Partisan Conflict in Nonverbal Communication

Mathias Rask[†] Frederik Hjorth[‡]

Conditionally accepted, *Political Science Research and Methods*

This version: March 20, 2025

In multiparty systems, parties signal conflict through communication, yet standard approaches to measuring partisan conflict in communication consider only the verbal dimension. We expand the study of partisan conflict to the nonverbal dimension by developing a measure of conflict signaling based on variation in a speaker's expressed emotional arousal, as indicated by changes in vocal pitch. We demonstrate our approach using comprehensive audio data from parliamentary debates in Denmark spanning more than two decades. We find that arousal reflects prevailing patterns of partisan polarization and predicts subsequent legislative behavior. Moreover, we show that, consistent with a strategic model of behavior, arousal tracks the electoral and policy incentives faced by legislators. All results persist when we account for the verbal content of speech. By documenting a novel dimension of elite communication of partisan conflict and providing evidence for the strategic use of nonverbal signals, our findings deepen our understanding of the nature of elite partisan communication.

[†]Corresponding author. PhD Student, Department of Political Science, Aarhus University

[‡]Associate Professor, Department of Political Science, University of Copenhagen

Introduction

On January 19, 2004, Vermont Governor Howard Dean doomed his presidential campaign. In an attempt to rally his supporters after a disappointing third place in the Iowa caucuses, Dean shouted a list of the primary states ahead followed by a loud, primal yell, now known as the 'Dean Scream'. Its subsequent internet virality is widely perceived as a contributing factor in ending Dean's presidential aspirations. The use of vocal style in politics can also be more deliberate. After becoming the United Kingdom's prime minister, Margaret Thatcher famously went through extensive voice coaching, dramatically altering her vocal style in order to present a more powerful persona. The 'Dean Scream' and Thatcher's engineered voice change illustrate the role of vocal style, and the nonverbal dimension more generally, in elite political communication.

Although nonverbal communication is broadly understood to matter in the abstract, researchers almost exclusively study the verbal dimension of elite political communication in practice (Damann et al. 2023). Most notably, a rich literature uses parliamentary speech as a window to party competition, particularly in systems where party cohesion masks within-party differences. This work typically builds on scaling methods using speech word counts (Slapin and Proksch 2008; Lauderdale and Herzog 2016; Hjorth et al. 2015), and more recently word embeddings (Rheault and Cochrane 2020) to capture parties' and individual legislators' ideological positions. Another line of work uses statistical models (Gentzkow et al. 2019) and machine learning classifiers (Peterson and Spirling 2018) to measure polarization from word choice. Finally, scholars have used sentiment and content analysis to uncover non-positional dimensions of partisan conflict (Proksch et al. 2019). While studies in this vein tap into different aspects of partisan conflict, e.g. positional vs. non-positional (Skytte 2021; Bjarnøe et al. 2023; Serra 2010; Jung and Tavits 2021), they share a focus on word choice, i.e. the verbal dimension of speech.

In this paper, we expand the study of elite partisan conflict to the nonverbal dimension. We do so by developing and validating a measure of nonverbal signaling of conflict based on variation in a speaker's emotional arousal, as indicated by changes in a speaker's vocal pitch relative to their own baseline. We test our approach using audio data from more than two decades of parliamentary debates in Denmark, the largest collection of natural audio in political science. To preview our findings, we find that emotional arousal measured using vocal pitch closely tracks lines of partisan polarization and that it predicts subsequent partisan behavior in the form of legislative voting. We also find that legislators use heightened arousal in more visible and high-profile debates and when addressing parties with greater policy bargaining leverage, both consistent with legislators using nonverbal communication strategically to signal to voter constituencies and to pressure pivotal parties.

We make three distinct contributions. First, we document a novel dimension by which partisan elites communicate conflict to citizens. This finding challenges the prevailing near-exclusive focus on verbal communication in existing research and highlights the need for multimodal studies of elite political communication.

Second, we extend the emergence of "audio as data" methodology to a novel domain. Earlier work has used audio data to study topics such as judicial decision making (Dietrich et al. 2019), oral court arguments (Knox and Lucas 2021), and gender representation (Dietrich et al. 2019; Rittmann 2023). However, we are the first to use audio data to study multiparty conflict.

Third, and more substantively, our findings shed new light on the strategic nature of elite nonverbal communication. To reiterate, we find that changes in vocal style are highly predictable from legislators' vote-seeking and policy-seeking motives, indicating that legislators use vocal style strategically to further their political objectives. Hence, our findings indicate that deliberate, strategic use of nonverbal communication (as in the case of Margaret Thatcher) is widespread in our empirical setting, and spontaneous, non-strategic use (as in the case of Howard Dean) is much less so. Notably, this conclusion cuts against some earlier work seeing vocal style as beyond the speaker's control (Dietrich et al. 2019). This difference is plausibly attributable to differences in speakers' institutional constraints, but our findings nevertheless add to our understanding of the nature of elite nonverbal communication. We revisit this question in the concluding section.

We proceed as follows: We elaborate on each of our contributions, situating them in the existing literature. We present our measurement approach, data, and measures for studying the nonverbal dimension of elite partisan conflict. We develop four hypotheses, which we evaluate in the results section. In the concluding section, we discuss the implications of our findings for future work.

Nonverbal Communication of Partisan Conflict

Political scientists have developed an array of tools for characterizing political conflict, particularly along partisan lines. The two most prominent are, arguably, roll calls and parliamentary speeches. The former has been used predominantly in Congressional scholarship using DW-NOMINATE scores to study positional polarization between Republicans and Democrats (e.g., McCarty et al. 2016) but also in multiparty contexts such as the European Parliament (e.g, Høyland 2010). Approaches based on parliamentary speech have documented rising partisan polarization reflected in word choice (Peterson and Spirling 2018; Gentzkow et al. 2019). While communication-based measures of partisan polarization accommodate the shortcomings of roll-call votes, existing measures have strictly relied on the verbal content of speech, i.e. the words used by partisan elites rather than the style in which they are used. Although a meaningful reduction from a methodological point of view, this narrow focus on verbal content limits our substantive understanding of the nature of partisan polarization in elite communication.

Although the focus on verbal content is substantively and analytically reasonable in many cases, it ignores a significant dimension of human communication. Most importantly, it strips away *nonverbal* elements of speech as an important marker of interpersonal conflict (Deutsch et al. 2011). Nonverbal speech includes aspects like intonation, volume, and accent, commonly referred to as paralinguistic cues (e.g. Scherer et al. 2003). This omission is remarkable given that the centrality of nonverbal communication in human and social interaction has been firmly established in thousands of linguistics and psychology studies.[1] This literature shows that listeners rely on speakers' vocal cues to make inferences about speakers' emotional state, intentions, and character traits (Zuckerman and Driver 1989; Laustsen et al. 2015; Banse and Scherer 1996; Owren and

[1] A Google Scholar search for "paralinguistic" returns $\approx$ 98,500 hits in December 2023.

Bachorowski 2007; Anderson et al. 2014; Scherer et al. 1984, 2003).[2]

We build on earlier work, including a substantial body of experimental evidence, that documents the distinct role played by nonverbal communication in shaping listeners' perceptions and evaluations. In a validation study, Cochrane et al. (2022) show that human coders consistently infer sentiment (i.e. positive vs negative) from text and audio clips, but arousal (i.e. how activated a speaker is) is not detected reliably from text, only from audio. Moving this fundamental insight to the political domain, a small number of studies examine how voters' evaluations of candidate traits are affected by voice characteristics, with lower-pitched candidates being rated as more competent and receiving more votes than higher-pitched candidates (Klofstad et al. 2012; Klofstad 2016; Tigue et al. 2012; Touati 1993; Cinar and Kıbrıs 2023). Recently, Damann et al. (2023) developed a framework to study the causal effects of multimodal political data sources, such as campaign speeches. Using the framework, they show that vocal delivery matters for voters' impression and evaluation of political candidates, even when verbal expressions are held constant.

This experimental body of work mostly considers between-speaker differences in voice characteristics in a small set of audio recordings, but a burgeoning literature is studying within-speaker changes in nonverbal expressions using massive audio collections. For example, Dietrich et al. (2019) and Rittmann (2023) show that changes in legislators' vocal pitch contain information about a legislator's issue engagement. Another line of work finds that political candidates strategically shift their rhetorical style to align with the demands of their audiences by lowering and heightening their phonetic articulation of vowels (Neumann 2019). Finally, nonverbal speech characteristics convey the attitudes of US Supreme Court Judges (Knox and Lucas 2021) and their subsequent voting behavior (Dietrich et al. 2019).

Voicing Partisan Conflict

We use these diverse sets of literature as our point of departure to theorize how legislators use nonverbal communication to signal partisan conflict. We focus on a particular aspect of nonverbal communication: the vocal dimension. To be sure, nonverbal communication also involves

[2]See Knox and Lucas (2021) for a review.

non-vocal features such as facial expressions, gestures, and body language (Boussalis et al. 2021; Neumann et al. 2022; Joo et al. 2019). Still, vocal communication makes up a significant part of nonverbal communication (Patel and Scherer 2013) and is particularly relevant in understanding inter-human conflict (Deutsch 1973).[3] We refer to the vocal dimension as *nonverbal speech* throughout and use *nonverbal signaling* to denote situations where nonverbal characteristics of a speech contain signals about partisan conflict.

We start from the simple observation that a higher voice is associated with conflict throughout social life. When having emotionally charged discussions with family members, debating politics with friends, or protesting in the streets, humans heighten their voices when they disagree. Intuitively, this happens since disagreement and conflict involve an emotional component that makes a speaker more emotionally activated, reflected in, e.g. a heightened voice. (We refer to 'voice in an auditory sense (e.g. loudness and pitch) and not in the metaphorical sense of being heard or represented (e.g. Mansbridge 1999)). In the following, we use the term 'emotional arousal' (or, interchangeably, simply 'arousal') to denote any perception of a speaker being more or less emotionally activated. We stress that 'emotional' in this context does not imply autonomic, i.e. arousal can potentially be employed strategically.

The association between a speaker's emotional arousal, whether strategic or not, and nonverbal conflict signaling is likely to hold in the political domain as well. In the following, we focus on this link as it manifests in speeches given in parliamentary debates, which legislators use primarily to signal the positions of the parties on issues that are up for discussion (Proksch and Slapin 2012; Bäck et al. 2021). Since most of the deliberations and negotiations over bills take place in committees and behind closed doors, the parliamentary debates generally serve to showcase policy positions and arguments publicly, and to highlight partisan differences (Laver et al. 2021), even when parties are not ideologically distinct (Kosmidis et al. 2019).

When should we expect the emotional arousal of a legislator to convey signals of partisan

[3] A multimodal approach combining text, audio, and video in studying political conflict is an important next step in unpacking the dimensions of political conflict. A multimodal approach has been pursued by Boussalis et al. (2021) in another domain, studying candidates' emotional displays and voters' reactions to them.

conflict? We focus on two theoretically plausible drivers of partisan conflict – *polarization* and *policy disagreement* – that may cause variation in a speaker's emotional arousal and discuss and hypothesize how each likely manifests in nonverbal communication. Although legislators can be aroused for other reasons than partisan conflict, this motive is likely to dominate other drivers (e.g., issue engagement) in certain parliamentary speeches, such as dyadic exhanges. We use the term 'dyadic exhanges' to refer to speeches in which the speaker addresses a single party by name or an individual legislator from that party. To fix terminology, we call the party of the former the *speaker party* and the latter the *target party*.

First, we expect arousal to reflect prevailing patterns of partisan polarization. Across dyadic exchanges, interactions between highly polarized pairs of speakers, as indicated by their partisanship, should be more conflictual on average. To the extent that partisan polarization is a driver of emotional arousal in the context of dyadic exchanges, this should then be reflected in legislators' nonverbal signals. In European multiparty systems, party competition is generally structured by multiparty blocs (Bale 2003), which is also true of the Danish case (Kosiara-Pedersen and Kurrild-Klitgaard 2018). As a result, we test the first hypothesis based on party blocs rather than individual parties, but the approach easily generalizes to non-bloc settings.[4] Hence, we expect:

> H_1: Emotional arousal is higher in speeches with outbloc target parties than in speeches with inbloc target parties.

Second, we consider policy disagreement a source of conflict. To form this expectation, we turn to the coalition literature, which has shown a close connection between intra-coalition conflict and the fate of bills: Bills on which coalition parties disagree are introduced later to the agenda relative to when coalition parties are united (Martin and Vanberg 2004), take longer to pass (Martin and Vanberg 2011), and are subject to greater scrutiny by non-coalition partners (Fortunato et al. 2019; Behrens et al. 2023; Wonka and Göbel 2016). Moreover, policy disagreement is known to be reflected in verbal expressions: The sentiment expressed by the opposition predicts whether

[4] The use of blocs to capture the ideology of legislators in a multiparty setting like the Danish parliament is also used by Laustsen and Petersen (2017).

the bills are passed unanimously (Proksch et al. 2019). While the government-opposition divide generally dominates legislative work (e.g. Hix and Noury 2016), conflict should also arise at the level of each bill (Proksch et al. 2019). Specifically, regardless of the ideological distance and the government-opposition status of the parties, interactions between speakers who disagree on a bill should be more conflictual on average than interactions between speakers who agree. To the extent that policy conflict is a driver of emotional arousal in dyadic exchanges in parliamentary debates on bills, this should be reflected in legislators' nonverbal signals. Hence, we expect:

> H_2: Emotional arousal is higher in speech directed at target parties with whom they disagree on a bill than in speech directed at target parties with whom they agree.

At this point, it can not be ruled out that H_2 is just a downstream cause of the stable bloc polarization tested formulated in H_1. To rule out that a relationship with policy disagreement does not merely reflect pre-existing polarization between parties, we add a test of H_2 including dyad fixed effects, which controls for differences in bloc affiliation and any other stable dyad-level differences in party alignment.

The Strategic Use of Nonverbal Signals

The first two hypotheses predict that legislators' nonverbal conflict signals, in the form of heightened emotional arousal, systematically reflect partisan polarization and policy disagreement. Still, they leave open the question of whether it is strategic. We now turn to this question, developing two hypotheses that test key observable implications of strategically used nonverbal conflict signals. We first explain why signaling conflict in parliamentary debate can exert pressure on a target party. We then develop specific expectations of which parties are strategically important to pressure and under which conditions.

To explain how nonverbal conflict signals can put pressure on target parties, we highlight the role of selective media uptake. News media are more likely to cover conflictual and emotional interactions (Schulz 2007; Dietrich et al. 2018; Gennaro and Ash 2023), and conflict often appears as a criterion in its own right in contemporary typologies of news selection criteria (Harcup and

O'Neill 2017). As a consequence, conflict signals increase the probability that an exchange in parliament will be picked up in news media. This heightened visibility can, in turn, put pressure on the target party to justify or potentially reconsider its position on the agenda item.

Crucially, legislators' ability to exploit the media channel varies across debates. The legislative calendar features a subset of *high-profile debates*, most notably the opening and closing debates in each parliamentary term. These debates are characterized by full attendance, a focus on principled policy debate rather than lawmaking, and most importantly significantly increased media coverage. Given the heightened availability of the media channel, we expect legislators to concentrate expressions of conflict in these debates strategically:

H_3: Emotional arousal is higher in high-profile than low-profile debates.

Our third hypothesis corresponds closely to the primary hypothesis in Osnabrügge et al. (2021) that legislators strategically employ emotive language in high-profile debates where legislative rhetoric is more likely to reach voters. This expectation is robustly supported by showing that the verbal content of speech varies systematically across debate types in the UK House of Commons. Our third hypothesis articulates our expectation that the strategic language use identified by Osnabrügge et al. extends to the nonverbal dimension. To isolate the distinct role of the nonverbal signals, we test H_3 using covariates capturing the verbal use of emotive language.

Strategic use of nonverbal communication has implications not only for when legislators signal partisan polarization but also for which partisan outgroups are targeted. Whereas H_3 captures legislators' vote-seeking motives (Mayhew 1974), our next expectation captures policy-seeking motives in the context of parliamentary debate. Given policy-seeking motives, we expect that legislators focus on target parties with the greatest influence on policy-making. Specifically, we expect parties to concentrate conflict signals on target parties with greater *bargaining leverage*, since they are most likely to ultimately affect policy, through either the existing governing majority or an alternative majority. This leads to our fourth and final hypothesis:

H_4: Emotional arousal is positively associated with the target party's bargaining leverage.

Following traditional conceptualizations (Shapley and Shubik 1954), we understand bargaining leverage as a function of parties' probability of participating in an alternative government, were one to form. We measure bargaining leverage using a recently developed approach (Kayser et al. 2023), (see 'Data and Methods' below).

Importantly, the prediction in H_4 cuts against what a theory of non-strategic nonverbal communication would predict. If nonverbal communication is primarily or entirely a non-strategic, affective reflex, nonverbal signals of conflict should be greater in speeches directed at more extreme targets, which are more likely to arouse anger (Webster 2021). In Appendix A, we show that bargaining leverage is higher for mainstream parties and lower for challenger parties at the ideological extremes. Hence, if nonverbal signaling of conflict were primarily an affective reaction to extreme parties, we should expect the opposite of the predicted association. Hypothesis 4 therefore aims to discriminate between the predictions of strategic vs. non-strategic theories of nonverbal communication.

We stress that these hypotheses are implicitly causal, i.e. they reflect our theoretical understanding of nonverbal communication as a reflection of partisan conflict. In some cases, the temporal order of variables rules out reverse causation (e.g., party affiliation temporally precedes nonverbal communication), but confounding remains a concern. As discussed below, we introduce a rich set of covariates to address confounding concerns. That said, causal identification is ultimately limited by the fact that policy conflict is not randomly assigned.

How Partisan Conflict is Reflected in Audio Data

Analyzing how partisan conflict is signaled in politicians' nonverbal communication is challenging. Whereas verbal measures such as negativity or scaling estimates can be derived from speech transcripts, nonverbal features of speeches are generally stripped away in transcription. Consequently, text-only transcripts are often ill-suited to study nonverbal dimensions of speech. Second, text-based measures of nonverbal communication rely on the coding procedures of data sources. For instance, Imre et al. (2023) develop a novel measure of 'coalition mood' based on applause patterns in parliamentary debates between coalition partners in Germany and Austria. While valuable, this measure depends on the availability of stenographic protocols marking nonverbal communication in party-to-party interactions.

To deal with these shortcomings, we turn to audio recordings of political speeches. Although recordings are widely and publicly available across many political institutions, they have received only scant attention from political scientists (though see Dietrich et al. 2019; Rittmann 2023; Neumann 2019; Knox and Lucas 2021). In addition to conveying the verbal content of speech (i.e. spoken words), audio contains information that goes beyond what we can infer from words alone, and most importantly, it conveys information on the emotional arousal of a speaker (Cochrane et al. 2022).

