

Fiva, Jon H.; Nedregård, Oda; Øien, Henning

**Working Paper**

## Polarization in Parliamentary Speech

CESifo Working Paper, No. 8818

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Fiva, Jon H.; Nedregård, Oda; Øien, Henning (2021) : Polarization in Parliamentary Speech, CESifo Working Paper, No. 8818, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/232415>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Polarization in Parliamentary Speech

*Jon H. Fiva, Oda Nedregård, Henning Øien*

## **Impressum:**

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email [office@cesifo.de](mailto:office@cesifo.de)

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: <https://www.cesifo.org/en/wp>

# Polarization in Parliamentary Speech

## Abstract

We study political polarization in a parliamentary setting dominated by strong parties. In addition to examining polarization along the left-right dimension, we consider political divergence between legislators belonging to the same political bloc. Are politicians' background characteristics unimportant when parties have powerful tools to discipline their rank-and-file? We investigate this question using legislative speech from the Norwegian Parliament and recently developed techniques for measuring group differences in high-dimensional choices. Across the background characteristics we consider — gender, age, urbanicity, and class background — we document substantial differences in speech, even when comparing legislators from the same party bloc and policy committee. Our results illuminate how individual legislators shape policymaking in party-centered environments.

JEL-Codes: D720.

Keywords: political polarization, text analysis, penalized logistic regression.

*Jon H. Fiva*  
*Department of Economics*  
*BI Norwegian Business School*  
*Oslo / Norway*  
*jon.h.fiva@bi.no*

*Oda Nedregård*  
*Department of Economics*  
*BI Norwegian Business School*  
*Oslo / Norway*  
*oda.nedregard@bi.no*

*Henning Øien*  
*Norwegian Institute for Public Health*  
*Oslo / Norway*  
*henning.oien@fhi.com*

January 2021

We are grateful to Jack Blumenau, Ali Cirone, Gary Cox, Elias Dinas, Yvonne Markaki, Carlo Prato, Carlos Sanz, Jesse Shapiro, Dan Smith, Martin Søyland, and Janne Tukiainen, for helpful comments and suggestions on an earlier draft. We thank Tuva Værøy for help with data collection.

# 1. Introduction

Political polarization appears to be on the rise across many established democracies (Boxell, Gentzkow and Shapiro, 2020; Draca and Schwarz, 2020). These trends are driven partly by more polarized electorates electing extremists into office.<sup>1</sup> However, to understand the drivers of political polarization one also needs to consider the political supply side. What candidates emerge in equilibrium (Hall, 2019)? To what extent can party leaderships force rank-and-file members to toe the party line (Ban, Moskowitz and Snyder, 2016; Canen, Kendall and Trebbi, 2020)?

Most of the existing literature on political polarization focuses on candidate-centered electoral systems. In this paper, we consider proportional representation (PR) systems, where the election of individual candidates depends strongly on their rank position on the ballot. In such systems, party elites employ a variety of strategies that can be used to discipline their rank-and-file incumbents (see, e.g., Cirone, Cox and Fiva, 2021; Hazan and Rahat, 2010; Kam, 2009). Party elites can, for example, promise future safe nominations to the loyal, while threatening to deselect the recalcitrant. Does strong party discipline make individual legislators irrelevant? If not, what are the consequences for political polarization within- and across parties?

We study whether politicians' social ties and group identities matter in a parliamentary system dominated by strong parties: Norway. As a window into *individual* legislators' priorities, we rely on legislative floor speech. The main purpose of parliamentary debates is to function as a forum for communication where legislators advocate their policy positions to their own parties, other parties, and voters (Grimmer, 2013; Martin and Vanberg, 2008; Proksch and Slapin, 2015). Since there is a low probability of implementing policies that have not been discussed, policy is bounded by the parliamentary discourse (Motolinia, 2020; Schmidt, 2008). Intraparty speech diversity is therefore infor-

---

<sup>1</sup>Autor et al. (2020) find, for example, that exposure of local labor markets to increased foreign competition from China has contributed to rising political polarization in the United States. Several researchers have also documented the importance of a changing media landscape for political polarization (e.g., Allcott et al., 2020; DellaVigna and Kaplan, 2007; Martin and Yurukoglu, 2017).

mative about the extent to which individual legislators shape policymaking (Lauderdale and Herzog, 2016).

To quantify within-party group differences, we rely on the penalized regression method for high dimensional data developed by Gentzkow, Shapiro and Taddy (2019).<sup>2</sup> Because the number of words a legislator could choose is large relative to the total amount of speech we observe, many words are said mostly by one group or the other purely by chance. Naïve estimators interpret such noise as evidence of polarization. The estimator proposed by Gentzkow, Shapiro and Taddy (2019) controls for this finite-sample bias by applying a lasso-type penalty on key model parameters.

We focus on four dimensions describing politicians’ background: gender, urbanicity, age, and class. Inspired by Gentzkow, Shapiro and Taddy (2019), we define within-bloc polarization as the ease with which a neutral observer could infer a legislators’ identity from a single word uttered in Parliament. Our polarization measure quantifies heterogeneity between legislators’ speech, and hence differs somewhat from other measures of polarization, such as ideology scores (e.g., Barber, 2016). Because parties might allocate legislators strategically to committees based on their descriptive background, we include committee fixed effects in our analysis.<sup>3</sup> By additionally controlling for parliamentary session and political bloc fixed effects, we isolate group differences in speech for legislators who belong to the same political bloc and same committee in each session. If parties make it difficult for rebels to express their views on the parliamentary floor, as argued by Proksch and Slapin (2015), the observable differences we find should be interpreted as lower bound estimates of the real levels of polarization in speech.

As a benchmark to assess within-bloc polarization, we first quantify polarization *across* the political blocs that dominate Norwegian politics; the left-leaning social democratic

---

<sup>2</sup>Gentzkow, Shapiro and Taddy (2019) define partisan differences in speech in a given session of the US Congress by “the ease with which an observer who knows the model could guess a speaker’s party based solely on the speaker’s choice of a single phrase”. They find that party polarization in Congress has increased dramatically since the mid-1990s.

<sup>3</sup>Heath, Schwindt-Bayer and Taylor-Robinson (2005) document, for example, a gender bias in committee assignment across several Latin American countries. Women are assigned disproportionately to committees that focus on women’s issues and social issues, and are often underrepresented on committees that deal with economics or foreign affairs. We find a similar pattern in Norway.

camp vs. the right-leaning conservative camp. We document substantial bloc polarization in legislative speech, which is moderately increasing over our sample period. After hearing a one-minute speech in parliament, a neutral observer (with knowledge of the true model) has, on average, about a 59 percent chance of correctly guessing the bloc identity of the speaker. We find similar, but somewhat more muted, differences across the four background characteristics we consider. Nonetheless, our results clearly demonstrate that politicians' background characteristics are important to understand what issues legislators choose to raise in parliament.

We use data from the national election studies as a guide to interpret the differences in speech between background pairs (e.g., men vs. women). We find a close resemblance between the words that our model identifies as polarizing and the issues that separate different types of voters in the survey data. For example, among politicians, as in the electorate, women seem to care more about family and welfare policies than their male colleagues (in the same policy committee). Men, on the other hand, appear to care relatively more about fiscal policies than women.<sup>4</sup> Similarly, we find that politicians that reside in the countryside do talk more about agriculture and transfer policies than urban politicians.

Overall, our findings support the image of politicians as rational actors that advocate for policies that maximize the expected utility of their social group. We find, for example, that politicians whose fathers have a white-collar occupation indeed tend to talk more about economic and business-related issues, while politicians with blue-collar backgrounds tend to talk more about industry and employment-related topics. These results could be interpreted in two broad ways. First, it may be that parties, who have formidable tools to discipline their rank-and-file, *allow* politicians' own preferences to shine through. If voters have diverse policy preferences, a vote-maximizing party might want politicians

---

<sup>4</sup>Several scholars have previously documented similar general differences in legislative speech between men and women. See, for example, Bäck, Debus and Müller (2014); Clayton, Josefsson and Wang (2017); Osborn and Mendez (2010); Blumenau (2021) for studies of Sweden, Uganda, the United States, and the United Kingdom, respectively. Less is known about the other background characteristics that we consider, especially class background and age (Gulzar, 2021).

with a particular background characteristic to promote these issues in the public debate. This is consistent with the logic behind models of multiparty spatial competition (e.g., Cox, 1990) or spatial models of entry deterrence (e.g., Schmalensee, 1978). Second, it may be that parties are *unable* to fully discipline their rank-and-file, in line with the commitment issues highlighted in the citizen-candidate framework (Alesina, 1988; Besley and Coate, 1998; Osborne and Slivinski, 1996).<sup>5</sup>

Our study contributes to the literature in three ways. First, rather than focusing exclusively on polarization along the left-right ideological dimension, as in most of the existing literature, we consider political divergence between legislators belonging to the same political bloc. A key advantage of our study is that our research design does not rely on pre-specified categories to distinguish legislators with different background characteristics. This allows us to keep the door open for potentially unexposed dimensions of polarization. This is particularly useful when examining polarization between legislators of different age groups, urbanicity and social background. Although generational and class disparities, and the urban-rural divide, usually highlighted as important in understanding recent developments in electoral polarization (e.g., Norris and Inglehart, 2019), empirical evidence documenting how these characteristics shape the behavior of legislators is relatively scant. We hope that our approach can inspire other researchers to study polarization by using tools similar to those we apply in this study.

Second, we contribute to the broader literature on how individual legislators shape policymaking in party-centered environments. Several recent papers use policy outcome data combined with as-good-as-random variation in political representation to quantify such effects. The findings from these studies are mixed.<sup>6</sup> Our text-based approach

---

<sup>5</sup>Several empirical studies from candidate-centered electoral environments find, in line with the citizen-candidate framework, that politicians' social ties and group identities, e.g., gender (Chattopadhyay and Duflo, 2004), ethnicity (Pande, 2003), and geographic ties (Carozzi and Repetto, 2016), matter for public policy. The findings in this literature is not, however, unequivocal, (see, e.g., Ferreira and Gyourko, 2014).

<sup>6</sup>Baskaran and Hessami (2019) find that having an additional woman in local councils in Bavaria tends to increase public childcare provision and leads to more frequent discussions of childcare policy. Bagues and Campa (2021) find, however, no effect of candidate gender quotas on policy outcomes in the context of Spanish local governments. Hyttinen et al. (2018) find that one more councilor employed by the public sector increases public spending in Finnish municipalities. Using data from Norway, Fiva, Halse and Smith (2021) find that legislators' hometowns receive more mentions in legislative speeches



complements existing studies that focus on aggregate-level policy outcomes and provides a better understanding of the forces that drive policy formation.

Finally, we contribute to an emerging literature on intraparty politics in list-based electoral systems. Much of this literature focuses on how parties allocate nominations and valuable positions to their members (e.g., Buisseret et al., 2019; Cox et al., 2020; Folke, Persson and Rickne, 2016; Fujiwara and Sanz, 2020; Meriläinen and Tukiainen, 2018) and how this shapes party stability (e.g., Buisseret and Prato, 2021; Cirone, Cox and Fiva, 2021; Matakos et al., 2018). Cirone, Cox and Fiva (2021), for example, argue that party elites create career paths within the party partly because they want to increase legislative cohesion. Empirically, it is complicated to quantify the extent to which party leaders control their rank-and-file members. Roll call votes in parliamentary systems suffer from a number of problems that prevent them from forming a reliable basis for estimating individual legislators' ideal points (Peterson and Spirling, 2018; Schwarz, Traber and Benoit, 2017). We believe that our study, based on legislative speech, delivers important insights about intraparty dynamics in party-centered environments.

## 2. Empirical case: Norway 1981–2020

### 2.1 *Election system*

Norwegian national elections are held every fourth year in September (1981, 1985, ..., 2017) using closed-list proportional representation. Each parliamentary session starts in the first week of October.<sup>7</sup> Seats are allocated in two rounds. First, regular seats are allocated at the district level using the Modified Sainte-Laguë method. Second, adjustment seats are given to parties that are underrepresented nationally after the first-tier seats have been allocated, provided that those parties reach an electoral threshold of 

---

relative to other municipalities, but no clear evidence that these municipalities get any special benefits in terms of central-to-local redistribution.

<sup>7</sup>Below, we define parliamentary sessions by the year in which they ended. For example, the 2020 session, refers to the session starting at October 2, 2019, and ending October 1, 2020.

4% of the national vote count (Fiva and Smith, 2017).<sup>8</sup> Candidate nominations and rank positions are formally determined by party conventions at the electoral district level. In our sample period (1981–2020), the parliament consisted of 155–169 members (MPs).

## 2.2 *The left-right dimension*

Historically, the Norwegian policy space has been well represented by a left-right dimension (Strøm and Leipart, 1993), where the main political divide went between the left-leaning social democratic bloc and the right-leaning conservative bloc. The left-wing bloc consists of two parties, the *Socialist Left Party* (SV) and the *Labour Party* (A). The right-wing bloc is more fragmented, and consists of the *Conservatives* (H), the *Liberals* (V), *Christian Democrats* (KrF), and the *Progress Party* (FrP). The *Centre Party* (SP)) has traditionally sided with the conservative bloc, but formed a government together with A and SV in the 2005–2013 period. Because the bloc affiliation of SP is unclear, we exclude legislators from this party from our main empirical analyses.<sup>9</sup> During our 40-year sample period, Norway was governed by minority governments for about 29 years (see Appendix Table A.1).<sup>10</sup>

## 2.3 *Party discipline*

Voting against one’s party on a whipped vote is the ultimate act of defiance (Proksch and Slapin, 2015). Like in most parliamentary systems, intraparty cohesiveness in roll-call voting is extremely high in Norway. In the most recent parliamentary session, for example, the seven main parties were individually united in 96% of cases.<sup>11</sup> Generally,

---

<sup>8</sup>Adjustment seats were introduced in 1989. From 1989 to 2001, eight second-tier seats could be allocated to parties in any district. Since 2005, one party in each district is allocated a second-tier seat (hence, 19 adjustment seats in total).

<sup>9</sup>In addition to the seven main parties mentioned, four minor parties have been represented by a single MP in one or more election periods during our sample period (*Future for Finnmark*, the *Coastal Party*, the *Green Party*, and the *Red Party*).

<sup>10</sup>The governments with a majority in parliament in our sample period are: Willoch (H) Jun 1983 — Sep 1985, Stoltenberg (A) Oct 2005 — Oct 2013, and Solberg Jan 2019 — Jan 2020).

<sup>11</sup>The sample includes roll call votes recorded by the electronic voting device of the *Storting*, and therefore excludes unanimous and some near-unanimous decisions. Appendix Figure A.1 illustrate that when a party is *not* united, it is typically because a small fraction of legislators broke with the party line. Appendix Figure A.2 shows that party discipline has been stable at a high level for decades.

parties only allow legislators to break party ranks on issues of strong constituency interest (e.g., roads) or moral beliefs (e.g., abortion), and only when they do not threaten the standing of the government (Rasch, 1999).

## 2.4 *Parliamentary committees*

There are currently 12 standing committees (*fagkomiteer*) in the Norwegian Parliament. These have responsibility for the majority of parliamentary proceedings. Committee size varies from 11 to 18 members. Representatives are proportionally assigned to each committee, according to their party's size in the parliament. The exception is *The Standing Committee of Scrutiny and Constitutional Affairs*, where all parties are represented.<sup>12</sup>

Figure 1 shows the number of legislators from each party by the policy area of the standing committee. We choose to describe groupings by policy area rather than committee name, because some committees change names, merge, or split during our sample period (see Appendix Table A.2). In most instances, each policy area captures a single committee, but there are a few exceptions. For example, "Foreign Affairs" and "Defence" were two separate committees in the 1993–2009 period.

Figure 1 shows that the largest parties in each political bloc, the Labour Party and the Conservatives, are typically represented by several legislators in all policy areas.<sup>13</sup> Most of the other parties, however, are typically represented by a single legislator in each policy area.

## 2.5 *Parliamentary speech*

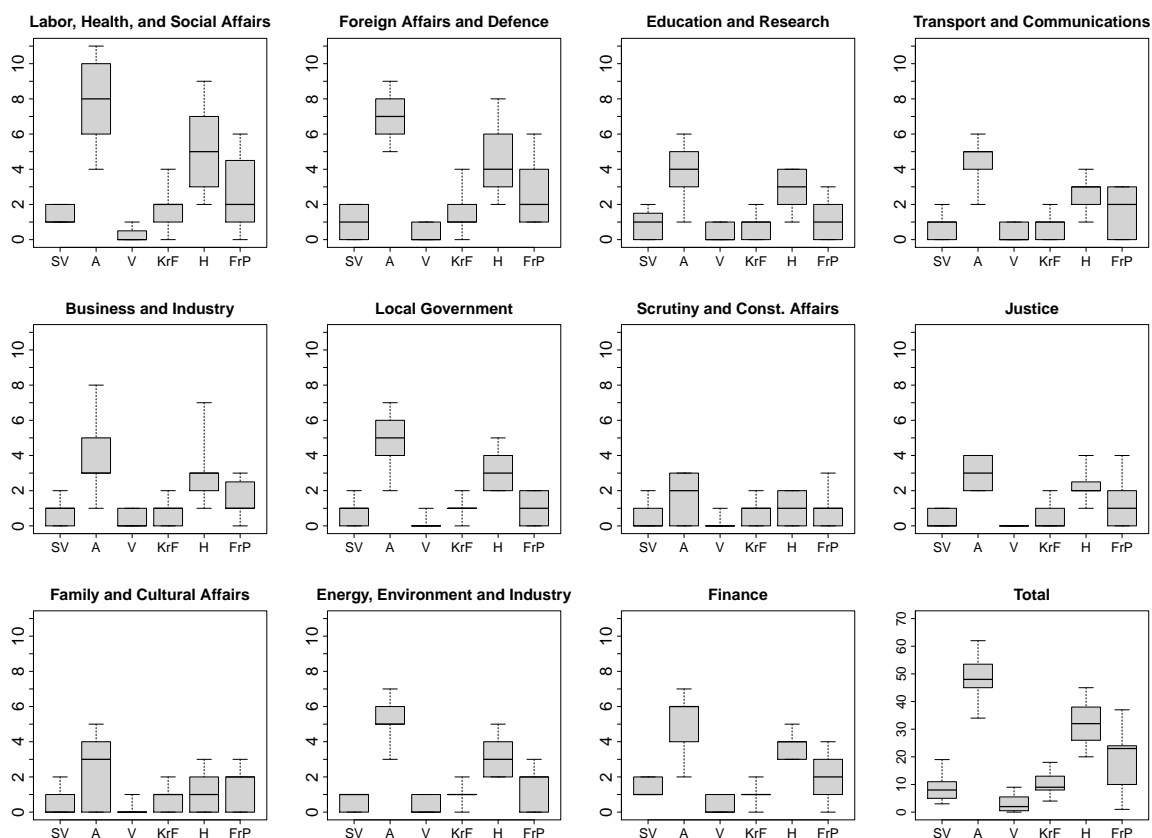
There are strict behavior rules in the Norwegian Parliament. All speeches must be addressed to the parliamentary president and should strictly concern the matter that is discussed. The tone should be formal and the audience is not allowed to call out or demon-

---

<sup>12</sup>If a party is not represented in all committees after the proportional assignment, the party can demand that its member in *The Standing Committee of Scrutiny and Constitutional Affairs* is also appointed to one of the other committees.

<sup>13</sup>Appendix Figure A.3 plots seat shares over time for each of the main parties and a residual "other" category.

Figure 1: Number of legislators from each party by policy area



*Note: This box-and-whisker plot shows the number of legislators from each main party by policy areas in the 1981–2020 period. Whiskers indicate the minimum and maximum values for each party. The Socialist Left Party (SV), the Labour Party (A), the Christian Democrats (KrF), the Liberal Party (V), the Conservative Party (H), and the Progress Party (FrP) constitute the main parties. Because some committees change names, merge, or split during the sample period we present descriptive statistics by policy area, rather than by individual committee. Appendix Table A.2 provides an overview of the committee structure in the Norwegian Parliament, and how committees map into policy areas. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days.*

strate other forms of rowdy disapproval or agreement. In contrast to many other parliamentary systems, such as Finland’s (Simola, 2020) and the United Kingdom’s (Proksch and Slapin, 2012), the speech length is strictly regulated by parliamentary rules in the Norwegian Parliament.<sup>14</sup>

MPs can ask questions of cabinet members in three ways, as explained by Søyland (2020). *Interpellations*, the most formal type of question in parliament, allow cabinet ministers one month time to prepare their response to written questions from MPs.<sup>15</sup> During *Ordinary Question Hour* (*Ordinær spørretime*; introduced in 1949), cabinet members answer written questions from MPs that are submitted at least six days prior to the debate (max three minutes). After the answer is given, the questioner and the responder are allowed two comments each of, at maximum, one minute. During *Oral Question Hour* (*Muntlig spørretime*; introduced in 1996) MPs may pose short oral questions for cabinet members to answer immediately (two minutes each).

