UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E ATUÁRIA
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Party Competition Under Dictatorships: Evidence From Congressional Speech in Brazil

Competição Partidária sob Ditaduras: Evidências a partir de Discursos Parlamentares no Brasil

Alan Funtowicz
Supervisor: Raphael Bottura Corbi

São Paulo
2024

Prof. Dr. Carlos Gilberto Carlotti Júnior
Reitor da Universidade de São Paulo

Profa. Dra. Maria Dolores Montoya Diaz
Diretora da Faculdade de Economia, Administração, Contabilidade e Atuária

Prof. Dr. Cláudio Ribeiro de Lucinda
Chefe do Departamento de Economia

Prof. Dr. Mauro Rodrigues Júnior
Coordenador do Programa de Pós-Graduação em Economia

ALAN FUNTOWICZ

Party Competition Under Dictatorships: Evidence From Congressional Speech in Brazil

Competição Partidária sob Ditaduras: Evidências a partir de Discursos Parlamentares no Brasil

Dissertação apresentada ao Programa de Pós-Graduação em Economia do Departamento de Economia da Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo, como requisito parcial para a obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Raphael Bottura Corbi

Versão Original

São Paulo

2024

Catalogação na Publicação (CIP)
Ficha Catalográfica com dados inseridos pelo autor

Funtowicz, Alan.
Party Competition Under Dictatorships: Evidence From Congressional Speech in Brazil / Alan Funtowicz. – São Paulo, 2024.
63 p.

Dissertação (Mestrado) – Universidade de São Paulo, 2024.
Orientador: Raphael Bottura Corbi.

1. Party competition. 2. Congressional speech. 3. Emotive rhetoric. 4. Natural Language Processing. 5. Dictatorships. I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

Alan Funtowicz

Party Competition Under Dictatorships: Evidence From Congressional Speech in Brazil
Competição Partidária sob Ditaduras: Evidências a partir de Discursos Parlamentares no Brasil

Work presented as a requirement for the degree of Master in Economics from the Institute of Economic Research at the University of São Paulo (IPE-USP).

Raphael Bottura Corbi
University of São Paulo

Gloria Gennaro
University College London

Lucas Martins Novaes
INSPER

Marcos Yamada Nakaguma
FGV-EESP

São Paulo
2024

Agradecimentos

Começo agradecendo as pessoas mais importantes para a realização deste trabalho: meus pais, Ingrid e Ronaldo. O apoio incondicional deles em todas as etapas da minha formação foi e continua sendo peça chave para minha trajetória pessoal e profissional. Agradeço aos meus irmãos, Andre e Renan, e a toda minha família pela força durante esse período. À minha avó Amália, por todo o carinho e preocupação em saber como andavam os estudos e o desenvolvimento da dissertação.

Aos meus amigos de infância, colégio, graduação, mestrado e os que encontrei no meio do caminho: este trabalho não teria sido possível sem as conversas e discussões (principalmente as não acadêmicas) que tivemos ao longo do tempo. Sou imensamente grato por ter vocês na minha vida.

Agradeço ao meu orientador, Raphael Corbi, por toda a dedicação e tempo despendidos a mim. É imensurável o quanto aprendi durante nossas reuniões às quinta-feiras nesses últimos dois anos. Agradeço ao professor Naercio Menezes por todo o apoio desde os tempos de graduação. Gostaria de agradecer também a todos os professores e funcionários do departamento de Economia da USP com quem tive contato pelos aprendizados durante os anos de graduação e mestrado.

Por fim, agradeço aos membros da banca por terem aceito o convite. Ao CNPq e à FIPE pelo apoio financeiro durante parte importante deste trabalho.

"Quando chegar o momento, esse meu sofrimento
Vou cobrar com juros, juro
Todo esse amor reprimido, esse grito contido
Este samba no escuro
Você que inventou a tristeza
Ora, tenha a fineza de desinventar
Você vai pagar e é dobrado
Cada lágrima rolada nesse meu penar"
- Francisco Buarque de Hollanda

Abstract

This paper examines the role of political competition and emotional rhetoric on congressional discourse during the Brazilian military dictatorship era. Leveraging historical archives of Congressional records, we curated a comprehensive database of parliamentary speeches from 1975 to 1986. Specifically, our analysis explores an exogenous political reform that increased party competition within the opposition during the redemocratization phase under the military regime. By combining theoretical frameworks drawn from the party competition literature with newly developed Natural Language Processing (NLP) techniques, our findings unveil a significant increase in the primary opposition party's use of predominantly negative emotional rhetoric. These results highlight the impact of political competition on nonpolicy attributes in party strategies and demonstrate how NLP methods can be integrated with models of party competition to study political speech.

Key-words: Party competition; Congressional speech; Emotive rhetoric; Natural Language Processing; Dictatorships.

Resumo

Este artigo examina o papel da competição política e da retórica emocional no discurso parlamentar durante a era da ditadura militar brasileira. Utilizando arquivos históricos de registros do Congresso, elaboramos um banco de dados abrangente de discursos parlamentares de 1975 a 1986. Especificamente, nossa análise explora uma reforma política exógena que aumentou a competição partidária dentro da oposição durante a fase de redemocratização sob o regime militar. Ao combinar a literatura teórica sobre competição partidária com técnicas recém-desenvolvidas de Processamento de Linguagem Natural (PLN), nossos resultados revelam um aumento significativo no uso de retórica emocional predominantemente negativa pelo principal partido de oposição. Esses resultados destacam o impacto da competição política em atributos de *nonpolicy* nas estratégias partidárias e demonstram como métodos de PLN podem ser integrados com modelos de competição partidária para estudar o discurso político.

Palavras-chave: Competição partidária; Discurso parlamentar; Retórica emocional; Processamento de Linguagem Natural; Ditaduras.

List of Figures

List of Tables

Contents

1 Introduction

A substantial body of literature analyzes the strategic behavior of political parties. The seminal work of Downs (1957) postulates that parties formulate their policies primarily to win elections rather than implement their ideologies, often converging towards the center of the political spectrum. For instance, parties may adopt pragmatic positions on immigration or healthcare policies to appeal to a broader electorate. Conversely, Stokes (1963) challenges the Downsian model by introducing the concept of valence issues, proposing that political competition involves not only positioning along a left-right spectrum but also managing universally valued objectives effectively. Valence refers to nonpositional issues where there is broad agreement among the electorate on desired outcomes, such as reducing corruption, requiring parties to demonstrate competence and integrity, rather than merely their stance on specific issues.[1]

Recent theoretical work highlights how valence influences candidates' strategies in elections ((Schofield (2003); Green and Hobolt (2008)). In particular, valence tends to accumulate in environments with low ideological polarization. In contrast, politicians with lower valence tend to adopt more polarized positions, regardless of whether their platforms are determined exogenously or endogenously (Ashworth and Bueno de Mesquita (2009)). Using data from British party manifestos and leaders' speeches, Kosmidis et al. (2019) found a negative empirical correlation between valence and polarization. This suggests that parties tend to use a more positive emotional tone when they are less polarized. However, there is limited evidence to support a direct causal effect of increased competition on valence accumulation.

This paper bridges the gap by leveraging an exogenous political competition shock in Brazil—the entry of new opposition parties—to examine its impact on valence, evaluated through emotive rhetoric. During the Brazilian military dictatorship (1964-1985), the government banned all parties except ARENA, supported by the military, and MDB, a moderate opposition party. This two-party system persisted until 1979, when new parties were legally allowed. Consequently, only opposition parties were formed, enhancing the competitive landscape for MDB against ARENA. In line with theoretical models of party competition, our results show positive and significant increases in the use of emotional rhetoric by politicians affiliated with the MDB after the reform, demonstrating that the party responded to the competition shock by increasing valence levels. We structure our argument in four steps, outlined below.