How Pitch Reflects Emotional Arousal

Our indicator of emotional arousal is based on changes in a speaker's vocal pitch, the perceptual analog of the fundamental frequency ($F0$) of a waveform (Rabiner and Schafer 2010).[5] The perception of the vocal pitch increases monotonically, but not linearly, with $F0$ such that a voice with a higher $F0$ is perceived as higher and vice versa. Each speaker has a baseline $F0$ which is largely explained by biological and physiological factors such as sex and height (Pisanski et al. 2016; Evans et al. 2006). However, a rich psychological literature on the vocal expression of emotions shows that variation in pitch is a robust and strong indicator of expressed and perceived emotional arousal, independently of the verbal content (Bänziger and Scherer 2005; Scherer et al. 2003; Banse and Scherer 1996).[6]

[5] The $F0$ is a physical property of any sound wave whether it arises from human speech, animal calls, explosions, or traffic noise. In the case of human speech, it is defined as "the number of vibrations per second made by vocal folds to produce a vocalization" (Tusing and Dillard 2000, 150).

[6] The link between pitch and emotional arousal is based on a continuous model of emotions where a human's emotional state can be placed in a two-dimensional space of valence (i.e. sentiment) and intensity (i.e. arousal). Pitch is a measure of the latter and is found to increase in both positive (joy/happy) and negative states (anger, fear, sad). Another model uses discrete emotions such as fear, joy, anger, etc. (Ekman 1992, 1999). However, while pitch is a

Politicians also use other nonverbal signals such as loudness, speech rate, or jeering to convey partisan and policy disagreement. We focus on the vocal pitch for three reasons. First, as already outlined, vocal pitch is a robust indicator of emotional arousal, which we expect to be higher in conflictual and polarizing contexts. Second, vocal pitch is shown to predict legislators' issue engagement and policy priorities in both presidential and parliamentary democracies (Dietrich et al. 2019; Rittmann 2023) and judges' vote intentions (Dietrich et al. 2019). Third, pitch estimation is less sensitive to recording quality than features such as loudness since it depends less on the spectral characteristics of a sound (Vainio et al. 2023). This is particularly important when analyzing nonverbal features of speech over time.

Conflict-driven Arousal in Dyadic Exchanges

To be sure, variation in pitch can reflect other motivations than partisan conflict. To mitigate this issue, we consider only a type of interaction in which partisan conflict motives are likely to dominate. Specifically, we consider dyadic exchanges in parliamentary proceedings, where higher emotional arousal expressed in a speech is more likely to indicate partisan conflict than, for example, issue engagement due to the nature of this type of interaction. The partisan nature of such interactions makes them a prime avenue for the expression of partisan conflict. Hence, when legislators target out-partisans in a dyadic exchange, a heightening pitch is, on average, more indicative of partisan conflict in that specific context. Conversely, when legislators heighten their pitch when mentioning their social groups, this is likely more indicative of issue engagement and group commitments (e.g., Dietrich et al. 2019).

Other Drivers of Emotional Arousal

While we expect conflict motives to dominate variation in arousal in dyadic exchanges, partisan conflict is not the only potential driver of arousal. We illustrate our measurement model in Figure 1 in the form of a directed acyclical graph (DAG).

reliable indicator of how strong (i.e. aroused) emotions are expressed (Banse and Scherer 1996), it is a challenging task to discriminate between discrete emotions from $F0$ contours. For this task, text-based measures achieve better results (Widmann 2021).

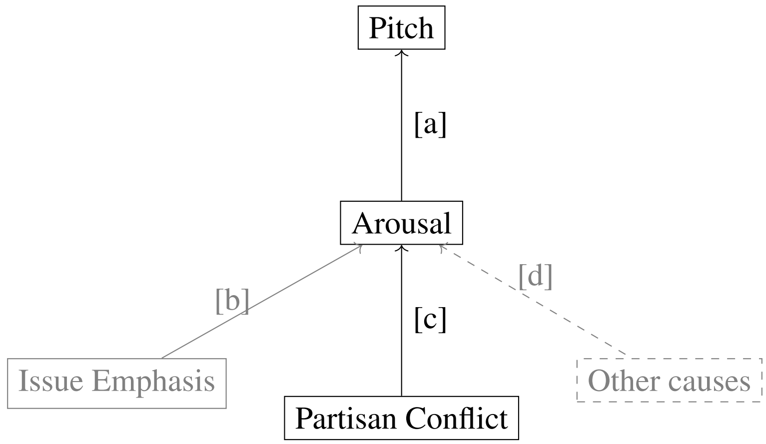


Figure 1: Directed acyclical graph illustrating documented and hypothesized links between pitch, arousal, conflict, and other causes. The link from partisan conflict is highlighted to illustrate that we expect this motive to dominate in the context of dyadic exchanges.

We rely on theory and evidence establishing pitch as a measure of emotional arousal (link [a]), and we expect arousal to be heightened under conditions of partisan conflict (link [c]). At the same time, prior work has found vocal pitch to be informative of other motivations. Most pertinently, Dietrich et al. (2019) and Rittmann (2023) use pitch as an indicator of issue emphasis in parliamentary speech. We illustrate this established link as [b] in Figure 1.

Other work uses pitch as an indicator for other types of motivations, namely voting intentions (Dietrich et al. 2019), and inter-party politics (Arnold and Küpfer 2024). For simplicity, we represent these and other potential drivers of heightened arousal as link [d] in Figure 1. While we theorize that legislators use arousal as a conflict signal in strategic ways (see the presentations of hypotheses 3 and 4), we note that this measurement model in itself does not hinge on that point, i.e. arousal can signal conflict for either strategic or non-strategic reasons.

Standardization of Vocal Pitch

We follow extant work and use a standardized measure at the speech level where the pitch is converted to standard deviations above or below a speaker's average (Rittmann 2023; Dietrich et al. 2019). This is done to parse out heterogeneity arising from speaker-specific voice differences such as physiological and biological factors, akin to the reason for including unit fixed effects in dealing with panel data (Rheault and Borwein 2019) and to take possible measurement error into account (Dietrich et al. 2019). As a consequence, the estimates reflect within-speaker changes in pitch. We compute the baseline of each speaker as the average pitch in all speeches given by a speaker in our corpus (see 'Data and Measures'). Further information about our pitch measure and the implication of standardization is shown in Appendix B.[7]

Yet, standardization is not without drawbacks. By removing speaker heterogeneity, we pay the cost of not being able to examine the role of stable speaker-level differences in vocal pitch. Most obviously, the gender gap in vocal pitch suggests that men and women face very different constraints and roles expectations regarding vocal style (Boussalis et al. 2021). In Appendix C, we show that we find no appreciable effect heterogeneities with respect to gender. The question of individual, including gendered, differences is revisited in the concluding section.

Validation

The link between vocal pitch and emotional arousal has been extensively validated in the psychological literature (see, e.g., SI of Dietrich et al. 2019), but not explicitly in the context of parliamentary speeches and as an expression of partisan conflict. To validate that vocal pitch is a valid measure arousal in such setting, we conduct three validation exercises and an analysis of the measurement error, which we present in additional detail in Appendix D. To summarize the results of the validation analysis, we show that (1) coders are able to consistently and reliably infer a legislator's emotional arousal from speech-level audio recordings, (2) speaker-standardized speech-level estimates of vocal pitch are strongly correlated with the manual arousal codings, and (3) that pitch is negatively associated with text sentiment, consistent with the assumption that variation in arousal in the context of dyadic exchanges reflects conflict.

[7]The unstandardized pitch distribution (panel F2a) has a bimodal shape arising from the physiological differences in the size of the vocal cords between men and women. If we were to interpret changes in absolute differences in pitch, the results would primarily reflect gender differences rather than differences in nonverbal signals of partisan conflict. Standardization effectively removes this heterogeneity, parsing out all time-invariant speaker-level characteristics.

Data and Measures

We validate our approach in the context of parliamentary debates in the Danish parliament, the Folketing. Audio recordings are available for a vast amount of political institutions (e.g. Barari and Simko 2023), but legislatures, in particular, maintain comprehensive archives with more than a thousand hours of recordings. We focus on Denmark since it is possible to obtain digitized recordings spanning more than two decades of debate, longer than any other archive to the best of our knowledge.

Text-Audio Corpus and Alignment

We collect all available recordings of plenary sessions in the Danish parliament from October 2000 to September 2022 covering six national elections (2001, 2005, 2007, 2011, 2015, and 2019), 28 parliamentary terms, and a total of 2,186 debates containing $850,357$ speeches. To obtain transcripts of the recordings, we rely on a combination of ParlSpeech V2 and manually scraped XML files. We follow extant work using parliamentary speeches (e.g., Castanho Silva et al. 2024; Dietrich et al. 2019) and remove shorter speeches from the corpus. Shorter speeches are typically procedural, interjections, and interruptions that carry little substantive information for study of partisan and policy conflict. For the main analysis, we use a threshold of 40 or more words but we show in Appendix E that the results are nearly numerically identical to the choice of threshold.[8] This leaves us with a total of $393,264$ transcribed speeches. As the next step, we align the transcripts with the corresponding audio using the `Python` library *speechannote*.[9] We were able to align text and audio for 96.3 pct. of our speeches (a total of $378,566$). We elaborate on how we construct and preprocess our data in Appendix F.

Legislative Votes

For our second hypothesis, we measure policy conflict based on disagreement in legislative voting. We obtain voting records from an enhanced version of ParlSpeech V2 (Rauh and Schwalbach 2020)

[8] We also remove speeches given by chairs as these contain no substantive information.

[9] The package is expected to be publicly available in medio 2025.

where speeches are linked to legislation. This data is available from November 2007 to September 2022. As a final step, we match each speech containing voting records with our preprocessed transcript using fuzzy string matching based on the Jaro-Winkler (JW) distance. We were able to match 80 pct. of the speeches in the enhanced ParlSpeech V2 to our transcript.

Party Dyads

Three of our four hypotheses regard dyadic exchanges in parliament, i.e. where the speaking legislator addresses a party either by name or in the form of an individual legislator from that party. To identify dyadic speeches, we rely on a dictionary approach containing the names of all parties and legislators (Schwalbach 2023). To be classified as a dyad, the speech must reference one party or one of more of its MPs. We refer to such dyadic exchanges as *party dyads* (or, interchangeably, simply 'dyads'). We identify dyads in 38.9. pct of all aligned speeches. Appendix G shows how the share of party dyads varies over time, and how this relates to the development of inbloc and outbloc dyads.[10]

We use the party dyads to define our main predictors in H_1 and H_2. For H_1, we use a binary measure based on a party's bloc affiliation. The Danish party system is characterized by two blocs, one denoted as left and the other as right. While each bloc contains substantial party differences, this has typically dominated party competition at the national level (Kosiara-Pedersen and Kurrild-Klitgaard 2018). For H_2, we also use a binary measure based on whether two parties voted together. Importantly, this is conceptually and empirically distinct from H_1 as parties vote across their bloc affiliation.[11] While the share of votes where parties agree is larger within blocs, parties vote between blocs on several occasions (approx. 39 pct.).

High-Profile Debates

Our third hypothesis H_3 predicts that legislators signal higher emotional arousal in debates that generate citizen and media attention. This prediction cuts across both dyadic and non-dyadic exchanges. While most debates are low-profile with a principled focus on law-making, the opening

[10]We illustrate the distribution of dyads disaggregated to the party level in Panel G2a in Figure G2 in Appendix G.

[11]See Panel G2b in Figure G2 in Appendix G for a visualization of the distribution of voting dyads.

and closing debates of parliamentary terms stand out. These debates last an entire day, often over twelve hours, and cover programmatic policy differences rather than specific legislation. Because of their formal status and ideologically charged content, opening and closing debates receive considerable attention from voters, either directly or indirectly through the extensive media coverage of the debates (Osnabrügge et al. 2021). Hence, we define opening and closing debates as high-profile debates here and use the remaining set of debates as the reference category. We collect the dates of these debates, match them to our dataset, and finally generate an indicator of whether a speech is given in a low- or high-profile debate (a total of 6,116 party dyads).

Target Bargaining Leverage

Our fourth hypothesis H_4 predicts that legislators use heightened emotional arousal to put pressure on parties with higher bargaining leverage. We measure this using a recent approach introduced in Kayser et al. (2023). Briefly put, the authors start from the premise that parties' leverage in the policy process ultimately arises from credible threats to leave the government or the ability to join an alternative government. Building on this notion, Kayser et al. calculate coalition inclusion probabilities (CIPs) using a predictive model with data on historical coalition patterns, party types and ideologies, election results, public opinion polls, and country-level institutional features as inputs. We use CIPs for Danish parties, which Kayser et al. provide at the monthly level starting in 1970. We rely on the static version of the CIP data, which does not rely on information from between-election opinion polls. In Appendix A we present the CIPs for each party from 2000-2019.

Verbal Covariates

To rule out that any observed relationship arises due to a strong correlation between verbal and nonverbal speech features, we generate three text-based measures. One reasonable concern is that our nonverbal measure of partisan conflict is encoded in speech sentiment. To account for this possibility, we first define a measure of sentiment capturing the relative use of positive and negative words in a speech using the Danish sentiment tool Sentida (Lauridsen et al. 2019). A separate concern is that our measure merely tracks emotive rhetoric (Osnabrügge et al. 2021). To address this, we follow the approach suggested by Gennaro and Ash (2022) and construct a continuous measure of emotionality at the speech level as our second measure. A final concern is that variation in our nonverbal measure is explained entirely by the topics discussed in party dyads. If polarizing and conflictual policy topics are discussed more in party dyads with an out-party target than in-party targets, an observed relationship could reflect topic selection rather than partisan conflict. If so, this suggests that issue engagement is a confounder. To account for this, we estimate a Structural Topic Model (Roberts et al. 2014) with $k = 40$ and include the resulting topics as fixed effects. For details on how each verbal covariate is constructed, see Appendix H.

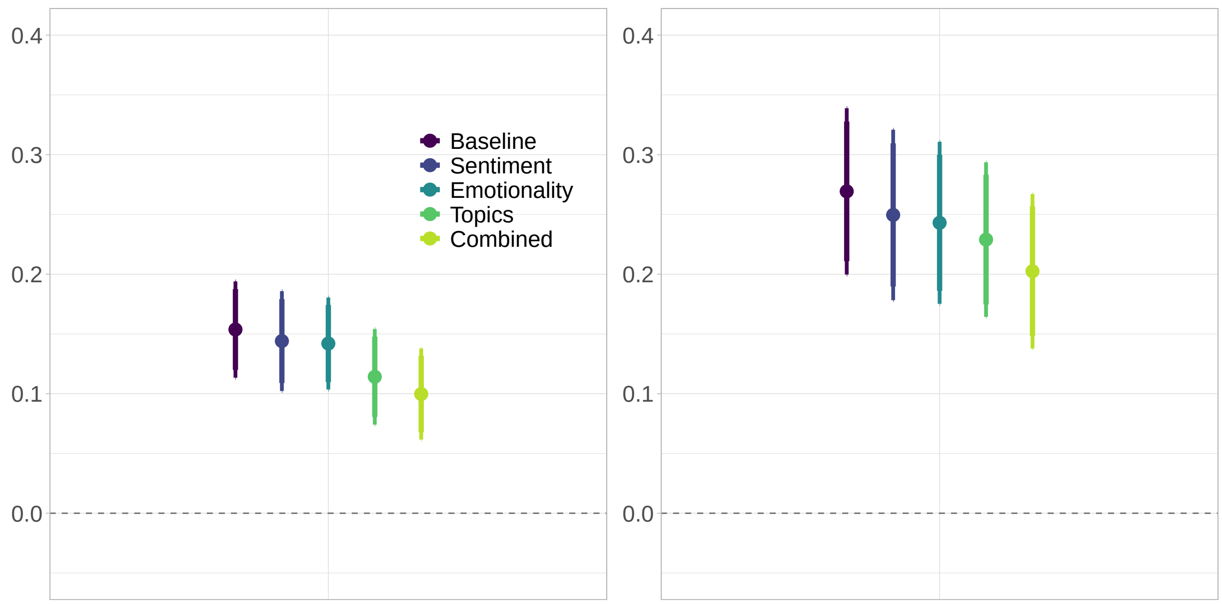
Results

We now present results from tests of each of the four hypotheses. To simplify exposition, we first jointly present the results for H_1 and H_2, then results for H_3 and H_4. For each hypothesis, we estimate a series of linear regression models using the OLS estimator. For H_1, H_2, and H_4, the data consists only of speeches classified as party dyads. The dependent variable in all models is the speaker-standardized speech-level vocal pitch during a speech i by each legislator j. We refer to this simply as 'pitch' or 'vocal pitch'. All regression tables are in Appendix I.

Hypotheses 1 and 2: Polarization and Legislative Voting

We present results for H_1 and H_2 in Figure 2. Panel 2a tests our first hypothesis predicting that legislators speak with a higher pitch in speeches between polarized pairs of legislators. We test this using an indicator of whether a speech is directed at an outbloc or inbloc legislator. The outbloc measure is a binary indicator that takes the value of 1 if the target and speaker blocs differ and 0 otherwise.

Across all five models, we estimate a positive and significant ($p < 0.001$) coefficient. On average, the estimated effect size indicates that legislators speak with a pitch roughly 0.15 standard deviations higher when talking to outbloc legislators than when talking to inbloc legislators. The relationship is robust to the inclusion of covariates capturing text sentiment, emotionality, or speech



(a) Partisan Polarization **(b)** Policy Conflict

Figure 2: Coefficients for partisan polarization (left panel) and policy conflict (right panel) with standardized pitch as the outcome. Predictors are whether the target party is outbloc (left panel) and whether the target party voted differently in a legislative vote on a specific bill. Standard errors are clustered at the dyad level (speaker party $\leftrightarrow$ target party). Thick and thin error bars are the model-specific 90 and 95 pct. confidence intervals respectively. Y-axes are held fixed across the two panels to maximize comparability.

topic fixed effects and is only slightly reduced when all three text measures are included. While sentiment and emotionality have virtually no impact on the estimate, speech topic fixed effects result in a minor decrease but remain highly significant ($p < 0.001$).

Turning to Panel 2b, we test whether changes in pitch also conveys signals of partisan conflict when the conflict concerns policy disagreement and not general polarization. To do this, our second hypothesis turns the focus to legislative bills, where we predict that nonverbal signals also signal a party's bill-level disagreement with other parties. If so, legislators should speak with a heightened pitch when addressing legislators from parties with whom they vote differently.