## 3. Data

### 3.1 570,000 speeches over four decades

We use speeches from the Norwegian Parliament covering the 1981–2020 period to construct our main data set (N=565,737).<sup>16</sup> For the 2004–2020 period our data set includes time stamps. Figure 2 shows the distribution of speech length for this sub-sample. Speech

---

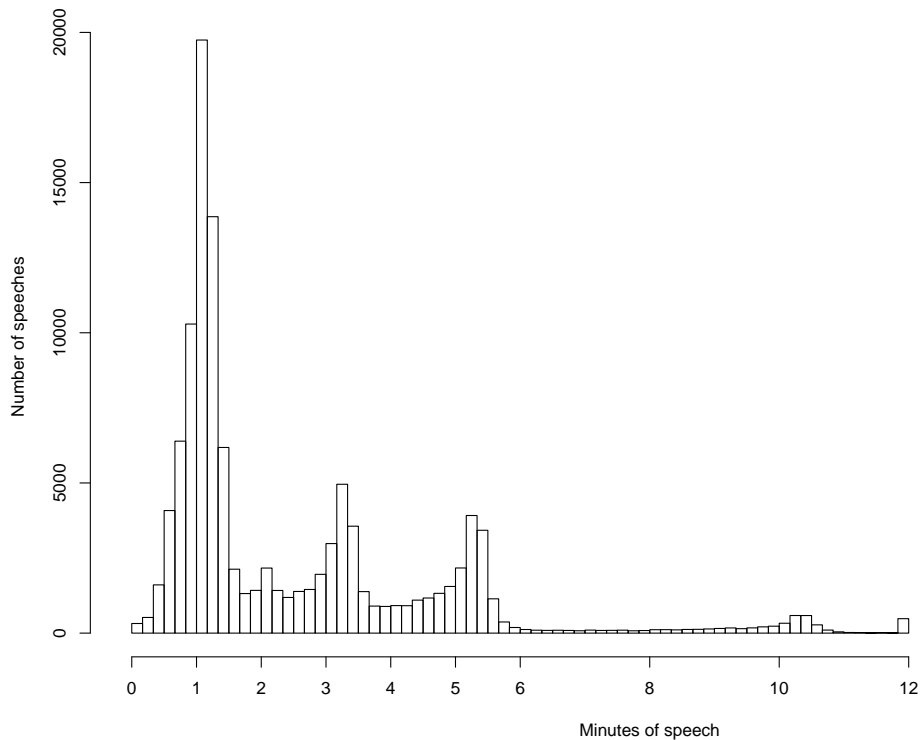
<sup>14</sup>The first speech of an ordinary debate (*Debattinnlegg*) is restricted to 15 minutes, while the second and third speeches are restricted to 10 and 3 minutes, respectively. Minor comments should not exceed 1 minute. Accounts by cabinet ministers (*redegjørelse*) should not exceed 1 hour. If the account is followed by a debate, one MP from each party is allowed 5 minutes to comment. MPs are not allowed to speak more than two times during a debate on each topic, but the parliamentary president can make exceptions. Examples of other countries that limit MP’s speaking time are Switzerland (Schwarz, Traber and Benoit, 2017) and Australia ([www.aph.gov.au/Parliamentary\\_Business/Chamber\\_documents/Senate\\_chamber\\_documents/standingorders/a00](http://www.aph.gov.au/Parliamentary_Business/Chamber_documents/Senate_chamber_documents/standingorders/a00)).

<sup>15</sup>In the ensuing legislative debate, the questioner is allowed 10 minutes to initiate the discussion, and the targeted cabinet member, other cabinet members, and the Prime Minister are allowed 10 minutes each to respond. The floor is open for other MPs to give one comment of no more than 5 minutes each.

<sup>16</sup>We are inspired by the *The Talk of Norway* (ToN) data set which covers parliamentary speech in the 1998 to 2016 period (N = 250,373) (Lapponi et al., 2018). We have extended this data set to cover October 1981 (the start of the 1982 parliamentary session) to March 2020.

length is measured as the time from the start of the speech to the beginning of the next speech, and therefore slightly overstates the actual speech length. The rules of conduct are clearly visible in the empirical distribution of speech length. There are clear spikes in the data just above one, three, five, and ten minutes.

Figure 2: Empirical Distribution of Speech Length



*Note: The figure shows the distribution of speech length in the 2004–2020 period. In this period, the data includes the timestamp at the beginning of every speech. Speech length is measured as the number of minutes from the start of the speech to the beginning of the next speech. Measured speech length slightly overstates the actual speech length by the amount of time it takes for the next speaker to start. However, because debate length is strictly controlled by the parliamentary rules of conduct, the time between speakers is generally minimal. This is clearly seen in the distribution. Speeches are clustered just above the one, three and five-minute mark. The bin size is set to ten seconds. Roughly 40 percent of the speeches do not have a timestamp. These speeches are mostly very short speeches about parliamentary procedure. We discard them. To improve graph readability we censor the speeches at 12 minutes.*

## Reading data

Deputy MPs substitute for MPs who are promoted to the cabinet or for some other reason are prevented from serving. To avoid having our results be affected by deputy MPs, who only have a limited period during which they are eligible to speak, we drop all speeches by deputy MPs (21,228 observations).

Since there are two official forms of written Norwegian, *bokmål* and *nynorsk*,<sup>17</sup> linguistic forms of words could be picked up as polarizing for reasons other than the dimensions that we want to capture. To overcome this bias we run a language identifier algorithm, which allows us to identify and exclude all speakers that predominantly use the minority language *nynorsk*, defined as having a majority of their speeches in this form (111 MPs). Subsequently, we remove remaining individual *nynorsk* speeches (25,763 observations).

We also drop speeches by presidents and vice presidents of the Parliament (170,157 observations), since these contain formalities and parliamentary proceedings that are of little relevance to the question we are studying. Lastly, we exclude party-independent MPs (54 observations) and MPs from minor parties (4125 observations). This is to restrict our analysis to parties that have existed for the entire time period that we are studying.

## Feature selection

Before standardizing the features, we eliminate names of all MPs and cabinet members in our sample. We then lemmatize all words to allow several versions of a word to be analyzed as one.<sup>18</sup> Lemmatization is better than stemming at discriminating between words with different meanings, and hence yields a more accurate image of speech.

To reduce the number of features to something manageable, a common first step is to strip out elements of the raw text other than words (Gentzkow, Kelly and Taddy, 2019). In line with this convention we remove all punctuation, numbers, symbols and parenthe-

---

<sup>17</sup>*Nynorsk* is used by a minority of the Norwegian population. Approximately 14 percent of Norwegian pupils had *nynorsk* as their main language in primary education in 2019/2020 (*Grunnskolenes Informasjonssystem* <https://gsi.udir.no/>).

<sup>18</sup>The lemmas are obtained using the Oslo-Bergen tagger (Johannessen et al., 2012).

ses. We keep words that are said more than five times in a parliamentary session, and eliminate words that are said less than twenty times across all sessions. Subsequently, we remove a set of extremely common words and filler words using a list of 462 stopwords.<sup>19</sup> We also eliminate 140 procedural words in parliamentary speech because they appear frequently and their use is likely not informative about the interparty differences that we wish to measure (Gentzkow and Shapiro, 2010; Gentzkow, Kelly and Taddy, 2019).<sup>20</sup> Next, we remove a list of words containing party names, party acronyms, parliamentary titles, and terms describing blocs of parties (131 words). This leaves us with a vocabulary of 20,866 unique lemmas. Appendix B shows how feature engineering and lemmatization affect two example speeches.

Because compound words are quite common in Norwegian, e.g., *velferdsstat* meaning “welfare state”, we rely on single words or unigrams as input into our analyses. Some previous work, such as Gentzkow, Kelly and Taddy (2019), rely on two-word phrases (bigrams).

## Document term matrix

We aggregate speech so that a document captures all speech by a given speaker in one parliamentary session. After cleaning, we have 578 unique legislators and 4 764 legislator-session observations. Appendix Table A.3 shows that, on average, a legislator speaks 46 times, and utters 4400 words in a session (after pre-processing).

### 3.2 *Characteristics of politicians*

As mentioned above, we consider within-bloc speech polarization for four background characteristics; gender, urbanicity, age, and class background. Most of this information is readily available from electoral lists and is organized by Fiva and Smith (2017). We manually supplement this data using biographies whenever necessary.

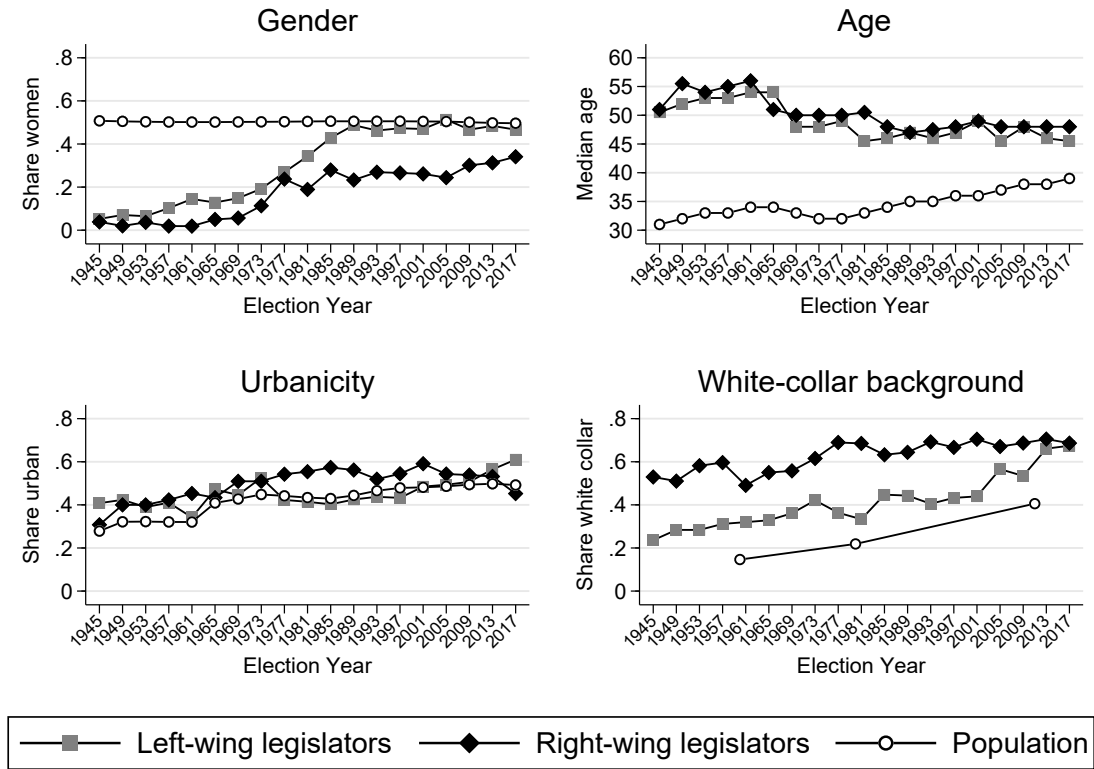
---

<sup>19</sup>We use the *Long Stopword List* from <https://www.ranks.nl/stopwords> (accessed March 30, 2020) translated from English to Norwegian.

<sup>20</sup>We collect the procedural words from <https://www.stortinget.no/no/Stottemeny/Ordbok/> (accessed March 30, 2020).



Figure 3: Descriptive Representation over Time



*Note: The figure shows how background characteristics of left-wing legislators (gray squares), right-wing legislators (black diamonds) and the population (white circles) evolved over the 1945-2017 period. The top-left panel plots the fraction of women in the legislature (population) by election year. The top-right panel plots the median age of legislators (citizens) by election year. The bottom-left panel plots the fraction of legislators (citizens) residing in municipalities with “town” status (using data from Fiva and Smith (2017)). The bottom-right panel plots the fraction of legislators whose father held a white-collar occupation (ISCO codes 1, 2, 3, 4, 5, and a residual category including self-employed and capitalists) (using biographical information from the Archive of Politicians) by election year. For this panel, we do not have a complete population counterpart readily available. Instead we rely on numbers from Modalsli (2017), which are based on father-son pairs identified in Norwegian censuses (1960 with fathers observed in 1910; 1980 with fathers observed in 1960; and 2011 with fathers observed in 1980).*

Figure 3 displays how descriptive representation in the Norwegian Parliament has evolved since 1945 for each political bloc. The top-left panel shows that the fraction of female legislators in the left-wing bloc increased dramatically during the 1970s and 1980s. Since about 1990, there is close to gender parity within the left-wing bloc. The fraction of women legislators in the right-wing bloc has increased more modestly during our sample period. In total, women are still underrepresented in parliament today.

An understudied aspect of political representation is the difference in age of the citizens and legislators (Gulzar, 2021). The top-right panel shows that Norwegian legislators are somewhat older than the general population (which includes children), but there is no clear difference across political blocs. In the most recent election, the median legislator age is 47.

To distinguish between politicians from urban and rural areas, we use each candidate’s municipality of residence (typically reported on the ballot). We code legislators (citizens) residing in a municipality with town status as urban. The bottom-left panel of Figure 3 shows that urban areas are represented in parliament in about proportion to their size in the population for both blocs. This is likely driven both by the electoral law, which distributes seats *across* districts partly based on their population size, and parties’ strong tendency to geographically balance their ticket *within* the district (Fiva, Halse and Smith, 2021).

To measure class background, we rely on information about fathers’ occupation from the *Archive of Politicians* at the Norwegian Centre for Research Data.<sup>21</sup> Using these biographical data, we classify fathers’ occupation using ISCO-08 (International Standard Classification of Occupations) codes. In our main analysis, we measure class background using two broad categories: Politicians whose father held a “blue-collar occupation” (ISCO codes 6–9; including farmers) and politicians whose father held a “white-collar occupation” (ISCO codes 1–5; including a residual “other” category). The bottom-right panel of

---

<sup>21</sup>An alternative measure of class background could be based on politicians’ own pre-office occupation. However, because most MPs come from white-collar jobs, we quickly run into problems with statistical precision when analyzing these data based on the methods presented below.

Figure 3 contrasts the development in legislators’ social background with corresponding numbers from Modalsli (2017), based on father-son pairs identified in Norwegian population censuses (1960 with fathers observed in 1910; 1980 with fathers observed in 1960; and 2011 with fathers observed in 1980). There is a positive trend in the share of fathers having a white-collar occupation for both left-wing legislators, right-wing legislators, and the general (prime-age male) population. The substantial level-difference between the three curves, suggest that even in the comparatively egalitarian case of Norway, elected politicians are a privileged elite.

In Appendix Table A.4 we present descriptive statistics for the speech data by legislator background characteristics. On average, women speak somewhat less than men (40 versus 50 speeches per session). We find similar average differences for the other background characteristics: the young speak slightly more than the old; urban legislators speak slightly more than rural legislators; and legislators with white-collar background speak slightly more than legislators with blue-collar background.

## 4. Methods

To quantify differences in speech patterns by legislators with different background characteristics, we build on the methods developed by Gentzkow, Shapiro and Taddy (2019). While Gentzkow, Shapiro and Taddy (2019) study differences in political speech *across* the two parties in the United States Congress, we are primarily interested in quantifying differences across legislators belonging to the *same political bloc*. In this section we explain how we extend the framework of Gentzkow, Shapiro and Taddy (2019) to achieve this goal.

### 4.1 *Measurement and estimation of within-bloc polarization*

Gentzkow, Shapiro and Taddy (2019) specify a multinomial model of speech where choice probabilities,  $\mathbf{q}_t^{D_i}(x_{it})$ , defined over the  $J$  available words for speaker  $i$  in session  $t$ , vary

by party ( $D_i \in \{Republican, Democrat\}$ ).<sup>22</sup> In our application,  $D_i$  is an individual bivariate characteristic of speaker  $i$  (e.g,  $D_i \in \{Woman, Man\}$ ). To estimate the choice probabilities for each background characteristic, we run separate multinomial logistic regressions where

$$q_{jt}^{D_i}(x_{it}) = \frac{e^{u_{ijt}}}{\sum_j e^{u_{ijt}}}, \quad (1)$$

$$u_{ijt} = \alpha_{jt} + x'_{it}\gamma_{jt} + \phi_{jt}1_{i \in \{D_i=1\}}.$$

Our coefficient of interest is  $\phi_{jt}$  which measures the difference in the propensity to use word  $j$  in session  $t$  by  $i \in \{D_t = 1\}$ . Because we are interested in measuring whether legislative speech is distinguishable according to legislators' background characteristics within political bloc over time, we include bloc and session fixed effects in  $x_{it}$ . In addition, we include committee fixed effects to rule out that potential differences are driven by allocation of committee membership.<sup>23</sup>

To quantify the divergence between  $q_t^{\{D_i=1\}}$  and  $q_t^{\{D_i=0\}}$ , we define polarization at  $\mathbf{x}$  as

$$\pi_t(\mathbf{x}) = \frac{1}{2}\mathbf{q}^{\{D_i=1\}}(\mathbf{x})\boldsymbol{\rho}_t(\mathbf{x}) + \frac{1}{2}\mathbf{q}^{\{D_i=0\}}(\mathbf{x})(1 - \boldsymbol{\rho}_t(\mathbf{x})), \quad (2)$$

where  $\rho_{jt}(x) = \frac{q_{jt}^{\{D_i=1\}}(x)}{q_{jt}^{\{D_i=1\}}(x) + q_{jt}^{\{D_i=0\}}(x)}$ , and is the posterior probability a neutral observer assigns to  $\{D_i = 1\}$  after hearing word  $j$ . The first term on the right-hand side of equation (2) is the product of this posterior probability, the propensity to use word  $j$  at  $\{D_i = 1\}$ , and the prior of  $\{D_i = 1\}$ . The second term on the right hand side is

<sup>22</sup>The multinomial model of speech can be described as follows (Gentzkow, Shapiro and Taddy, 2019): The speech by legislator  $i$  in session  $t$  can be represented by a  $J$ -vector of word counts,  $\mathbf{c}_{it}$ . The element  $j$  in  $\mathbf{c}_{it}$  is then the number of times legislator  $i$  used the word  $j$  in session  $t$ . The model represents a simplified view of speech generation, and assumes that the speeches are generated by independent draws from a multinomial distribution. The generation of speech  $\mathbf{c}_{it}$  can then be represented by a vector of choice probabilities,  $\mathbf{q}_{it}^{D_i} = \{q_{it1}, \dots, q_{itJ}\}$ , defined over the  $J$  available words in the parliamentary corpus, i.e., legislator  $i$  chooses  $\mathbf{c}_{it}$  of size  $m_{it} = \sum_j \mathbf{c}_{it}$  by drawing  $m_i$  words according to  $\mathbf{q}_{it}^{D_i}$ . The model can be summarized as follows:  $\mathbf{c}_{it} \sim MN(m_i, \mathbf{q}_{it}^{D_i})$ .

<sup>23</sup>If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days to construct the committee fixed effects.

the probability of  $\{D_i = 0\}$  after hearing word  $j$  times the propensity to use word  $j$  for  $\{D_i = 0\}$ . Polarization, as defined in equation 2, is the expected posterior at  $\mathbf{x}$ . The mean probability at  $\mathbf{x}$  of guessing the correct value of  $D_i$  in session  $t$  after hearing word  $j$  weighted by the propensity of using word  $j$ . The average polarization, which is the value of polarization we report in the results section, is the average of Equation (2) across the values of  $\mathbf{x}$ ,  $\bar{\pi}_t$ .<sup>24</sup>

The model is estimated using the distributed multinomial regression (*dmr*) developed in Taddy (2015). This method approximates the logit likelihood implied in (1) by  $J$ -independent Poisson likelihoods.<sup>25</sup> In high-dimensional speech data, small sample bias is a problem because the number of words in the vocabulary is large relative to the number of speakers. As a consequence, many words are said mostly by one party or another purely by chance. Without adjusting for this bias, the finite-sample noise will be attributed to polarization. We use the preferred method in Gentzkow, Shapiro and Taddy (2019) to account for small sample bias by including a Lasso penalty on the coefficient of interest.