First, by scraping historical Congressional Records, we construct a new database

[1]For example, positional issues include abortion and the death penalty, where opinions are typically divided. In contrast, valence issues include the care of the elderly and education. Valence can also encompass candidate competence, campaigning skills, and character traits (Kosmidis et al. (2019)).

comprising the majority of speeches delivered in the Lower Chamber (*Câmara dos Deputados*) from 1975 to 1982. This database, combined with newly developed Natural Language Processing (NLP) techniques, enables us to investigate the changes in emotional rhetoric following the entry of the new parties. Second, following Lauderdale and Herzog (2016), we calculate elected deputies' individual ideological positions (ideal point estimates) using expressed disagreement in political discourse. The results indicate that the parties opposing the dictatorship were more ideologically aligned, supporting the interpretation by some scholars that the military allowed the formation of more parties to deliberately fragment the opposition.[2]

Third, we calculate an emotion score for each speech using the method developed by Gennaro and Ash (2022). Descriptive data show variations in emotionality across demographics and, more importantly, a significant increase in emotional rhetoric by opposition members following the end of the two-party system. Finally, we conduct regressions using a difference-in-differences framework to quantify this effect. This analysis rules out the influence of year, individual characteristics, speech length, or compositional effects (e.g., deputies switching parties). Furthermore, by analyzing the negative-to-positive emotional words ratio, we find that the opposition employed more negative emotional language in Congress following the reform.

Our first contribution involves the use of quantitative text analysis to study Congressional speeches, building on the emerging literature that employs new NLP methods to explore various dimensions of political communication. Recently, significant advances have been made in the theoretical literature utilizing new methods to measure differences in speeches across groups, which can be applied to legislative debates (Lauderdale and Herzog (2016); Gentzkow et al. (2019a); Gentzkow et al. (2019b); Ash and Hansen (2023); Grimmer et al. (2022)). This has been followed by empirical works applying these methods to study topics such as differences in emotionality (Gennaro and Ash (2022)), sociodemographic characteristics (Fiva et al. (2023)), and metaphor use (Picard and Stammbach (2022)). We contribute to this body of work by demonstrating how the NLP literature can be integrated with established theoretical models of party competition.

This paper also contributes to the political competition literature that analyzes policy and nonpolicy attributes in parties' rhetoric. Osnabrügge et al. (2021) find evidence that emotive rhetoric in parliamentary speeches is an important nonpolicy tool that legislators strategically use to appeal to voters. Kosmidis et al. (2019) uses data from British party manifestos and leaders' speeches to evaluate the relationship between valence, defined as the message's positive emotional tone, and party competition. The study finds a negative correlation between valence and polarization, suggesting that parties employ a more positive emotional tone when they are less ideologically polarized

[2]See, for example, Napolitano (2018) and Batistella (2021).

and when voter preferences are uncertain. This strategic use of positive emotional tone as valence indicates that parties differentiate themselves in low-polarization contexts or amidst voter uncertainty.

Most of the literature discussed above is more descriptive. Here, we provide empirical evidence showing the impact of a competition shock on valence, measured as overall emotional rhetoric, in political speech. Leveraging new Congressional speech data from the authoritarian period in Brazil, we explore an arguably exogenous competition shock—the entrance of new oppositional parties—to study the effects on valence, giving results a causal flavor.

Finally, we contribute by studying the determinants of legislative political speech under a military dictatorship. Gandhi et al. (2020) examine different authoritarian regimes and find that the legislative branch often plays an important role in shaping policy, even in antidemocratic settings. In our context, Desposato (2001) models legislative behavior using roll-call voting data from the Brazilian dictatorship. His findings show that deputies considered pressures from both the military and voters, sometimes going against the regime if it meant career progress, even under the threat of losing their mandate. Our findings add to this literature by demonstrating that rhetoric may be used as an instrument by the opposition in antidemocratic environments to differentiate themselves from deputies aligned with the regime.

The remainder of this article is organized as follows: Section 2 presents the historical background with a description of the military dictatorship in Brazil and the parties' dynamics; Section 3 describes a model of political competition that takes into account both policy and nonpolicy attributes; Section 4 explains the text analysis methods utilized; Section 5 presents the data and main results, and Section 6 provides some concluding remarks.

2 Historical Background

In this section, we summarize the historical background and institutional setting of our study.

2.1 The Brazilian authoritarian period

In the second half of the twentieth century, several South American countries, including Paraguay, Brazil, Bolivia, Peru, Argentina, Chile, and Uruguay, experienced dictatorial governments within the context of the Cold War. This shift was a response to left-wing uprisings and an attempt to maintain the status quo, often supported by the United States. Brazil was the second country to adopt an authoritarian government, a significant event given its status as the largest and most populous country in the region.

The authoritarian regime in Brazil had the backing of local elites, including most of the military high command, business organizations, large landowners, part of the high hierarchy of the Catholic Church, and the majority of the middle class. The opposition in civil society comprised students, trade unionists, and advocates of land reform. This opposition was not cohesive; some factions were more radical and desired a communist regime, while others preferred social-democratic reforms (Pereira (2020)).

On March 31, 1964, the democratic government of João Goulart was overthrown, with the dictatorship formally starting on April 1, 1964. Although not a communist, Goulart was perceived as indecisive in a scenario where inflation reached 80 percent annually, and the left was pressing for social reforms. With the support of the military, sectors of civil society, the media, and governors of the richest states, Congress declared the presidency vacant in Goulart's absence and removed many elected legislators. On April 11, 1964, in an attempt to lend a legalistic appearance to the coup, the first president of the authoritarian regime, Marshal Castelo Branco, was elected by the remaining Congress members (Napolitano (2018)).

The heterogeneous alliance among the supporters of the coup did not last long due to conflicts over the military's perceived role in politics. Some sectors within the army argued that politicians from all spectra caused institutional instability and advocated for a central role for the military. Consequently, a high-ranking general always occupied the presidency during the dictatorship, concentrating power. Nevertheless, the new government consistently tried to maintain a facade of legitimacy (Batistella (2021)), ruling through a series of Institutional Acts (*Atos Institucionais*) that restricted but did not extinguish the Legislative and Judiciary branches and kept direct elections for local governments, the Lower and Upper Houses.

The authoritarian period in Brazil is often classified into three distinct phases:[3]

- *A legalist disguise for the dictatorship* (1964-1968): The military dictatorship was established under the guise of legality, revoking some of Goulart's progressive laws and suppressing opposition. Harsh economic measures and restricted labor rights were imposed. Elections were manipulated to ensure military control, leading to significant repression of students and workers. Institutional Acts granted the president extensive powers, bypassing democratic norms.

- *Years of Lead* (1969-1978): This phase was marked by heightened repression, characterized by institutional acts that concentrated power with the president and military courts. Significant human rights abuses occurred, including torture and political assassinations. Economic difficulties and international pressures eventually led to gradual political liberalization efforts under President Geisel, who sought to ease censorship and restore some civil liberties.

- *Political Reopening* (1979-1985): This is the main background of our study. This phase saw gradual political reopening under President Figueiredo, who aimed to reduce the dictatorship's authoritarian aspects without fully relinquishing power. Opposition movements, including labor strikes and the amnesty campaign, grew stronger. Despite limited reforms and partial amnesty, the government faced mounting pressure, leading to the eventual transition to civilian rule with the election of Tancredo Neves and the end of the military regime.

We focus on the *Political Reopening* period, detailing significant legislative changes that initially restricted political freedoms and later restored them. The following subsection also explores the dynamics between ARENA and MDB during the two-party system, emphasizing the impact of the 1979 reform on MDB.