As for the first hypothesis, we find a positive and significant ($p < 0.001$) coefficient across all five models. The relationship is almost invariant to the inclusion of verbal covariates across the board. Once again, sentiment and emotionality have no impact on the result, but the inclusion of speech topic fixed effects slightly reduces the estimate. Yet as for H_1, it remains highly significant

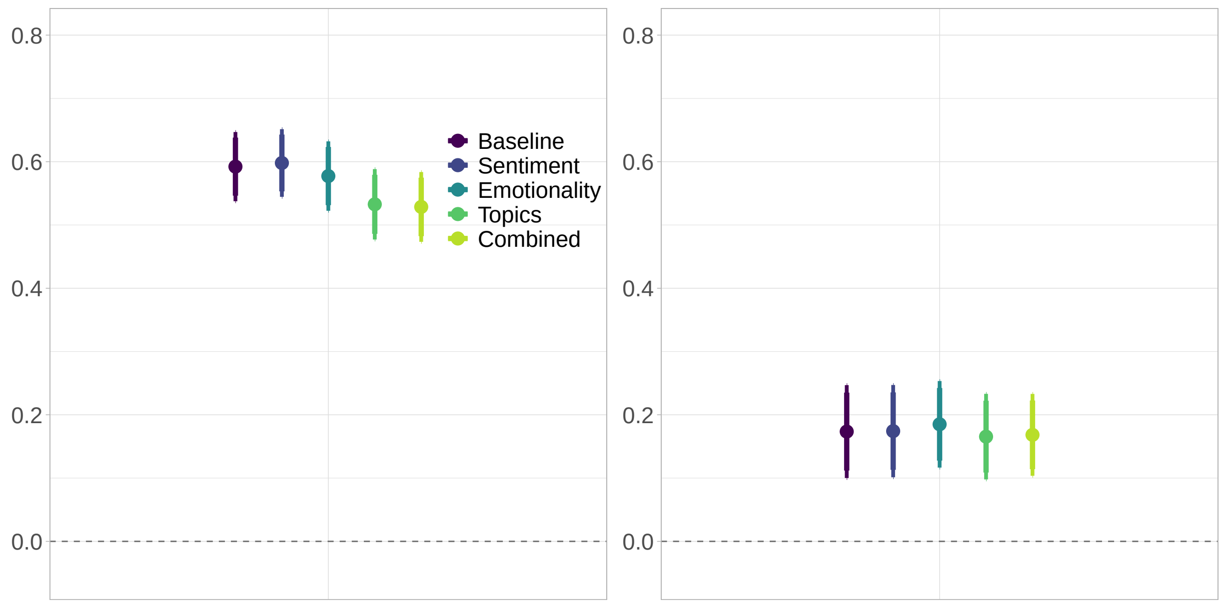
($p < 0.001$). The estimated coefficient is twice the size of the coefficients for H_1 (0.2) in the full model. Together, the results strongly indicate that legislators use nonverbal communication to signal partisan and policy conflict and that it accounts for a distinct dimension of elite partisan polarization compared to what is captured by verbal expressions alone. The coefficients for policy agreement are roughly twice the size for bloc polarization. This suggests that nonverbal signals of partisan conflict primarily reflect disagreement over policy, at least in the parliamentary arena.

To verify that the results in Panel 2b do not simply reflect bloc differences, we also test H_2 using models including dyad fixed effects. We present these models in Table J1 in Appendix J. The coefficient for policy disagreement is robustly significant across these models and retains the entire magnitude (0.2025 compared to 0.2021) of the coefficients shown in Panel 2b. Hence, even considering only variation within the same party dyads, nonverbal conflict signals strongly predict policy disagreement. As an additional robustness check, we use legislative vote margin to measure policy disagreement. In Appendix J, we show that results for H_2 are robust to using this alternative measure. Lastly, to explore the potentially confounding role of time-specific factors, we show results by year and weekday in Appendix J, finding no significant heterogeneity with respect to time.

Hypotheses 3 and 4: Debate Type and Target Bargaining Leverage

We now turn to our third and fourth hypotheses, which are observable implications of a strategic model of nonverbal communication. We use the same specifications with the only change being the predictor in each test. The results are in Figure 3.

Panel (a) reports the results for our third hypothesis, predicting that legislators signal more partisan conflict in debates that attract substantial attention from the general public. We test this by regressing vocal pitch on an indicator of whether the debate is high-profile (= 1) or low-profile (= 0) across both dyadic and non-dyadic exchanges. The estimates reported are the coefficients on this indicator, which capture the average difference between high- and low-profile debates. Consistent with the reasoning in Osnabrügge et al. (2021), we expect a positive estimate if legislators use vocal style strategically. Across the five models, this is also what we find. The coefficient is



(a) Electoral: High- vs. Low-Profile Debates. **(b)** Policy: Target Party Bargaining Leverage.

Figure 3: Coefficients for debate type (left panel) and bargaining leverage (right panel) with standardized pitch as the outcome. The predictors are whether the speech is given in a high- compared to a low-profile debate (left panel) and the policy bargaining leverage of the target party (right panel). Standard errors are clustered at the dyad level (speaker party $\leftrightarrow$ target party). Thick and thin error bars are the model-specific 90 and 95 pct. confidence intervals respectively. Y-axes are held fixed across the two panels to maximize comparability.

large, positive, strongly statistically significant ($p < 0.001$), and largely invariant to the inclusion of verbal covariates. Once again, the estimate is invariant to sentiment and emotionality but is slightly smaller when topics are included in the model. These results also mirror the findings in Osnabrügge et al. (2021), where the relationship is only slightly reduced. In the full model, legislators speak with a .51 standard deviation higher pitch in high-profile debates than in low-profile debates. We interpret this as indicative of a strategic model of nonverbal communication as high-profile debates reach a much larger electoral audience.

Turning to Panel (b) and our final and fourth hypothesis, we find that nonverbal signaling of conflict, as indicated by changes in pitch, correlates positively with a party's bargaining leverage. In the bivariate specification, moving across the full range of bargaining leverage is associated with the standardized pitch rising by nearly 0.168 standard deviations. As for the other hypotheses, the relationship is highly robust to including controls for verbal content, and the coefficient in the specification with the full set of controls remains virtually unchanged. Compared to Hypothesis 1-3, topics have no impact on the relationship. Together, the results for H_3 and H_4 are consistent with a strategic model of nonverbal signaling of conflict, suggesting that legislators deliberately use nonverbal communication to further their electoral or policy goals. In substantive terms, the results across the four hypotheses show that nonverbal communication captures a distinct dimension of partisan and policy conflict, and indicate that legislators use nonverbal communication strategically.

All the estimates above are expressed in standard deviations. To contextualize our results, our baseline estimates across tests correspond to around 9 percent (for the smallest observed differences) to around 33 percent (for the largest observed differences) of the difference in pitch between men and women in our sample. All baseline estimates exceed the estimated 'just noticeable difference' in pitch of around 5 Hz (Liu 2013), i.e. the minimal difference that is discernible to the human ear. In other words, the differences we observe are noticeable, and for the largest estimates amount to around a third of the range in pitch between the average man and woman.

Conclusion and Discussion

Elite partisan conflict is an important feature of democratic systems, yet our understanding of how elites communicate partisan conflict to citizens is limited, partly because existing approaches only consider the verbal content of communication. To expand the study of partisan conflict in elite communication to the nonverbal domain, we examine how elite partisan conflict is reflected in legislators' nonverbal signals. Analyzing audio data from two decades of debates in the Danish parliament, we find that partisan conflict is systematically reflected in a legislator's nonverbal speech signals and that these signals predict subsequent legislative voting. Furthermore, we find evidence consistent with a strategic use of nonverbal communication: Legislators react more strongly to outbloc targets in high-profile debates and when addressing parties with greater bargaining leverage. Importantly, these associations remain largely unchanged when we account for the verbal content of speech, which strongly suggests that nonverbal communication accounts for a distinct dimension of elite communication of partisan conflict.

Some caveats are in order. One set of caveats pertains to our measurement approaches. Our outcome, speaker-standardized vocal pitch, is a somewhat crude measure that may not capture every aspect of nonverbal signaling of conflict. We expect that using a richer set of audio features can provide a more nuanced and fine-graned measure of nonverbal signaling and increase the precision of the measure. Relatedly, conflict signals in a speech are in practice likely to be concentrated in particular sentences or even single words. Our choice to average pitch across the entire speech effectively glosses over this variation. Future research could improve on this aggregation problem, for example by making use of novel methods for multimodal alignment which enable linkage between text and audio data at the word level (Arnold and Küpfer 2024).

A second set relates to our measures of the verbal content of speech. We have employed a rich set of text covariates, but if these measures do not fully capture the relevant dimension in speech verbal content, our control strategy correspondingly fails to fully account for the role of the verbal content of speech. We see the development of approaches to more directly compare the role of nonverbal communication to that of verbal content as an important avenue for future research.

A third set of caveats relates to external validity. The evidence presented here comes from the Danish parliamentary system, which is characterized by, among other features, high party cohesion, relatively high party system fractionalization, and a low level of partisan polarization. Without evidence from other contexts, it is uncertain how well these findings travel to party systems with other characteristics. This is also true of previous work on audio in politics, which mainly relies on the United States as a case (although see Rittmann 2023). On this front, we see promise in the largely non-language-specific nature of nonverbal communication. Whereas text data approaches require a cross-language approach to extend to other contexts, a given measurement strategy based on audio features alone could in principle be directly applied in novel contexts without accounting for language changes (Scherer et al. 2001) at least within non-tonal communities. Like Danish, the majority of Indo-European languages are non-tonal, meaning that variation in intonation does not change the meaning of words.

These caveats notwithstanding, our findings hold important implications for the study of political communication and representation. First, our findings expand the set of known features by which elites communicate partisan conflict to citizens. Specifically, we expand this set to the nonverbal domain, studied through legislators' vocal pitch. This in turn implies that any assessment of partisan polarization among elites based on verbal content alone is incomplete: Even in the hypothetical absence of conflict in the verbal content of speech, elites may still signal conflict by nonverbal means. As a consequence, efforts to encourage elites to engage in more civil, bipartisan, and conciliatory behavior should consider both verbal and nonverbal dimensions of communication.

Second, the indication in our findings that vocal style is employed strategically adds nuance to our understanding of the intentionality of nonverbal communication. Consider for example Dietrich et al. (2019), who work from the assumption that changes in vocal inflections are "beyond their conscious communication" (238). Consistent with this assumption, the authors find that vocal pitch in US Supreme Court Justices' oral arguments predicts voting over and above what can be predicted from other observables. Comparing our empirical setting to that in Dietrich et al., institutional constraints are likely an important moderator. Most obviously, US Supreme Court Justices are lifetime appointees and thus by institutional design immune from reelection incentives. In contrast, we study a parliamentary setting where reelection motives loom large. Our findings indicate that in a competitive parliamentary setting, nonverbal communication is not beyond the realm of conscious communication. Regardless of the role of these plausible institutional moderators, our findings imply that nonverbal communication cannot be assumed to be an 'honest' window into the speakers' true emotional state and that such an assumption would have to be justified in any specific application (see also Rittmann 2023).

Our findings also have methodological implications for experimental political science. At present, survey experimental treatments in political science overwhelmingly consist of text vignettes designed to convey the stimulus of interest (though see Damann et al. 2023). This treatment mode dominates even though the nonverbal dimension of political communication carries significant information, which is not reflected in a text vignette but could be captured in e.g. audio or video snippets. Our findings underscore that researchers have to consider the use of multimodal treatments in survey experimental design.

Fourth and finally, our findings have substantive implications for how individual differences in vocal style condition political representation. Our theoretical framework implies that citizens perceive higher pitch to reflect increased emotional arousal, and our findings and validation analysis are consistent with this dynamic. While our analysis parses out stable individual differences by standardizing pitch within legislators, it is worth considering how a given legislator's communication is affected by their vocal pitch (e.g., Cinar and Kıbrıs 2023). For example, are legislators with a higher baseline vocal pitch generally perceived as signaling more partisan conflict? If citizens fail to adequately correct for elites' vocal style when interpreting their speech, it could lead to biased elite perceptions. Such misperceptions could underpin for example the perception of women in politics as more emotional, a trait ascription often used to dismiss women candidates (Campbell 1994). The role of trait inferences, and potential misperceptions, based on vocal style and nonverbal communication more broadly, is an important topic for future research.

References

Anderson, Rindy C, Casey A Klofstad, William J Mayew, and Mohan Venkatachalam. 2014. "Vocal fry may undermine the success of young women in the labor market". *PloS one* 9 (5): e97506.

Arnold, Christian and Andreas Küpfer. 2024. "Alignment helps make the most of multimodal data".

Bäck, Hanna, Marc Debus, and Jorge M Fernandes. 2021. *The politics of legislative debates*. Oxford University Press.

Bale, Tim. 2003. "Cinderella and her ugly sisters: the mainstream and extreme right in Europe's bipolarising party systems". *West European Politics* 26 (3): 67–90.

Banse, Rainer and Klaus R Scherer. 1996. "Acoustic profiles in vocal emotion expression.". *Journal of personality and social psychology* 70 (3): 614.

Bänziger, Tanja and Klaus R Scherer. 2005. "The role of intonation in emotional expressions". *Speech communication* 46 (3-4): 252–267.

Barari, Soubhik and Tyler Simko. 2023. "Localview, a database of public meetings for the study of local politics and policy-making in the united states". *Scientific Data* 10 (1): 135.

Behrens, Lion, Dominic Nyhuis, and Thomas Gschwend. 2023. "Political ambition and opposition legislative review: Bill scrutiny as an intra-party signaling device". *European Journal of Political Research*.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. "quanteda: An r package for the quantitative analysis of textual data". *Journal of Open Source Software* 3 (30): 774–774.

Bjarnøe, Camilla, James Adams, and Amber Boydstun. 2023. ""Our Issue Positions are Strong, and Our Opponents' Valence is Weak": An Analysis of Parties' Campaign Strategies in Ten Western European Democracies". *British Journal of Political Science* 53 (1): 65–84.

Boussalis, Constantine, Travis G Coan, Mirya R Holman, and Stefan Müller. 2021. "Gender, candidate emotional expression, and voter reactions during televised debates". *American Political Science Review* 115 (4): 1242–1257.

Campbell, Sue. 1994. "Being dismissed: The politics of emotional expression". *Hypatia* 9 (3): 46–65.

Castanho Silva, Bruno, Danielle Pullan, and Jens Wäckerle. 2024. “Blending in or standing out? gendered political communication in 24 democracies”. *American Journal of Political Science*.

Cicchetti, Domenic V. 1994. “Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology.”. *Psychological assessment* 6 (4): 284.

Cinar, Asli Ceren and Özgür Kıbrıs. 2023. “Persistence of voice pitch bias against policy differences”. *Political Science Research and Methods*: 1–15.

Cochrane, Christopher, Ludovic Rheault, Jean-François Godbout, Tanya Whyte, Michael W-C Wong, and Sophie Borwein. 2022. “The automatic analysis of emotion in political speech based on transcripts”. *Political Communication* 39 (1): 98–121.

Damann, Taylor J, Dean Knox, and Christopher Lucas. 2023. “A causal framework for multimodal speech”.

Deutsch, Morton. 1973. *The resolution of conflict: Constructive and destructive processes*. Yale University Press.

Deutsch, Morton, Peter T Coleman, and Eric C Marcus. 2011. *The handbook of conflict resolution: Theory and practice*. John Wiley & Sons.

Dietrich, Bryce, Dan Schultz, and Tracey Jaquith. 2018. “This Floor Speech Will Be Televised: Understanding the Factors that Influence When Floor Speeches Appear on Cable Television”. Technical report, Working Paper.

Dietrich, Bryce J, Ryan D Enos, and Maya Sen. 2019. “Emotional arousal predicts voting on the US Supreme Court”. *Political Analysis* 27 (2): 237–243.

Dietrich, Bryce J, Matthew Hayes, and Diana Z O’Brien. 2019. “Pitch perfect: Vocal pitch and the emotional intensity of congressional speech”. *American Political Science Review* 113 (4): 941–962.

Ekman, Paul. 1992. “Are there basic emotions?”.

Ekman, Paul. 1999. “Basic emotions”. *Handbook of cognition and emotion* 98 (45-60): 16.

Evans, Sarah, Nick Neave, and Delia Wakelin. 2006. “Relationships between vocal characteristics and body size and shape in human males: an evolutionary explanation for a deep male voice”. *Biological psychology* 72 (2): 160–163.

Fortunato, David, Lanny W Martin, and Georg Vanberg. 2019. “Committee chairs and legislative review in parliamentary democracies”. *British Journal of Political Science* 49 (2): 785–797.

Gennaro, Gloria and Elliott Ash. 2022. “Emotion and reason in political language”. *The Economic Journal* 132 (643): 1037–1059.

Gennaro, Gloria and Elliott Ash. 2023. “Televised Debates and Emotional Appeals in Politics: Evidence from C-SPAN”. *Center for Law & Economics Working Paper Series* 2023 (01).

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.

Harcup, Tony and Deirdre O’Neill. 2017. “What is news? News values revisited (again)”. *Journalism studies* 18 (12): 1470–1488.

Hix, Simon and Abdul Noury. 2016. “Government-opposition or left-right? The institutional determinants of voting in legislatures”. *Political Science Research and Methods* 4 (2): 249–273.

Hjorth, Frederik, Robert Klemmensen, Sara Hobolt, Martin Ejnar Hansen, and Peter Kurrild-Klitgaard. 2015. “Computers, coders, and voters: Comparing automated methods for estimating party positions”. *Research & Politics* 2 (2): 2053168015580476.

Høyland, Bjørn. 2010. “Procedural and party effects in European Parliament roll-call votes”. *European Union Politics* 11 (4): 597–613.

Imre, Michael, Alejandro Ecker, Thomas M Meyer, and Wolfgang C Müller. 2023. “Coalition mood in european parliamentary democracies”. *British Journal of Political Science* 53 (1): 104–121.

Joo, Jungseock, Erik P Bucy, and Claudia Seidel. 2019. “Automated coding of televised leader displays: Detecting nonverbal political behavior with computer vision and deep learning”.

Jung, Jae-Hee and Margit Tavits. 2021. “Valence attacks harm the electoral performance of the left but not the right”. *The Journal of Politics* 83 (1): 277–290.

Kayser, Mark A, Matthias Orlowski, and Jochen Rehmert. 2023. “Coalition inclusion probabilities: a party-strategic measure for predicting policy and politics”. *Political Science Research and Methods* 11 (2): 328–346.

Klofstad, Casey A. 2016. “Candidate voice pitch influences election outcomes”. *Political psychology* 37 (5): 725–738.

Klofstad, Casey A, Rindy C Anderson, and Susan Peters. 2012. "Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women". *Proceedings of the Royal Society B: Biological Sciences* 279 (1738): 2698–2704.

Knox, Dean and Christopher Lucas. 2021. "A dynamic model of speech for the social sciences". *American Political Science Review* 115 (2): 649–666.

Kosiara-Pedersen, Karina and Peter Kurrild-Klitgaard. 2018. "Change and stability in the Danish party system". In *Party system change, the European crisis and the state of democracy*, pp. 63–79. Routledge.

Kosmidis, Spyros, Sara B Hobolt, Eamonn Molloy, and Stephen Whitefield. 2019. "Party competition and emotive rhetoric". *Comparative Political Studies* 52 (6): 811–837.

Lauderdale, Benjamin E and Alexander Herzog. 2016. "Measuring political positions from legislative speech". *Political Analysis* 24 (3): 374–394.

Lauridsen, Gustav Aarup, Jacob Aarup Dalsgaard, and Lars Kjartan Bacher Svendsen. 2019. "SENTIDA: A new tool for sentiment analysis in Danish". *Journal of Language Works-Sprogvidenskabeligt Studentertidsskrift* 4 (1): 38–53.

Laustsen, Lasse and Michael Bang Petersen. 2017. "Perceived conflict and leader dominance: Individual and contextual factors behind preferences for dominant leaders". *Political Psychology* 38 (6): 1083–1101.

Laustsen, Lasse, Michael Bang Petersen, and Casey A Klofstad. 2015. "Vote choice, ideology, and social dominance orientation influence preferences for lower pitched voices in political candidates". *Evolutionary Psychology* 13 (3): 1474704915600576.

Laver, Michael, H Back, M Debus, and JM Fernandes. 2021. *Analysing the Politics of Legislative Debate'*. Oxford University Press, Oxford.

Liu, Chang. 2013. "Just noticeable difference of tone pitch contour change for english-and chinese-native listeners". *The Journal of the Acoustical Society of America* 134 (4): 3011–3020.