The estimator is then given by the following minimization problem<sup>26</sup>:

---

<sup>24</sup> $\bar{\pi}_t = \frac{1}{N_t} \sum_i \pi_i(x_{it})$ , where where  $N_t$  is the number of legislators in session  $t$ .

<sup>25</sup>High-dimensional choice models can be computationally infeasible to estimate using traditional methods. In *dmr* the logit likelihood is approximated by a Poisson likelihood assuming that the word counts are independently Poisson distributed with mean  $e^{\log(m_{it})+u_{ijt}}$ , where  $m_{it}$  is the speech length of speaker  $i$  in session  $t$ . Taddy (2015) shows that these assumptions implies that the negative log-likelihood function for every word  $j$  is proportional to:

$$l(\alpha_{jt}, \gamma_{jt}, \phi_{jt}) = \sum_t \sum_i [m_{it} \exp(\alpha_{jt} + x'_{it} \gamma_{jt} + \phi_{jt} 1_{i \in \{D_t=1\}}) - c_{ijt}(\alpha_{jt} + x'_{it} \gamma_{jt} + \phi_{jt} 1_{i \in \{D_t=1\}})]$$

The estimation can therefore be done by fitting  $J$ -independent Poisson regressions, which is a huge computational advantage because the regressions can be done entirely in parallel.

<sup>26</sup>The Lasso penalty,  $\lambda_j |\phi_{jt}|$ , shrinks the polarization coefficients toward zero, and some of the coefficients are set exactly equal to zero. This yields a sparse solution, which limits the problem of small sample bias (Gentzkow, Shapiro and Taddy, 2019; Gentzkow, Kelly and Taddy, 2019). The  $\lambda_j$  determines the degree of shrinkage. The optimal value of  $\lambda_j$  is found by starting with a high  $\lambda_j$  that shrinks all the polarization coefficients to zero, and then subsequently decreasing  $\lambda_j$  until a corrected version of the Akaike's information criterion is minimized (Taddy, 2015). As recommended by Gentzkow, Shapiro and Taddy (2019), we also include a constant penalty  $\psi = 10^{-5}$  on the other coefficients, which is necessary to ensure convergence. The minimization of the penalized log-likelihood function in 3 is done by using the *dmr* package in R (Taddy, 2015).

$$\hat{\alpha}_{jt}, \hat{\gamma}_{jt}, \hat{\phi}_{jt} = \arg \min_{\alpha_{jt}, \gamma_{jt}, \phi_{jt}} l(\alpha_{jt}, \gamma_{jt}, \phi_{jt}) + N \sum_t [\psi(|\alpha_{jt}| + \|\gamma_{jt}\|_1) + \lambda_j |\phi_{jt}|] \quad (3)$$

## 4.2 Magnitudes

The average polarization  $\bar{\pi}_t$  represents the posterior that a neutral observer assigns to a speaker’s true identity (e.g., gender) after hearing a single word. Political opinions are not usually expressed in single words. To represent the polarization of typical speeches we therefore compute the polarization by number of words. To do this we follow Gentzkow, Shapiro and Taddy (2019), and quantify the informativeness of speech by speech length and session using Monte Carlo simulations. For each speaker  $i$  and session  $t$ , we draw words according to the estimated choice probabilities in Equation (2) and compute the average polarization for each session and length of speech.

## 4.3 Validation

As previously mentioned, the number of words a legislator could choose is large relative to the total amount of speech we observe. As a consequence, many words are said mostly by one type of legislator purely by chance. The estimator we use controls for this bias by applying a lasso penalty (see Equation (3)), but is still biased in finite samples.

To quantify the bias in finite samples, we rely on a permutation test in which we randomly reassign  $D_i \in \{0, 1\}$  to speakers and then re-estimate the model on the resulting data. We do this 100 times. In such “random” series,  $q_t^{\{D_i=1\}} = q_t^{\{D_i=0\}}$  by construction, so the true value of  $\pi_t(\mathbf{x})$  is 1/2 in all  $t$ . The deviation from 1/2 provides a valid measure of finite-sample bias under the permutation.

# 5. Results

The Norwegian Parliament, like almost all other parliaments across the world, is far from a mirror image of the population it represents (see Figure 3). To what extent

does descriptive representation matter for the parliamentary debate? Do women and men, representing the same party on the same committee, choose different words when speaking in parliament?

Before analyzing polarization across groups of *legislators*, we take a step back and ask: What are the key issues that separate *voters* in our sample period? To answer this question we use data from the ten waves of the *Norwegian National Election Survey* (1981-2017) ( $N = 20,303$ ).

### 5.1 *Voter surveys*

The top-left panel of Figure 4 reports coefficients from a linear probability model where the dependent variable is equal to one (zero) if the survey respondent is affiliated with a left-wing (right-wing) party. The independent variables capture whether the survey respondent mentions the relevant policy area when asked to name one or two issues which had particular influence on how they voted. This panel shows that left-wing voters, relative to right-wing voters, consider redistribution, employment, and childcare, to be important policy areas. Right-wing voters, on the other hand, care relatively more about the economy ('business', 'taxation', 'inflation', and 'economy'), family issues ('abortion', 'family'), immigration, and defense.

There are also considerable differences across voter background characteristics, although, naturally, they are less pronounced than for party affiliation.<sup>27</sup> Men, relative to women, are concerned with many of the same issues as survey respondents affiliated with right-wing parties. The main exception is abortion, which female respondents are much more likely to mention as one of their two key policy areas. Women seem to care relatively more about welfare policies ('children', 'elder care', 'health care' and 'education').

Like women, the young care about children and education, but also nuclear energy and the environment. The old care relatively more about pensions and eldercare.

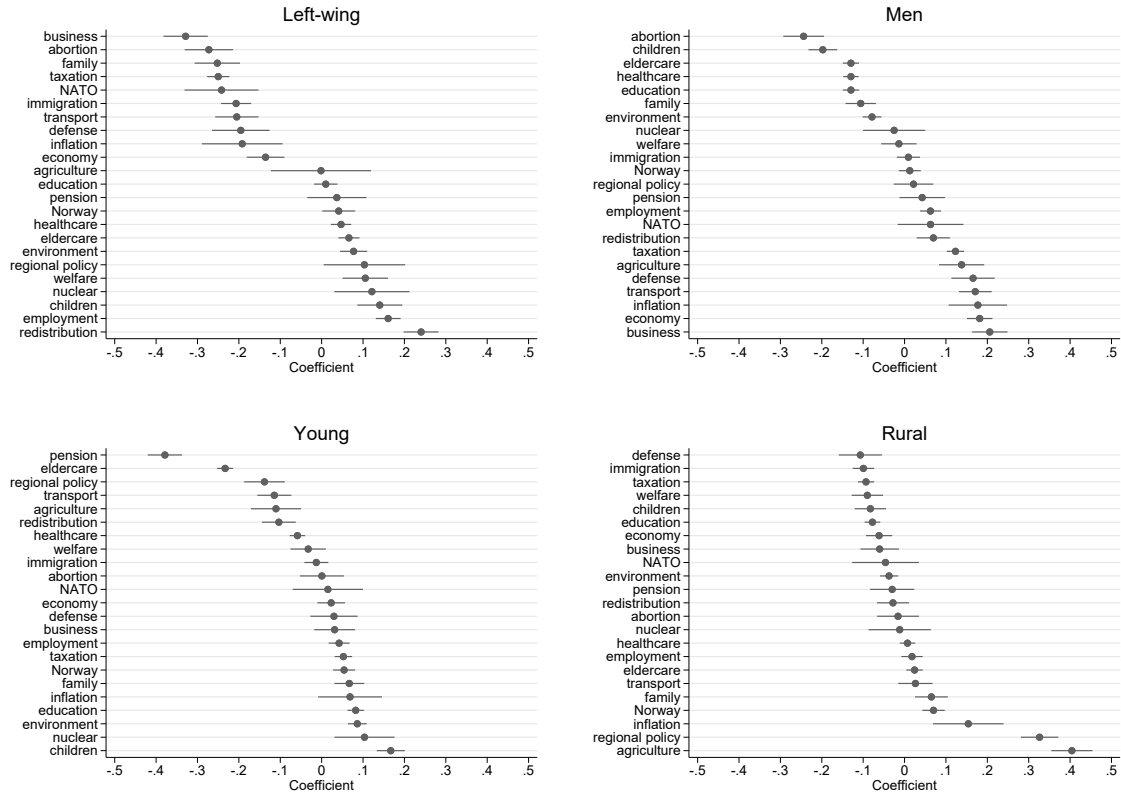
Voters residing in rural areas differ markedly from voters residing in urban areas when

---

<sup>27</sup>The  $R^2$  in the party affiliation specification is 0.11. For the background characteristics specifications, the  $R^2$  are 0.08, 0.06, and 0.04.

it comes to regional policy and agriculture, while many of the other differences are more muted.<sup>28</sup>

Figure 4: Voter preferences measured in surveys: 1981–2017



*Note: This figure reports coefficients and corresponding 95% confidence intervals for four linear probability models. The dependent variables are equal to one if the survey respondent is (i) affiliated with one of the two left-wing parties (rather than one of the four right-wing parties) ( $N = 10,237$ ), (ii) male ( $N = 20,303$ ), (iii) young ( $< 47$  years) ( $N = 20,303$ ), and (iv) residing in municipality without town status (“rural”) ( $N = 19,444$ ). The independent variables are dummies capturing whether the survey respondent mentions the policy area when asked the following question “Can you name one or two issues which had particular influence on the way you voted?”. Data from the National Election Survey 1981–2017.*

## 5.2 Across-bloc polarization

The left-hand panel of Figure 5 plots the average polarization (defined by Equation (2)) across party blocs for each year in our sample period. The gray shaded area represents the

<sup>28</sup>The survey does not include direct questions about class background (or parents occupation). For this reason, we cannot explore the fourth background dimension that we consider in section 5.3. Appendix Figure A.4 reports results when splitting our survey sample in two: 1981–1997 and 2001–2017.



average polarization in hypothetical data in which each speaker’s party bloc is randomly assigned with the probability that the speaker is right-wing (left-wing). The upper and lower bounds on the light gray shaded area corresponds to the 5th and the 95th highest polarization scores across the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th highest polarization scores. For each year in our sample period, the observed polarization lies outside the placebo distribution, providing strong statistical evidence of bloc polarization in legislative speech.

In the first two decades of our sample period, the estimated  $\bar{\pi}$  falls in the range 0.502 – 0.503, but then starts trending upwards.<sup>29</sup> The year 2019 is the year with the highest estimated polarization, with a  $\bar{\pi}$  of 0.506. In most of that parliamentary session, Erna Solberg led a majority center-right government (Appendix Table A.1). Another salient jump in estimated polarization occurs after the majority left-wing government (*Stoltenberg II*) replaced the minority center-right government (*Bondevik II*) after the 2005 election.

When interpreting Figure 5, the reader should keep in mind that  $\bar{\pi}$  is the posterior that a neutral observer expects to assign to a speaker’s true party bloc after hearing a *single word* from the vocabulary (defined in section 3.1). As a rough comparison, Gentzkow, Shapiro and Taddy (2019) report a  $\bar{\pi}$  of about 0.502 – 0.504 for the 1870–1990 period in the United States Congress, which later increases to about 0.510. Simola (2020) finds lower levels of polarization in the Finnish Parliament. For the period after 2000, she finds a  $\bar{\pi}$  of about 0.502 – 0.504.<sup>30</sup> In section 5.4, we consider the informativeness of legislative speech by speech length.

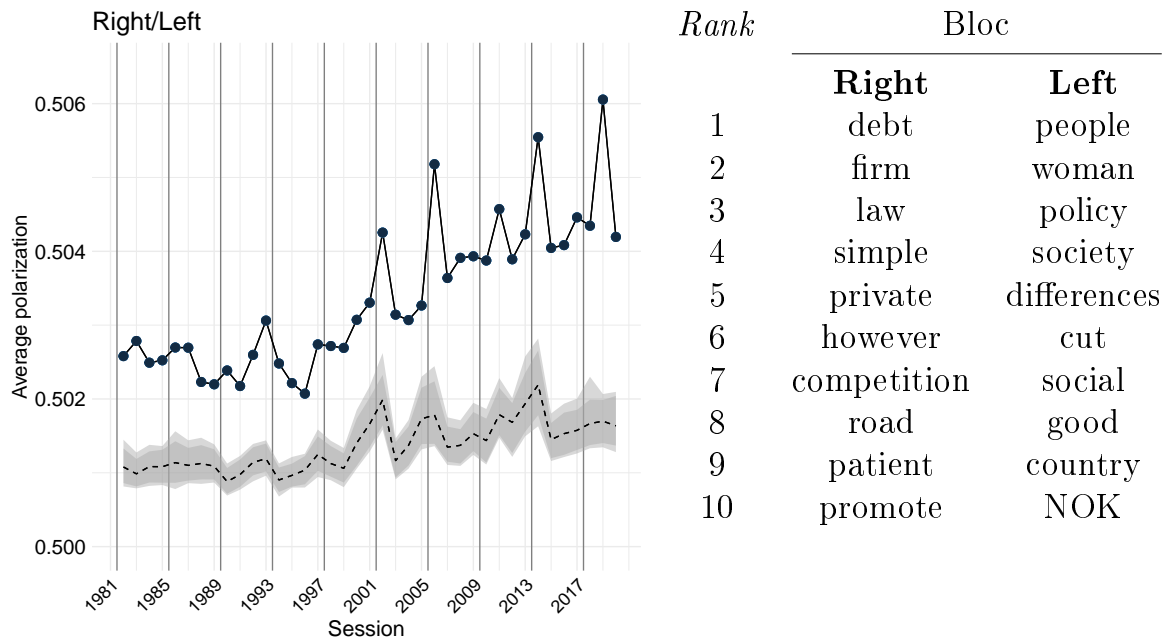
The right-hand panel of Figure 5 shows the 10 most polarizing words for each bloc across the 1981–2020 period. As in Gentzkow, Shapiro and Taddy (2019), we find that the top words align closely with the policy positions and narrative strategies of parties. Right-

---

<sup>29</sup>Boxell, Gentzkow and Shapiro (2020) measure time trends in affective polarization across nine OECD countries. They find that all countries *except* Germany, Norway, and Switzerland exhibit a positive linear trend after 2000.

<sup>30</sup>There are several reasons why our estimates are not directly comparable to the ones reported by Gentzkow, Shapiro and Taddy (2019) and Simola (2020). For example, both of these studies have available more text in each session than we do and they rely on bigrams rather than unigrams.

Figure 5: Across-bloc polarization over time



*Note: In the left-hand panel the black points correspond to the average bloc polarization of speech for each session in the period 1981–2020 after controlling for legislators’ committee assignment. The bars indicate elections. The gray shaded area represents the average polarization in hypothetical data in which each speaker’s party bloc is randomly assigned with a probability that the speaker is right-wing. We construct 100 hypothetical data sets and compute the average polarization in each session. The upper and lower bounds of the light gray shaded area correspond to the 5th and the 95th highest polarization scores across the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th highest polarization scores. The dashed line corresponds to the mean polarization for each session across the distribution of placebo estimates. In the right-hand panel we provide a list of the 10 words with the highest relative utility for each bloc.*

leaning legislators, like right-leaning voters, appear concerned about the performance of the private sector. Our model identifies “firm”, “private” and “competition” as key right-leaning words. Left-leaning legislators, on the other hand, focus on “people”, “women”, and “society”. Some key policy issues for left-leaning voters (established in Figure 4), such as redistribution (“distribution”, “fair”, “inequality”), employment (“job”, “work life”, “employee”) and childcare (“kindergarten”), do not feature in the top-ten list, but can be found in the top-fifty (see Appendix Table A.5).

### 5.3 Within-bloc polarization

Figure 6 presents within-bloc polarization in parliament over time across four background dimensions: gender, age, urbanicity, and class background. In almost all parliamentary

sessions, we can rule out that observed differences in speech are driven by random variations, since observed differences (black points) fall outside the placebo distribution (gray shaded area). The level of within-bloc polarization of approximately 0.502, is comparable, or perhaps slightly below, the across-bloc polarization that we find before the turn of the century.

Table 1 also shows that the words that separate background pairs (belonging to the same party and committee) reflect meaningful disparities between the groups. For example, female legislators appear to care relatively more about family policies, social policies and schooling (see also Appendix Table A.6).<sup>31</sup> Rural legislators talk about regional policy (‘municipalities’, ‘district’), agriculture, and sparsely populated areas like *Finnmark*. Urban legislators appear more concerned about the challenges that big cities, such as *Oslo*, face (see also Appendix Table A.7).

We find that politicians whose fathers have a white-collar occupation indeed tend to talk more about economic and business-related issues, while politicians with blue-collar backgrounds tend to talk more about industry and employment-related topics.<sup>32</sup> These findings align well with existing studies from the United Kingdom (e.g., O’Grady, 2019) United States (e.g., Carnes, 2013) and Latin America (e.g., Carnes and Lupu, 2015) which show that legislators’ social class shapes attitudes, values, and behavior in office.

For legislators’ age, there is no striking correspondence between the top-ten words reported in Table 1 and the survey evidence from Figure 4. However, based on the top-fifty most polarizing words for the age dimension (see Appendix Table A.8), we clearly see that young legislators talk more about childcare (“children”, “kindergarten”) and education (“school”, “teacher”, “student”). Older legislators, on the other hand, appear to be more concerned about health care (“patient”, “hospital”). However, we see no evidence, based

---

<sup>31</sup>In a survey of the literature, O’Brien and Piscopo (2019) (p.54) argue that scholars typically categorize women’s interest as those (i) issues that directly affect women *as women* (e.g., reproductive health), (ii) issues connected to women’s traditional role as caregivers (e.g., children), and (iii) issues tied to the social sphere more broadly (e.g., health care and education).

<sup>32</sup>There is also a clear tendency for blue-collar politicians to raise issues related to agriculture and regional policy. On the other hand, white-collar politicians appear to be relatively more concerned about international issues, such as development aid, multilateral organizations, and refugees (Appendix Table A.9).

on the top-fifty lists, that older legislators talk more about pensions, nor that the young talk more about the environment, as one might have expected after seeing the survey evidence.<sup>33</sup>

For gender, age, and urbanicity, we detect no clear trends in within-party polarization over time. Class background, however, seem to be increasingly polarizing over time (bottom-right panel of Figure 6).

#### 5.4 *Magnitudes*

In Figure 7, we plot the average expected posterior across two time periods, by the length of speech, for each dimension we consider.<sup>34</sup> In the first half of our sample period (1982-2000), we find that the probability that a neutral observer correctly guesses the speaker’s bloc affiliation is about 57 (64) percent after hearing one minute (three minutes) of speech.<sup>35</sup> In line with Table 5, we find evidence suggesting that it becomes easier to predict a legislator’s bloc affiliation over time. In the second half of our sample period (2001–2020), we estimate that the probability that a neutral observer correctly guesses the speaker’s bloc affiliation is about 61 (70) percent after hearing one-minute (three minutes) of speech. When interpreting these numbers one should keep in mind that parliamentary speeches might reflect more cohesion within a party than actually exists because party leaders make it difficult for rebels to express their views on the floor (Proksch and Slapin, 2015).

