernandes, Jorge M. 2016. "Intra-party delegation in the Portuguese legislature: assigning committee chairs and party coordination positions." *The Journal of Legislative Studies* 22(1):108–128.

Giannetti, Daniela and Andrea Pedrazzani. 2016. "Rules and speeches: How parliamentary rules affect legislators' speech-making behavior." *Legislative Studies Quarterly* 41(3):771–800.

Grimmer, Justin and Brandon M Stewart. 2013. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis* 21(3):267–297.

Hansen, Martin Ejnar. 2011. "A random process? Committee assignments in Dáil Éireann." *Irish Political Studies* 26(3):345–360.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2009. *The elements of statistical learning.* 2nd ed. New York: Springer-Verlang.

Herzog, Alexander and Kenneth Benoit. 2015. "The most unkindest cuts: speaker selection and expressed government dissent during economic crisis." *The Journal of Politics* 77(4):1157–1175.

Joachims, Thorsten. 1998. Text categorization with support vector machines: Learning with many relevant features. Vol. 1398 Machine Learning: ECML-98. ECML 1998. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence) Berlin, Heidelberg: Springer.

Kam, Christopher J. 2009. *Party discipline and parliamentary politics.* Cambridge: Cambridge University Press.

Kiewiet, D. Roderick and Mathew D. McCubbins. 1991. *The logic of delegation.* Chicago: University of Chicago Press.

Krehbiel, Keith. 1991. *Information and legislative organization.* Ann Arbor: University of Michigan Press.

Landis, J. Richard and Gary G Koch. 1977. "The measurement of observer agreement for categorical data." *Biometrics* (33):159–174.

Laver, Michael, Kenneth Benoit and John Garry. 2003. "Extracting policy positions from political texts using words as data." *American Political Science Review* 97(2):311–331.

Leston-Bandeira, Cristina. 2009. "Dissent in a Party-Based Parliament The Portuguese Case." *Party Politics* 15(6):695–713.

Li, Tao. 2007. "The messenger game: Strategic information transmission through legislative committees." *Journal of Theoretical Politics* 19(4):489–501.

Mair, Peter and Jacques Thomassen. 2010. "Political representation and government in the European Union." *Journal of European Public Policy* 17(1):20–35.

Maltzman, Forrest and Lee Sigelman. 1996. "The politics of talk: Unconstrained floor time in the US House of Representatives." *The Journal of Politics* 58(3):819–830.

Martin, Lanny W. and Georg Vanberg. 2008. "Coalition government and political communication." *Political Research Quarterly* 61(3):502–516.

Martin, Shane. 2011. "Electoral institutions, the personal vote, and legislative organization." *Legislative Studies Quarterly* 36(3):339–361.

May, John D. 1973. "Opinion structure of political parties: the special law of curvilinear disparity." *Political Studies* 21(2):135–151.

Mayhew, David R. 1974. *Congress: The electoral connection.* New Haven: Yale University Press.

Mickler, Tim A. 2017. "Who gets what and why? Committee assignments in the German Bundestag." *West European Politics* pp. 1–23.

Müller, Wolfgang C. 2000. "Political parties in parliamentary democracies: Making delegation and accountability work." *European Journal of Political Research* 37(3):309–333.

Proksch, Sven-Oliver and Jonathan B. Slapin. 2012. "Institutional foundations of legislative speech." *American Journal of Political Science* 56(3):520–537.

Proksch, Sven-Oliver and Jonathan B. Slapin. 2014. *The politics of parliamentary debate*. Cambridge: Cambridge University Press.

Quinn, Kevin M, Burt L Monroe, Michael Colaresi, Michael H Crespin and Dragomir R Radev. 2010. "How to analyze political attention with minimal assumptions and costs." *American Journal of Political Science* 54(1):209–228.

Rasch, Bjørn Erik. 1999. Electoral systems, parliamentary committees, and party discipline: The Norwegian Storing in a comparative perspective. In *Party discipline and parliamentary government*, ed. S. Bauwler, D. Farrell and R. Katz. Columbus, OH: Ohio State University Press pp. 121–140.

Robertson, David Bruce. 1976. *A theory of party competition*. London: John Wiley & Sons.

Saalfeld, Thomas and Kaare W. Strøm. 2014. Political parties and legislators. In *The Oxford Handbook of Legislative Studies*, ed. S. Martin T. Saalfeld and K. Strøm. Oxford: Oxford pp. 371–398.

Sartori, Giovanni. 1987. *The theory of democracy revisited*. Chatham, N.J: Chatham House Publishers.

Shepsle, Kenneth A. 1978. *The giant jigsaw puzzle: Democratic committee assignments in the modern House*. Chicago: University of Chicago Press.

Shepsle, Kenneth A and Barry R Weingast. 1981. "Political preferences for the pork barrel: A generalization." *American Journal of Political Science* pp. 96–111.

Strøm, Kaare. 1998. "Parliamentary committees in European democracies." *The Journal of legislative studies* 4(1):21–59.

Volkens, Andrea, Pola Lehmann, Theres Matthieẞ, Nicolas Merz and Sven Regel. 2016. "The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2016b.".

Wang, Sida and Christopher D Manning. 2012. Baselines and bigrams: Simple, good sentiment and topic classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*. Association for Computational Linguistics pp. 90–94.

Yordanova, Nikoleta. 2009. "The rationale behind committee assignment in the European Parliament: Distributive, informational and partisan perspectives." *European Union Politics* 10(2):253–280.

Yu, Bei, Stefan Kaufmann and Daniel Diermeier. 2008a. "Classifying party affiliation from political speech." *Journal of Information Technology & Politics* 5(1):33–48.

Yu, Bei, Stefan Kaufmann and Daniel Diermeier. 2008b. Exploring the characteristics of opinion expressions for political opinion classification. In *Proceedings of the 2008 international conference on Digital government research*. Digital Government Society of North America pp. 82–91.