Mansbridge, Jane. 1999. "Should blacks represent blacks and women represent women? A contingent" yes"". *The Journal of politics* 61 (3): 628–657.

Martin, Lanny W and Georg Vanberg. 2004. "Policing the bargain: Coalition government and parliamentary scrutiny". *American Journal of Political Science* 48 (1): 13–27.

Martin, Lanny W and Georg Vanberg. 2011. *Parliaments and coalitions: The role of legislative institutions in multiparty governance*. Oxford University Press.

Mauss, Iris B and Michael D Robinson. 2009. “Measures of emotion: A review”. *Cognition and emotion* 23 (2): 209–237.

Mayhew, David R. 1974. *Congress: The electoral connection*. Yale university press.

McCarty, Nolan, Keith T Poole, and Howard Rosenthal. 2016. *Polarized America: The dance of ideology and unequal riches*. MIT Press.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. “Distributed representations of words and phrases and their compositionality”. *Advances in neural information processing systems* 26 .

Neumann, Markus. 2019. “Hooked with phonetics: The strategic use of style-shifting in political rhetoric”. In *Annual Meeting of the American Political Science Association. Washington, DC*.

Neumann, Markus, Erika Franklin Fowler, and Travis N Ridout. 2022. “Body Language and Gender Stereotypes in Campaign Video”. *Computational Communication Research* 4 (1).

Nielsen, Finn Årup. 2011. “A new anew: Evaluation of a word list for sentiment analysis in microblogs”. *arXiv preprint arXiv:1103.2903*.

Osnabrügge, Moritz, Sara B Hobolt, and Toni Rodon. 2021. “Playing to the gallery: Emotive rhetoric in parliaments”. *American Political Science Review* 115 (3): 885–899.

Owren, Michael J and Jo-Anne Bachorowski. 2007. “Measuring emotion-related vocal acoustics”. *Handbook of emotion elicitation and assessment*: 239–266.

Park, Tae Jin, Naoyuki Kanda, Dimitrios Dimitriadis, Kyu J Han, Shinji Watanabe, and Shrikanth Narayanan. 2022. “A review of speaker diarization: Recent advances with deep learning”. *Computer Speech & Language* 72 : 101317.

Patel, S. and K. R. Scherer. 2013. “Vocal behavior”. In J. A. Hall and M. L. Knapp (Eds.), *Nonverbal communication*, pp. 167–204. De Gruyter Mouton.

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.

Pisanski, Katarzyna, Valentina Cartei, Carolyn McGettigan, Jordan Raine, and David Reby. 2016. “Voice modulation: a window into the origins of human vocal control?”. *Trends in cognitive sciences* 20 (4): 304–318.

Proksch, Sven-Oliver, Will Lowe, Jens Wäckerle, and Stuart Soroka. 2019. "Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches". *Legislative Studies Quarterly* 44 (1): 97–131.

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

Rabiner, Lawrence and Ronald Schafer. 2010. *Theory and applications of digital speech processing*. Prentice Hall Press.

Rask, Mathias. 2023. "PolAnnotate: Matching Audio to Transcripts". *Working Paper*.

Rauh, Christian and Jan Schwalbach. 2020. "The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies".

Rheault, Ludovic and Sophie Borwein. 2019. "Multimodal techniques for the study of a ect in political videos". Technical report, Working Paper.

Rheault, Ludovic and Christopher Cochrane. 2020. "Word embeddings for the analysis of ideological placement in parliamentary corpora". *Political Analysis* 28 (1): 112–133.

Rittmann, Oliver. 2023. "Legislators' Emotional Engagement with Women's Issues: Gendered Patterns of Vocal Pitch in the German Bundestag". *Forthcoming in British Journal of Political Science*.

Roberts, Margaret E, Brandon M Stewart, and Dustin Tingley. 2019. "Stm: An r package for structural topic models". *Journal of Statistical Software* 91 : 1–40.

Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand. 2014. "Structural topic models for open-ended survey responses". *American journal of political science* 58 (4): 1064–1082.

Rodriguez, Pedro L and Arthur Spirling. 2022. "Word embeddings: What works, what doesn't, and how to tell the difference for applied research". *The Journal of Politics* 84 (1): 101–115.

Scherer, Klaus R, Rainer Banse, and Harald G Wallbott. 2001. "Emotion inferences from vocal expression correlate across languages and cultures". *Journal of Cross-cultural psychology* 32 (1): 76–92.

Scherer, Klaus R, Tom Johnstone, and Gundrun Klasmeyer. 2003. *Vocal expression of emotion.*, Volume 40. Oxford University Press.

Scherer, Klaus R, D Robert Ladd, and Kim EA Silverman. 1984. "Vocal cues to speaker affect: Testing two models". *The Journal of the Acoustical Society of America* 76 (5): 1346–1356.

Schulz, Ida. 2007. "The journalistic gut feeling". *Journalism Practice* 1 (2): 190–207.

Schwalbach, Jan. 2023. "Talking to the populist radical right: A comparative analysis of parliamentary debates". *Legislative Studies Quarterly* 48 (2): 371–397.

Serra, Gilles. 2010. "Polarization of what? A model of elections with endogenous valence". *The Journal of Politics* 72 (2): 426–437.

Shapley, Lloyd S. and Martin Shubik. 1954. "A method for evaluating the distribution of power in a committee system". *American Political Science Review* 48 (3): 787–792.

Skytte, Rasmus. 2021. "Dimensions of elite partisan polarization: Disentangling the effects of incivility and issue polarization". *British Journal of Political Science* 51 (4): 1457–1475.

Slapin, Jonathan B and Sven-Oliver Proksch. 2008. "A scaling model for estimating time-series party positions from texts". *American Journal of Political Science* 52 (3): 705–722.

Tigue, Cara C, Diana J Borak, Jillian JM O'Connor, Charles Schandl, and David R Feinberg. 2012. "Voice pitch influences voting behavior". *Evolution and Human Behavior* 33 (3): 210–216.

Touati, Paul. 1993. "Prosodic aspects of political rhetoric". In *ESCA workshop on prosody*.

Tusing, Kyle James and James Price Dillard. 2000. "The sounds of dominance. Vocal precursors of perceived dominance during interpersonal influence". *Human Communication Research* 26 (1): 148–171.

Vainio, Martti, Antti Suni, Juraj Šimko, and Sofoklis Kakouros. 2023. "The power of prosody and prosody of power: An acoustic analysis of finnish parliamentary speech". *arXiv preprint arXiv:2305.16040*.

Webster, Steven W. 2021. "The role of political elites in eliciting mass-level political anger". In *The Forum*, Volume 19, pp. 415–437. De Gruyter.

Widmann, Tobias. 2021. "How emotional are populists really? Factors explaining emotional appeals in the communication of political parties". *Political psychology* 42 (1): 163–181.

Wonka, Arndt and Sascha Göbel. 2016. "Parliamentary scrutiny and partisan conflict in the Euro crisis. The case of the German Bundestag". *Comparative European Politics* 14 : 215–231.

Zuckerman, Miron and Robert E Driver. 1989. "What sounds beautiful is good: The vocal attractiveness stereotype". *Journal of nonverbal behavior* 13 (2): 67–82.

Online Appendix

Contents

A Bargaining Leverage and Extremity

Figure A1 shows Coalition Inclusion Probabilities (CIPs) for parties in *Folketinget* from 1998 to 2019, the range of data available in the most recent data available from `https://coalition-leverage.org/` as of September 2023. The data are from the static version of the CIP data, i.e. without using information from between-election polling. As shown, parties at the ideological extremes such as Red/Green Alliance (EL), Danish People's Party (DF), and New Right (NB) all consistently rank at the very bottom.

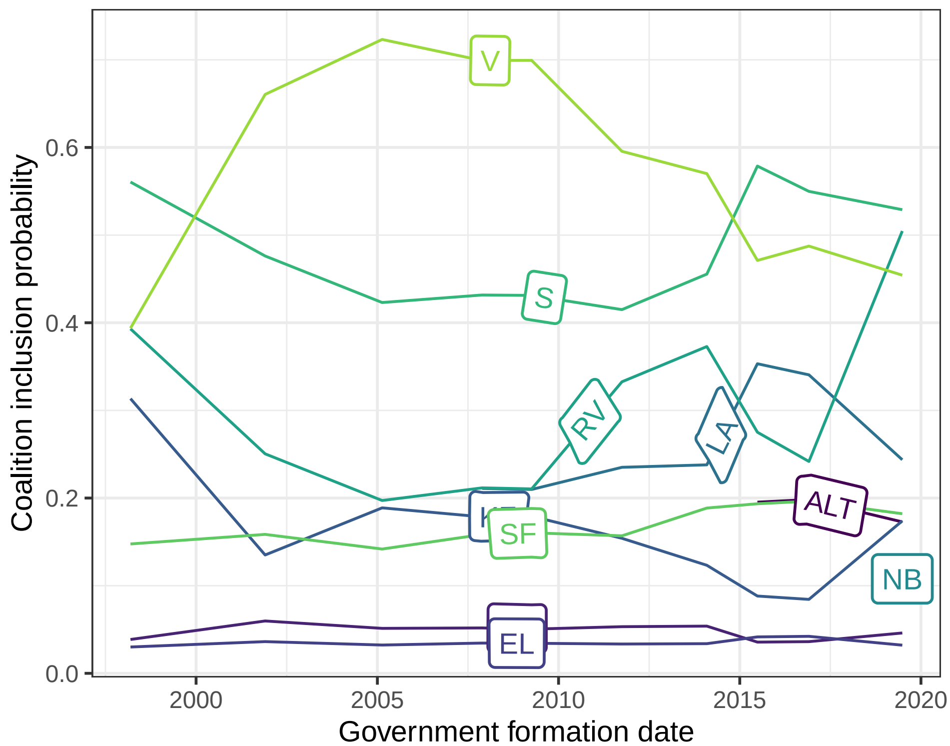
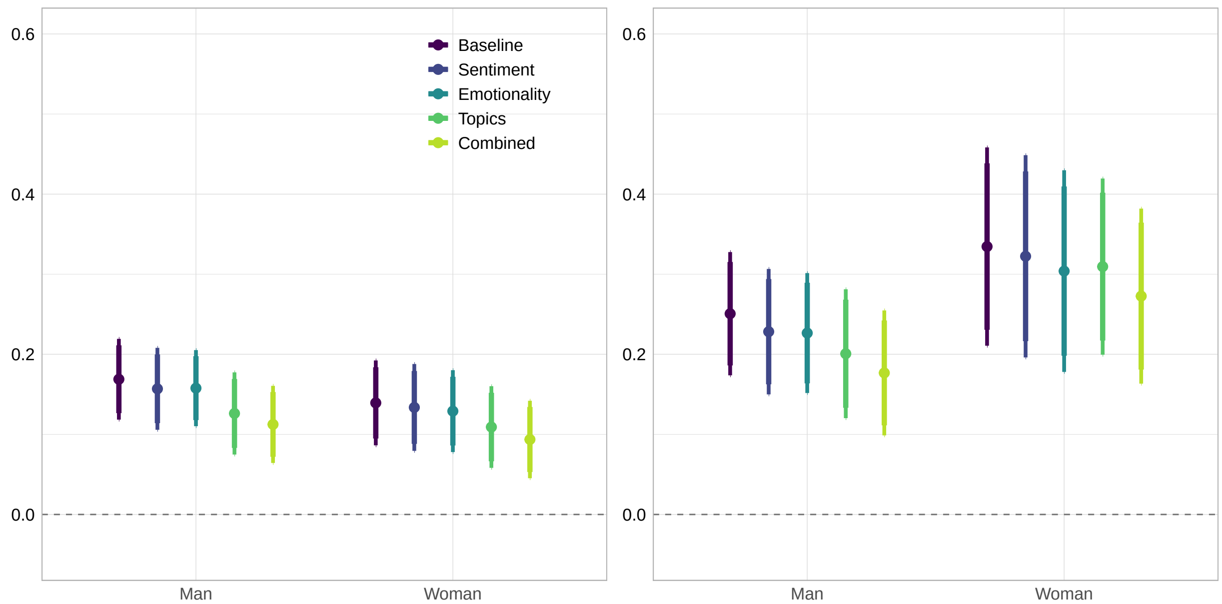


Figure A1: Coalition Inclusion Probabilities (CIPs) for parties in *Folketinget*, 1998-2019.

B Pitch Measurement

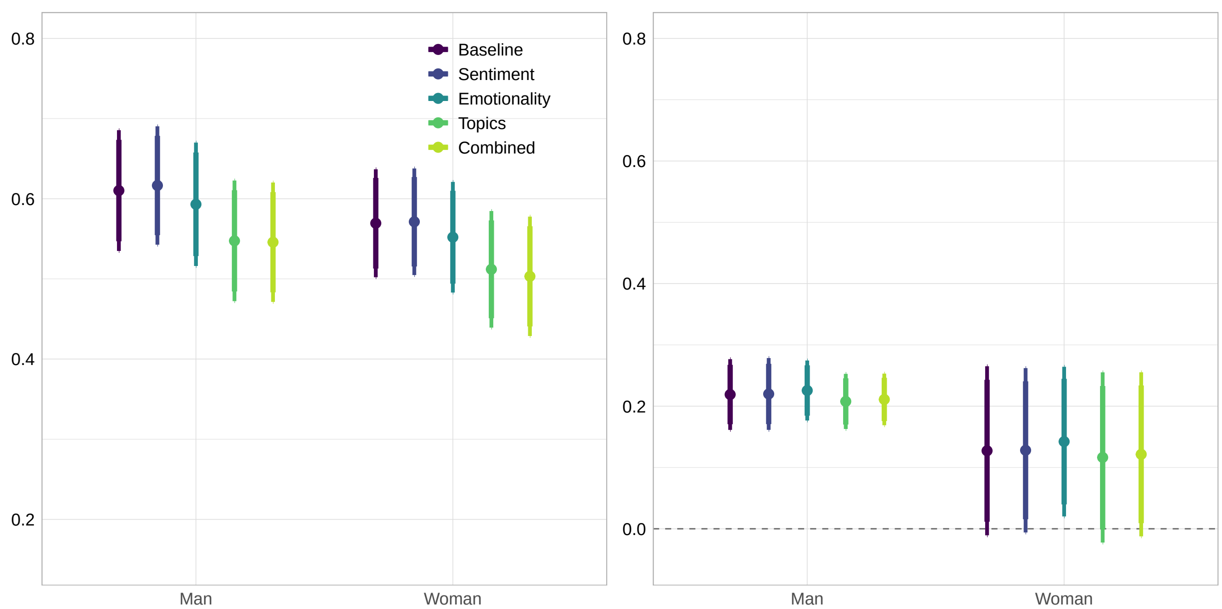
We computed the pitch using the `R` package *communication* developed in Knox and Lucas (2021). The library estimates the $F0$ of an audio signal with two separate algorithms and operates on 25 ms windows with a 12.5 ms overlap. Each window is then parsed through a hamming function to counter the spectral leakage caused by the windowing. For a 60-second audio file, this results in $4,800$ total windows where each audio feature is computed within each window. We consider a window as having a valid pitch estimate if both $F0$ algorithms consider the window as voiced (i.e. a non-zero estimate), otherwise, the window is considered as non-voiced. Our final pitch estimate for a single speech is the average of the voiced windows. We also compared our estimates to the *Praat* software used by Dietrich et al. (2019) using sex-specific settings with no change in the results.

C Gender Effects



(a) Partisan Polarization **(b)** Policy Conflict

Figure C1: Coefficients for partisan polarization (left panel) and policy conflict (right panel) estimated separately for men and women. Standardized pitch is used as the outcome and predictors are whether the target party is outbloc (left panel) and whether the target party voted differently in a legislative vote on a specific bill (right panel). Standard errors are clustered at the dyad level (speaker party ↔ target party). Thick and thin error bars are the model-specific 90 and 95 pct. confidence intervals respectively. Y-axes are held fixed across the two panels to maximize comparability.



(a) Electoral: High- vs. Low-Profile Debates. **(b)** Policy: Target Party Bargaining Leverage.

Figure C2: Coefficients for debate type (left panel) and bargaining leverage (right panel) estimated separately for men and women. Standardized pitch is used as the outcome and predictors are whether the speech is given in a high- compared to a low-profile debate (left panel) and the policy bargaining leverage of the target party (right panel). Standard errors are clustered at the dyad level (speaker party $\leftrightarrow$ target party). Standard errors are clustered at the level of dyads (left panel) and target party (right panel) respectively. Thick and thin error bars are the model-specific 90 and 95 pct. confidence intervals respectively. Y-axes are held fixed across the two panels to maximize comparability.

D Validation of Pitch

In this appendix, we present our validation analysis of using pitch as measure of emotional arousal as a signal of partisan conflict. We present the results of three distinct validation exercises before we conclude by inspecting measurement error in using pitch to measure arousal.

D.1 Validation Exercise 1: Binary Arousal

The first validation exercise investigates the extent to which: (1) coders consistently can infer a legislator's emotional arousal from speech-level audio-only clips and (2) standardized vocal pitch of the speaker aligns with a manual labels of a binary arousal coding. To evaluate this, two coders were asked to independently evaluate whether a speaker is animated (=1) or subdued (=0) on a random selection of 100 speeches.[12] Speeches were sampled uniformly at random from the population of speeches containing dyadic exchanges and which lies between the 25th and 75th percentile in length.

The percentage agreement between the two coders is 87 pct. with a Krippendorff's $\alpha = 0.62$. The relatively low intercoder reliability is probably a result of the binary coding procedure, which inevitably introduces boundary cases in the labeling. The alignment between speaker-standardized vocal pitch and the binary emotional arousal codings is shown in Figure D1. The figure shows that the average pitch is higher when both coders label a speaker as being aroused, and lowest when neither perceives the speaker as such. The difference in vocal pitch in SDs between a speech in which a speaker is classified as emotionally aroused and subdued is 0.78. This result is consistent with the notion that changes in vocal pitch from a speaker's own baseline, to a certain extent, capture a speaker's emotional arousal in dyadic exchanges during parliamentary debates.

Although the binary coding validation results are promising, it is also clear that the measure contains substantial noise, as evident by the overlap between the distributions of the three coder conditions. One possible explanation for this noise is the unit of analysis. The empirical analysis, and consequently the validation, is done at the speech level. These vary in length, which might

[12] Note that the binary coding sums to 99 and not 100 because one of the sampled speeches lacked aligned audio.