We estimate that the probability of correctly guessing the speaker’s gender, age, urbanicity status, or background category, is similar in magnitude to the estimated across-bloc polarization in the first half of our sample period. A one-minute speech gives a neutral observer about 56–58 percent chance of correctly guessing the relevant back-

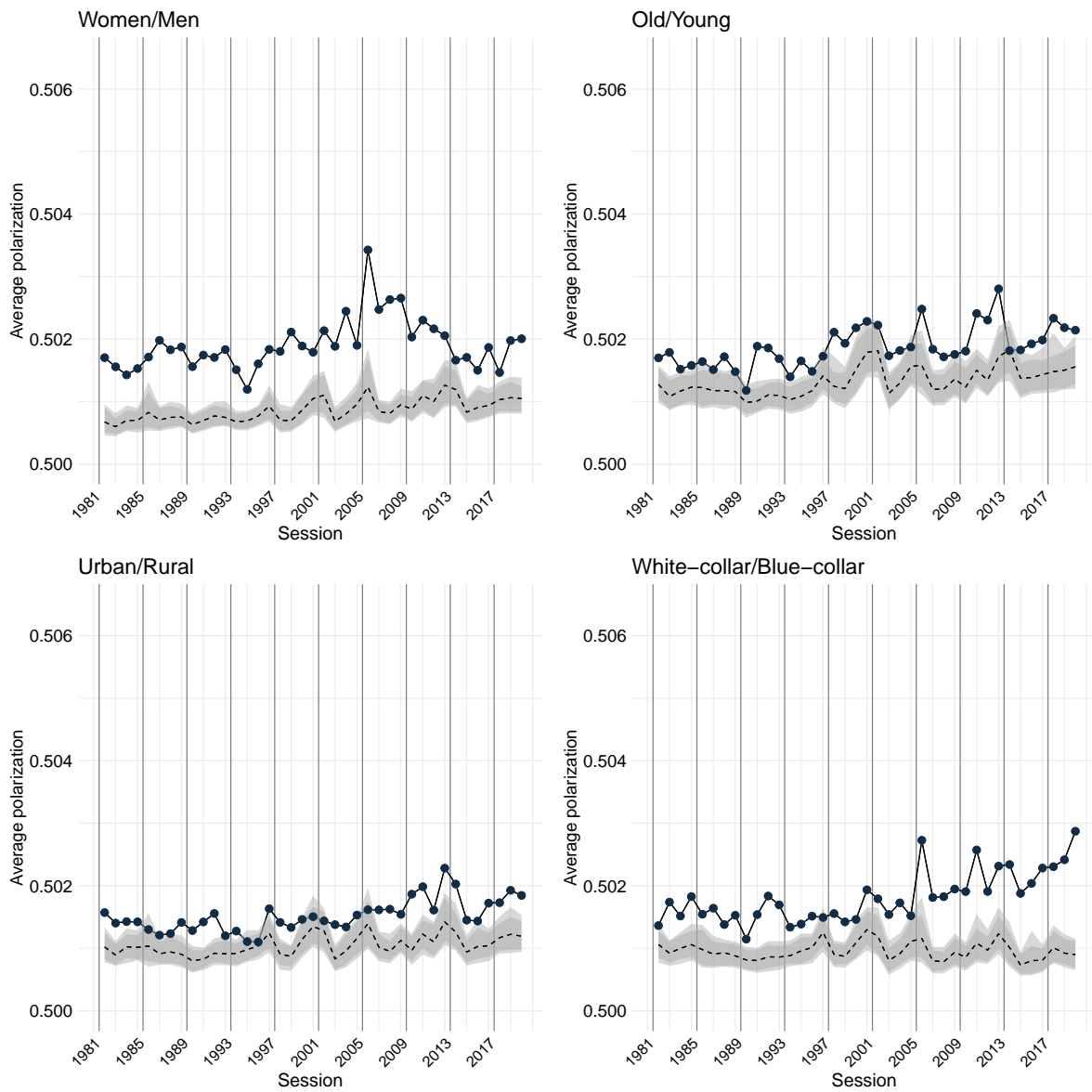
---

<sup>33</sup>The existing empirical evidence on the effect of legislators’ age is relatively scarce. Poole (2007) found that US Congress members’ voting records demonstrate a high degree of continuity across a politician’s lifetime suggesting that age is not important.

<sup>34</sup>In Appendix Figure A.5 we show the results for individual sessions.

<sup>35</sup>After pre-processing our data, one minute of speech is roughly 36 words. To calculate the median number of words per minute of speech, we use data from the 2004–2020 period, which, as previously mentioned, includes a time stamp for each speech. Before pre-processing, one minute of speech is roughly 150 words (see section 3.1 and Appendix B).

Figure 6: Within-bloc polarization over time



*Note: This figure displays polarization of legislative speech for four dimensions (given in the sub-panel headings) in the period 1981–2020, controlling for bloc and the legislator’s committee assignment. The models are estimated separately for each background characteristic. Cabinet members are treated as a separate committee. The black points correspond to the average polarization of speech in each session and the bars indicate elections. The gray shaded area represents the average polarization in hypothetical data in which each speaker’s identity is randomly assigned. We construct 100 hypothetical data sets and compute the average polarization in each session. The dashed line corresponds to the mean polarization for each session across the distribution of placebo estimates. The upper and lower bounds of the light gray shaded area correspond to the 5th and the 95th highest polarization scores across the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th highest polarization scores.*

Table 1: Most polarizing words for each background dimension

<i>Rank</i>	Gender		Age	
	<b>Women</b>	<b>Men</b>	<b>Old</b>	<b>Young</b>
1	children	Norwegian	collaboration	Norway
2	woman	relations	Nordic	policy
3	labour	debt	debt	party
4	parent	party	employment	choose
5	young	political	mention	people
6	kindergarten	lay	emphasize	school
7	measure	decrease	development	money
8	child prot. service	association	country	Norwegian
9	family	Norway	billion	job
10	municipality	post	area	trade

<i>Rank</i>	Urbanicity		Father's occupation	
	<b>Urban</b>	<b>Rural</b>	<b>White</b>	<b>Blue</b>
1	Norwegian	municipality	Norway	debt
2	Oslo	children	children	view
3	Norway	relations	policy	emphasize
4	unequal	district	school	positive
5	director	agriculture	Norwegian	treatment
6	self	nok	kindergarten	situation
7	international	service	pupil	district
8	city	Finnmark	teacher	self
9	police	opportunity	of course	view
10	concern	industry	gladly	agriculture

*Note: This table displays the most polarizing words of legislative speech for four dimensions: gender, age, urbanicity, and class. Each model is estimated separately and controls for the speaker's party and committee assignment.*

ground characteristic. There is some indication that within-polarization is increasing over time for all four background characteristics, but this tendency is weaker than for across-bloc polarization. Consistent with Table 6, we find the increase over time to be most pronounced for occupational background.

## 5.5 *Sensitivity checks*

Our baseline analyses suggest that legislator background characteristics matter for legislative speech. In the following we explore the sensitivity of this result to various modelling choices.

First, we repeat our baseline analyses without controlling for party bloc affiliation. In other words, we compare legislators with different background characteristics (from the same committee) both across and within blocs. If parties orchestrate political speech, we expect that women who are particularly distinctive in talking about social issues will not differ much from the men in the same party, but they will differ considerably from the men in the opposite bloc.<sup>36</sup> One might also expect that social background would appear more polarizing in such analyses because parties traditionally recruit candidates from different pools of people (see Figure 3).

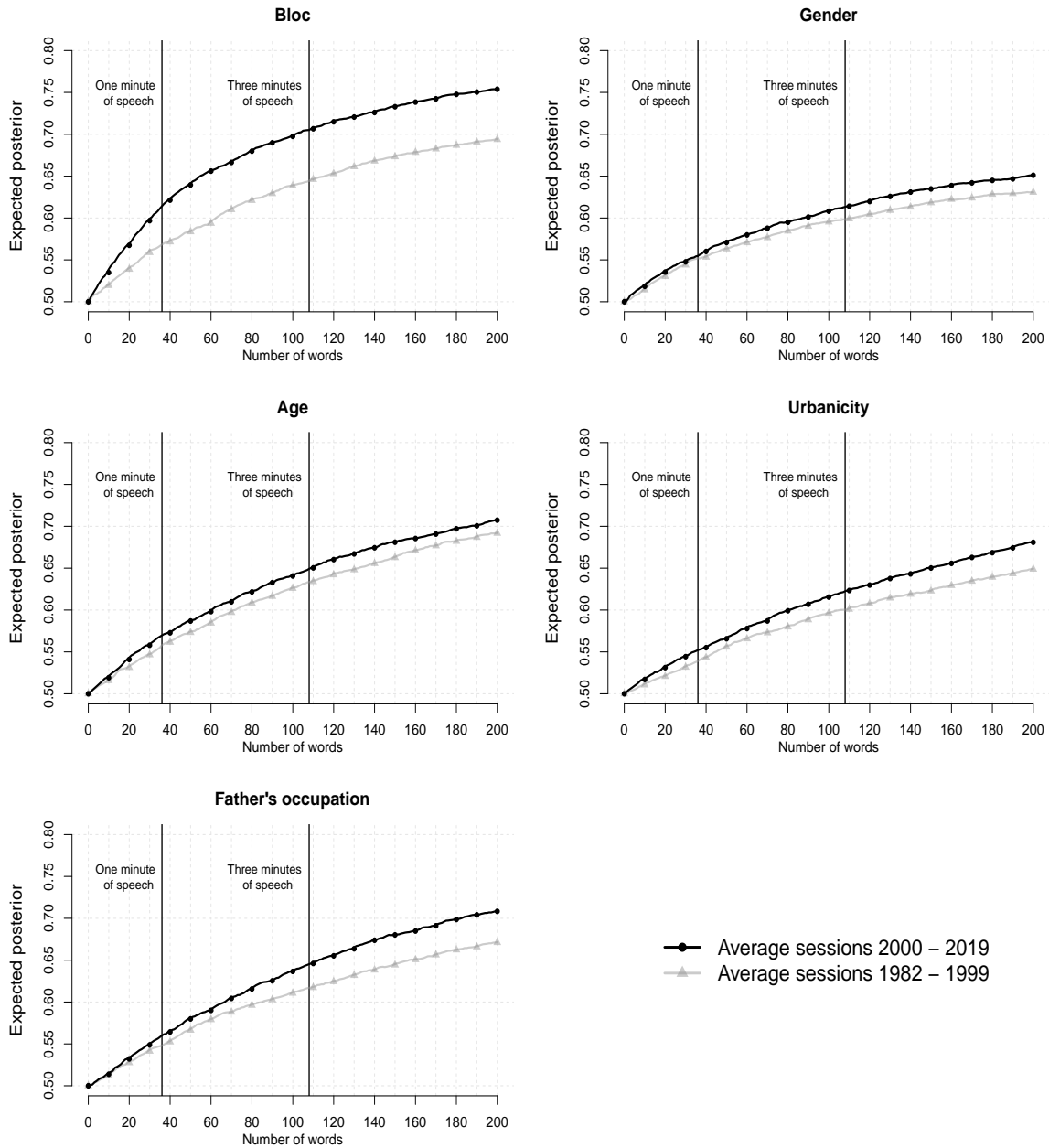
Appendix Figure A.6 shows, however, that the results for gender, age, urbanicity, and social background, are remarkably similar to the baseline analyses. This is also the case if we replace party bloc fixed effects (left vs. right) with more fine-grained party fixed effects (Appendix Figure A.7).

Next, we explore the role of standing committees in the parliament. Individual committees are unlikely to be mirror images of the parliament both because of demand and supply forces. On the demand side, parties might keep an eye on background characteristics of legislators when orchestrating the composition of committees. On the supply

---

<sup>36</sup>Moreover, women on the left and women on the right might both talk about “children”, but differ in the words they use to describe the appropriate policies toward children (e.g., kindergarden versus cash-for-care (see, e.g., Bettinger, Hægeland and Rege, 2014)). If this is the case, we expect to find that different words show up as polarizing with and without bloc fixed effects. This is not the case. Kindergarden is, for example, ranked sixth both with and without bloc fixed effects (full results omitted for brevity).

Figure 7: Polarization by Speech Length



*Note: The figure shows average polarization as a function of number of words after data pre-processing (see section 3.1 for details). The expected posterior is calculated by drawing 200 words for each speaker  $i$  and session  $t$ , given characteristics  $x_{it}$  using the estimated choice probabilities. The expected posterior as a function of number of words is calculated in the following way: Within each session, we calculate the average probability that a neutral observer assigns to the speaker's true identity after hearing the first word. Then we use this probability as the prior, when we calculate the posterior probability after hearing an additional word. This is continued until 200 words are spoken. The average lines in the figure are found by taking the average over the session-specific, expected posterior. The vertical line marks the median number of words per minute of speech. This number is calculated by dividing the number of words by the number of minutes for each speaker-session observation, and taking the median of this ratio across observations. The median one-minute speech is calculated using data from the 2004–2020 period, which contain the starting time of each speech.*



side, politicians with different backgrounds may be interested in different policy areas and may therefore self-select into different committees.

In the data, we find substantial imbalance in the composition of committees when it comes to gender (Appendix Figure A.8), but less so for the other background characteristics (Appendix Figure A.9 – A.11). For example, women are underrepresented in the *Committee of Finance* and overrepresented in the *Committee of Family and Cultural Affairs* throughout our sample period. This imbalance is particularly strong in the 2005–2009 election period. In these parliamentary sessions, the *Committee of Finance* and the *Committee of Family and Cultural Affairs* consisted of 14% and 75% women, respectively.

Our analyses control for committee assignment via fixed effects. In other words, we are comparing differences in speech patterns across women and men, who belong to the same committee and party bloc. If we drop the committee fixed effects, we reach qualitatively the same conclusions as in our baseline analysis. As one might expect, the differences in the estimates for gender polarization occur when committee assignment is least balanced across genders. The most salient difference is, however, that the placebo distributions widen without committee fixed effects (Appendix Figure A.12).

## 6. Conclusion

Many countries have become increasingly polarized along the left-right ideological dimension in recent decades. In the United States, for example, differences in language used by Democrats vs. Republicans in the 2010s is greater than ever before. In this paper, we document a similar, but more muted trend in left-right polarization in the Norwegian Parliament.

In addition to examining polarization along the left-right dimension, we consider political divergence between legislators belonging to the same political bloc. In party-centered environments, like Norway, one might expect parties to fully discipline their rank-and-file members, and therefore to limit within-bloc polarization to a minimum. This is not what

we observe in our speech data. Conversely, we find strong evidence that elected officials' identities and social ties provide important information for understanding how they engage in parliamentary debate. Our results demonstrate how political divergence between legislators reflect the same dimensions that have been highlighted as explanations for the recent developments in polarization in the electorate (e.g., Norris and Inglehart, 2019).

Our findings can be interpreted in two ways. First, parties may want to strategically allocate policy issues to legislators with a certain background characteristic. If voters support parties in response to the speeches that reach them, and we view "speech space" as a multi-dimensional space whose dimensions correspond to characteristics that voters value, then parties will likely want their members to spread out in the space, as in models of multiparty spatial competition (e.g., Cox, 1990). Second, the intraparty speech diversity that we document could reflect party weakness. Voters might punish parties for having divergent speeches, but party elites are unable to discipline their rank-and-file because of the commitment issues highlighted in the citizen-candidate framework (Alesina, 1988; Besley and Coate, 1998; Osborne and Slivinski, 1996). Future research should seek to distinguish between what is driven by parties' inability to stifle politicians' individual preferences and what can be attributed to strategic considerations at the party level.

## References

- Alesina, Alberto. 1988. "Credibility and Policy Convergence in a Two-Party System with Rational Voters." *American Economic Review* 78:796–806.
- Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer and Matthew Gentzkow. 2020. "The Welfare Effects of Social Media." *American Economic Review* 110(3):629–676.
- Autor, David, David Dorn, Gordon Hanson and Kaveh Majlesi. 2020. "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure." *American Economic Review* 110(10):3139–3183.
- Bäck, Hanna, Marc Debus and Jochen Müller. 2014. "Who Takes the Parliamentary Floor? The Role of Gender in Speech-making in the Swedish Riksdag." *Political Research Quarterly* 67(3):504–518.
- Bagues, Manuel and Pamela Campa. 2021. "Can Gender Quotas in Candidate Lists Empower Women? Evidence from a Regression Discontinuity Design." *Journal of Public Economics* forthcoming.
- Ban, Pamela, Daniel Moskowitz and James M Snyder. 2016. "Leadership Power in Congress, 1890-2014: Evidence from PAC Contributions and Newspaper Coverage." Available at SSRN: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2765092](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765092).
- Barber, Michael J. 2016. "Ideological donors, contribution limits, and the polarization of American legislatures." *The Journal of Politics* 78(1):296–310.
- Baskaran, Thushyanthan and Zohal Hessami. 2019. "Competitively Elected Women as Policy Makers." CESifo Working Paper No. 8005.
- Besley, Timothy and Stephen Coate. 1998. "Sources of Inefficiency in a Representative Democracy: A Dynamic Analysis." *American Economic Review* 88:139–156.
- Bettinger, Eric, Torbjørn Hægeland and Mari Rege. 2014. "Home with Mom: The Effects of Stay-at-home Parents on Children's Long-run Educational Outcomes." *Journal of Labor Economics* 32(3):443–467.
- Blumenau, Jack. 2021. "The Effects of Female Leadership on Women's Voice in Political Debate." *British Journal of Political Science* forthcoming.
- Boxell, Levi, Matthew Gentzkow and Jesse M Shapiro. 2020. "Cross-country trends in affective polarization." NBER Working Paper No. 26669.
- Buisseret, Peter and Carlo Prato. 2021. "Competing Principals? Legislative Representation in List Proportional Representation Systems." *American Journal of Political Science* forthcoming.
- Buisseret, Peter, Olle Folke, Carlo Prato and Johanna Rickne. 2019. "Party Nomination Strategies in List Proportional Representation Systems." Available at SSRN: <https://ssrn.com/abstract=3425692>.

- Canen, Nathan, Chad Kendall and Francesco Trebbi. 2020. "Unbundling Polarization." *Econometrica* 88(3):1197–1233.
- Carnes, Nicholas. 2013. *White-collar government: The hidden role of class in economic policy making*. University of Chicago Press.
- Carnes, Nicholas and Noam Lupu. 2015. "Rethinking the Comparative Perspective on Class and Representation: Evidence from Latin America." *American Journal of Political Science* 59(1):1–18.
- Carozzi, Felipe and Luca Repetto. 2016. "Sending the Pork Home: Birth Town Bias in Transfers to Italian Municipalities." *Journal of Public Economics* 134:42–52.
- Chattopadhyay, Raghavendra and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72(5):1409–1443.
- Cirone, Alexandra, Gary W Cox and Jon H Fiva. 2021. "Seniority-based Nominations and Political Careers." *American Political Science Review* forthcoming.
- Clayton, Amanda, Cecilia Josefsson and Vibeke Wang. 2017. "Quotas and Women's Substantive Representation: Evidence from a Content Analysis of Ugandan Plenary Debates." *Politics & Gender* 13(2):276–304.
- Cox, Gary W. 1990. "Centripetal and Centrifugal Incentives in Electoral Systems." *American Journal of Political Science* 34(4):903–935.
- Cox, Gary W, Jon H Fiva, Daniel M Smith and Rune J. Sørensen. 2020. "Moral Hazard in Electoral Teams." CESifo Working Paper No. 8357.
- DellaVigna, Stefano and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics* 122(3):1187–1234.
- Draca, Mirko and Carlo Schwarz. 2020. "How Polarized are Citizens? Measuring Ideology from the Ground-up." Available at SSRN: <https://ssrn.com/abstract=3154431> or <http://dx.doi.org/10.2139/ssrn.3154431>.
- Ferreira, Fernando and Joseph Gyourko. 2014. "Does Gender Matter for Political Leadership? The Case of US mayors." *Journal of Public Economics* 112:24–39.
- Fiva, Jon H., Askill H. Halse and Daniel M. Smith. 2021. "Local Representation and Voter Mobilization in Closed-list Proportional Representation Systems." *Quarterly Journal of Political Science* forthcoming.
- Fiva, Jon H. and Daniel M. Smith. 2017. "Norwegian Parliamentary Elections, 1906-2013: Representation and Turnout Across Four Electoral Systems." *West European Politics* 40(6):1373–1391.
- Folke, Olle, Torsten Persson and Johanna Rickne. 2016. "The Primary Effect: Preference Votes and Political Promotions." *American Political Science Review* 110(3):559–578.