2.2 Political parties during the dictatorship

The most significant Institutional Acts for our study, examined in greater detail, are IA-2 (1965), which banned political parties, and IA-5 (1968), which intensified repression and significantly curtailed free speech. In 1978 and 1979, respectively, IA-5 was revoked, and Law No. 6767 allowed for the formation of more political parties, thereby reestablishing multipartidarism. We focus on the period from 1975 to 1982, characterized by gradual political reopening and culminating in the 1982 elections, which featured multiple parties and heightened competition. Figure A.2 shows the timeline of events related to our study.

[3]See detailed information (in Portuguese) here.

Institutional Act No. 2, enacted in October 1965, banned the thirteen existing political parties, instituted indirect presidential elections, and significantly expanded executive power over constitutional, legislative, and budgetary matters. The act further empowered military courts to try civilians accused of crimes against national security. Additionally, the president was granted the authority to dissolve Congress and other legislative bodies, remove congressmen, and declare a state of siege for 180 days without prior approval (Napolitano (2018)). This, along with Complementary Act No. 20, enacted in November 1965, which required a minimum of 120 deputies and 20 senators to form a new party, effectively imposed a two-party system that lasted until 1979.

We now discuss the dynamic between ARENA (*National Renewal Alliance*) and MDB (*Brazilian Democratic Movement*) during the two-party system period. Despite being the official opposition party, MDB members adopted a moderate stance that did not significantly challenge the dictatorship. In the 1970s, a well-known anecdote suggested that MDB was the party that said "Yes," while ARENA was the party that said "Yes, Sir," reflecting their lack of opposition to the military regime (Batistella (2021)). During this period, some voters who were critical of the dictatorship preferred to nullify their vote rather than support MDB.

However, it is important to note the internal dispute within MDB between radicals, who advocated for a harder stance, and moderates, who believed a weaker opposition was better than no opposition. The military utilized institutional and complementary acts to influence elections and control both the Lower and Upper Chambers, creating a delicate balance that MDB needed to maintain to avoid prohibition.

The repeal of IA-5 in October 1978 marked a significant milestone in civil society's pursuit of democracy. Regarded as the most repressive institutional act, it empowered the President to suspend Congress, dismiss public officials, confiscate property, and suspend political rights without judicial review, thereby significantly curbing civil liberties and political freedoms. Despite the period of gradual political reopening, the military government did not desire a change as radical as abolishing the decree. However, pressures from liberal political groups, the Catholic Church, middle-class sectors, social movements, and students compelled the government to concede and abolish it.

In November 1979, the political liberalization progressed with Congress's approval of Law No. 6767, which marked the end of the two-party system. Batistella (2021) states that this aimed to fragment the opposition (namely, MDB), as the newly established parties were predominantly moderate or left-wing, and ARENA (now renamed to PDS - *Social Democratic Party*) remained the only right-wing party. Table 2 shows a significant migration of MDB members to other parties in the period following this reform, with minimal defection from ARENA members.

After 1979, MDB was renamed to PMDB (*Party of the Brazilian Democratic Movement*), and there were four new parties:

- PP (*Popular Party*): the largest party besides ARENA and MDB, composed of former moderate MDB members and ARENA dissidents;
- PTB (*Brazilian Labor Party*): centrist, composed of labor leaders;
- PDT (*Democratic Labor Party*): left-wing, composed of labor and union leaders;
- PT (*Worker's Party*): left-wing, composed of union leaders.

The reform constituted a significant competition shock that was more pronounced for MDB. Figure 1 shows a significant migration of 63 deputies from MDB to new parties in 1980, while ARENA lost 20 of its members to PP. All parties, except ARENA, were pro-democracy, aiming to reorganize the country after years of authoritarian rule. In 1981, discussions emerged about uniting all center and left-wing parties into one. Although this did not materialize, it indicates that democratic parties at the time shared a common goal and were vying for space as the main opposition to the dictatorship-endorsed party.

Voters in 1975-1982. Mettenheim (1990) provides a comprehensive analysis of voters' behavior in Brazil from 1974 to 1982, arguing that the emergence of competitive party politics was a primary driver of voter alignment during this period. Based on electoral survey data, he identifies significant increases in political knowledge related to the liberalization of diverse information sources, such as newspapers. Figure 2 illustrates that important Congressional speeches were salient to the electorate through the main newspapers at that time, *Folha* and *Estadão*. This characterization of Congressional speeches as a communication channel between politicians and voters is crucial for interpreting legislative debates as instruments for garnering votes, particularly in light of the absence of an official Congressional TV channel until its founding in 1998.

3 A simple model of party competition

In this section, we present a theoretical framework for analyzing the impact of a party competition shock on the use of valence, measured as emotional rhetoric. Our analysis is based on the model by Ashworth and Bueno de Mesquita (2009), detailing its main features and how it adapts to our case. The core idea, justified and discussed in subsequent sections, is that the reform allowing the entrance of new parties represented a major competition shock to MDB, since these parties were pro-democracy and competed for space in the opposition. According to the model, both groups had an incentive to accumulate valence, as it serves as a tool for differentiation and attracting votes.

Model overview

We model candidate behavior in elections where candidates first choose policy platforms and then invest in valence. A key insight is that when platforms converge, such as in our setting, the incentives for valence accumulation are heightened, leading to intense competition over nonpolicy factors. Our analysis focuses on the behavior of two parties: MDB and "New parties", which we consider to be a single group due to their limited size and shared objective of stronger opposition to the dictatorship.

The voter chooses one candidate and cares about two attributes: policy platform $x \in \mathbb{R}$ and valence $v \in \mathbb{R}^+$. The voter's payoff function is defined as $v - (x^* - x)^2$, where x^* is the voter's ideal point (unknown *ex-ante* and commonly believed to be uniformly distributed). If indifferent, the voter randomizes 50-50. This means the voter's satisfaction increases with higher valence and decreases with greater deviation from their ideal point. We denote the two candidates by A and B. For simplicity, let a member of MDB be candidate A and a member of one of the new parties be candidate B, with the assumption that $x_A \leq x_B$. A candidate c chooses both x and v to maximize:

$$R \Pr(\text{c wins}) - v_c$$

where R represents the ratio of the benefit from winning office to the cost of accumulating valence.[4] The timing of the game is as follows:

- Candidates simultaneously choose (x_A, x_B).
- Candidates observe (x_A, x_B) and choose valences (v_A, v_B).
- x^* is determined by nature.
- The voter observes $(x_A, x_B, v_A, v_B, x^*)$ and chooses one of the candidates.

Voting

Given the voter's payoff function, she prefers A to B if:

$$v_A - (x^* - x_A)^2 > v_B - (x^* - x_B)^2, \tag{1}$$

prefers B if the inequality is reversed and is indifferent if the payoffs are equal.

[4] While such models usually consider valence as campaign spending, one could consider emotive rhetoric in Congress as costly due to time constraints for speeches, potential sanctions for deviating from party positions, and possible criticism from other Congress members.

Case 1: $x_A = x_B$

This is the case of convergent platforms. If $v_A = v_B$ (i.e., valences are equal), Equation (1) becomes an equality, so the voter is indifferent and randomizes 50-50. If $v_A \neq v_B$, the voter chooses the candidate with higher valence with a probability of 1.

Case 2: $x_A \neq x_B$

The optimal rule is a cutoff rule[5], that can be defined as:

$$\text{voter chooses } A \iff x^* < \hat{x}(x_A, x_B, v_A, v_B)$$

Where $x^* < \hat{x}(x_A, x_B, v_A, v_B)$ is obtained solving for x^* in Equation (1). Isolating $v_A - v_B$:

$$v_A - v_B > (x^* - x_A)^2 - (x^* - x_B)^2$$

Substituting $(x^* - x_A)^2 - (x^* - x_B)^2 = (x_A - x_B)(2x^* - x_A - x_B)$, dividing both sides by $2(x_A - x_B)$ and isolating x^* gives us:

$$x^* < \underbrace{\frac{x_B + x_A}{2} + \frac{v_A - v_B}{2(x_B - x_A)}}_{\hat{x}} \tag{2}$$

The main implication of Equation (2) (and of the voting subgame) is that a valence advantage has a greater impact on the cutoff when the platforms are closer together.