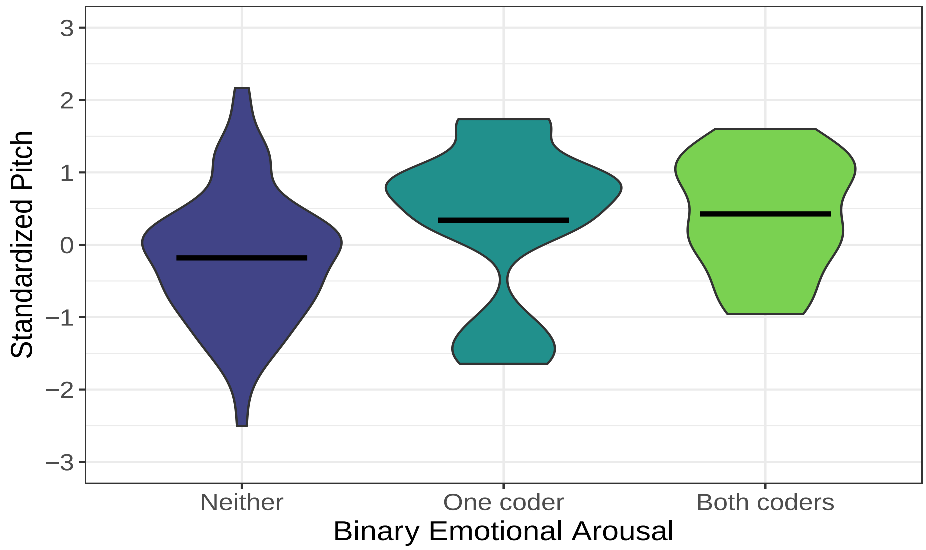


Figure D1: Distribution of speaker-standardized (speech-level) pitch estimates across three coding conditions for 100 randomly selected speeches using a binary arousal coding. The horizontal black bars denote the average standardized pitch in each condition. 'Neither' means that none of the two coders labeled a speech as activated, 'One coder' means that one of two labeled a speech as activated, and 'Both coders' means that both labeled a speech as activated. Coder agreement 87 pct. and Krippendorff's $\alpha = 0.62$.

cause coders to focus on different parts of a speech, thus creating noise in the measure. Using this line of reasoning, Cochrane et al. (2022) use sentences as the unit of analysis in investigating intercoder reliability codings of emotional arousal in audiovisual recordings since "sentences (...), is the smallest natural unit to convey meaning in speech" (p. 101).

Another possible explanation concerns the level used in the arousal coding. Using a binary labeling procedure inevitably introduces measurement error since it creates a larger proportion of boundary cases and because the link between vocal pitch and emotional arousal is continuous in theory (Mauss and Robinson 2009). To test this, we implement a second validation exercise below using a continuous labeling of arousal.

D.2 Validation Exercise 2: Continuous Arousal

The second validation exercise has the same goals as the first but uses a continuous labeling procedure, investigating the extent to which: (1) coders consistently can infer a legislator's emotional arousal from speech-level audio-only clips and (2) standardized vocal pitch of the speaker aligns with a manual labels of a continuous arousal coding. To evaluate this, two coders were given the same instruction as used by Cochrane et al. (2022):

> *On a scale from 0-10, where 0 indicates that the speaker was very subdued, 5 indicates that they were in a normal state of calm, and 10 indicates that the speaker was very animated, please indicate the emotional state of the speaker.*

Based on this labeling procedure, two coders independently labeled a random selection of 100 speeches. Speeches were sampled uniformly at random from the population of speeches containing dyadic exchanges and which lies between the 25th and 75th percentile in length.[13]

The percentage agreement between the two coders in the continuous coding is similar to the percentage of the binary coding with 87 pct. but has Krippendorff's $\alpha = 0.85$ and an intraclass coefficient (ICC) on 0.87. This is considered excellent reliability (Cicchetti 1994, p. 286). This suggests that the low intercoder reliability in the first coding is due to the binary labeling procedure and not to the general perception of a speaker's emotional arousal. The alignment between speaker-standardized vocal pitch and the continuous emotional arousal codings is shown in Figure D2 separated by coder. The figure shows that the pitch estimates strongly align with the manual arousal labels for each of the coders with Pearson correlation coefficients of 0.89 and 0.87, respectively. This shows that changes in vocal pitch from a speaker's own baseline systematically reflect a speaker's emotional arousal in dyadic exchanges during parliamentary debates.

To provide further details on the manual labels, we summarize the distribution of activated and non-activated speeches using binary and continuous coding, respectively, and their relevant reliability metrics in Table D1. Although the coding agreement is similar (87 pct.), the distributions

[13]This sampling also added the criteria that a speech must not have been labeled before, discarding the 100 sampled speeches from the binary coding.

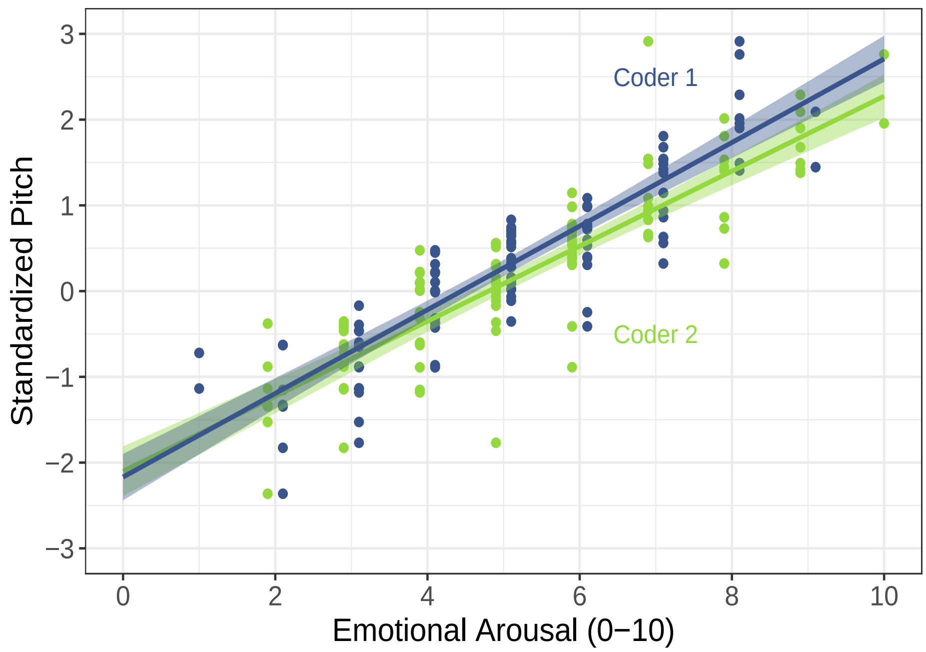


Figure D2: Alignment between continuous arousal coding and speaker-standardized (speech-level) pitch estimates. ICC=0.85.

vary as expected. In continuous coding, more speeches are labeled as activated (using a binary cutoff of 5 on the 0-10 scale) compared to binary coding. Furthermore, using the cutoff point of 5 to binarize the continuous labels, the difference in vocal pitch in SDs between a speech in which a speaker is classified as emotionally aroused and subdued is 1.45, nearly twice the difference of the binary coding. This further reinforces the validity of the link between arousal and pitch in dyadic exchanges.

	Coding	
	Binary (Exercise 1)	Continuous (Exercise 2)
Coder agreement		
Neither	71	50
One coder	13	13
Both coders	15	37
Intercoder reliability		
Agreement pct.	87 pct.	87 pct.
Krippendorff's α	0.62	0.85
Cohen's κ	0.62	–
ICC	–	0.87

Table D1: Distribution of labels for the binary (exercise 1) and continuous (exercise 2) coding and intercoder realiability metrics. For the distribution of the continuous coding, a speech is classified as activated if arousal is labeled as 5 (the midpoint of the 0-10 scale) or larger. The category 'Neither' means that the two agree that the speech is nonactivated, 'One coder' means that one of the two coders have labeled a speech as activated, and 'Both coders' means that the two coders agree that a speech is activated. Note that the binary coding sums to 99 and not 100 because one of the sampled speeches lacked aligned audio. The agreement percentage is computed using the binary representation of the continuous coding also with 5 as the cutoff.

D.3 Validation Exercise 3: Pitch and Partisan Conflict in Dyadic Exchanges

The third validation exercise investigates the correlation between speaker-standardized speech-level vocal pitch and text sentiment to validate the usefulness of our context of dyadic exchanges. Using pitch to measure conflict-driven arousal, we should expect pitch to be larger in dyadic speeches in which legislators use more negative language. Since parliamentary debates primarily serve the role of showcasing policy positions, they are *per se* more about partisan conflict than substantial negotiation (Laver et al. 2021). Hence, we expect a negative relationship between pitch and text sentiment. To evaluate this, we compare speaker-standardized speech-level estimates of vocal pitch with the sentiment of each speech computed using the Danish sentiment measuring tool Sentida (Lauridsen et al. 2019).

The result is reported in Figure D3, visualizing the bivariate relationship between the pitch estimates and text sentiment for dyadic speeches. The relationship is negative ($\hat{\beta} = -0.23$) and strongly statistically significant ($t = 19$, $p < 0.001$). Whilst still negative ($\hat{\beta} = -0.17$), the correlation is weaker in non-dyadic speeches. We take this as firm evidence that pitch tracks expressions of conflict in parliamentary debates and particularly closely in dyadic speeches. Importantly, however, the sentiment does not explain pitch as the correlation is still fairly weak, suggesting that pitch captures a distinct dimension of political speech.

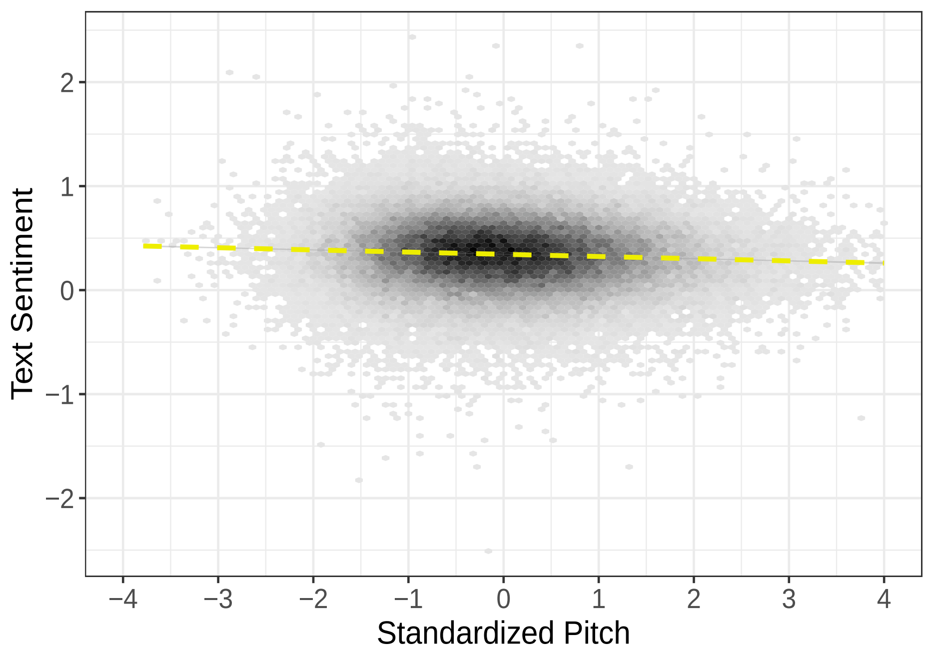


Figure D3: Correlation between speech sentiment and pitch for dyadic speeches, $\hat{\beta} = -0.23$ (t = 19, p < 0.001).

D.4 Understanding Measurement Error in Pitch

We lastly consider residual measurement error in pitch. We do so with a qualitative reading of the cases with the largest residuals relative to the new annotation exercise, i.e. the cases where perceived arousal deviates most from standardized pitch. We illustrate this case selection in Figure Figure D4.

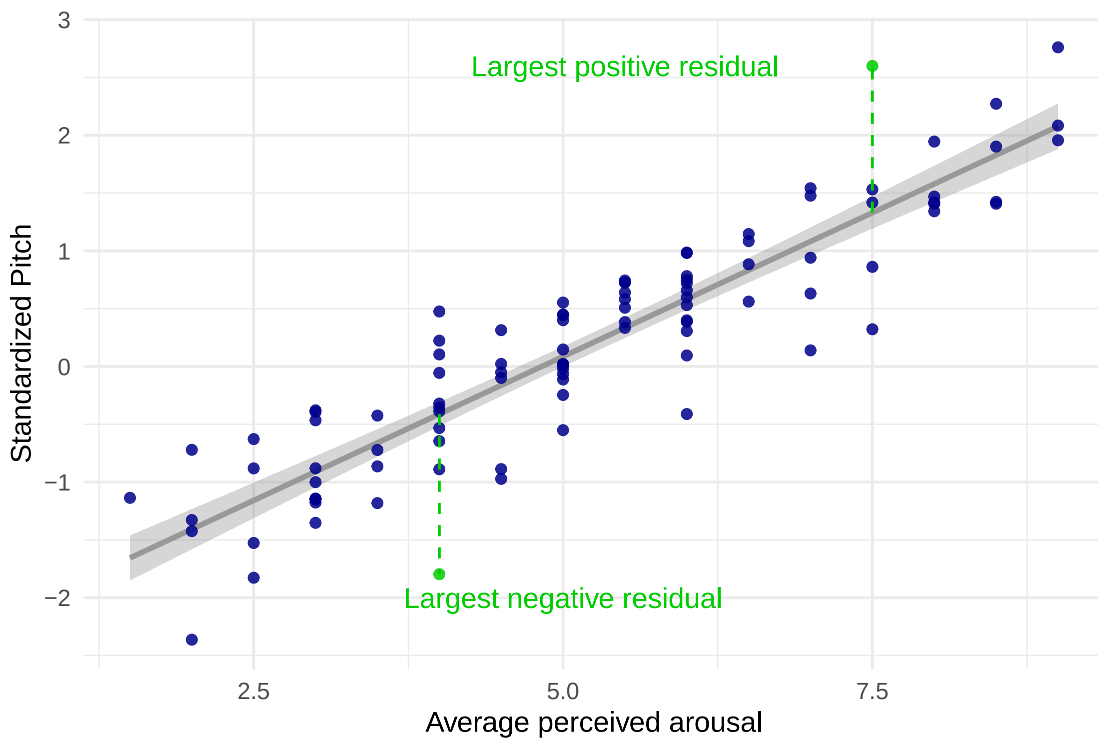


Figure D4: Average perceived arousal from Validation Exercise 2 (x-axis) plotted against standardized pitch at the speech level (y-axis). The two highlighted points represent the largest positive and negative residuals when regressing pitch on average perceived arousal.

In Table D2, we present the original and translated versions of these two residuals. The positive residual text (top) is a speech by an MP from the Liberal Party targeting an MP from the Socialist People's Party. The negative residual text (bottom row) is a speech by an MP from the Conservatives targeting an MP from the Social Liberals.

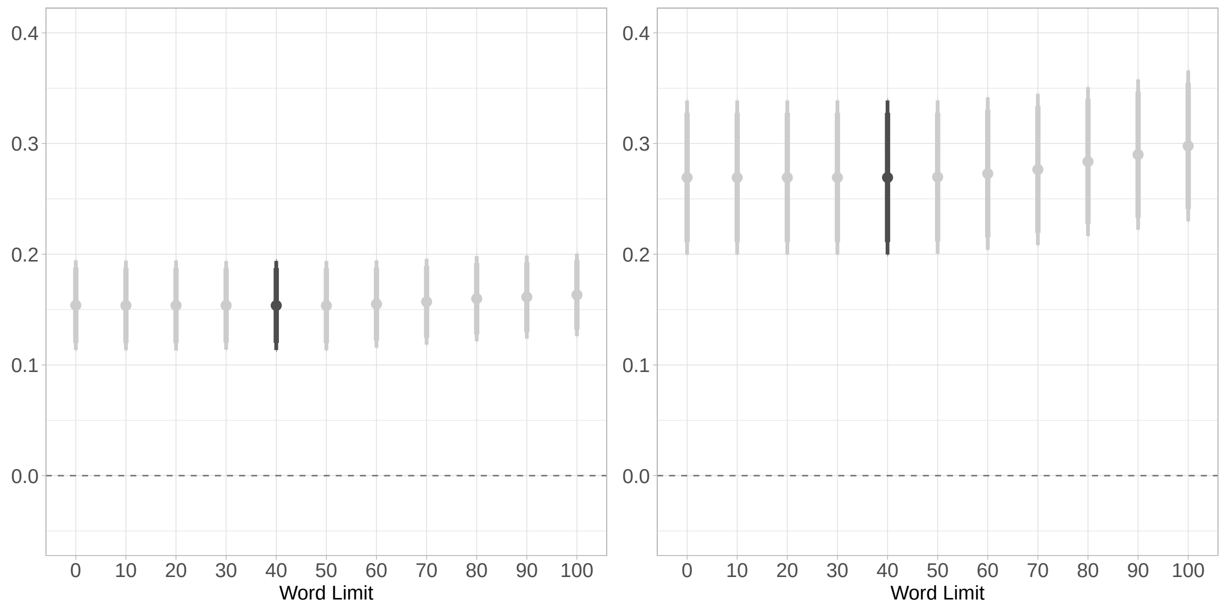
Based on qualitative inspection, we see no plausible confounders driving these cases: they are not about theoretically distinct topics, and represent typical party dyads. In fact, in both cases a mainstream right legislator is targeting a mainstream left legislator.

Moreover, a qualitative reading aligns with the assigned arousal scores: the positive residual (which has lower than predicted perceived arousal) is in fact relatively subdued, whereas the negative residual (which has higher than predicted perceived arousal) has a quite aggressive tone. This reading is consistent with the residual variation in standardized pitch reflecting non-systematic measurement error.

Table D2: Original and translated texts of largest positive and negative residuals.