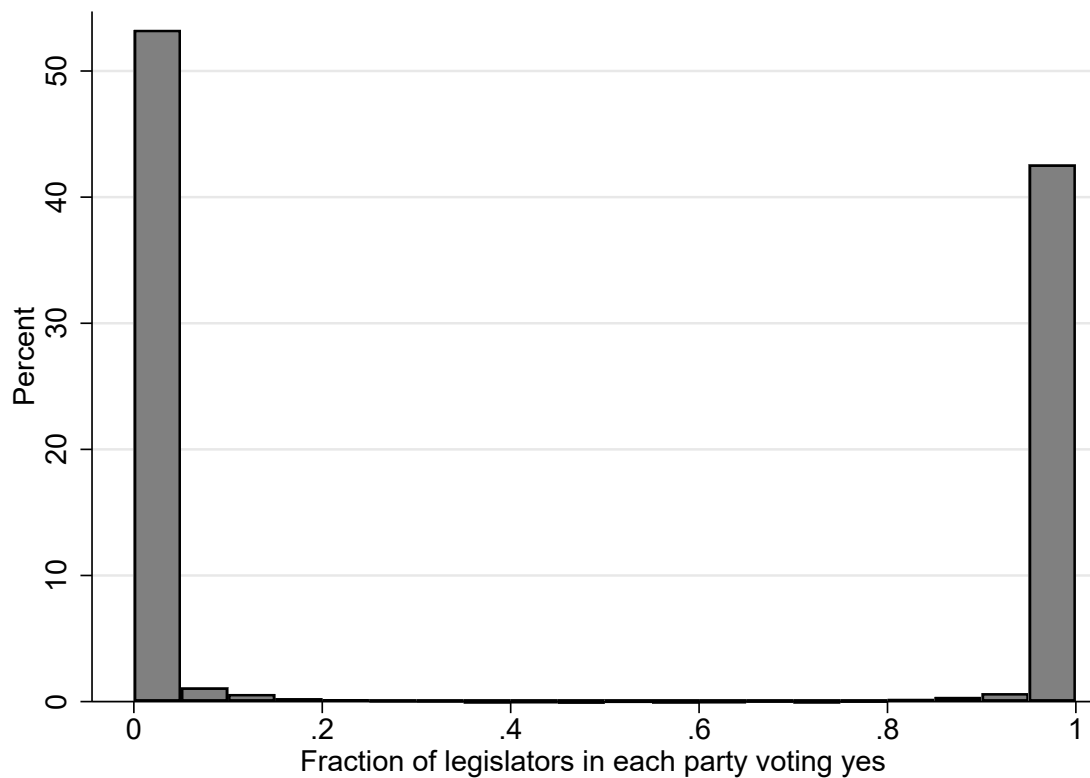
- Fujiwara, Thomas and Carlos Sanz. 2020. “Rank Effects in Bargaining: Evidence from Government Formation.” *The Review of Economic Studies* 87(3):1261–1295.
- Gentzkow, Matthew, Bryan Kelly and Matt Taddy. 2019. “Text as Data.” *Journal of Economic Literature* 57(3):535–74.
- Gentzkow, Matthew and Jesse M Shapiro. 2010. “What Drives Media Slant? Evidence from US Daily Newspapers.” *Econometrica* 78(1):35–71.
- Gentzkow, Matthew, Jesse M Shapiro and Matt Taddy. 2019. “Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech.” *Econometrica* 87(4):1307–1340.
- Grimmer, Justin. 2013. *Representational Style in Congress: What Legislators Say and Why it Matters*. Cambridge University Press.
- Gulzar, Saad. 2021. “Who Enters Politics and Why?” *Annual Review of Political Science* forthcoming.
- Hall, Andrew B. 2019. *Who Wants to Run?: How the Devaluing of Political Office Drives Polarization*. University of Chicago Press.
- Hazan, Reuven Y. and Gideon Rahat. 2010. *Democracy Within Parties: Candidate Selection Methods and Their Political Consequences*. Oxford, UK: Oxford University Press.
- Heath, Roseanna Michelle, Leslie A. Schwindt-Bayer and Michelle M. Taylor-Robinson. 2005. “Women on the Sidelines: Women’s Representation on Committees in Latin American Legislatures.” *American Journal of Political Science* 49(2):420–436.
- Hyytinen, Ari, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen and Janne Tukiainen. 2018. “Public employees as politicians: Evidence from close elections.” *American Political Science Review* 112(1):68–81.
- Johannessen, Janne Bondi, Kristin Hagen, André Lynum and Anders Nøklestad. 2012. OBT+ stat. A combined rule-based and statistical tagger. In *Exploring Newspaper Language. Corpus compilation and research based on the Norwegian Newspaper Corpus*, ed. Gisle Andersen. John Benjamins Publishing Company pp. 51–65.
- Kam, Christopher J. 2009. *Party Discipline and Parliamentary Politics*. Cambridge University Press.
- Lapponi, Emanuele, Martin G. Søyland, Erik Velldal and Stephan Oepen. 2018. “The Talk of Norway: A Richly Annotated Corpus of the Norwegian Parliament, 1998–2016.” *Language Resources and Evaluation* 52(3):873–893.
- Lauderdale, Benjamin E and Alexander Herzog. 2016. “Measuring Political Positions from Legislative Speech.” *Political Analysis* pp. 374–394.
- Martin, Gregory J. and Ali Yurukoglu. 2017. “Bias in Cable News: Persuasion and Polarization.” *American Economic Review* 107(9):2565–2599.

- Martin, Lanny W and Georg Vanberg. 2008. "Coalition government and political communication." *Political Research Quarterly* 61(3):502–516.
- Matakos, Konstantinos, Riikka Savolainen, Orestis Troumpounis, Janne Tukiainen and Dimitrios Xeftaris. 2018. "Electoral institutions and intraparty cohesion." *VATT Institute for Economic Research Working Papers* 109.
- Meriläinen, Jaakko and Janne Tukiainen. 2018. "Rank Effects in Political Promotions." *Public Choice* 177(1-2):87–109.
- Modalsli, Jørgen. 2017. "Intergenerational mobility in Norway, 1865–2011." *The Scandinavian Journal of Economics* 119(1):34–71.
- Motolinia, Lucia. 2020. "Electoral Accountability and Particularistic Legislation: Evidence from an Electoral Reform in Mexico." *American Political Science Review* pp. 1–17.
- Norris, Pippa and Ronald Inglehart. 2019. *Cultural backlash: Trump, Brexit, and authoritarian populism*. Cambridge University Press.
- O'Brien, Diana Z and Jennifer M Piscopo. 2019. The Impact of Women in Parliament. In *The Palgrave Handbook of Women's Political Rights*. Springer pp. 53–72.
- O'Grady, Tom. 2019. "Careerists Versus Coal-Miners: Welfare Reforms and the Substantive Representation of Social Groups in the British Labour Party." *Comparative Political Studies* 52(4):544–578.
- Osborn, Tracy and Jeanette Morehouse Mendez. 2010. "Speaking as Women: Women and Floor Speeches in the Senate." *Journal of Women, Politics & Policy* 31(1):1–21.
- Osborne, Martin J. and Al Slivinski. 1996. "A Model of Political Competition with Citizen-Candidates." *The Quarterly Journal of Economics* 111(1):65–96.
- Pande, R. 2003. "Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India." *American Economic Review* 93(4):1132–1151.
- Peterson, Andrew and Arthur Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems." *Political Analysis* 26(1):120–128.
- Poole, Keith T. 2007. "Changing Minds? Not in Congress!" *Public Choice* 131(3-4):435–451.
- 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. 2015. *The Politics of Parliamentary Debate: Parties, Rebels, and Representation*. Cambridge University Press.

- Rasch, Bjørn Erik. 1999. Electoral Systems, Parliamentary Committees, and Party Discipline: the Norwegian Storting in a Comparative Perspective. In *Party Discipline and Parliamentary Government*, ed. David M. Farrell Shaun Bowler and Richard S. Katz. Ohio State University Press Columbus pp. 121–140.
- Schmalensee, Richard. 1978. “Entry Deterrence in the Ready-to-eat Breakfast Cereal Industry.” *The Bell Journal of Economics* pp. 305–327.
- Schmidt, Vivien A. 2008. “Discursive Institutionalism: The Explanatory Power of Ideas and Discourse.” *Annual review of political science* 11.
- Schwarz, Daniel, Denise Traber and Kenneth Benoit. 2017. “Estimating Intra-Party Preferences: Comparing Speeches to Votes.” *Political Science Research and Methods* 5(2):379–396.
- Simola, Salla. 2020. “A Century of Partisanship in Finnish Political Speech.” Unpublished manuscript available at <https://sites.google.com/site/sallasimolaecon/home/research>.
- Strøm, Kaare and Jørn Y. Leipart. 1993. “Policy, Institutions, and Coalition Avoidance: Norwegian Governments, 1945-1990.” *The American Political Science Review* 87:870–887.
- Søyland, Martin. 2020. “Parliamentary Debates: A Computational Social Science Approach.” PhD thesis at the University of Oslo.
- Taddy, Matt. 2015. “Distributed Multinomial Regression.” *The Annals of Applied Statistics* 9(3):1394–1414.

## Appendix A: Supplementary figures and tables

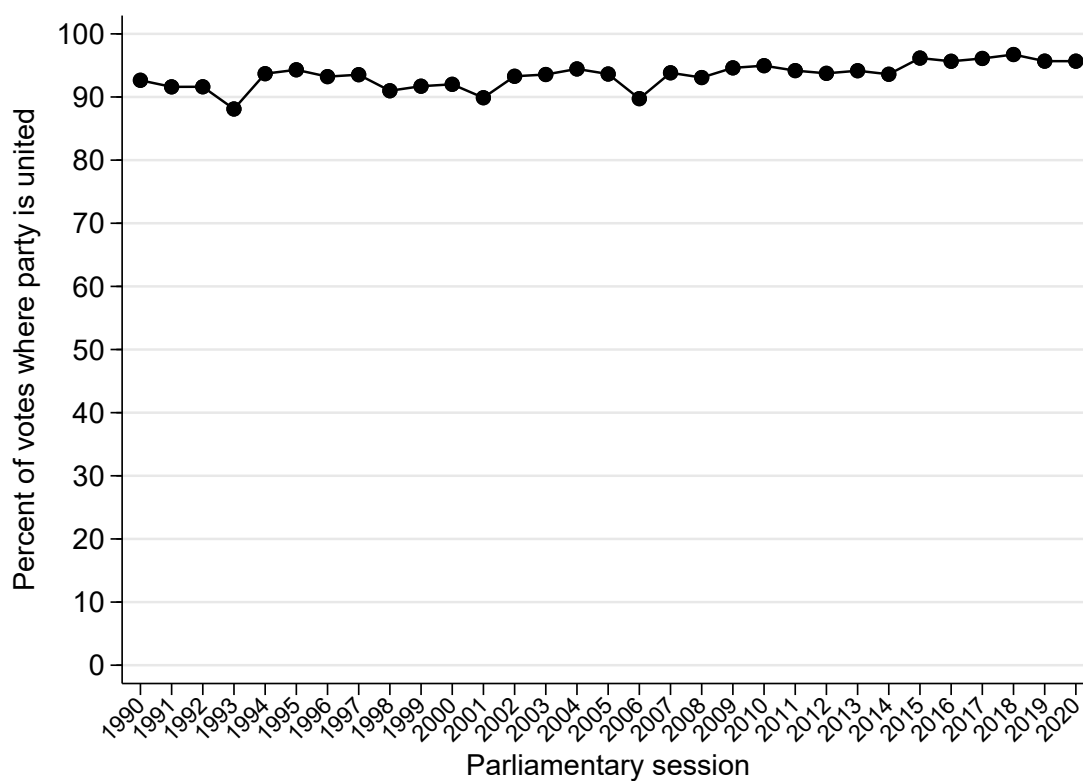
Figure A.1: Analysis of roll call votes 1990-2020



*Note: The figure shows the percent of legislators of each party voting yes to a proposal. The sample includes roll call votes recorded by the electronic voting device of the Storting, excluding unanimous and near-unanimous decisions, in the 1990-2020 period. The unit of observation is the party-vote ( $N = 154,878$ ). The main parties are: the Socialist Peoples' Party/Socialist Left Party (SV), the Labour Party (A), the Centre Party (SP), the Christian Democrats (KrF), the Liberal Party (V), the Conservative Party (H), and the Progress Party (FrP).*

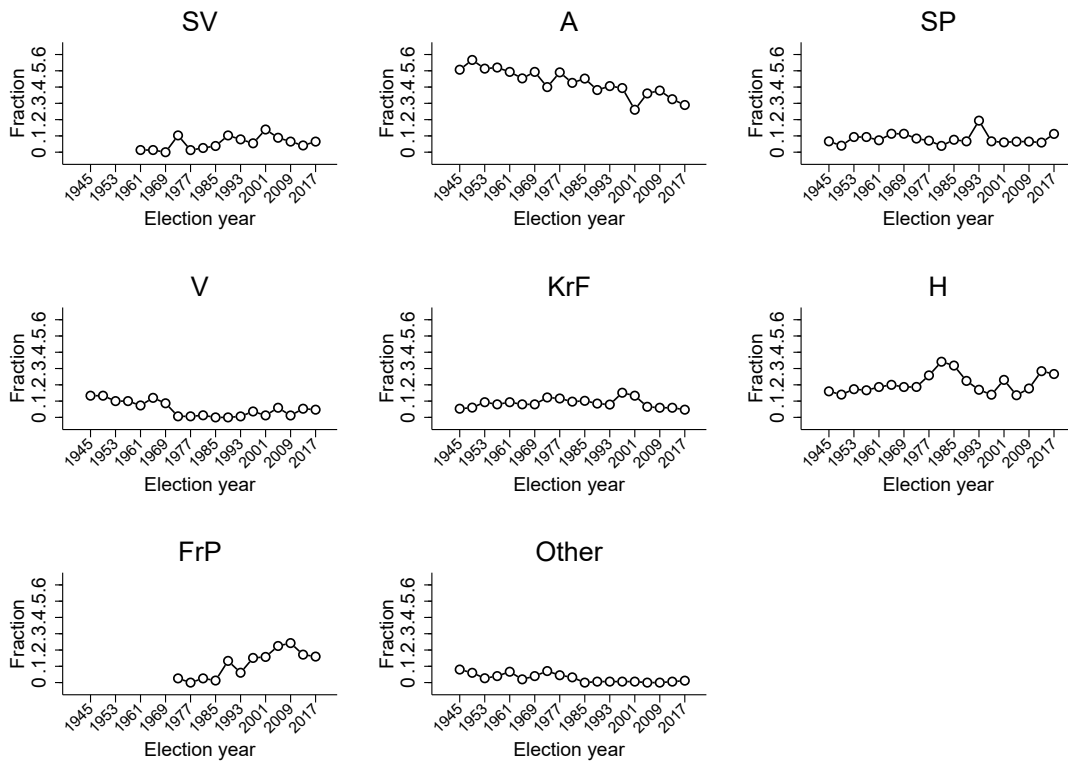


Figure A.2: Party discipline measured by roll-call votes 1990-2020



*Note: The figure shows the fraction of votes where a party is united by parliamentary session. The sample includes roll call votes recorded by the electronic voting device of the Storting, excluding unanimous and near-unanimous decisions, in the 1990-2020 period. The unit of observation is the party-vote ( $N = 154,878$ ). The main parties are: the Socialist Peoples' Party/Socialist Left Party (SV), the Labour Party (A), the Centre Party (SP), the Christian Democrats (KrF), the Liberal Party (V), the Conservative Party (H), and the Progress Party (FrP).*

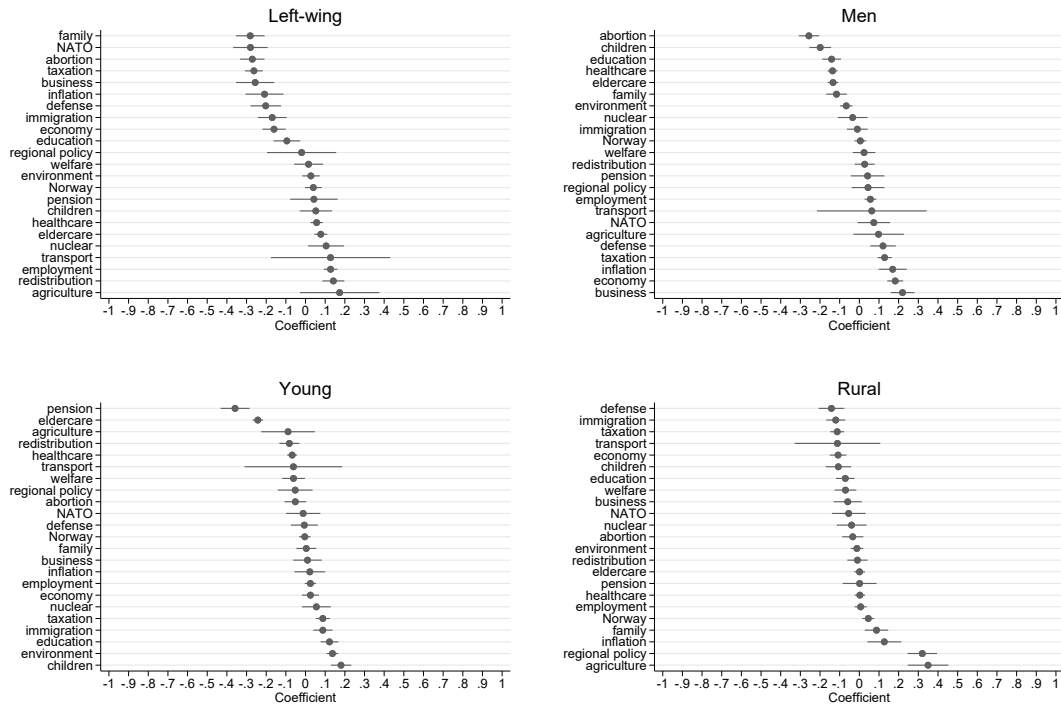
Figure A.3: Parties' Seat Shares by Election Year



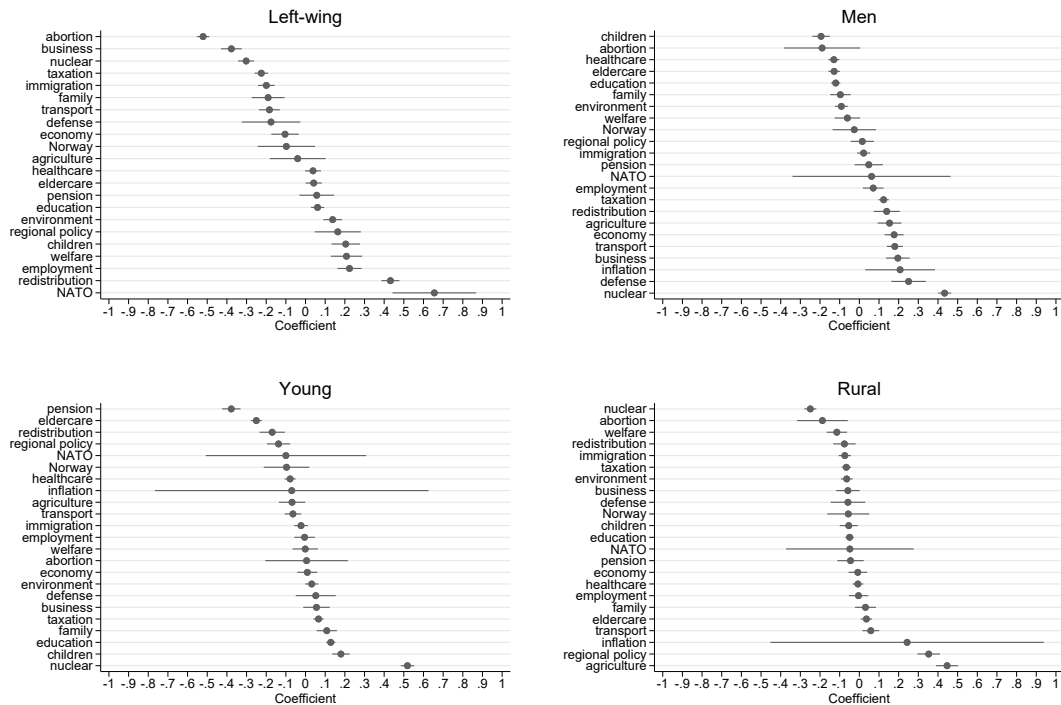
*Note: Figure shows the parties' seat shares by election year. The main parties are: the Socialist Peoples' Party/Socialist Left Party (SV), the Labour Party (A), the Centre Party (SP), the Christian Democrats (KrF), the Liberal Party (V), the Conservative Party (H), and the Progress Party (FrP). "Other" is a residual category for non-main parties.*