Valence

Rearranging the terms in Equation (1), A wins the election if the following condition is satisfied:

$$v_A > \underbrace{2(x_B - x_A)x^* + x_A^2 - x_B^2 + v_B}_{\equiv \mathcal{H}} \tag{3}$$

That is, there exists a "hurdle" $\mathcal{H}$ that candidate A needs to exceed to win. When the variance of $\mathcal{H}$ increases, the marginal benefit of valence (i.e., the increase in the probability of winning the election) decreases. Since the variance of $\mathcal{H}$ includes $2(x_B - x_A)x^*$, it

[5]This is due to the payoff function increasing in both x^* and x in a way that their combined effect is greater when both variables are high.

increases with polarization. Consequently, the marginal benefit of valence decreases with polarization. This establishes the main features of the valence subgame:

- Candidates have strong incentives for valence accumulation with fully convergent platforms.
- Candidates have no incentive to accumulate valence with highly divergent platforms.
- Candidates have some incentive to accumulate valence with moderately divergent platforms but less than in the case of full convergence.

Platforms

The final step in finding an equilibrium is when parties choose their platforms, considering the outcomes of the previous subgames. Recall from Equation (2) that when platforms are closely aligned, even minor differences in valence significantly influence the voter's decision. In the extreme case of identical platforms, even a minimal valence advantage would be decisive, leading to intense competition over valence and ultimately nullifying payoffs. However, since parties anticipate this, they avoid complete platform convergence.

The payoff function for choosing platforms depends on the relative positions of the candidates' platforms, the benefit-to-cost ratio for accumulating valence, and the distribution of voter preferences. In equilibrium, the platforms are chosen to balance the benefits of appealing to the median voter with the costs of engaging in valence competition. This results in a level of platform divergence where candidates are not identical but still positioned to attract a broad voter base while minimizing excessive spending on valence. Thus, the equilibrium involves candidates diverging just enough to make valence competition manageable, ensuring that their expected payoffs remain positive.

Competition shock with similar platforms

We propose that the entrance of new players on the left-wing spectrum in 1980 represented an exogenous competition shock, wherein parties were closely aligned in platforms and, therefore, had the incentive to accumulate valence. As outlined in the model, when platforms are closely aligned, even small differences in valence significantly influence voter decisions. This creates intense competition over valence as parties strive to gain even a slight edge. This, combined with greater freedom of speech, could help explain the strong shift in the use of emotional language by MDB members during the two years following the introduction of new parties into the political arena.

Endogenous versus exogenous platforms. The presented model assumes that platforms are endogenously determined. However, suppose platforms are exogenous (i.e., politicians do not anticipate changes in valence and, consequently, do not adjust their platforms beforehand). In that case, Equation (3) still delineates the voters' strategy, wherein an exogenous decrease in platform divergence amplifies incentives for valence. Since parties cannot adjust platforms, all adjustments will be in valence in that scenario. If they could change platforms, the effects of the exogenous shock could be softened, with some absorbed in platforms and some in valence. In our specific case, modeling platforms as exogenous does not alter the model's primary implications; in fact, it intensifies the effects on valence accumulation.

4 Natural Language Processing methods

Our main variables of interest are created using established and newly developed NLP methods. First, we calculate annual individual ideal point estimates (θ) for each congress member using the *Wordshoal* model from Lauderdale and Herzog (2016). We then calculate an emotion score for each speech, following the procedure outlined in Gennaro and Ash (2022). For each speech, we compute the ratio of negative to positive emotional words using our pre-defined emotion dictionary as a seed. Lastly, we apply the Latent Dirichlet Allocation (LDA) model to identify topics for each speech.

For all these steps, standard text-processing techniques are applied: words are tokenized; stopwords, punctuation, capitalization, and digits are removed; and each token is then stemmed with the Portuguese *Snowball* Stemmer.

4.1 Ideal point estimates

Estimates of parties' and candidates' policy positions (ideal points) are typically derived from roll-call votes (Poole and Rosenthal (1997)). While reliable, this method can be limited by the availability of individual data or a lack of variation due to strong party discipline (i.e., politicians voting with their party to avoid sanctions despite personal disagreements). Recently, the literature has expanded to include the analysis of legislative speech to estimate ideological positions. Lauderdale and Herzog (2016) proposes a method for estimating expressed disagreement using legislative speech. Specifically, the *Wordshoal* model quantifies political opinions by analyzing speeches in legislative debates.

First, it examines the frequency of different words used in each debate using the *Wordfish* model (Slapin and Proksch (2008)). The information is then combined to identify overall political speech patterns using Bayesian factor analysis. This method

accounts for changes in word meaning between debates, enabling a clearer understanding of each politician's stance on key issues based on their speech. Their analyses reveal that speech scores strongly reflect government-opposition dynamics and significant intra-party variations. Additionally, a second dimension distinguishing establishment from anti-establishment parties is identified. The model also shows a larger increase in speech polarization compared to roll-call votes, highlighting pronounced polarization in recent years, while roll-call polarization had already been higher for longer.

Our analysis estimates ideal points (θ) individually for each speaker in each year. We consider legislative sessions with more than five speeches and more than two different speakers. Figure 3 shows the mean of θ for each year-politician, aggregated by party. Two important observations are the growing ideological distance between ARENA and MDB after 1978 and the increased ideological proximity between MDB and the new parties after 1980. While this aligns with election competition theories predicting a policy convergence towards the center to attract the median voter, it is noteworthy that there is a significant ideological shift following the formation of new parties. This indicates that the new parties aligned more closely with MDB than with ARENA. This is understandable in the context of the dictatorship, as most congressmen forming these new parties were former MDB members.

Figures 4 and 5 present the individual ideal point estimates and the party-aggregated probability density function for ideal point estimates for the two-year periods before and after the reform that allowed more parties into the political system, respectively. After 1978, the ideological divergence became more evident at the individual level, with a higher concentration of MDB and new party politicians exhibiting lower θ values, whereas ARENA members tend to have higher θ values, as indicated by the blue and red masses in the figures. The probability density functions also suggest that the divergence between MDB and ARENA became more prominent in 1979, indicated by the increased distance between the two lines. Members of the new parties appear to align more closely with MDB, as their PDF is slightly left-skewed for this group.

4.2 Emotion score and ratio of negative to positive words

We use the emotion score proposed by Gennaro and Ash (2022), with minor adaptations to the Portuguese language. The main idea of their measure is to use a *word embedding* approach, transforming words into vectors where the direction carries semantic meaning. They build two main dimensions: an emotion pole and a reason pole. Each word has a relative emotionality measure, corresponding to its distance from the emotion pole relative to the reason pole. This approach can be applied to entire documents, such as congressional speeches, to measure their relative distance to both poles, thereby creating

a measure of emotionality for each speech.

As in Gennaro and Ash (2022), we use the 'Cognitive Processing' and 'Affective Processing' categories from the Linguistic Inquiry and Word Count (LIWC)[6] in Portuguese to select lists of words related to emotion, A, and reason, C. Words that appear in both dictionaries, contain digits or consist of multiple terms are excluded. We then apply the *Word2Vec* algorithm from Mikolov et al. (2013) to learn the word embeddings, capturing the co-occurrence patterns of words in congressional speeches. Each word w is represented by a vector $\mathbf{w}$, and each speech is represented by a vector $\mathbf{d}_i$, which is the average of the vectors $\mathbf{w}$ of the words that appear in the speech. Each list, A and C, has its own vector, $\mathbf{A}$ and $\mathbf{C}$, representing the averages of the vectors for the words in the lists, $w \in A$ and $w \in C$, respectively.