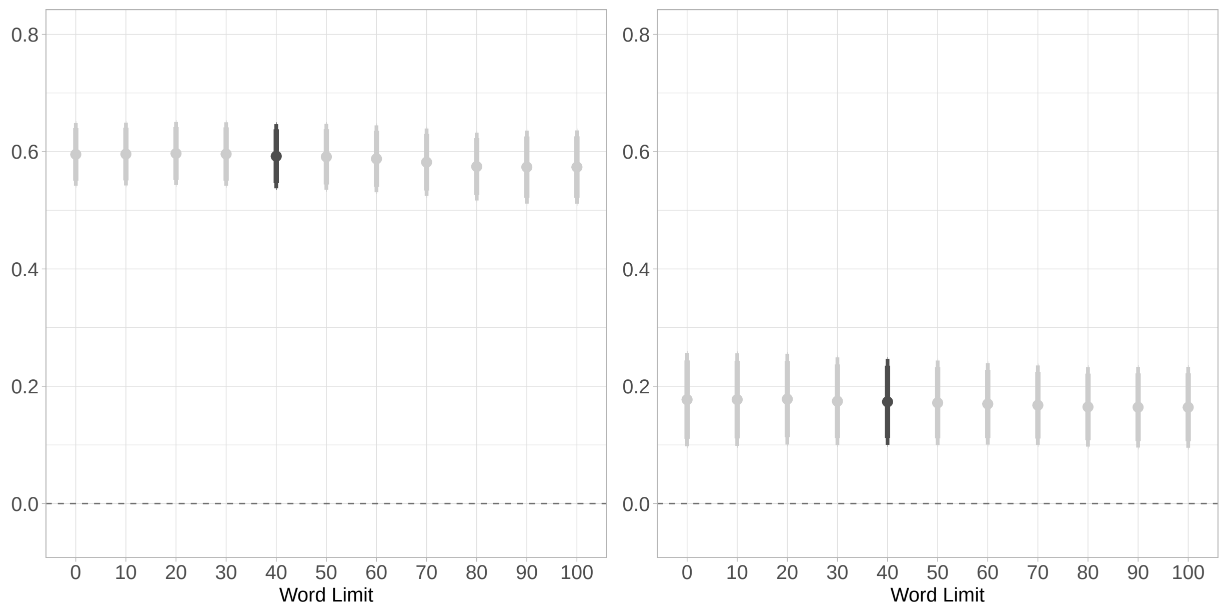
Residual	Original text	Translated text
Positive	"Hvis jeg forstår hr. Steen Gade korrekt, går spørgsmålet også på, hvorvidt det med en resolution kan sikres, at der i al fremtid vil kunne opnås enighed om, hvordan FN skal agere. Der må jeg blot konstatere, at hvad angår Kosovo, nåede man jo ikke frem til nogen løsning i FN's Sikkerhedsråd. Der valgte man jo uden et FN-mandat og med kraftige advarsler fra Kofi Annan som generalsekretær at foretage en aktion. Det kan også blive nødvendigt fremover. Det er da selvfølgelig, fuldstændig som udenrigsministeren allerede har gjort rede for i dag, vigtigt, at man har en drøftelse i FN's Sikkerhedsråd om de principper og de ting, der ligger til grund, hvis man vil anvende magt. Det er vi da helt enige om. Men derfor er der da stadig væk en forskel på at have en drøftelse om det i Sikkerhedsrådet og så skulle lave en resolution, som vi ved ikke har nogen gang på jorden."	"If I understand Mr Steen Gade correctly, the question is also about whether a resolution can ensure that agreement can be reached on how the UN should act in the future. All I have to say is that, as far as Kosovo is concerned, no solution was reached in the UN Security Council. Without a UN mandate and with strong warnings from Kofi Annan as Secretary-General, it was decided to take action. It may also be necessary in the future. As the Foreign Minister has already explained today, it is of course important to have a discussion in the UN Security Council about the principles and the things that form the basis for the use of force. We are in complete agreement on that. However, there is still a difference between having a discussion about it in the Security Council and then having to draft a resolution, which we know has no effect on the ground."
Negative	"Jeg kunne aldrig drømme om at møde op i EU og så blive gjort til grin. Og derfor vil jeg sige, at hr. Morten Østergaards forslag jo er fuldstændig vanvittigt. Jeg kunne da ikke drømme om at møde op og komme med påstande om noget, som jeg på regeringens vegne mange, mange gange her i Folketinget har hørt er blevet fuldstændig afvist. Jeg synes, det europæiske samarbejde skal bruges til seriøse ting og til ting, hvor der er konkrete sager, man kan tage hånd om, og derfor må jeg sige, at den der slags ideer synes jeg ville være endnu en paradeforestilling, som ikke har noget som helst andet formål end at kaste sig ud i en masse påstande, der slet ikke kan bevises."	"I would never dream of showing up in the EU and then being made a fool of. And that is why I would say that Morten Østergaard's proposal is completely insane. I would never dream of showing up and making claims about something that I, on behalf of the government, have heard completely rejected many, many times here in the Danish Parliament. I think that European cooperation should be used for serious things and for things where there are concrete issues that can be dealt with, and therefore I have to say that I think that this kind of idea would be another parade show that has no purpose whatsoever other than to throw itself into a lot of claims that cannot be proven at all."

E Word Limits



(a) Partisan Polarization **(b)** Policy Conflict

Figure E1: Coefficients for partisan polarization (left panel) and policy conflict (right panel) across different word limits using the "Baseline" specification. Standardized pitch is used as the outcome. Bars and dots in black denote the word limit used to retain speeches in the main results. Bars and dots in grey denote additionally used word limits. Predictors are whether the target party is outbloc (left panel) and whether the target party voted differently in a legislative vote on a specific bill. Standard errors are clustered at the dyad level (speaker party ↔ target party). Thick and thin error bars are the model-specific 90 and 95 pct. confidence intervals respectively. Y-axes are held fixed across the two panels to maximize comparability.



(a) Electoral: High- vs. Low-Profile Debates. **(b)** Policy: Target Party Bargaining Leverage.

Figure E2: Coefficients for debate type (left panel) and bargaining leverage (right panel) across different word limits using the "Baseline" specification. Standardized pitch is used as the outcome. Bars and dots in black denote the word limit used to retain speeches in the main results. Bars and dots in grey denote additionally used word limits. The predictors are whether the speech is given in a high- compared to a low-profile debate (left panel) and the policy bargaining leverage of the target party (right panel). Standard errors are clustered at the dyad level (speaker party $\leftrightarrow$ target party). Thick and thin error bars are the model-specific 90 and 95 pct. confidence intervals respectively. Y-axes are held fixed across the two panels to maximize comparability.

F   Data Description

Description of Audio

We collected all available recordings of parliamentary debates in the Danish parliament from October 2000 to September 2022 covering six national elections (2001, 2005, 2007, 2011, 2015, and 2019), 28 parliamentary terms, and a total of $2,282$ debates. As far as we know, this is the largest collection of natural audio data used in political science. Of the $2,282$ debates, we were able to download a recording for $2,260$ debates. We downloaded recordings for 2000-2009 from the Danish Royal Library and for 2010-2022 from Folketinget's own website. This was done using two custom-written scrapers in `Python`, which extracted the `.m3u8`-file of each debate. We then used the open-source software `ffmpeg` to convert the `.m3u8` stream to a local `.mp4`-file, which was further converted to an audio file (`.wav`) to conserve space.[14] For the resulting `.wav`-files we used a sampling rate of $16,000$ Hz (i.e. $16,000$ samples each second), a single channel (mono), and 16-bits (i.e. sample values $x \in \{-32767, 32768\}$). We removed recordings from our collection, which only contained speech from a chair or a single legislator.

Description of Transcript

Our research question requires a joint text-audio analysis. Specifically, our definition of party dyads requires that we obtain information on the text of speech, and also information about the party affiliation of the speaking legislator and the mentioned politicians or parties in the text. To obtain this, we exploited that each recording has a corresponding transcripts, which contains information on both the text of speech and speaker names. For this, we relied on a combination of ParlSpeech V2 (Rauh and Schwalbach 2020) and manually scraped XML files. We used the former from 2000-2018 and the latter for the period 2019-2022.

We were able to match a transcript to a recording for $2,186$ out of $2,260$ (≈ 97 pct.). ParlSpeech V2 structures speeches by date and speech number, but does not contain a debate identifier. Hence, dates where multiple debates took place are not matched as this resulted in two recordings for

[14]We used an intermediate step (from `.m3u8` to `.mp4`) because the quality of the audio were better preserved by the `.m3u8` to `.mp4` conversion.

one transcript. This is not impossible to resolve, but we refrained from manually matching the remaining 3 pct. due to the high match percentage. In total, our $2,186$ debates contained a total of $850,363$ speeches with each debate containing an average of 389 speeches with a standard deviation of 249. Without speech from chairs, the average is 217 and the standard deviation 127.

Alignment

After collecting and matching audio recordings to transcripts, the next step was to align the audio signal with the corresponding speech in each transcript. To do this, we used the workflow proposed by Rask (2023) and the associated `Python` library `speechannote`, which automatically generates annotations for audio recordings with almost human-level accuracy.[15] The annotations contain the start and end time of each word in a speech, speaker names, and the text of each speech. The approach combines speaker diarization, automatic speech recognition (ASR), and speaker identification into a common workflow relying entirely on open-source and pretrained models, resulting in a fully automated pipeline.

Speaker diarization is a task originating in computer science and that deals with segmenting audio recordings containing human speech according to the uniqueness of voices. The output from a diarization system is a list of speech segments with timestamps each with an assigned speaker label. This process works entirely unsupervised and is shown to generalize across different speech settings (Park et al. 2022). Political speech is a forgiving task for speaker diarization, as it more formal and regulated by rules than, for instance, a conversation.

After diarization, each speech segment is annotated with the speaker name using a fuzzy string matching approach also developed in Rask (2023). Importantly, this enables speaker recognition as a weakly-supervised learning task, which automatizes the workflow. The weakly-supervised identification is done by compiling reference audio for each speaker by scoring ASR output and a target using a similarity metric (e.g. Jaccard or cosine). The logic is that many audio recordings contain corresponding transcripts, which can be viewed as a “fuzzy target” since a transcript is a written, but non-verbatim record, of the speeches in the recording. The compiled reference audio

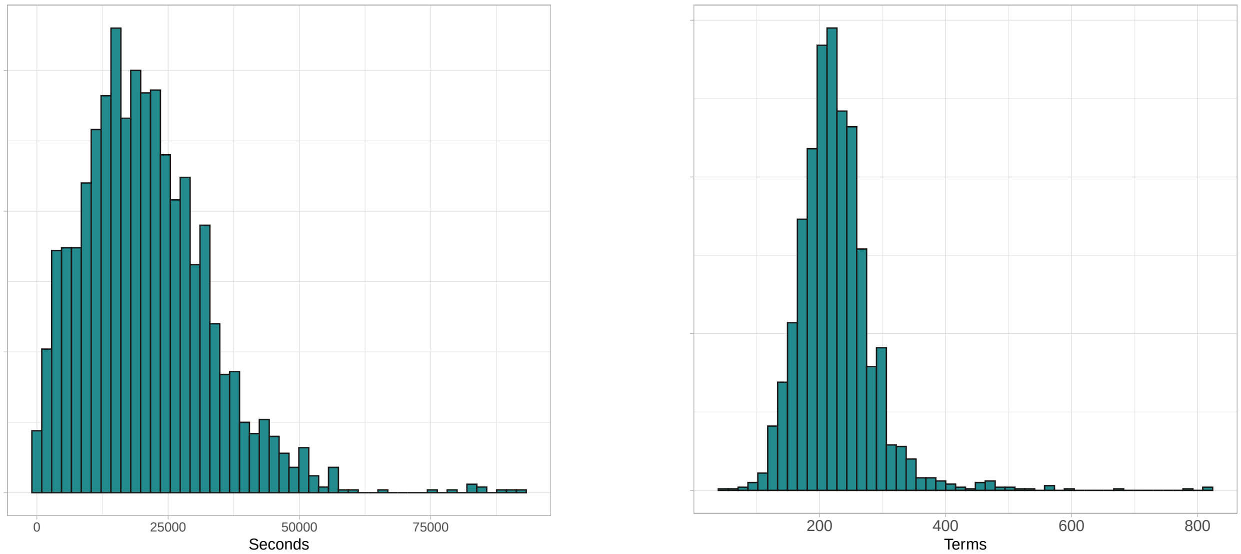
[15] The library is still under development but will be publicly available at GitHub in primo 2024.

is then used to identify the speaker of each segment using a simple threshold.

As the final step, the audio signal is aligned with the text. This can be done by matching each speech segment to a single speech in the recording's corresponding transcript or by using the ASR output from each speech segment. For this paper, we used non-verbatim transcripts since they contain standardized references to other legislators and parties. For instance, if a legislator references another legislator without using the proper honorific, this is corrected in the transcript. As a result, identifying party dyads is more efficient in transcripts than in ASR output. To match a speech segment to the transcripts, we applied ASR on each segment and then compared it with the text contained in the transcript using cosine similarity. A speech segment was then matched to a speech transcript using fuzzy string matching. We were able to match a speech segment to a speech transcript for 96.3 pct. of the speeches resulting in $378,566$ aligned speeches.

Summary of Text-Audio Corpus

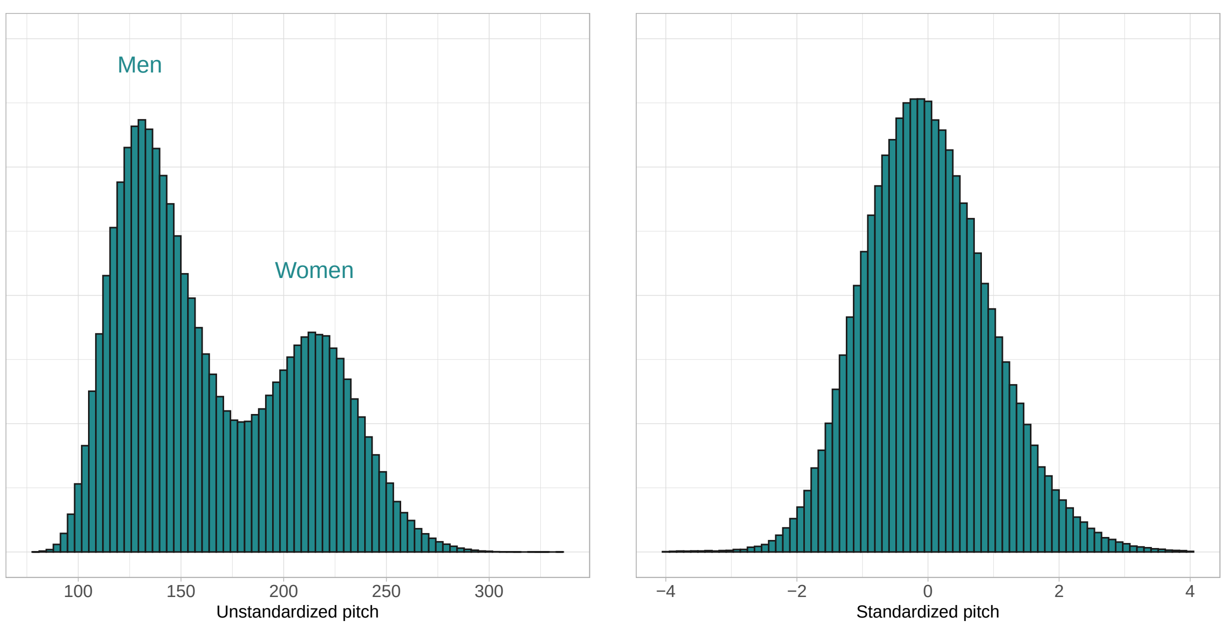
After preprocessing and aligning, we were left with a total of $378,566$ aligned speeches distributed across $2,148$ debates. The average duration of each debate recording is $\approx 21,300$ seconds with a standard deviation of $12,095$. Converted to hours minutes, and seconds, this corresponds to an average of 5 hours, 55 minutes, and 1 seconds with a standard deviation of 3 hours, 21 minutes, and 35 seconds. As seen in Figure F1a, the distribution is right-skewed due to the presence of a few very long recordings (e.g. opening and closing debates).



(a) Average duration of recordings.

(b) Average terms in transcripts.

Figure F1: Distribution of the average duration and terms for $2,148$ debates used in the empirical analysis. The distributions are plotted as histograms with 50 bins. The data is aggregated to the debate-level such that it is the debate-wise average. Panel F1b only shows the distribution for debates with an average length of shorter than $1,000$ words for presentation purposes.



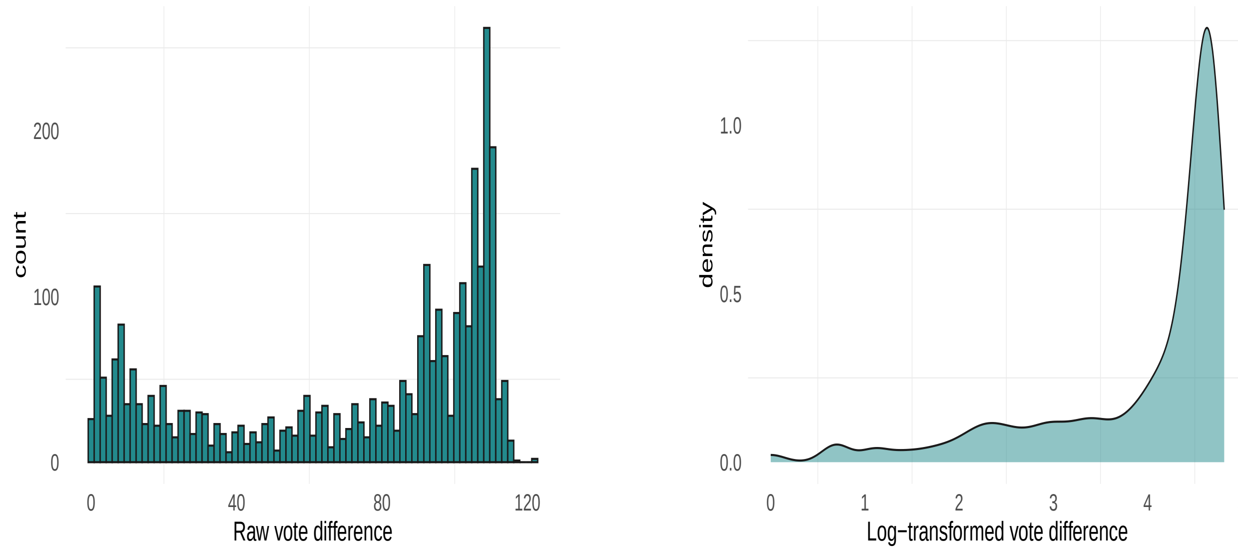
(a) Unstandardized vocal pitch.

(b) Standardized vocal pitch.

Figure F2: Distribution of speech-level $F0$ before and after standardization.

Table F1: Summary table for a selected set of variables in the dataset.

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max
Outbloc	2	0	0.59	0.49	0.00	1.00	1.00
Out-vote	3	79	0.36	0.48	0.00	0.00	1.00
Vote difference	120	79	47.74	39.25	0.00	34.00	122.00
CIP	82	0	0.28	0.23	0.03	0.20	0.72
Sentiment	141640	0	0.32	0.28	−1.69	0.33	1.90
Emotionality	146973	0	1.00	0.13	0.52	1.00	1.46
Std. Pitch (average)	146993	0	0.04	0.98	−4.98	−0.01	4.93
Pitch (average)	146993	0	166.64	43.90	81.45	153.49	326.16
Pitch (modulation)	146993	0	34.68	9.29	1.16	34.15	87.25
Pitch change (average)	146993	0	−1.03	0.48	−4.19	−0.97	1.13
Loudness (average)	146993	0	60.78	6.96	35.47	62.61	84.49
Loudness (modulation)	146993	0	5.84	0.78	1.19	5.87	11.49



(a) Raw vote difference.

(b) Log-transformed vote difference.

Figure F3: Distribution of the vote difference between 'yes' and 'no' across $3,174$ legislative votes. The vote difference is transformed to the absolute difference between 'yes' and 'no' votes such that 0 means there were equally many for and against. Panel F3a shows the raw vote difference and Panel F3b shows the natural logarithm of the vote difference with 1 added as a constant to make sure the logarithm is mathematically defined.