Figure A.4: Voter preferences measured in surveys: 1981-1997 and 2001-2017

### 1981-1997

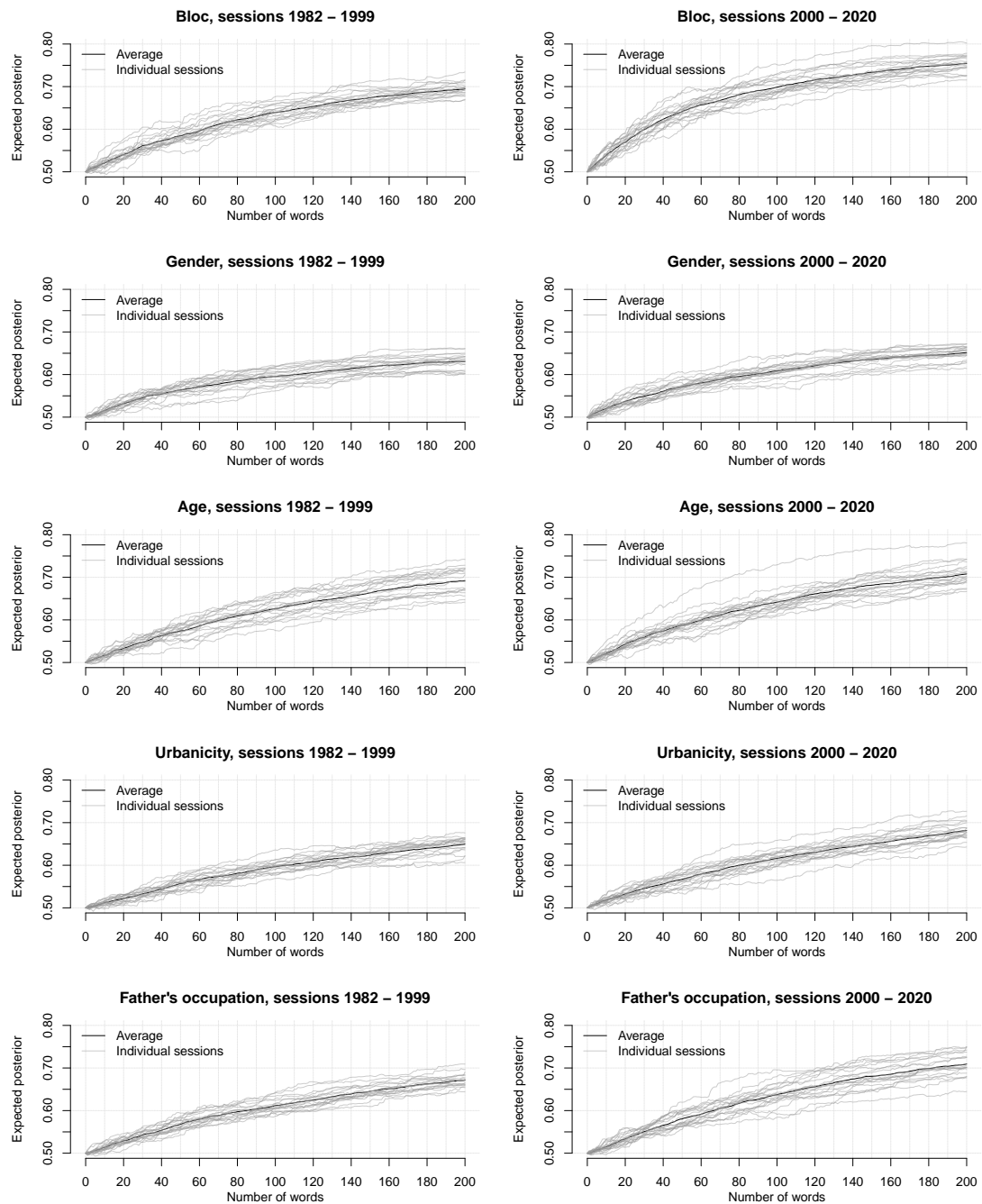


### 2001-2017



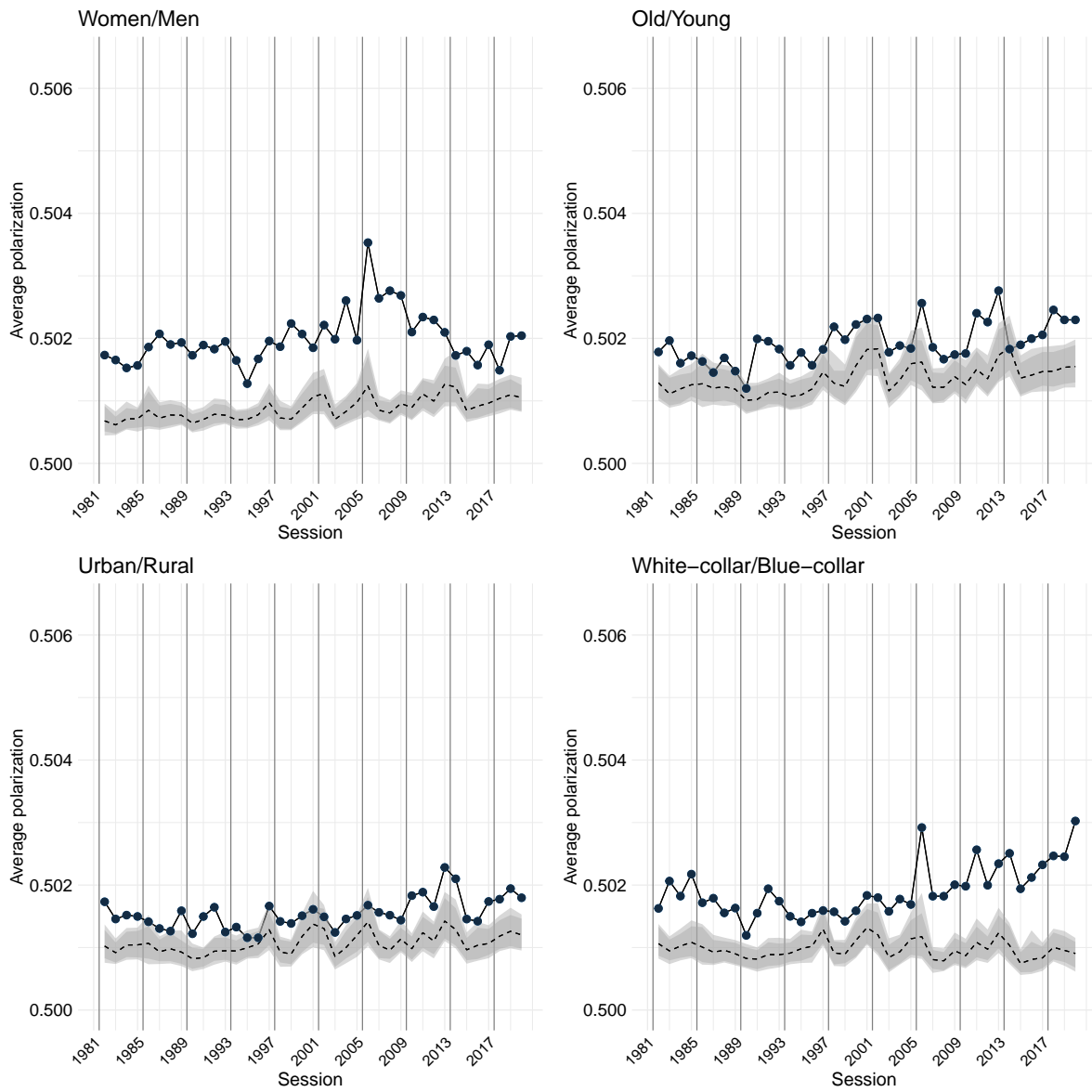
Note: This figure reports coefficients and corresponding 95% confidence intervals for four linear probability models, as in Figure 4. The top panel uses data from the National Election Survey 1981-1997 (five waves). The bottom panel uses data from the National Election Survey 2001-2017 (five waves).

Figure A.5: Polarization by speech length for individual sessions



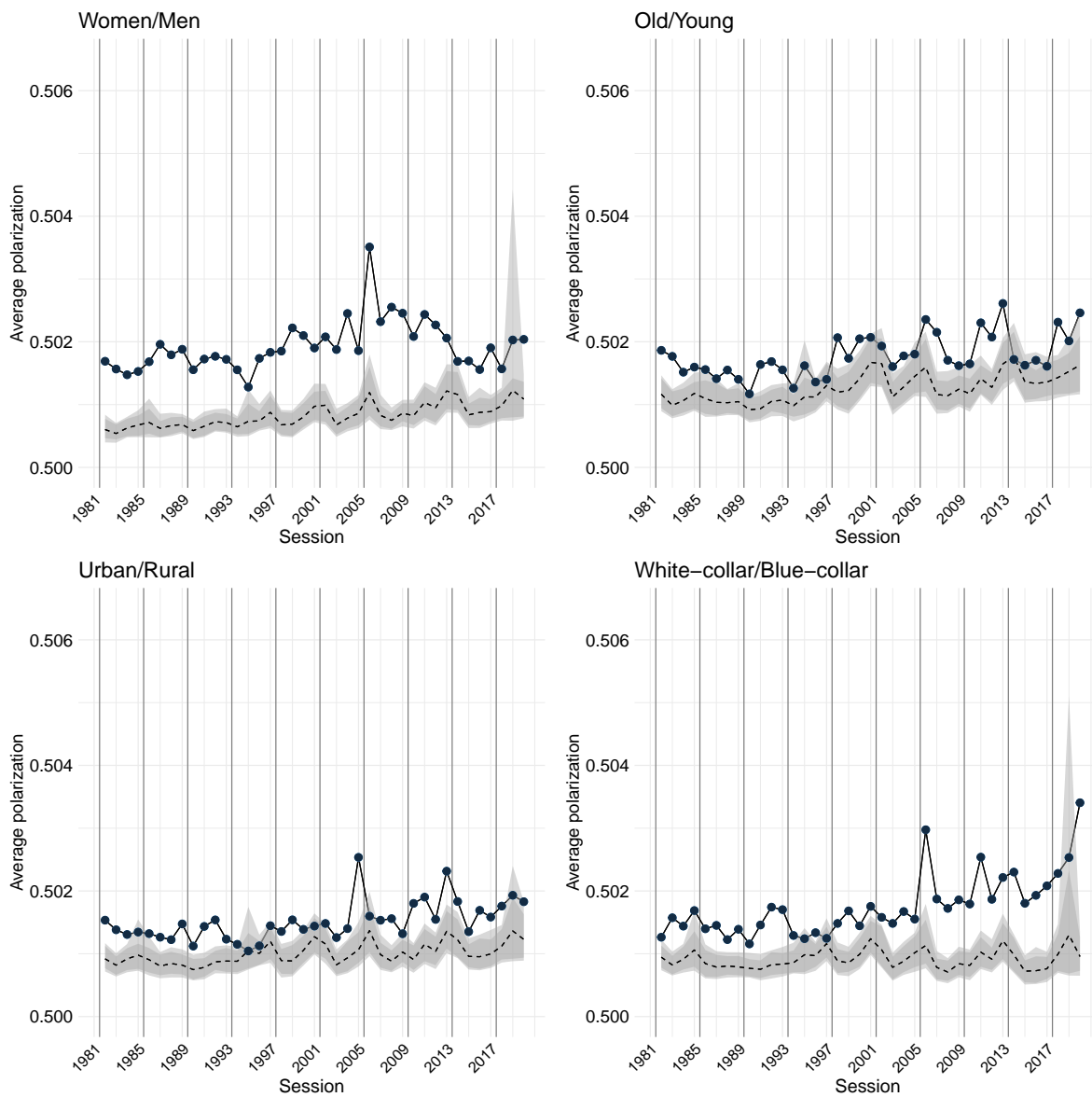
*Note: The figure shows average gender, age, urbanicity and social background polarization as a function of number of words for individual sessions and an average across all sessions. The expected posterior is computed by drawing 200 words for each speaker  $i$  and session  $t$ , given characteristics  $x_{it}$  using the estimated choice probabilities. The expected posterior is calculated in the following way: Within each session, we calculate the average probability that a neutral observer assigns to the speaker's true identity after hearing the first word. Then we use this probability as the prior, when we calculate the posterior probability after hearing an additional word. This is continued until 200 words are spoken. The average lines in the figure are found by taking the average across session-specific expected posterior.*

Figure A.6: Analysis of background polarization when controlling for committee assignment, but not party or bloc



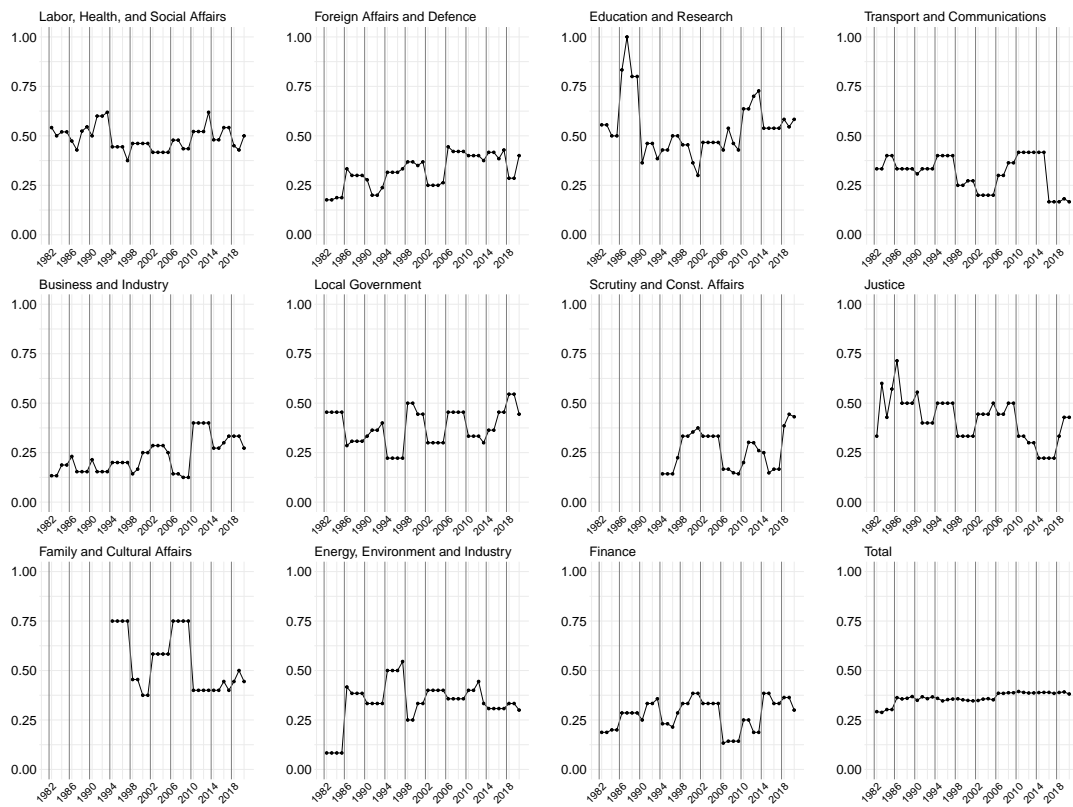
*Note: This figure displays polarization of legislative speech for four dimensions (given in the sub-panel headings) in the period 1981–2020, controlling for the legislator’s committee assignment. The models are estimated separately for each background characteristic. Speeches from legislators who do not sit in one of the 12 main Parliamentary committees are disregarded. The black points correspond to the average polarization of speech in each session and the bars indicate elections. The gray shaded area represents the average polarization in hypothetical data in which each speaker’s identity is randomly assigned. We construct 100 hypothetical data sets and compute the average polarization in each session. The dashed line corresponds to the mean polarization for each session across the distribution of placebo estimates. The upper and lower bounds of the light gray shaded area correspond to the 5th and the 95th highest polarization scores across the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th highest polarization scores.*

Figure A.7: Analysis of background polarization when controlling for committee assignment and party affiliation (rather than bloc affiliation)



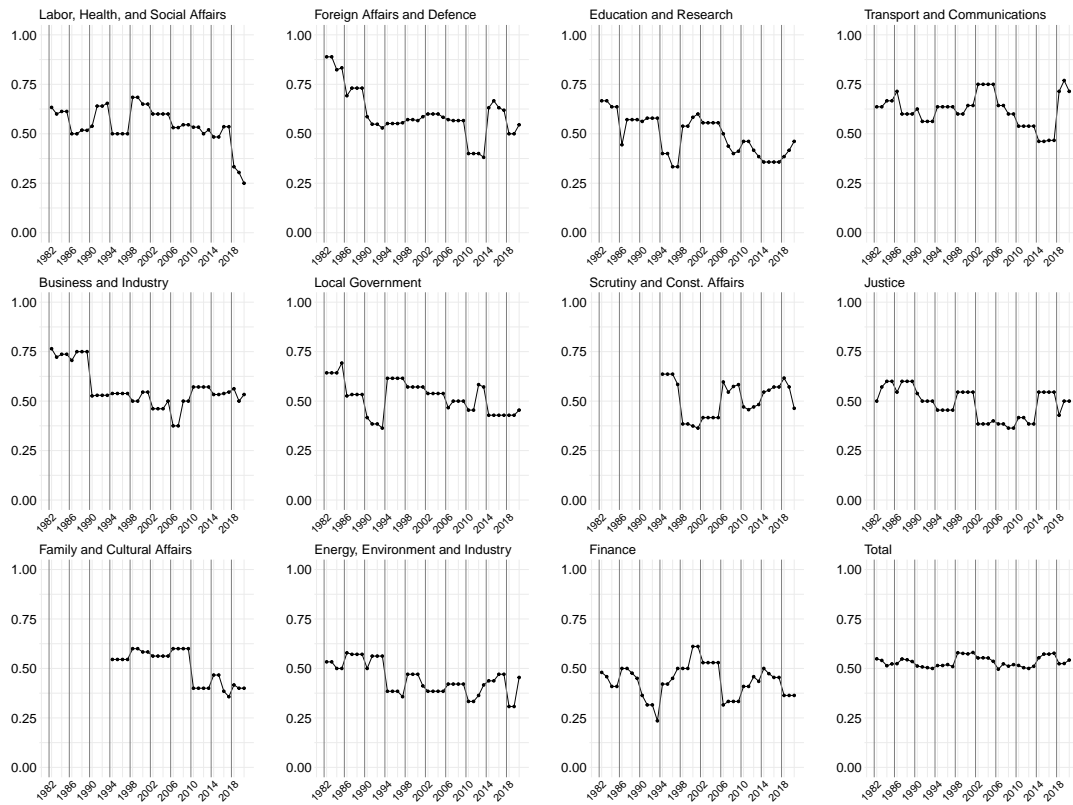
*Note: This figure displays polarization of legislative speech for four dimensions (given in the sub-panel headings) in the period 1981–2020. The models are estimated separately for each background characteristic. The black points correspond to the average polarization of speech in each session after controlling for legislators’ committee assignment and party. The bars indicate elections. The gray shaded area represents the average polarization in hypothetical data in which each speaker’s identity is randomly assigned. We construct 100 hypothetical data sets and compute the average polarization in each session. The dashed line corresponds to the mean polarization for each session across the distribution of placebo estimates. The upper and lower bounds of the light gray shaded area correspond to the 5th and the 95th highest polarization scores across the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th highest polarization scores.*

Figure A.8: Fraction of female politicians by policy area and parliamentary session



*Note: The figures show the fraction of female politicians, by policy area and parliamentary session. The black vertical lines indicate elections. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days. For a detailed description of which committees are contained in each policy area, see Appendix Table A.2.*

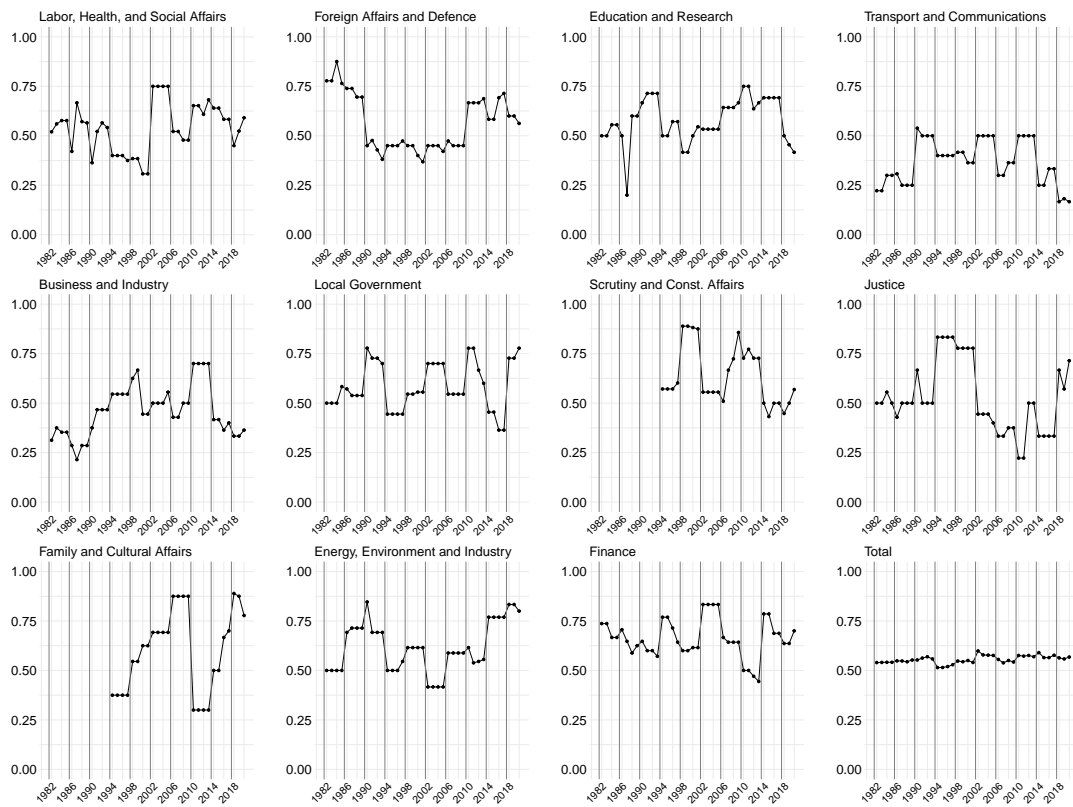
Figure A.9: Fraction of old politicians by policy area and parliamentary session