Each speech is then scaled along the emotion and cognition dimensions through the following measure:

$$Y_i = \frac{\text{sim}(\mathbf{d}_i, \mathbf{A}) + 1}{\text{sim}(\mathbf{d}_i, \mathbf{C}) + 1}, \tag{4}$$

Where $\text{sim}(\mathbf{d}_i, \mathbf{A})$ is the cosine similarity between the speech and the emotion pole, $\text{sim}(\mathbf{d}_i, \mathbf{C})$ is the cosine similarity between the speech and the cognition pole, and the addition of 1 in both the numerator and denominator smooths outliers. It is important to emphasize that a speech equally balanced between emotion and cognition has a value of $Y_i = 1$. Changes in the emotion score reflect shifts in emotion *relative* to cognition rather than changes in emotion alone.

Figure 6 shows the average party emotion score. Before 1980, emotion levels for ARENA and MDB were relatively constant, fluctuating together. After 1979, coinciding with the reform that allowed more parties into the political system, emotion spiked for MDB, surpassing ARENA for the first time. Additionally, the emotion levels for the new parties were higher than ARENA but lower than MDB.

In Figure 7, we show that there was not only an increase in the average emotion score post-reform for MDB members but also a significant shift in the entire distribution. This indicates that the change was not localized to the most emotional speakers but rather reflects a behavioral change within the majority of its members.

Ratio of negative to positive words. Our next exercise is to investigate whether this shift in emotion was mainly due to positive or negative emotions. We disentangle 'negative' and 'positive' emotions by separating these categories from our list of emotional words, A, defined at the beginning of this section. Here, we follow a traditional dictionary approach and use the ratio of negative to positive emotional words in each speech rather than word embeddings.

[6] A set of categorized dictionaries validated by linguists.

Figure 8 shows a notable rising trend in using negative emotional words by MDB members starting in 1978. ARENA members also show a smaller increase in the use of negative emotional words. There is a high base level of negative emotional words for the new parties, but a decreasing trend is observed from 1980 to 1982. This may be associated with the relaxation of free speech restrictions that began to take effect during the final chapter of the authoritarian period, notably marked by the repeal of Institutional Act No. 5 in 1978.

4.3 Topic modelling

To separate the speeches into different topics, we follow Blei et al. (2003) and run the Latent Dirichlet Allocation (LDA) model, an unsupervised algorithm that identifies clusters of words frequently appearing together. LDA assumes that each document is a mixture of a small number of topics and that each word in the document is attributable to one of these topics. The model outputs a set of topics, each represented by a distribution of words, and provides the topic distribution for each document, allowing us to understand the main themes present in the speeches.

To prepare the speeches for topic modeling, the text is first cleaned by removing non-alphanumeric characters, numbers, and one- and two-letter words. Punctuation, stopwords, and specific terms such as "president", "sir", and "speaker" are also removed. Next, the cleaned text is converted into a document-feature matrix (DFM), where each row represents a speech, and each column represents a term. The DFM is then trimmed by eliminating infrequent or overly common terms across documents[7], ensuring that the remaining terms are more representative of the distinct topics in the speeches. Finally, the LDA model is run on the processed DFM with 20 topics using the Gibbs sampling method.[8]

Topics are primarily used as controls in some of the regression tables, but we also analyze their distribution across different parties before and after the reform. Table B.3 shows that before the 1980 reform, the main topics addressed in Congress were the Legal System, Legislative Debate, Transportation, and Agribusiness. In contrast, after the reform, Politics and Democracy, Regional Development, and Defense became some of the dominant topics. Table B.4 presents the average emotion score of topics by party and period. Before the 1980 reform, topics such as Defense, Labor Rights, Politics and Democracy had higher emotional scores for both ARENA and MDB. After the reform, the emotion scores increased notably for topics like Media, Politics and Democracy, and Parliamentary Discourse across all parties. Emotion levels for Defense remained high,

[7] Only terms that appear at least five times and in no more than 85% of speeches are kept.
[8] The topics are then manually assigned based on the top thirty words that define them.

particularly for ARENA and MDB, while other topics like Health and Education, Finance, and Transportation showed relatively stable emotion scores throughout the period.

5 Data and Empirical Evidence

In this section, we present the main sources of data used and descriptive statistics, followed by empirical evidence suggesting that the entrance of new parties propelled valence competition. To advance claims on causality, we analyze the results using a difference-in-differences approach. The law permitting new parties is considered an exogenous competition shock for MDB relative to ARENA, triggering an increase in negative emotive rhetoric. Empirically separating the effects of increased competition from the loosening of free speech restrictions is not possible. We argue that the timing of these events strongly suggests that competition plays a major role. This is justified by the increase in polarization and the ratio of negative to positive emotional words before the reform, whereas the measure of emotive rhetoric only spikes in 1980, the year after the reform.

5.1 Congressional speeches

Congressional speech data are scraped from historical archives stored on the *Câmara dos Deputados* website. The website's search section contains information (date, phase, summary, speaker, publication page) for all speeches given since 1946 but not the speeches themselves. This served as a reference for the scraping process (see Column "Total" in Table A.1). After gathering this information, all Lower Chamber Diaries (*Diário da Câmara dos Deputados*) are downloaded. Since the speaker's name precedes all speeches and the precise page of the journal on which the speech is located is provided, it is possible to scrape them using regular expressions that match the name to the beginning of the speech and end at the start of another speech by a different speaker.

Table A.1 shows the result of the scraping process. Situations where the PDF has a reading problem, or speeches are not found constitute less than four percent of the total. The regular method, in which we match the full speaker name to the one in the PDF, is successful more than 80% of the time. If we extend the method to match names that are at least 80% similar (Fuzzy), the success rate is nearly 100%. As the speaker is not guaranteed to be the same in both the reference table and the PDF, a conservative approach is taken by using only those that match the full name.

Table 1 shows the number of deputies and speeches in our sample from 1975 to 1982. Almost all elected deputies delivered at least one speech during their mandates. The total number of deputies changed in 1979 following the elections held in 1978. The

number of MDB and ARENA members was fairly equal until 1979, although ARENA consistently maintained a majority. In 1980, there was a notable shift: while the number of ARENA members remained stable, MDB lost many seats, which were then occupied by members of new parties. However, MDB members typically delivered more speeches, approximately 1,000 more than ARENA, representing around 20% of total speeches in a year. This dynamic also shifted in 1980 when ARENA took the lead, as the opposition was now divided between MDB and members of the new parties.

The final step of the scraping process involves validating the speeches. Figure A.1 plots the 100 most frequent terms for each party and period. While a simple analysis of word clouds does not provide visual evidence of different speech patterns between party members, it does highlight frequently used words, such as "government" (*governo*), "state" (*estado*), "national" (*nacional*) and "Brazil".

5.2 Individual characteristics

Individual characteristics of the deputies are used to assess how emotionality varies by demographics and as control variables in the main regression tables. This information is sourced from the *Dados Abertos* API on the *Câmara dos Deputados* website[9]. Information on gender, age, profession, and education is available. Panel A of Table 3 presents individual characteristics by party and period. Our sample comprises mainly highly educated, middle-aged males working in legal and public service sectors. This is observed both across parties and periods.

Panel B shows the number of speeches per politician and speech length. Before the reform, MDB members had a significantly higher average number of speeches. After 1980, although MDB members still had a higher average, the difference diminished. This higher average did not translate into longer speeches, as the mean and median speech lengths were similar for both ARENA and MDB, pre- and post-reform, and for other parties after the reform.