G Party Dyads

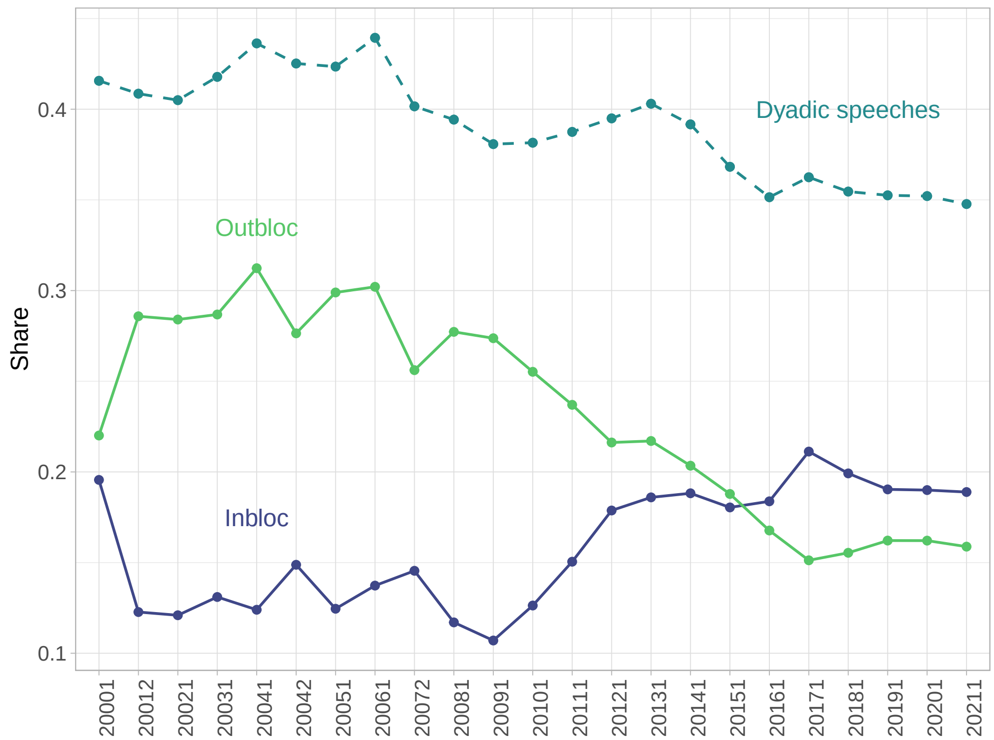
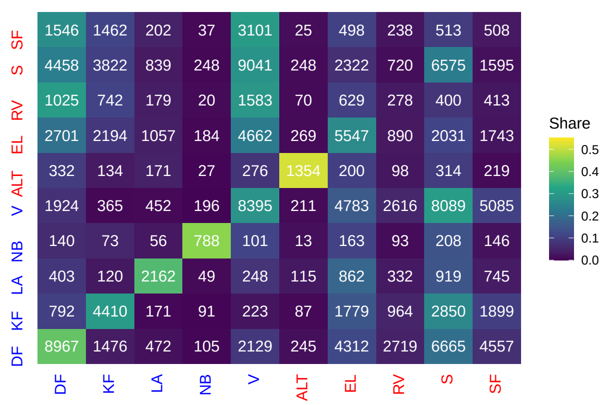
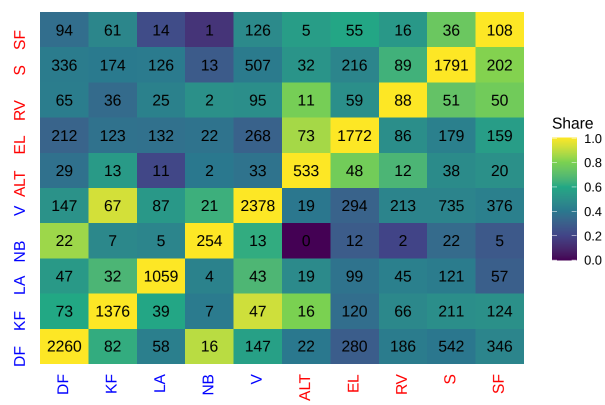


Figure G1: Share of dyadic speeches over time. The figure plots the number of dyadic speeches normalized by the speech count in every parliamentary term. Dyadic speeches are defined using the dictionary-based approach outlined in the main text . It also shows the number of speeches directed at inbloc and outbloc dyads normalized by the total number of speeches for each term. We only include parliamentary terms with at least 9000 speeches in the figure.



(a) Speech dyads



(b) Vote dyads

Figure G2: Heatmap of party dyads between the 10 parties. Colors denote the left and right bloc. Panel G2a shows the distribution of party dyads at the speech level scaled by the number of speeches directed from speaker party *i* to target party *j* normalized by the total number of speeches given by *i*. Numbers within each dyad represent the total number of speeches for each dyad. Panel G2b shows the distribution of party dyads at the vote level scaled by the share of votes that party *i* and *j* agree upon. Numbers within each dyad represent the total number of speeches that mentions the party dyad in bill debates. Note that the shares and counts are not the full distribution of how parties vote together but only on bills where we also identify dyadic speeches. We show the dyadic distribution since this is the variation we are studying in the empirical analysis.

H Construction of Verbal Covariates

Sentiment

To measure sentiment, we used the Danish analysis tool Sentida, which is available in both `R` and `Python`. Sentida is a sophisticated rule-based system that makes use of adverb modifiers, exclamation marks, and negations when scoring the sentiment of each word. We computed a sentiment score at the speech level and normalized the score by speech length (number of words) to take varying lengths of speeches into account. As a result, our sentiment measure is a mean across all words in the speech. The average sentiment was 0.32 with a standard deviation of 0.29 (see Table F1).

Emotionality

To measure emotionality, we used the approach put forward by Gennaro and Ash (2022). Unlike traditional dictionary-based measures, the measure is continuous and less sensitive to the presence of individual words in a dictionary. The method scales speeches by combining word embeddings with two dictionaries containing words associated with each pole. We define one pole for affective language and one pole for neutral language. If pitch merely tracks emotive language, this would imply that the affective pole tracks high pitch values and the neutral poles low pitch values.

To construct the poles, we start by collecting a dictionary containing valence scores for Danish words using AFINN (Nielsen 2011). The dictionary contains a total of $3,552$ words with a mean score of -0.62, a standard deviation of 2.12, and a minimum and maximum value of -5 and 5, respectively. We define affective words as words having an absolute valence score of 3 or more and neutral words as words having an absolute valence score of 1. The full distribution of the valence scores is shown in Figure H1 colored by affective, neutral, and undefined words. Based on our definitions, we define two lists of seeds with 771 affective and 517 neutral words.

We then fit locally trained word embeddings based on the Word2Vec algorithm (Mikolov et al. 2013) as implemented in the `Python` library gensim. We fit the model on our corpus speeches after removing speech by chair and speeches with less than 40 words (a total of $393,264$ spanning 28 parliamentary terms). Since word embeddings learn by the context in which each word occurs, we

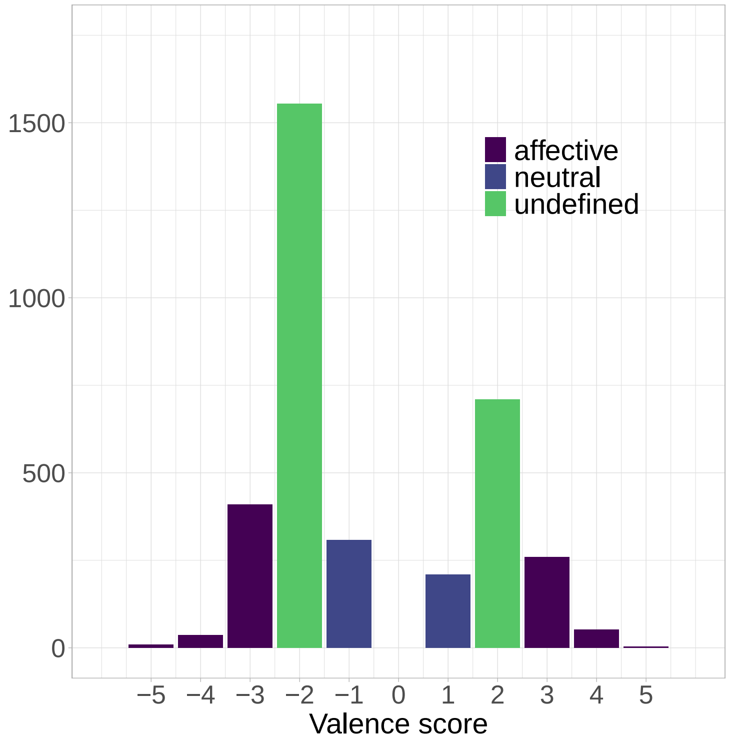


Figure H1: Distribution of AFINN valence scores colored by affective and neutral definitions.

followed the convention and used minimal preprocessing. We removed digits, punctuation, and Danish stopwords (as defined by SpaCy), and converted our words to lowercase. We also removed words with two characters or less. After preprocessing, our vocabulary used to fit the model had $68,694$ unique words. We use 300-dimensional vectors, a window size of 20, a minimum word count of 10, downsampling of high-frequency words with threshold 0.001, and training for 10 epochs. We also tested other choices, but the results were virtually identical (Rodriguez and Spirling 2022). After training, we have a 300-dimensional vector representation of each word in our vocabulary.

We then used the word embeddings to scale speeches with an emotionality score based on our seed words. First, we define an affective and neutral dimension by taking the element-wise average of the vectors for each seed word. This gives us two new vectors, **A** and **N**, which are the average of word vectors **w** for words w in the affective $w \in A$ and neutral seed list $w \in N$ respectively.[16]

[16]Our affective dimension is the average of 469 word vectors and the neutral dimension is the average of 439 word vectors. The drops from 771 and 517 happen since not all seed words are in our vocabulary.

Second, we represent each speech as a vector using the same procedure as for the affective and neutral vectors. We denote each speech by a vector $\mathbf{s}_i$, which is the average of the word vectors for the words w in speech i. In the end, we have a 300-dimensional vector representation for each pole (affective and neutral), and for each speech i. Following Gennaro and Ash (2022), we then define our measure as:

$$E_i = \frac{\text{sim}(\mathbf{s_i}, \mathbf{A}) + b}{\text{sim}(\mathbf{s_i}, \mathbf{N}) + b}$$

where $\text{sim}(\mathbf{a}, \mathbf{b}) = (\mathbf{a} \cdot \mathbf{b}) / (||\mathbf{a}|| \cdot ||\mathbf{b}||)$ denotes the cosine similarity between vectors **a** and **b**, while b is a smoothing parameter set to $b = 1$. The measure scales each speech by its emotionality in a continuous space $E_i \in \mathbb{R}$ with $E_i = 1.0$ meaning that a speech is equally affective and neutral. Hence, $E_i > 1.0$ indicates that a speech is more emotive than neutral and vice versa for < 1.0 values. Our measure of emotionality is only weakly correlated with sentiment ($\rho = -0.092$) showing that the two measures capture separate dimensions of political speech (Gennaro and Ash 2022). Our emotionality measure has a mean of 1.00 and a standard deviation of 0.12. According to our measure, this means that Danish politicians evenly balance emotive and neutral language in their speeches.

Topics

To measure topics, we use a Structural Topic Model (Roberts et al. 2014) implemented in the stm package in `R` (Roberts et al. 2019). We estimated models using $k \in \{35, 40, \ldots, 55\}$ with close to identical results. Before fitting the models, we preprocessed our speeches using the quanteda package (Benoit et al. 2018). We removed speaker names, punctuation, numbers, and stopwords from the text, converted it to lowercase, and applied stemming. After tokenizing the text, we kept only tokens that contained three or more characters and which occurred ten or more times and in at least five speeches across the corpus. Finally, we also removed frequently occurring tokens. After inspecting the models, we decided to use $k = 40$ in the empirical analysis. This was a result of both model diagnostics and human validation. The latter showed that $k = 40$ contained more coherent and fine-grained topics than its rivals. For the $k = 40$, we were able to identify 35 coherent topics (87.5 pct.), which we labeled accordingly. In Figure H2, we show the results of bivariate regressions with z-standardized pitch as the dependent variable and a binary topic indicator as the predictor.

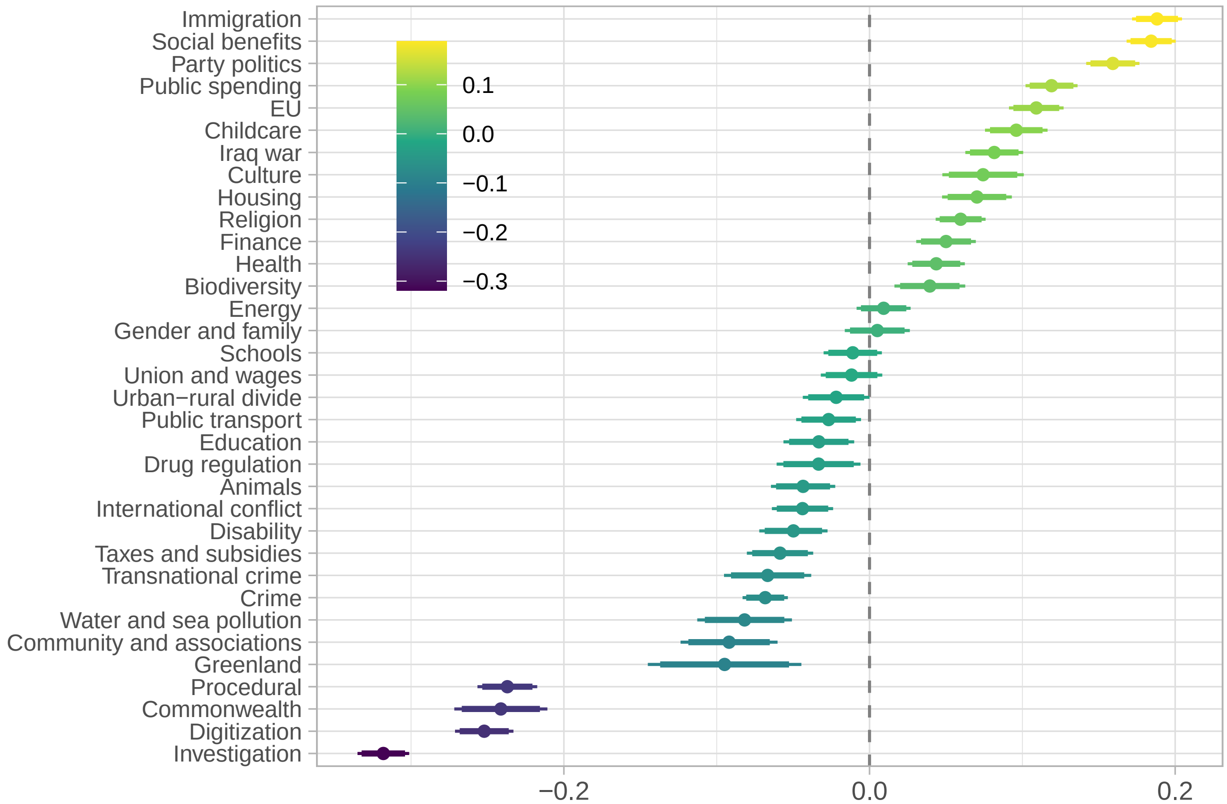


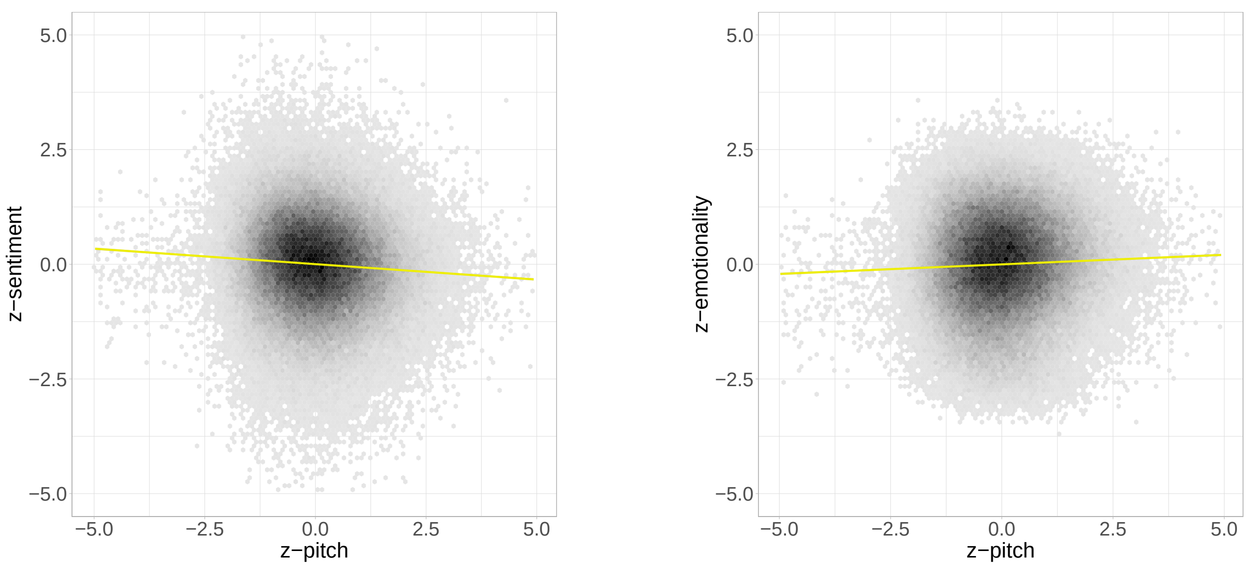
Figure H2: Correlation between topics and z-standardized pitch for a STM with $k = 40$ with 35 labeled topics. Topic labels are manually defined based on topic keywords (FREX, Lift, and Score). Larger values show that topics, on average, are associated with a higher pitch. Thick and thin error bars represent 90 and 95 pct. confidence intervals respectively.

Correlations

Table H1: Correlation between textual measures and pitch

	Sentiment	Emotionality
Estimate	−0.064***	0.040***
	(0.003)	(0.003)
N	147 169	147 148
R^2	0.00	0.00

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001



(a) z-standardized sentiment

(b) z-standardized emotionality

Figure H3: Correlation between z-standardized pitch and z-standardized sentiment and emotionality. The correlations are from simple bivariate linear regressions based on dyadic speeches.

Table H2: Mean and rank of pitch, sentiment, and emotionality by topic. Values are z-standardized to have zero mean and unit variance. Table is sorted by pitch rank.

	Mean			Rank		
	Pitch	Sentiment	Emotionality	Pitch	Sentiment	Emotionality
Immigration	1.674	0.487	0.990	1	8	9
Social benefits	1.640	0.400	-0.088	2	10	18
Party politics	1.444	-0.072	1.261	3	18	5
Public spending	1.110	-0.310	0.033	4	22	16
EU	1.031	-0.108	0.048	5	19	15
Childcare	0.931	-0.634	-0.666	6	26	27
Iraq war	0.806	1.851	1.523	7	3	2
Culture	0.756	-0.399	1.231	8	24	7
Housing	0.718	0.091	-0.846	9	13	29
Religion	0.617	0.673	1.734	10	6	1
Finance	0.544	0.087	-0.376	11	14	20
Health	0.490	0.474	-0.448	12	9	21
Biodiversity	0.459	-0.646	-0.540	13	27	24
Energy	0.202	-1.219	-0.603	14	30	25
Gender and family	0.169	-0.151	1.408	15	21	3
Schools	0.034	-1.267	0.223	16	31	12
Union and wages	0.028	0.168	-1.020	17	12	31
Urban-rural divide	-0.058	-0.909	-0.284	18	29	19
Public transport	-0.099	-0.032	-0.014	19	16	17
Education	-0.154	-1.365	-0.535	20	32	23
Drug regulation	-0.157	2.346	0.722	21	2	10
Animals	-0.239	0.642	-0.456	22	7	22
International conflict	-0.241	0.802	1.359	23	5	4
Disability	-0.293	-0.143	0.087	24	20	14
Taxes and subsidies	-0.366	0.032	-0.785	25	15	28
Crime	-0.429	2.908	1.250	26	1	6
Transnational crime	-0.441	0.807	0.129	27	4	13
Water and sea pollution	-0.570	0.190	-1.562	28	11	33

Community and associations	-0.658	-1.450	-0.908	29	33	30
Greenland	-0.688	-0.338	1.068	30	23	8
Procedural	-1.854	-0.658	-0.617	31	28	26
Commonwealth	-1.928	-1.713	0.449	32	34	11
Digitization	-1.977	-0.505	-2.628	33	25	34
Investigation	-2.501	-0.038	-1.142	34	17	32

I Regression Tables

Table I1: Regression table for H1

	Dependent Variable: Standardized Pitch				
	A	B	C	D	E
Intercept	−0.0481***	0.0248+	−0.2178**		
	(0.0139)	(0.0136)	(0.0796)		
Text sentiment		−0.2085***			−0.2134***
		(0.0160)			(0.0160)
Emotionality			0.1772*		0.0822
			(0.0711)		(0.0835)
Outbloc target	0.1537***	0.1440***	0.1420***	0.1141***	0.0997***
	(0.0205)	(0.0213)	(0.0196)	(0.0204)	(0.0194)
Topic FE	✗	✗	✗	✓	✓
N	147 273	147 273	147 252	122 469	122 448
R^2	0.01	0.01	0.01	0.02	0.02

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors clustered at the dyad level.