*Note: The figures show the fraction of politicians above 47 years old, by policy area and parliamentary session. The black vertical lines indicate elections. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days. For a detailed description of which committees are contained in each policy area, see Appendix Table A.2.*

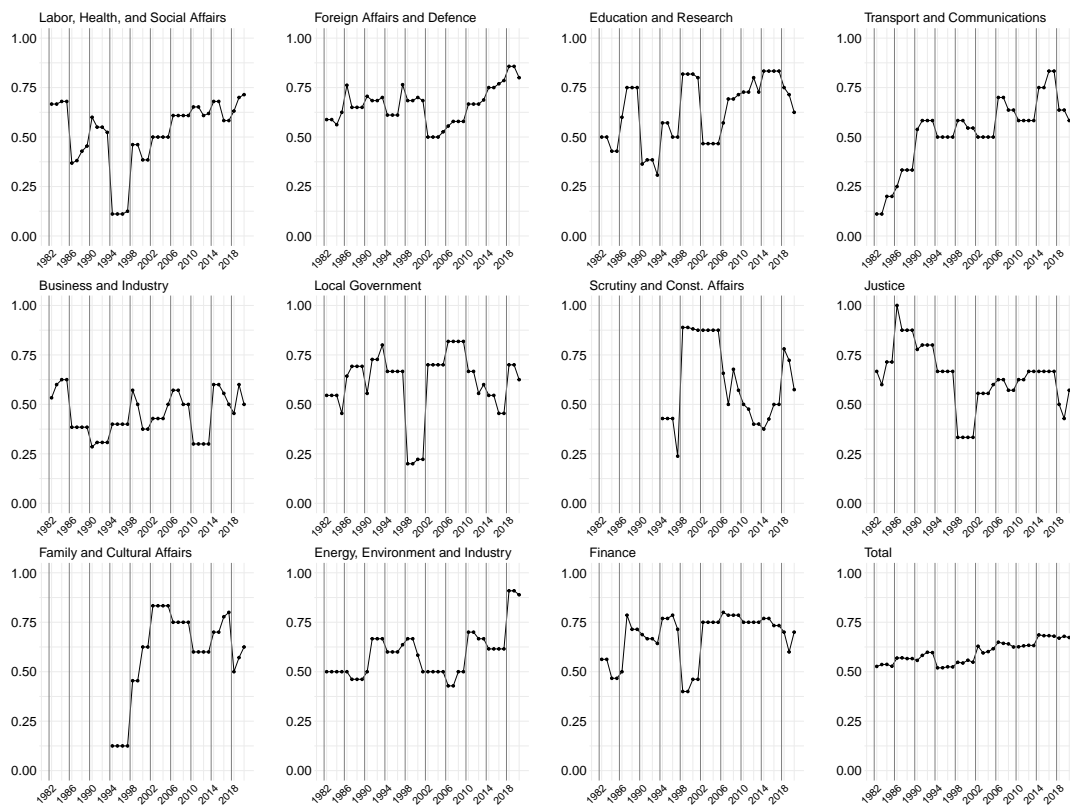


Figure A.10: Fraction of urban politicians by policy area and parliamentary session



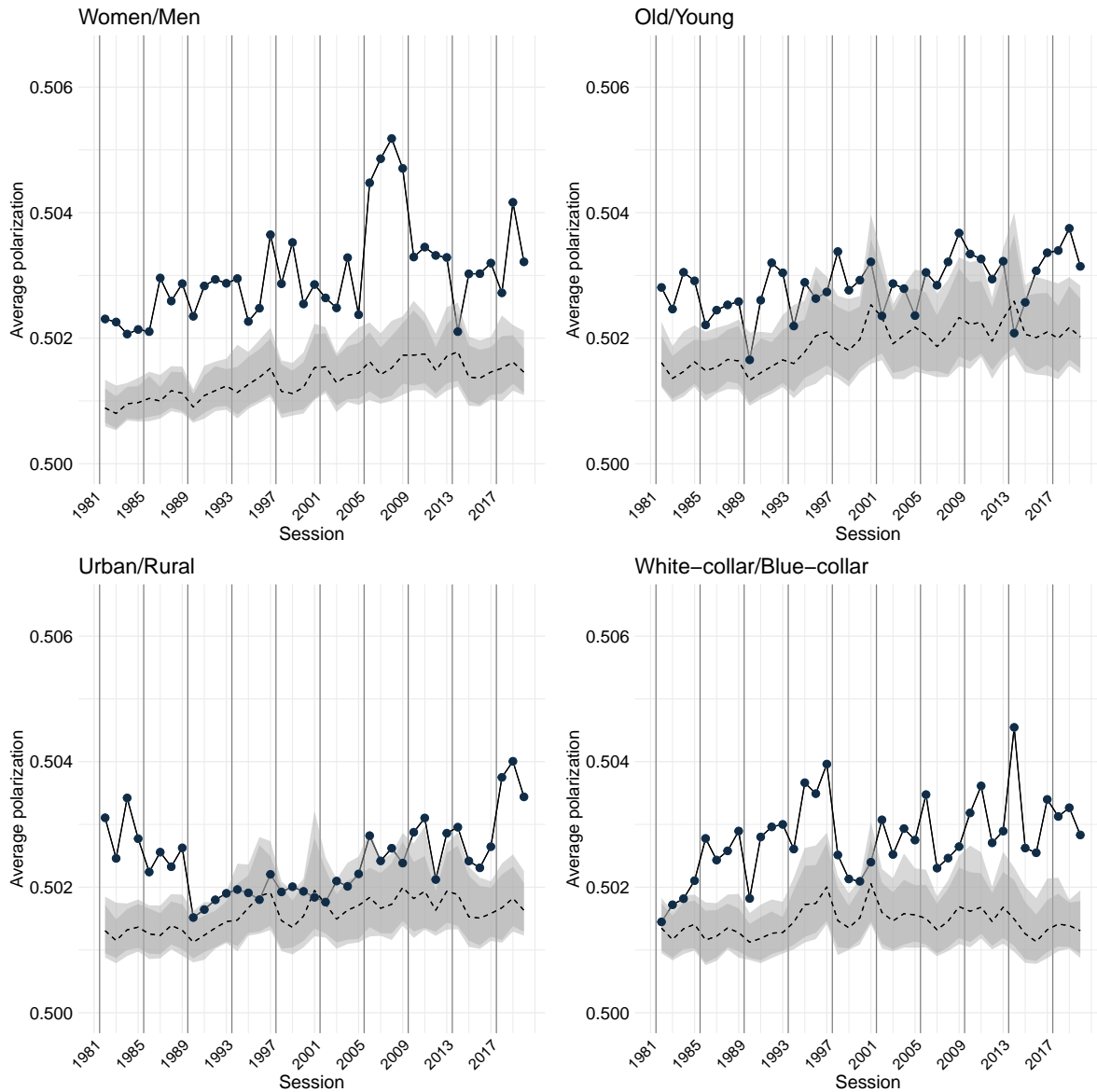
*Note: The figures show the fraction of politicians living in a municipality with a city status, by policy area and parliamentary session. The black vertical lines indicate elections. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days. For a detailed description of which committees are contained in each policy area, see Appendix Table A.2.*

Figure A.11: Fraction of politicians with white-collar background by policy area and parliamentary session



*Note: The figures show the fraction of politicians whose fathers worked in white-collar jobs, by policy area and parliamentary session. The black vertical lines indicate elections. If a politician switches committees during a parliamentary session, we use the committee where he/she spent most days. For a detailed description of which committees are contained in each policy area, see Appendix Table A.2.*

Figure A.12: Analysis of background polarization when controlling for bloc affiliation, but not committee



*Note: This figure displays polarization of legislative speech for four dimensions (given in the sub-panel headings) in the period 1981–2020. The models are estimated separately for each background characteristic. The black points correspond to the average polarization of speech in each session after controlling for legislator’s bloc affiliation. The bar indicate elections. The gray shaded area represents the average polarization in hypothetical data in which each speaker’s identity is randomly assigned. We construct 100 hypothetical data sets and compute the average polarization in each session. The dashed line corresponds to the mean polarization for each session across the distribution of placebo estimates. The upper and lower bounds of the light gray shaded area correspond to the 5th and the 95th highest polarization scores across the placebo distributions. The dark gray shaded area represents the corresponding 10th and 90th highest polarization scores.*

Table A.1: Norway's Governments 1981–2020

Time period	Prime minister	Parties	Parl. basis	Appointment reason	Resignation reason
Oct 1981 – Jun 1983	Kåre Willoch (H)	H	Minority	General elections	-
Jun 1983 – Sep 1985	Kåre Willoch (H)	H, KrF, SP	Majority	Government expansion	-
Sep 1985 – May 1986	Kåre Willoch (H)	H, KrF, SP	Minority	-	Government crisis
May 1986 – Oct 1989	Gro H. Brundtland (A)	A	Minority	Government crisis	General elections
Oct 1989 – Nov 1990	Jan P. Syse (H)	H, KrF, SP	Minority	General elections	Government crisis
Nov 1990 – Oct 1996	Gro H. Brundtland (A)	A	Minority	Government crisis	Change prime minister
Oct 1996 – Oct 1997	Thorbjørn Jagland (A)	A	Minority	Change prime minister	General elections
Oct 1997 – Mar 2000	Kjell M. Bondevik (KrF)	KrF, SP, V	Minority	General elections	Government crisis
Mar 2000 – Oct 2001	Jens Stoltenberg (A)	A	Minority	Government crisis	General elections
Oct 2001 – Oct 2005	Kjell M. Bondevik (KrF)	KrF, H, V	Minority	General elections	General elections
Oct 2005 – Oct 2013	Jens Stoltenberg (A)	A, SV, SP	Majority	General elections	General elections
Oct 2013 – Jan 2018	Erna Solberg (H)	H, FrP	Minority	General elections	-
Jan 2018 – Jan 2019	Erna Solberg (H)	H, FrP, V	Minority	Government expansion	-
Jan 2019 – Jan 2020	Erna Solberg (H)	H, FrP, V, KrF	Majority	Government expansion	-
Jan 2020 –	Erna Solberg (H)	H, V, KrF	Minority	Government reduction	-

*Note: The parties are Socialist Left Party (SV), the Labour Party (A), the Centre Party (SP), the Christian Democrats (KrF), the Liberal Party (V), the Conservative Party (H), and the Progress Party (FrP). Source [www.regjeringen.no](http://www.regjeringen.no).*

Table A.2: Committee structure in the Norwegian Parliament, 1981–2020

Policy area	English committee name	Norwegian committee name	Time period
Business and Primary Industry	Agriculture	Landbrukskomiteen	1981–1993
Business and Primary Industry	Maritime and Fisheries	Sjøfart og fiskerikomiteen	1981–1993
Business and Primary Industry	Business	Næringskomiteen	1993–2020
Education and Research	Church and Education	Kirke- og undervisningskomiteen	1981–1993
Education and Research	Church, Education and Research	Kirke-, utdannings- og forskningskom.	1993–2017
Education and Research	Education and Research	Utdannings- og forskningskomiteen	2017–2020
Energy, Environment and Industry	Energy and Industry	Energi- og industrikomiteen	1981–1993
Energy, Environment and Industry	Energy and the Environment	Energi- og miljøkomiteen	1993–2020
Family and Cultural Affairs	Family, Culture and Admin.	Familie-, kultur- og administrasjonskom.	1993–2005
Family and Cultural Affairs	Family and Culture	Familie- og kulturkomiteen	2005–2020
Finance	Finance	Finanskomiteen	1981–2020
Foreign Affairs and Defence	Foreign Affairs and Constitution	Utenriks- og konstitusjonskomiteen	1981–1993
Foreign Affairs and Defence	Defence	Forsvarskomiteen	1981–2009
Foreign Affairs and Defence	Foreign Affairs	Utenrikskomiteen	1993–2009
Foreign Affairs and Defence	Foreign Affairs and Defence	Utenriks- og forsvarskomiteen	2009–2020
Justice	Justice	Justiskomiteen	1981–2020
Labor, Health, and Social Affairs	Social Affairs	Sosialkomiteen	1981–2005
Labor, Health, and Social Affairs	Consumer and Administration	Forbruker- og administrasjonskomiteen	1981–1993
Labor, Health, and Social Affairs	Labor and Social Affairs	Arbeids- og sosialkomiteen	2005–2020
Labor, Health, and Social Affairs	Health and Social Affairs	Helse- og omsorgskomiteen	2005–2020
Local Government	Local Govern. and Environ. Protect.	Kommunal- og miljøvernkomiteen	1981–1993
Local Government	Local Government	Kommunalkomiteen	1993–2005
Local Government	Local Govern. and Administration	Kommunal- og forvaltningskomiteen	2005–2020
Scrutiny and Constitutional Affairs	Scrutiny and Constitutional Affairs	Kontroll- og konstitusjonskomiteen	1993–2020
Transport and Communications	Transport	Samferdselskomiteen	1981–2005
Transport and Communications	Transport and Communications	Transport- og kommunikasjonskomiteen	2005–2020

*Note: This table shows the committee structure in the Norwegian Parliament in our sample period (1981–2020). Our baseline analysis includes committee fixed effects. The aggregation of committees to policy areas is used for descriptive analyses only (e.g., Figure 1). Data source: [www.regjeringen.no](http://www.regjeringen.no).*

Table A.3: Summary statistics

<i>Variable</i>	<b>Median</b>	<b>Mean</b>	<b>St.Dev.</b>	<b>Min</b>	<b>Max</b>	<b>N</b>
Number of words	3563	4390.1	3343.17	86	40302	4768
Number of speeches	35	45.86	38.39	1	391	4768
Words per speech	99.51	106.7	37.34	21.83	384	4768
Minutes of speech	87.73	110.36	83.6	2.47	755.75	1479
Words per minute	36.16	36.31	5.94	2.59	132.7	1479

*Note: This table shows the median, mean, standard deviation, minimum, and maximum for the variables listed in Column 1. The last column shows the number of speaker-session observations. There are 4,768 speaker-session observations, and 578 unique speakers in the period 1981–2020. Speech length is calculated using the 2004–2020 period. We removed two extreme values of over 1,000 words per minute. The summary statistics for minutes of speech and words per minute are based on 1,479 speaker-sessions, and 258 unique speakers. All variables are based on the sample after data pre-processing (see section 3.1 for details.)*

Table A.4: Summary statistics by speaker characteristics.

Variable	Median	Mean	St.Dev.	Min	Max	N	Median	Mean	St.Dev.	Min	Max	N
<i>Bloc</i>	Right						Left					
Number of words	4106.0	4850.5	3207.4	129.0	26726.0	1759.0	3228.0	4000.3	3008.3	86.0	27466.0	2223.0
Number of speeches	38.0	47.8	35.8	1.0	306.0	1759.0	31.0	41.5	35.9	1.0	303.0	2223.0
Words per speech	105.6	113.0	37.8	28.6	384.0	1759.0	101.8	108.5	37.9	21.8	352.2	2223.0
Minutes of speech	112.7	134.1	95.4	5.4	755.8	517.0	80.0	98.2	79.5	2.5	723.5	664.0
Words per minute	36.4	36.6	5.2	2.6	55.4	517.0	36.3	36.4	5.5	22.6	99.9	664.0
<i>Gender</i>	Women						Men					
Number of words	3150.0	3856.6	2897.6	86.0	26132.0	1733.0	3867.0	4694.7	3537.3	106.0	40302.0	3035.0
Number of speeches	29.0	39.6	35.6	1.0	306.0	1733.0	39.0	49.5	39.5	1.0	391.0	3035.0
Words per speech	102.5	110.8	38.9	21.8	352.2	1733.0	97.7	104.3	36.2	26.5	384.0	3035.0
Minutes of speech	80.0	104.3	92.7	2.5	755.8	605.0	94.8	114.6	76.4	4.2	509.5	874.0
Words per minute	37.1	37.5	6.3	24.8	132.7	605.0	35.4	35.5	5.5	2.6	99.9	874.0
<i>Age</i>	Old						Young					
Number of words	3332.5	4066.8	3054.4	106.0	27466.0	2546.0	3881.0	4760.5	3611.3	86.0	40302.0	2222.0
Number of speeches	32.0	42.5	36.3	1.0	313.0	2546.0	40.0	49.7	40.3	1.0	391.0	2222.0
Words per speech	99.7	107.5	38.8	21.8	384.0	2546.0	99.3	105.8	35.6	32.8	307.4	2222.0
Minutes of speech	87.0	108.0	80.8	5.4	723.5	820.0	88.8	113.2	87.0	2.5	755.8	659.0
Words per minute	35.9	36.2	6.4	2.6	132.7	820.0	36.4	36.4	5.3	22.6	85.6	659.0
<i>Rurality</i>	Urban						Rural					
Number of words	3782.0	4662.1	3702.8	86.0	40302.0	2653.0	3393.0	4048.9	2791.7	134.0	26132.0	2115.0
Number of speeches	37.0	48.5	42.4	1.0	391.0	2653.0	33.0	42.5	32.4	1.0	254.0	2115.0
Words per speech	99.2	107.6	39.2	21.8	384.0	2653.0	99.9	105.6	34.9	28.6	333.0	2115.0
Minutes of speech	91.9	117.4	94.4	2.5	755.8	809.0	85.1	101.8	67.3	4.2	482.7	670.0
Words per minute	36.1	36.5	6.4	2.6	132.7	809.0	36.2	36.1	5.4	22.6	99.9	670.0
<i>Occupation</i>	White						Blue					
Number of words	3777.5	4648.1	3618.6	106.0	40302.0	2822.0	3351.0	4053.5	2880.5	86.0	26726.0	1881.0
Number of speeches	38.0	48.4	40.1	1.0	391.0	2822.0	32.0	42.5	35.9	1.0	306.0	1881.0
Words per speech	98.5	106.0	37.7	26.5	384.0	2822.0	101.2	107.9	36.8	21.8	333.0	1881.0
Minutes of speech	88.0	107.0	70.7	4.2	509.5	912.0	90.4	118.6	102.9	2.5	755.8	534.0
Words per minute	36.5	36.5	5.6	2.6	85.6	912.0	35.5	36.0	6.6	23.1	132.7	534.0

Note: This table shows summary statistics across overlapping characteristics. It shows the median, mean, standard deviation, minimum, and maximum for the variables listed in Column 1. There are 4,768 speaker-session observations, and 578 unique speakers in the period 1981–2020. Speech length is calculated using the 2004–2020 period. All variables are based on the sample after data pre-processing (see section 3.1 for details.)