5.3 Outcome variables

Panel C of Table 3 shows summary statistics for the outcome variables specified in Section 4. The average of θ is smaller for both MDB and ARENA after the reform. While the average of θ for new parties is higher than that for ARENA, the distribution's right tail is similar to that for MDB (0.681 and 0.671, respectively). For the emotion score, there is a 4.3% increase for MDB post-reform, whereas for ARENA, it corresponds to a 1.4%

[9]See Câmara dos Deputados Open Data.

increase. The ratio of negative to positive words also shows a more pronounced increase for MDB, from 0.964 to 1.35, compared to ARENA, which increased from 0.671 to 0.779.

5.4 Variations in emotionality across demographics

In our first empirical exercise, we assess how the emotion score and the ratio of negative to positive emotional words vary by individual characteristics. This provides a more in-depth data description and allows us to engage with the previous literature. Concretely, we test the following specification:

$$Y_{ist} = \gamma_1 * Gender + \gamma_2 * Age + \gamma_3 * Party + \gamma_4 * Education + \gamma_5 * Profession + \lambda_t + \varepsilon_{ist} \quad (5)$$

Where Y_{is} is the outcome variable Y, for individual i, and speech s. Reference values for gender, age group, party, educational level, and profession are, respectively, male, 23-35, ARENA, high school (completed or not), and public service. These variables are the same ones depicted in Panel A of Table 3. Standard errors are clustered at the politician level, and all columns include year-fixed effects (λ_t). Additionally, topic-fixed effects are included to assess whether there is intra-topic variation in the outcomes.

The results are displayed in Table 4. Female speakers exhibit higher levels of both emotional expression and the ratio of negative to positive emotional words compared to their male counterparts. Politicians affiliated with MDB and parties other than ARENA show elevated levels of emotion and greater use of negative emotional words, aligning with their overall opposition status during that period. Older individuals tend to be more emotional than younger ones, while higher-educated individuals display lower levels of emotionality. Politicians whose occupations are related to agriculture, arts and media, and religious and community services exhibit higher levels of emotion than those in public service. Interestingly, individuals in healthcare and those with religious backgrounds tend to use less negative emotional rhetoric.

Our findings dialogue with the empirical literature examining how individual characteristics influence rhetoric in legislative debates. There are no major disagreements between our findings and the existing empirical evidence. Gennaro and Ash (2022) show that Democrats tend to use more emotional language than Republicans in the US Congress, and members from historical minorities—Blacks, Hispanics, Catholics, and Jews—use more emotional language than Whites and Protestants. Picard and Stammbach (2022) employ the same emotionality measure to study the use of metaphors, finding a positive relation between the two. Black legislators exhibit higher metaphor usage, while Native Americans show lower levels. Fiva et al. (2023) use the method described in Gentzkow

et al. (2019b) and find that politicians' sociodemographics significantly influence the issues they address in parliamentary speeches, with female legislators tending to discuss family and education policies more than their male counterparts, who focus more on international collaboration and public control.

5.5 Emotional language after the reform

Our main analysis focuses on the use of emotive rhetoric by MDB members, comparing the pre- and post-reform periods. Despite suggestive descriptive evidence indicating a significant rise in the use of emotion following the introduction of new parties, we quantify how much of this increase is attributable to being an MDB member. This analysis builds on a difference-in-differences setting, given the clearly defined intervention period marked by the December 1979 reform that allowed the entrance of new oppositional parties.

MDB members are the *treated* group, while ARENA members are the *untreated* group. The sample for these exercises includes only members from these two parties, as new parties emerged only in the post-reform period.[10] The identifying assumption is that, in the absence of the reform, the potential use of emotive rhetoric would have followed the same trends after 1979 for both ARENA and MDB. We estimate the shifts in emotionality after the reform with the following regression:

$$Y_{its} = \lambda_t + \sum_{\tau=-4}^{-1} \gamma_\tau D_{i\tau} + \sum_{\tau=1}^{3} \delta_\tau D_{i\tau} + x_{ist} + \theta_s + \varepsilon_{ist} \tag{6}$$

where Y_{its} is the outcome variable for individual i at year t and speech s. The term λ_t denotes year-fixed effects, capturing common shocks or trends at time t. $D_{i\tau}$ is an indicator variable equal to 1 if individual i is affiliated with MDB τ years after the reform, and 0 otherwise. γ_τ captures the effect of being affiliated with MDB at different times pre-reform and δ_τ measures the impact of MDB affiliation at different times post-reform. x_{ist} represents a vector of individual characteristics (gender, age group, profession, education status) and θ_s controls for speech length. Finally, ε_{ist} is the error term, capturing unobserved factors affecting the outcome variable. Standard errors are clustered at the politician level.

The main results, indicated in Table 5, suggest a significant increase in the use of emotion and negative words after the reform. The result is persistent for the use of negative emotional words, with positive and significant coefficients for the three years after the reform. On the other hand, although positive, the coefficient related to the emotion measure for 1981 is non-significant when we include the full set of controls. The

[10] This choice was made to ensure a clear identification of treated and control groups. Results are robust to including members of the new parties in the sample.

coefficients for 1980 and 1982 remain positive and significant. Overall, these results are consistent with the parallel trends assumption, as all coefficients preceding the reform are not significant.

We also run the pre- (1975-1979) and post- (1980-1982) reform periods grouped as in a standard difference-in-differences regression with two periods. Table 6 indicates a significant positive effect of being an MDB member after the reform, both in the emotion measure and in the ratio of negative to positive emotional words. This result is robust to the inclusion of our full set of controls.

Our results are related to those of Kosmidis et al. (2019)[11], indicating a negative relation between emotive rhetoric and ideological polarization. Magnitude-wise, being a member of MDB in the post-reform period is associated with a 0.022 increase in the emotion score, which averages 0.952. The empirical literature using this same measure reports similar magnitudes: Gennaro and Ash (2022) finds that being a member of the opposition in the US Congress is associated with a 0.020 increase in the emotion score. Our findings indicate a 29.8% increase in the ratio of negative to positive emotional words, suggesting that MDB members significantly used more negative rhetoric after the reform.

Figure 9 displays the γ_τ and δ_τ coefficients estimated in Equation (6) with year 0 marking the law's approval permitting multiple parties. This motivates the argument that our main results can be interpreted primarily as the consequence of increased party competition rather than solely loosening free speech restrictions. The abolishment of IA-5, which restored significant political freedom, occurred in October 1978, corresponding to year -1 in the plot. If increased freedom were the primary driver of the shift in emotional rhetoric, we would expect to observe a rise as early as 1979, as seen with both the ideal point estimates (Figure 3) and the ratio of negative to positive emotional words (Figure 8), which the evidence does not support.

The elections held in 1978 resulted in a new legislature by early 1979, thereby alleviating concerns about compositional effects, as the composition of the Lower Chamber remained consistent from 1979 onwards. In Table 7, we perform a robustness check to address this concern, including only those politicians who did not switch parties throughout the entire study period. The results remain largely unchanged, though there is a modest reduction in the coefficient size for 1980.

Figure 3 shows a significant increase in ideological divergence between ARENA and MDB in 1979. Given that this measure is related to speech disagreement, we might conclude that (i) the end of IA-5 likely played a significant role in increasing free speech in Congress, and (ii) the new legislature's composition may have been more polarized due

[11] However, magnitudes cannot be directly compared due to differences in research design and variable choices.

to the members' preferences and/or characteristics. Since the spike in emotional rhetoric occurred in 1980 and persisted until the new elections in 1982, primarily among opposition parties, we understand that the increased political freedom beginning in 1978 provided the necessary conditions for changes in emotional rhetoric, but it was the heightened competition that amplified it.