Table I2: Regression table for H2

	Dependent Variable: Standardized Pitch				
	A	B	C	D	E
Intercept	−0.0392+	0.0639**	−0.3920***		
	(0.0207)	(0.0219)	(0.1046)		
Text sentiment		−0.2809***			−0.2788***
		(0.0236)			(0.0255)
Emotionality			0.3748***		0.1557
			(0.0956)		(0.1090)
Out-vote	0.2693***	0.2496***	0.2430***	0.2289***	0.2025***
	(0.0355)	(0.0364)	(0.0346)	(0.0330)	(0.0330)
Topic FE	✗	✗	✗	✓	✓
N	31 181	31 181	31 180	26 962	26 961
R^2	0.02	0.02	0.02	0.04	0.04

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors clustered at the dyad level.

Table I3: Regression table for H3

	Dependent Variable: Standardized Pitch				
	A	B	C	D	E
Intercept	0.0089	0.0875***	−0.2451***		
	(0.0168)	(0.0173)	(0.0676)		
Text sentiment		−0.2483***			−0.2482***
		(0.0141)			(0.0156)
Emotionality			0.2544***		0.1823*
			(0.0676)		(0.0859)
High-publicity debate	0.5922***	0.5980***	0.5773***	0.5326***	0.5285***
	(0.0279)	(0.0273)	(0.0280)	(0.0284)	(0.0281)
Topic FE	✗	✗	✗	✓	✓
N	118 767	118 767	118 748	98 531	98 512
R^2	0.02	0.02	0.02	0.03	0.04

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors clustered at the dyad level.

Table I4: Regression table for H4

	Dependent Variable: Standardized Pitch				
	A	B	C	D	E
Intercept	−0.0062	0.0674**	−0.3409***		
	(0.0208)	(0.0205)	(0.0378)		
Text sentiment		−0.2292***			−0.2234***
		(0.0155)			(0.0193)
Emotionality			0.3327***		0.2035***
			(0.0288)		(0.0418)
Target bargaining	0.1735**	0.1742**	0.1850***	0.1655***	0.1683***
	(0.0374)	(0.0372)	(0.0349)	(0.0345)	(0.0330)
Topic FE	✗	✗	✗	✓	✓
N	147 219	147 219	147 198	122 420	122 399
R^2	0.00	0.01	0.00	0.02	0.02

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors clustered at the dyad level.

J Alternative Specifications

Time Variation

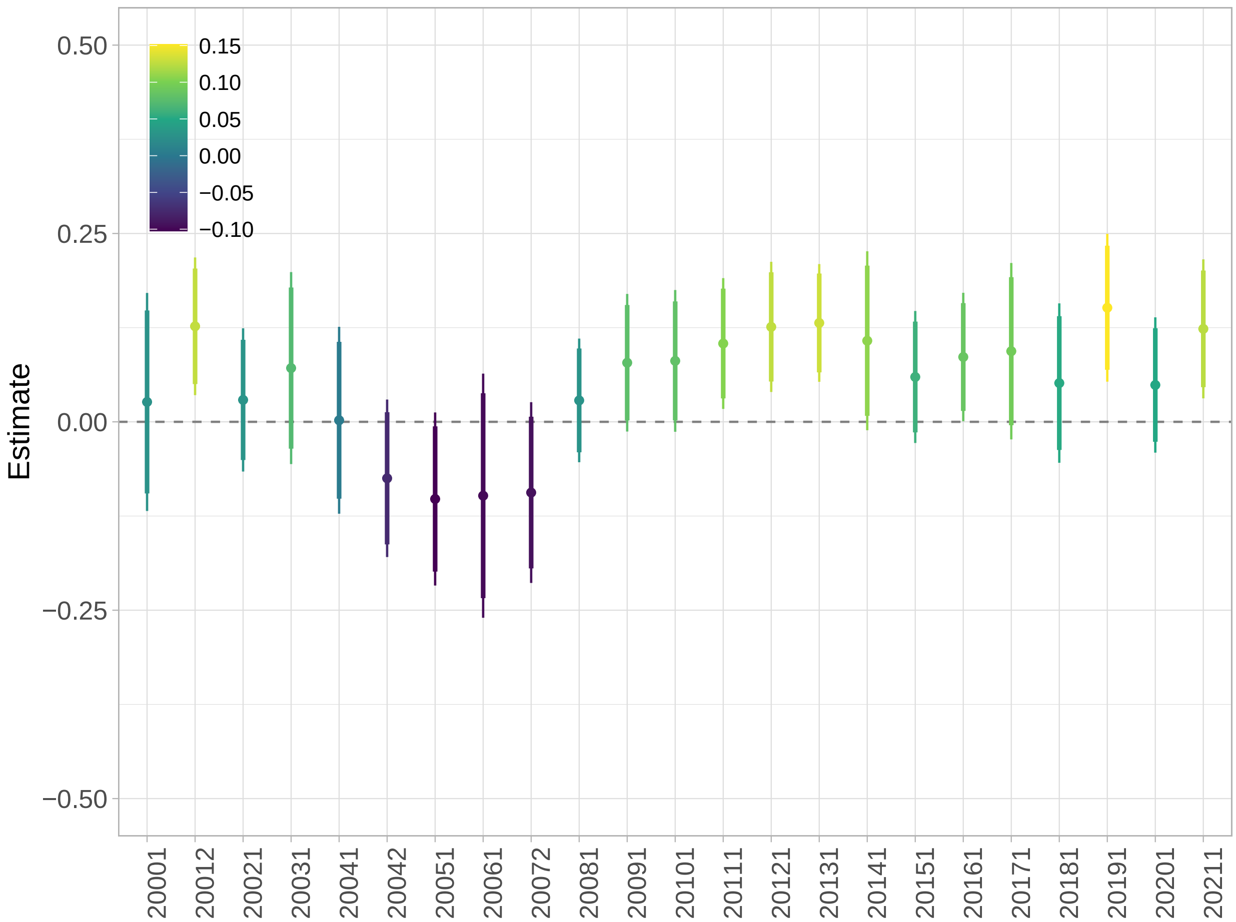
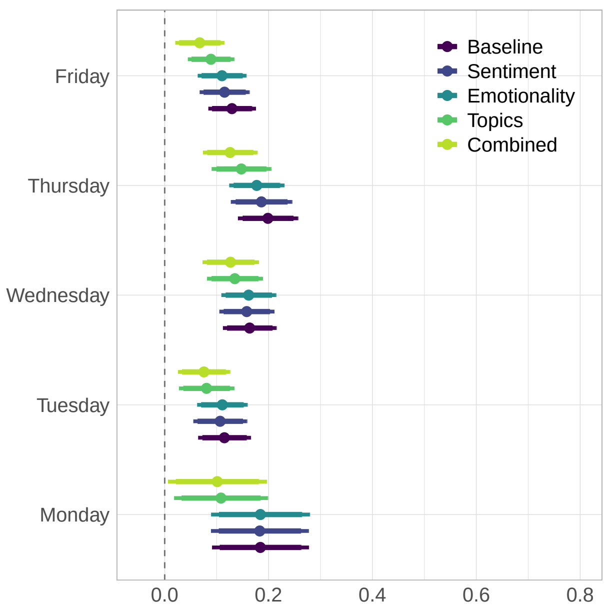
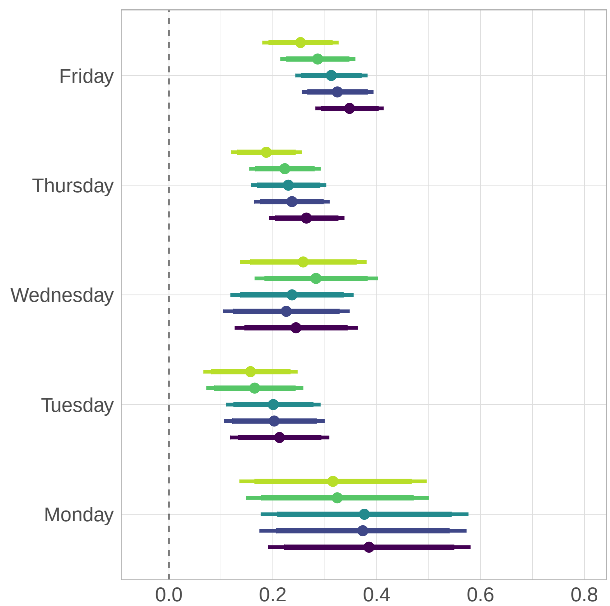


Figure J1: Estimates on outbloc target tested in H_1 estimated for each year in the data. Each estimate is from a regression model with sentiment, emotionality, and topic fixed effects as covariates ('Combined'). Standard errors are clustered by dyad. Thick and thin error bars represent 90 and 95 pct. confidence intervals respectively.



(a) Partisan Polarization



(b) Policy Conflict

Figure J2: Coefficients for partisan polarization (left panel) and policy conflict (right panel) estimated separately for each weekday. Each estimate subsets the data to speeches from Monday, Tuesday, Wednesday, Thursday, or Friday. Standardized pitch is used as the outcome. Standard errors are clustered at the dyad level (speaker party $\leftrightarrow$ target party). Thick and thin error bars represent 90 and 95 pct. confidence intervals respectively. X-axes are held fixed across the two panels to maximize comparability.

Testing H2 Including Dyad Fixed Effects

Table J1: Regression table for H2 with dyad FE

	Dependent Variable: Standardized Pitch				
	A	B	C	D	E
Text sentiment		−0.2714***			−0.2847***
		(0.0234)			(0.0254)
Emotionality			0.3582***		0.2143*
			(0.0701)		(0.0845)
Out-vote	0.2129***	0.2047***	0.2143***	0.2089***	0.2021***
	(0.0346)	(0.0346)	(0.0346)	(0.0319)	(0.0320)
Topic FE	✗	✗	✗	✓	✓
Dyad FE	✓	✓	✓	✓	✓
N	31 181	31 181	31 180	26 962	26 961
R^2	0.03	0.04	0.03	0.05	0.06

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors clustered at the dyad level.

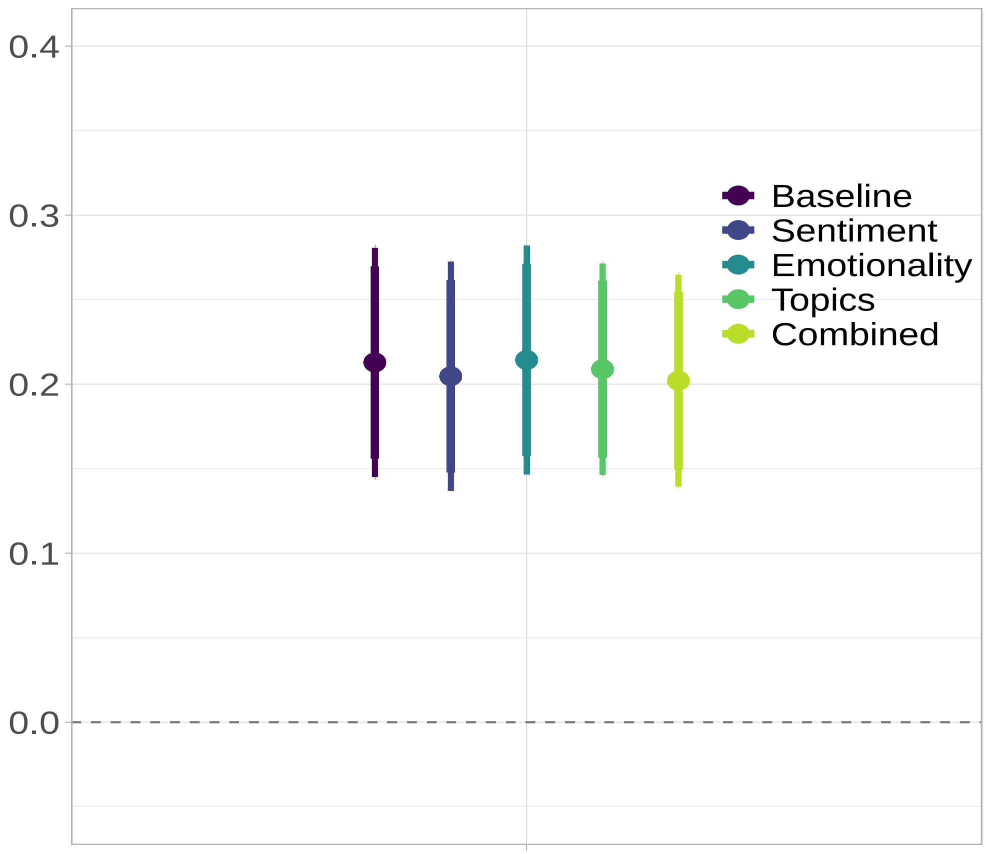


Figure J3: Coefficients on policy conflict tested in H_2. The model is identical to the results in Panel 2b in Figure 2 in the main text but includes additional dyad fixed effects.

Contested Votes

Here, we test H_2 using the closeness of votes as the predictor instead of vote disagreement. For each vote, we measure closeness as the absolute value of the vote margin with lower values implying closer votes. For each law, we divide the total of 'yes' and 'no' votes by the number of vote calls. Since each amendment is subject to vote, a single law can contain many votes. We can not link speeches to specific amendments, and hence we normalize it by the total number of amendments for each law. Otherwise, more amendments would be systematically related to policy conflict by construction.

Like the vote disagreement measure, this is a post-debate proxy. If legislators' nonverbal signals arise from policy disagreement, this should show up in a negative relationship between vote margin and pitch. That is, the pitch should decrease with increasing vote margins, reflecting more overall legislative agreement. We show the results with and without dyad fixed effects in Table J2-J3 and Figure J4. As expected, we find a negative and significant ($p < 0.001$) result across all five models, also when including dyad fixed effects. Without dyad fixed effects, an increase in vote margin reduces the pitch by 0.102 standard deviations. With dyad fixed effects, legislators speak with 0.093 standard deviations lower pitch.

Table J2: Regression table for log-transformed vote difference.

	Dependent Variable: Standardized Pitch				
	A	B	C	D	E
Intercept	0.4745***	0.5462***	0.0536		
	(0.0323)	(0.0335)	(0.1028)		
Text sentiment		−0.2717***			−0.2780***
		(0.0248)			(0.0245)
Emotionality			0.4111***		0.2205+
			(0.1052)		(0.1145)
Vote difference (log)	−0.1258***	−0.1193***	−0.1187***	−0.1089***	−0.1016***
	(0.0076)	(0.0077)	(0.0071)	(0.0075)	(0.0074)
Topic FE	✗	✗	✗	✓	✓
N	30 655	30 655	30 654	26 522	26 521
R^2	0.03	0.03	0.03	0.05	0.05

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

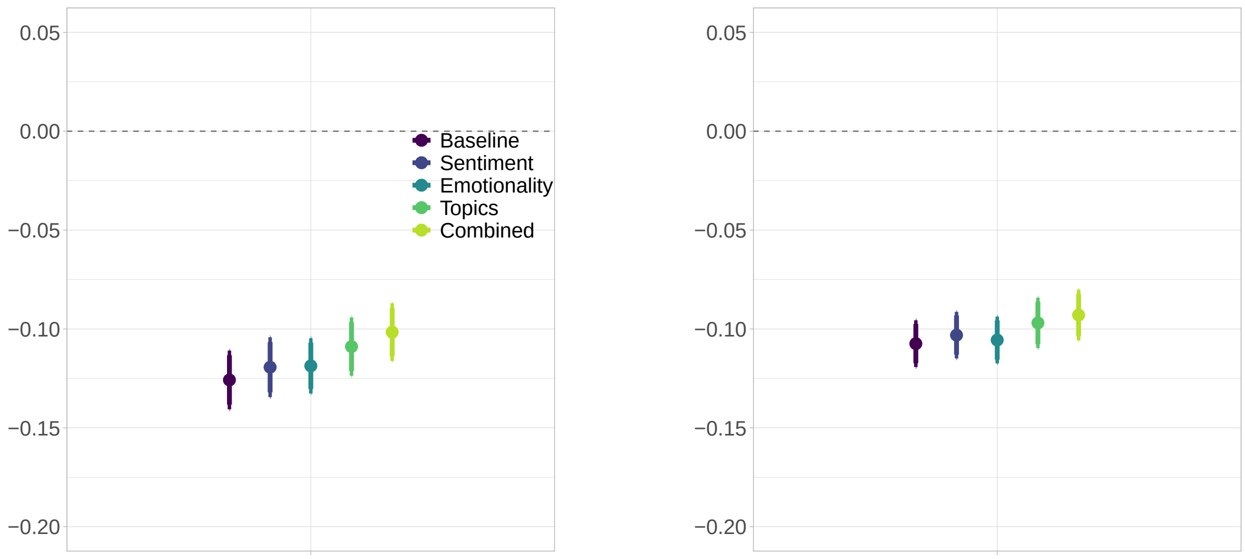
Standard errors clustered at the dyad level.

Table J3: Regression table for log-transformed vote difference with dyad FE

	Dependent Variable: Standardized Pitch				
	A	B	C	D	E
Text sentiment		−0.2464***			−0.2724***
		(0.0239)			(0.0236)
Emotionality			0.2691**		0.1563
			(0.0852)		(0.0950)
Vote difference (log)	−0.1074***	−0.1031***	−0.1056***	−0.0969***	−0.0930***
	(0.0060)	(0.0060)	(0.0060)	(0.0065)	(0.0065)
Topic FE	✗	✗	✗	✓	✓
Dyad FE	✓	✓	✓	✓	✓
N	30655	30655	30654	26522	26521
R^2	0.04	0.05	0.05	0.06	0.07

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Standard errors clustered at the dyad level.



(a) Dyad FEs ✗ **(b)** Dyad FEs ✓

Figure J4: Coefficients for policy conflict measured with the vote difference between ‘yes’ and ‘no’ votes on standardized pitch. Vote difference is measured as the (natural) logarithm of the absolute difference between ‘yes’ and ‘no’ votes. Panel J4a corresponds to the standard models used in the main analysis. Panel J4b uses the same standard models but with additional dyad fixed effects. Standard errors are clustered at the dyad level (speaker party ↔ target party). Thick and thin error bars are the model-specific 90 and 95 pct. confidence intervals respectively. Y-axes are held fixed across the two panels to maximize comparability.