Table A.5: Most polarizing words across blocs

Rank	Most polarizing for right				Most polarizing for left			
	English	Norwegian	#Right	#Left	English	Norwegian	#Right	#Left
1	debt	gjeld	208	188	people	folk	63	89
2	firm	bedrift	59	44	woman	kvinne	23	40
3	law	lov	75	63	policy	politikk	67	81
4	simple	enkelt	65	54	society	samfunn	56	69
5	private	privat	57	44	differences	forskjell	22	33
6	however	imidlertid	36	29	cut	kutte	13	25
7	competition	konkurransen	22	13	social	sosial	19	31
8	road	vei	52	38	good	bra	29	35
9	patient	pasient	29	17	country	land	163	174
10	promote	fremme	58	49	NOK	kr	98	109
11	tax	avgift	21	13	view	syn	41	49
12	business	næringsliv	51	39	money	penge	59	66
13	positive	positiv	50	42	school	skole	65	68
14	norwegian	norsk	208	193	conservative	borgerlig	7	26
15	entail	medføre	19	13	development	utvikling	76	85
16	reduce	redusere	58	53	initiative	tiltak	79	89
17	gladly	gjærne	41	32	international	internasjonal	46	52
18	party	parti	80	75	job	jobb	37	43
19	challenge	utfordring	56	47	situation	situasjon	84	94
20	police	politi	29	22	political	politisk	72	80
21	decrease	fall	57	50	rich	rik	8	15
22	project	prosjekt	39	32	community	felleskap	9	16
23	hope	håpe	36	30	entail	bety	55	62
24	problem	problem	83	82	trade	handel	33	35
25	use	benytte	20	14	youth	ungdom	13	21
26	solution	løsning	47	41	employee	ansatt	23	31
27	present	fremlegge	9	5	year	år	222	233
28	express	uttrykk	32	29	distribution	fordeling	8	14
29	exactly	nettopp	62	54	million	mill	50	60
30	based (on)	basere	20	15	education	utdanning	24	30
31	car driver	bilist	4	1	self	sjøl	4	19
32	generally	generell	28	23	responsibility	ansvar	63	69
33	tax	skatt	20	16	hear	høre	51	56
34	behalf	vegne	18	13	worklife	arbeidsliv	17	25
35	answer	svar	31	31	income	inntekt	26	33
36	by	innen	31	25	economic	økonomisk	70	81
37	pupil	elev	28	22	Trøndelag	Trøndelag	11	16
38	person	person	24	19	serious	alvorlig	29	33
39	voluntary	frivillig	16	11	young	ung	21	26
40	growth	vekst	28	25	discussion	diskusjon	14	19
41	family	familie	22	17	kindergarten	barnehage	17	24
42	politician	politiker	22	17	county	fylke	21	32
43	tax payer	skattebetaler	4	1	industry	industri	23	29
44	bureaucracy	byråkrati	7	3	unequal	ulik	49	54
45	happily	glede	17	12	fair	rettferdig	6	11
46	freedom of choice	valgfrihet	6	3	underline	understreke	46	52
47	propose	fremsette	8	4	man	mann	10	15
48	stimulate	stimulere	12	8	fine	fin	7	12
49	citizen	borger	6	2	work	arbeid	104	119
50	interesting	interessant	29	23	need	nødt	12	16

*Note: This table shows the 50 most polarizing words by bloc from our baseline model which includes parliamentary committee and parliamentary session fixed effects. For each word, we also report the number of occurrences per 100,000 words in the raw data (before feature selection and without controls).*



Table A.6: Most polarizing words across gender

Most polarizing for women					Most polarizing for men			
Rank	English	Norwegian	#Women	#Men	English	Norwegian	#Women	#Men
1	children	barn	156	49	Norwegian	norsk	172	215
2	woman	kvinne	65	14	relations	forhold	110	130
3	labor	arbeid	133	100	debt	gjeld	184	207
4	parent	forelder	37	13	party	parti	63	84
5	young	ung	40	16	political	politisk	65	80
6	kindergarten	barnehage	37	12	lay	ligge	73	86
7	initiative	tiltak	102	74	decrease	fall	45	58
8	child prot. service	barnevern	22	6	association	sammenheng	40	53
9	family	familie	32	14	Norway	Norge	179	185
10	municipality	kommune	157	107	post	innlegg	50	60
11	increase	øke	158	142	of course	selvfølgelig	31	41
12	self	sjøl	17	7	expression	uttrykk	20	35
13	education	utdanning	42	19	interesting	interessant	20	30
14	service	tilbud	53	30	policy	politikk	65	76
15	school	skole	95	53	director	direktør	12	36
16	pupil	elev	38	19	understand	forstå	28	36
17	adolescence	ungdom	25	12	point	peke	26	34
18	man	mann	20	8	in	inne	30	37
19	competence	kompetanse	37	20	understanding	oppfatning	15	27
20	young	unge	11	3	of course	selvsagt	17	28
21	mental	psykisk	20	7	fundament	grunnlag	39	49
22	life	liv	37	22	starting point	utgangspunkt	26	31
23	violence	vold	17	6	certain	viss	18	29
24	prevent	forebygge	18	8	out	ut	177	181
25	research	forskning	35	23	point	poeng	10	17
26	care	omsorg	17	7	type	type	25	29
27	equality	likestilling	12	4	register	registrene	21	28
28	responsibility	ansvar	77	60	etc	osv	11	16
29	girl	jente	9	2	NOK	kr	93	107
30	abuse	overgrep	11	5	gladly	gjerner	31	40
31	teacher	lærer	29	13	available	foreligge	16	26
32	job	jobb	53	34	corporation	selskap	21	37
33	old	gammal	34	23	stance	standpunkt	10	19
34	mother	mor	8	2	stand	stå	138	144
35	right	rettighet	26	14	billion	milliard	24	33
36	workplace	arbeidsplass	38	34	point	punkt	21	29
37	adult	voksen	11	4	situation	situasjon	81	91
38	working	jobbe	23	13	reality	realitet	12	20
39	patient	pasient	34	19	conclusion	konklusjon	9	15
40	health	helse	24	12	discuss	diskutere	24	25
41	handicapped	funksjonshemmet	11	4	create	skape	59	63
42	light	lys	29	23	argument	argument	11	15
43	group	gruppe	39	28	NRK	nrk	8	9
44	still	fremdeles	12	7	principled	prinsipiell	10	16
45	training	opplæring	17	8	mention	nevne	45	52
46	sick	syk	13	5	one's opinion	skjønn	10	14
47	knowledge	kunnskap	31	18	exactly	nettopp	57	60
48	weekday	hverdag	12	6	answer	svar	30	32
49	institution	institusjon	24	15	talk	snakk	44	45
50	cash-for-care	kontantstøtte	9	4	connection	forbindelse	45	53

*Note: This table shows the 50 most polarizing words by gender from our baseline model which includes political bloc, parliamentary committee, and parliamentary session fixed effects. For each word, we also report the number of occurrences per 100,000 words in the raw data (before feature selection and without controls).*

Table A.7: Most polarizing words across urbanicity

Rank	Most polarizing for urban				Most polarizing for rural			
	English	Norwegian	#Urban	#Rural	English	Norwegian	#Urban	#Rural
1	Norwegian	norsk	212	187	municipality	kommune	115	135
2	Oslo	Oslo	40	31	children	barn	80	86
3	Norway	Norge	190	173	relations	forhold	121	127
4	unequal	ulik	55	45	district	distrikt	17	28
5	director	direktør	32	23	agriculture	landbruk	10	20
6	self	sjøl	11	8	NOK	kr	97	110
7	international	internasjonal	53	43	service	tilbud	34	41
8	city	by	17	14	Finmark	Finmark	8	15
9	police	politi	26	26	opportunity	mulighet	93	99
10	concern	dreie	24	19	industry	næring	24	33
11	Bergen	Bergen	9	7	resource	ressurs	32	37
12	prison	fengsel	8	8	mill	mill	49	61
13	tie	knytte	37	34	family	familie	19	21
14	hospital	sykehus	24	22	Akershus	Akershus	2	8
15	corporation	selskap	33	29	settlement	bosetting	5	10
16	political	politisk	80	68	county	fylke	20	32
17	criminal	kriminell	7	5	however	imidlertid	31	36
18	patient	pasient	24	23	register	registrere	23	30
19	European Council	europaråd	4	2	year	år	221	234
20	starting point	utgangspunkt	30	28	Nordland	Nordland	3	7
21	human rights	menneskerett.	10	7	regional policy	distriktpol.	5	8
22	pensioner	pensjonist	7	6	church	kirke	10	11
23	ship	skip	6	5	Trøndelag	Trøndelag	12	15
24	fund	fond	9	7	PST	PST	67	71
25	woman	kvinne	30	29	Hedmark	Hedmark	2	6
26	nuclear weapon	atomvåpen	5	3	priority	satsing	24	27
27	health service	helsevesen	12	10	research	forskning	27	27
28	Drammen	Drammen	2	1	collaboration	samarbeid	53	52
29	of course	selvfølgelig	39	36	solution	løsning	42	47
30	public transport	kollektivtrafikk	3	3	agriculture pol.	landbrukspol.	4	7
31	country	land	171	164	mental	psykisk	10	12
32	housing	bolig	15	12	county council	fylkeskommune	17	21
33	transition	omstilling	11	11	farmer	bonde	6	8
34	exactly	nettopp	60	57	point	peke	29	35
35	national	nasjonal	36	36	task	oppgave	38	42
36	shipping	skipsfart	4	3	parent	forelder	20	21
37	talk	snakk	46	43	centrist	sentrumsparti	4	5
38	maritime	maritim	4	4	energy	kraft	18	20
39	green	grønn	10	8	contribute	bidra	64	66
40	Burma	Burma	1	0	service	tjeneste	27	30
41	money	penge	62	61	clear	tydelig	22	24
42	crime	kriminalitet	11	13	young	ung	23	24
43	free	fri	24	21	area	areal	3	6
44	party	parti	80	74	defence	forsvar	28	36
45	pension	pensjon	7	6	reduce	redusere	54	58
46	poor	fattig	12	10	Nordic	nordisk	20	20
47	private	privat	52	50	Husbanken	Husbanken	4	5
48	district	bydel	2	1	agriculture	jordbruk	4	8
49	Telemark	Telemark	2	3	addicted	avhengig	16	18
50	inmate	innsatt	3	2	meaning	mening	24	26

Note: This table shows the 50 most polarizing words by urbanicity status from our baseline model which includes political bloc, parliamentary committee, and parliamentary session fixed effects. For each word, we also report the number of occurrences per 100,000 words in the raw data (before feature selection and without controls).

Table A.8: Most polarizing words across age groups

Rank	Most polarizing for old				Most polarizing for young			
	English	Norwegian	#Old	#Young	English	Norwegian	#Old	#Young
1	collaboration	samarbeid	62	43	Norway	Norge	176	191
2	Nordic	nordisk	26	14	policy	politikk	61	84
3	debt	gjeld	209	191	party	parti	70	85
4	employment	arbeid	118	103	choose	velge	40	51
5	mention	nevne	56	43	people	folk	67	80
6	emphasize	understreke	54	43	school	skole	62	71
7	development	utvikling	86	74	money	penge	56	67
8	country	land	175	161	Norwegian	norsk	197	206
9	billion	mill	58	50	job	jobb	36	44
10	area	område	107	100	trade	handel	31	37
11	business	næringsliv	46	45	type	type	24	32
12	research	forskning	30	24	view	syn	40	49
13	patient	pasient	30	18	political	politisk	71	79
14	NOK	kr	102	103	politician	politiker	17	23
15	available	foreligge	26	20	entail	bety	53	62
16	treatment	behandling	66	56	children	barn	80	86
17	point	peke	36	27	decrease	fall	51	57
18	expression	uttrykk	33	28	only	eneste	23	29
19	Østfold	Østfold	6	3	relation	forhold	120	127
20	business	virksomhet	35	30	cut	kutte	3	4
21	competence	kompetanse	29	22	stand	stå	139	145
22	firm	bedrift	52	54	tie	knytte	34	38
23	Nordic region	norden	7	4	manage	klare	13	18
24	hospital	sykehus	27	19	introduce	innføre	23	30
25	regional	regional	15	10	discuss	diskutere	22	27
26	basis	grunnlag	48	44	experience	oppleve	32	37
27	European Council	europaråd	4	2	situation	situasjon	87	90
28	workplace	arbeidsplass	36	35	pupil	elev	23	27
29	director	direktør	29	28	ask/quiet	still	37	40
30	association	forbindelse	53	47	kindergarten	barnehage	17	23
31	project	prosjekt	40	32	difference	forskjell	24	29
32	management	forvaltning	14	11	concrete	konkret	27	33
33	predator control	felle	31	27	challenge	utfordring	51	53
34	seafarers	sjøfolk	4	2	discussion	diskusjon	14	18
35	develop	utvikle	33	26	work	jobbe	14	18
36	church	kirke	12	9	worklife	arbeidsliv	18	23
37	agriculture	landbruk	16	12	contribute	bidra	63	66
38	knowledge	kunnskap	24	20	growth	vekst	23	30
39	defence	forsvar	42	21	Oslo	oslo	35	38
40	European	europaisk	14	11	vote/right	stemme	28	31
41	illness	sykdom	10	6	teacher	lærer	17	20
42	expansion	utbygging	21	17	tax	skatt	14	23
43	county council	fylkeskommune	21	16	scheme	opplegg	13	21
44	Hedmark	Hedmark	4	3	simple	enkel	27	31
45	however	imidlertid	36	30	point	poeng	12	17
46	international	internasjonal	51	47	conservative	borgerlig	12	18
47	Northern territory	nordområde	5	3	challenge	utfordre	9	12
48	county	fylke	28	22	housing	bolig	11	16
49	value creation	verdiskaping	10	8	glad	glad	40	42
50	sami	samisk	6	6	private	privat	49	54

*Note: This table shows the 50 most polarizing words by age groups from our baseline model which includes political bloc, parliamentary committee, and parliamentary session fixed effects. For each word, we also report the number of occurrences per 100,000 words in the raw data (before feature selection and without controls).*

Table A.9: Most polarizing words across father’s occupation

Rank	Most polarizing for white-collar				Most polarizing for blue-collar			
	English	Norwegian	#White	#Other	English	Norwegian	#White	#Other
1	Norway	norge	195	164	debt	gjeld	199	201
2	children	barn	81	84	view	syn	41	51
3	policy	politikk	77	66	emphasize	understreke	46	53
4	school	skole	70	58	positive	positiv	44	50
5	Norwegian	norsk	208	192	treatment	behandling	58	67
6	kindergarten	barnehage	19	20	situation	situasjon	87	91
7	pupil	elev	28	19	district	distrikt	18	28
8	teacher	lærer	21	13	self	sjøl	3	22
9	of course	selsagt	27	19	view	lys	24	28
10	gladly	gjerne	40	32	agriculture	landbruk	11	20
11	increase	øke	149	144	understanding	oppfatning	22	25
12	worklife	arbeidsliv	20	20	patient	pasient	20	31
13	municipality	kommune	116	135	area	område	101	108
14	private	privat	53	48	association	sammenheng	48	52
15	human	menneske	47	45	answer	svar	30	34
16	vote	stemme	32	26	defence	forsvar	30	33
17	police	politi	27	23	development	utvikling	77	84
18	promote	fremme	56	52	priority	satsing	24	27
19	money	penge	63	60	especially	særdeles	4	7
20	people	folk	74	74	collaboration	samarbeid	53	52
21	Israel	Israel	5	1	if	dersom	44	48
22	firm	bedrift	55	49	job	jobb	38	43
23	sami	samisk	6	6	county	fylke	22	31
24	EF	EF	11	7	connection	forbindelse	48	54
25	cash-for-care	kontantstøtte	5	6	note	merknad	29	33
26	USI	USI	14	10	challenge	utfordring	52	52
27	child prot. services	barnevern	10	12	available	foreligge	22	24
28	woman	kvinne	28	32	Hedmark	Hedmark	2	6
29	stroke	slag	21	18	industry	industri	23	30
30	refugee	flyktning	9	7	old	gammal	25	29
31	fund	fond	9	7	forest	skog	4	7
32	row	rekke	42	39	create	skape	61	62
33	NRK	NRK	9	8	however	imidlertid	33	33
34	development aid	bistand	15	10	ideal	ideell	5	7
35	Bergen	bergen	9	7	agriculture	jordbruk	4	7
36	party	parti	81	73	nor	ei	4	8
37	quota	kvote	4	4	naturally	naturligvis	4	7
38	pay	betale	30	27	case	tilfelle	31	34
39	gas power plant	gasskraftverk	7	5	scheme	ordning	47	52
40	energy	energi	11	8	Nordland	nordland	4	7
41	exactly	nettopp	61	55	business	næring	25	32
42	economy	økonomi	38	34	activity	aktivitet	16	18
43	simply	simpelthen	2	0	solution	løsning	44	44
44	whether	hvorvidt	9	8	young	unge	4	7
45	parent	forelder	21	20	postal service	postverk	2	4
46	church	kirke	11	9	Svalbard	Svalbard	4	5
47	atom weapon	atomvåpen	5	4	Telemark	Telemark	2	4
48	market	marked	27	23	wolf	ulv	2	3
49	corporation	selskap	31	32	relations	forhold	123	126
50	point	peke	32	31	claim	hevde	13	15

*Note: This table shows the 50 most polarizing words by occupational background from our baseline model which includes political bloc, parliamentary committee, and parliamentary session fixed effects. For each word, we also report the number of occurrences per 100,000 words in the raw data (before feature selection and without controls).*

## Appendix B: Example speeches

To illustrate the feature selection process described in section 3.1 we provide two examples. First, consider the following speech by Per Sandberg (Progressive Party) held April 24, 2013.<sup>37</sup> This speech lasted 64 seconds and consisted of 147 words. After pre-processing, we are left with 48 words indicated in boldface in the excerpt below:

*Jeg blir litt **overrasket** når justisministeren **påpeker** at man har **redusert køen**. Ja, det er **enkelt** å **redusere køen** når man **slipper de kriminelle ut** og gir dem **strafferabatt**, for da blir det **ledig kapasitet** til å fylle opp med nye. Det er også **betenkelig** at justisministeren **peker** på **økt kapasitet** i nordregionen fordi man **åpner** for **elektronisk soning**. Det er jo ikke en **økt kapasitet** i fengselsinstitusjonene, det er bare en **økt kapasitet** gjennom å **øke antall kriminelle** som er ute på **åpen soning**. **Situasjonen i fengslene** begynner å bli **alvorlig**. Det **pekes** på at man i mye større grad må bruke verneveste og **skjold** i hverdagen. Man **finner** i større grad **ulike våpen**, **hjemmesnekrede våpen**, i **fengslene**. **Situasjonen er alvorlig**. Så mitt siste spørsmål til ministeren er: Vil ministeren nå **garantere** for at det ikke skjer **uheldige episoder** på grunn av **nedbemanning** i norske fengsler?*

As explained in section 3.1, we lemmatize all words to allow several versions of a word to be analyzed as one. Here is the Sandberg speech after pre-processing and lemmatization:

*overrasker påpeke redusere kø enkel redusere kø slipp kriminell ut strafferabatt ledig kapasitet betenkelig peke øke kapasitet åpne elektronisk soning øke kapasitet øke kapasitet øke antall kriminell åpen soning situasjon fengsel alvorlig peke skjold hverdag finn ulik våpen våpen fengsel situasjon alvorlig garantere uheldig episode nedbemanning norsk fengsel*

Translated to English:

*surprise point reduce queue simple reduce queue drop criminal out penalty discount free capacity questionable point increase capacity open electronic imprisonment increase capacity increase capacity increase number criminals open imprisonment situation prison serious point shield everyday find different weapon weapon prison situation serious guarantee unfortunate episode downsizing Norwegian prison*

---

<sup>37</sup>A video of this speech is available at <https://www.stortinget.no/no/Hva-skjer-pa-Stortinget/Videoarkiv/Arkiv-TV-sendinger/?mbid=/2013/H264-full/Storting/04/24/Stortinget-20130424-095410.mp4&msid=862&meid=9419>.

As a second example, we use a 33-second-long speech from Henrik Asheim (Conservatives) from March 18, 2015.<sup>38</sup> This speech consisted of 58 words before pre-processing. After pre-processing, we are left with 25 words, which we again indicate in boldface in the excerpt below:

*Mitt spørsmål til kunnskapsministeren lyder som følger: **Nasjonale** prøver er et viktig **styringsverktøy** i **skolen**. Samtidig er det slik at **stadig flere skoler** tar i bruk **iPad** i **opplæringen**. Dette **stiller nye krav** til **gjennomføringen** av **nasjonale** prøver og **kartleggingsprøver**. **Enkelte skoler** har meldt at det ikke er mulig å **gjennomføre** disse prøvene på **iPad**, noe som **vanskeliggjør gjennomføringen** av prøvene for **elever** og **lærere**. Hvordan vil statsråden sørge for at det blir **teknisk** mulig å **gjennomføre nasjonale** prøver på **iPad**?*

After pre-processing and lemmatization:

*lyd nasjonal styringsverktøy skole stadig skole iPad opplæring stille krav gjennomføring nasjonal kartleggingsprøve enkelt skole gjennomføre iPad vanskeliggjøre gjennomføring elev lærer teknisk gjennomføre nasjonal iPad*

Translated to English:

*sound national "management tool" school ever school iPad training set requirements implementation "mapping test" single school implement iPad complicate implementation student teacher technically implement national iPad*

---

<sup>38</sup>A video of this speech is available at <https://www.stortinget.no/no/Hva-skjer-pa-Stortinget/Videoarkiv/Arkiv-TV-sendinger/?mbid=/2015/H264-full/Storting/03/18/Stortinget-20150318-095457.mp4&msid=9817&meid=9639>