5.6 Additional checks

Additional robustness checks were conducted to further assess our findings. As indicated in Equation (4), our measure of emotion for each speech is the ratio of the affective score (i.e., the distance between the speech vector and the affective pole) to the cognition score. For instance, an overall increase in this measure could result from a decrease in the cognition score while the affective score remains constant. Table B.1 demonstrates that this is not the case. During the years in which we found significant results for the increase in emotion, the coefficients for the affective score are larger in magnitude. This indicates that the increase in emotionality is indeed driven by speeches among MDB members becoming more emotional at a faster rate than they are becoming more rational, as the coefficients for this outcome are also negative.[12]

Table B.2 shows the same specification as Equation (7) with the extended sample up to 1986. A significant positive increase in emotional rhetoric is observed in 1983 with the new legislature that entered after the 1982 elections. Interestingly, the ratio of negative to positive emotional words decreases from 1984 to 1986, coinciding with the end of the dictatorship. Given the commencement of another legislature in 1984, further analysis is required to distinguish between potential causes of this result, such as changes in the composition of the Lower Chamber.

[12]Because these scores are distances, a negative sign indicates an approximation to the affective and cognitive poles.

6 Conclusion

In this paper, we demonstrated how novel NLP methods can be applied to study party competition. By combining techniques that utilize Congressional speech to estimate ideological positions and emotive rhetoric during the Brazilian dictatorship with a difference-in-differences strategy, we found a significant increase in the use of emotional speech by MDB members, who faced increased competition following a reform that allowed more parties into the political system.

Importantly, this increase was driven by negative emotional language, indicating a heightened level of animosity from the opposition. Although it is not possible to completely disentangle the effects of increased political freedom, our analysis provides plausible reasons to suggest that the results were primarily driven by the increase in competition. Robustness checks indicate that the results persist even after restricting the sample to politicians who never switched parties, thereby mitigating concerns related to compositional effects.

Our results align with both theoretical and applied party competition literature, where nonpolicy features are modeled as valence. In this context, when parties are closely aligned ideologically, valence is used to differentiate them and appeal to voters. This article contributes to the empirical literature by examining the impact of a competition shock on valence, offering arguably causal evidence that emotive rhetoric serves as a key differentiation instrument for politicians in legislative debates.

The findings highlight innovative research possibilities through the use of unstructured data and novel text analysis methods. Extensions of this work could examine a longer time period and utilize roll-call voting[13] as an alternative measure of ideology. Additionally, the extended emotion time series showed a sharp increase in 2008, coinciding with the introduction of broader TV coverage across Brazilian municipalities, suggesting a potential impact of media exposure on emotional rhetoric.

Benefiting from advances in Artificial Intelligence and computational methods, we were able to compile a database comprising the majority of speeches from historical archives available online.[14] As methods such as Optical Character Recognition develop, it should be possible to scrape Congressional speeches dating back to 1946, opening many opportunities for further research using Natural Language Processing. As a cost-effective, reproducible, and reliable method, this can also be valuable to developing countries that lack abundant resources for data collection.

[13]For Brazil, this is available starting in 2001.

[14]We plan to make the scraping process available to the general public as soon as possible.

References

Abney, R., Adams, J., Clark, M., Easton, M., Ezrow, L., Kosmidis, S., and Neundorf, A. (2011). When does valence matter? heightened valence effects for governing parties during election campaigns. *Party Politics*, 19(1):61–82.

Ash, E. and Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, 15(Volume 15, 2023):659–688.

Ashworth, S. and Bueno de Mesquita, E. (2009). Elections with platform and valence competition. *Games and Economic Behavior*, 67(1):191–216.

Batistella, A. (2021). A ditadura militar e o bipartidarismo: casuísmos e um simulacro de democracia. *Revista de Pesquisa Histórica - CLIO*, 39(2).

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Buisseret, P. and Weelden, R. V. (2022). Polarization, valence, and policy competition. *AER: Insights*, 4(3):341–352.

Clark, M. (2009). Valence and electoral outcomes in western europe, 1976–1998. *Electoral Studies*, 28(1):111–122.

Clark, M. (2013). Understanding parties' policy shifts in western europe: The role of valence, 1976–2003. *British Journal of Political Science*, 44(2):261–286.

Clark, M. and Leiter, D. (2013). Does the ideological dispersion of parties mediate the electoral impact of valence? a cross-national study of party support in nine western european democracies. *Comparative Political Studies*, 47(2):171–202.

Denter, P. (2021). Valence, complementarities, and political polarization. *Games and Economic Behavior*, 128:39–57.

Desposato, S. W. (2001). Legislative politics in authoritarian brazil. *Legislative Studies Quarterly*, 26(2):287–317.

Dolezal, M., Ennser-Jedenastik, L., Müller, W. C., and Winkler, A. K. (2013). How parties compete for votes: A test of saliency theory. *European Journal of Political Research*, 53(1):57–76.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy*, 65(2):135–150.

Ezrow, L. (2007). The variance matters: How party systems represent the preferences of voters. *The Journal of Politics*, 69(1):182–192.

Figueredo, F. C. d., Mueller, B., and Cajueiro, D. O. (2022). A natural language measure of ideology in the brazilian senate. *Revista Brasileira de Ciência Política*, (37).

Fiva, J. H., Nedregård, O., and Øien, H. (2023). Group identities and parliamentary debates. Working paper.

Gandhi, J., Noble, B., and Svolik, M. (2020). Legislatures and legislative politics without democracy. *Comparative Political Studies*, 53(9):1359–1379.

Gennaro, G. and Ash, E. (2022). Emotion and Reason in Political Language. *The Economic Journal*, 132(643):1037–1059.

Gentzkow, M., Kelly, B., and Taddy, M. (2019a). Text as data. *Journal of Economic Literature*, 57(3):535–74.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2019b). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, 87(4):1307–1340.

Green, J. and Hobolt, S. B. (2008). Owning the issue agenda: Party strategies and vote choices in british elections. *Electoral Studies*, 27(3):460–476.

Grimmer, J., Roberts, M. E., and Stewart, B. M. (2022). *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press, Princeton, NJ.

Kosmidis, S., Hobolt, S. B., Molloy, E., and Whitefield, S. (2019). Party competition and emotive rhetoric. *Comparative Political Studies*, 52(6):811–837.

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

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

Mettenheim, K. V. (1990). The brazilian voter in democratic transition, 1974-1982. *Comparative Politics*, 23(1):23–44.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Jordan, M. I., LeCun, Y., and Solla, S. A., editors, *Advances in neural information processing systems*, pages 3111–3119.

Napolitano, M. (2018). The brazilian military regime, 1964–1985. *Oxford Research Encyclopedias, Latin American History*.

Osnabrügge, M., Hobolt, S. B., and Rodon, T. (2021). Playing to the gallery: Emotive rhetoric in parliaments. *American Political Science Review*, 115(3):885–899.

Pereira, A. W. (2020). *Modern Brazil: A Very Short Introduction*. Oxford University Press.

Pereira, L. A. R. (2021). *Electoral Competition and Platform Choice: A Computational Linguistics Approach Based on Brazilian Data*. PhD thesis, INSPER - Instituto de Ensino e Pesquisa, São Paulo, SP.

Persson, T. and Tabellini, G. (2000). *Political Economics: Explaining Economic Policy*. MIT Press, Cambridge, MA.

Picard, L. and Stammbach, D. (2022). Political metaphors in u.s. governor speeches. Working Paper 13/2022, Center for Law & Economics.

Poole, K. T. and Rosenthal, H. (1997). *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press, New York.

Power, T. J. and Zucco, C. (2009). Estimating ideology of brazilian legislative parties, 1990-2005: A research communication. *Latin American Research Review*, 44(1):218–246.

Ridenti, M. (2018). Censura e ditadura no brasil, do golpe à transição democrática, 1964-1988. *Opinião Pública*, 14(2):432–453.

Schofield, N. (2003). Valence competition in the spatial stochastic model. *Journal of Theoretical Politics*, 15(4):371–383.

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

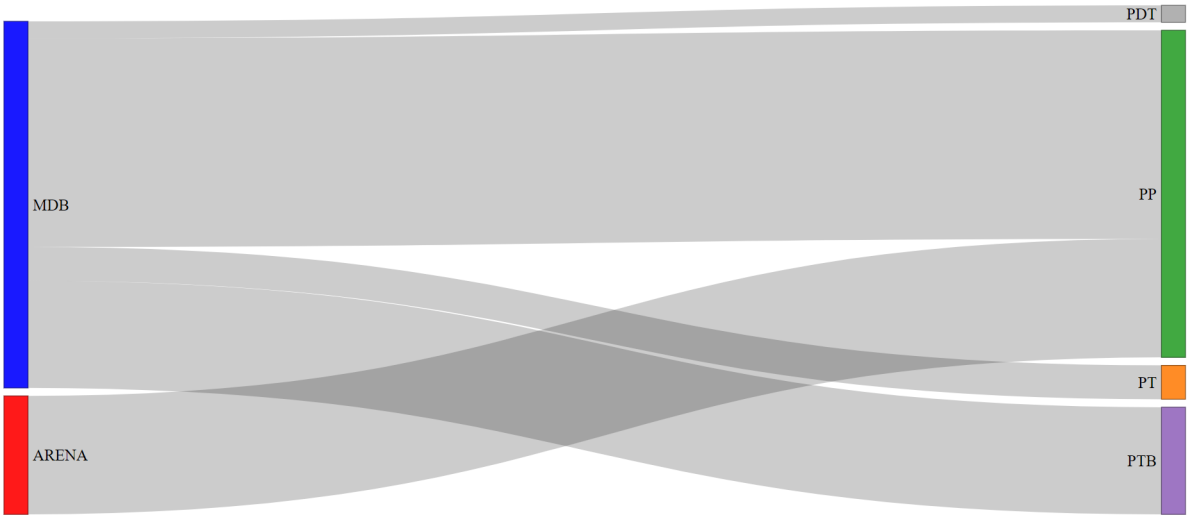
Slapin, J. B. and Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3):705–722.

Stokes, D. E. (1963). Spatial models of party competition. *American Political Science Review*, 57(2):368–377.

Wittman, D. (1977). Candidates with policy preferences: A dynamic model. *Journal of Economic Theory*, 14(1):180–189.

Figures

Figure 1: Outflow of deputies in 1980



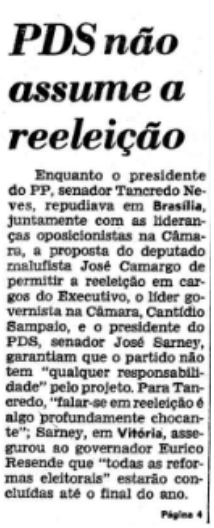
Notes: The figure displays the outflow of deputies from MDB and ARENA in 1980 when more parties were allowed into the political system. The first node (MDB to PDT) represents a size of 2; the second (MDB to PP), a size of 38; the third (MDB to PT), a size of 6; the fourth (MDB to PTB), a size of 17; and the sixth (ARENA to PP), a size of 20.

Figure 2: Media outlets

(a)

(b)

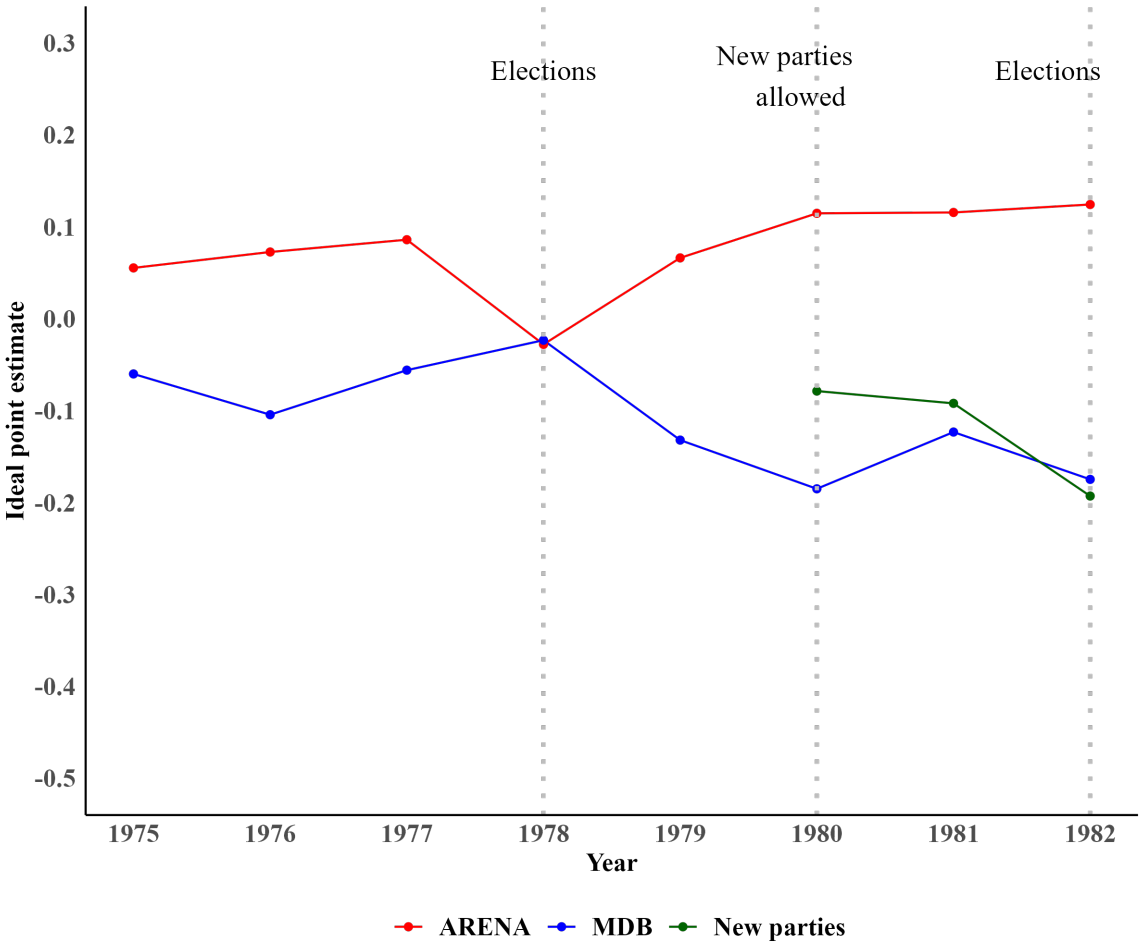
(c)



(d)

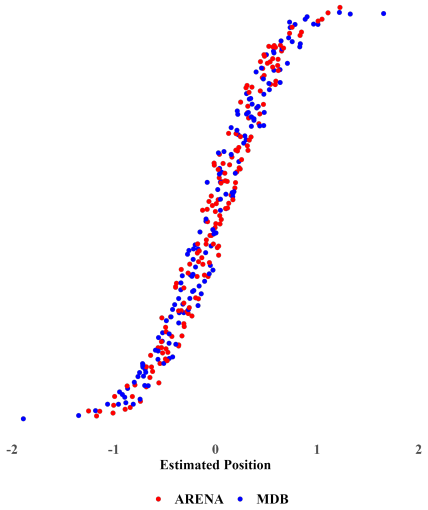
Notes: The images show covers of the journals *Folha de São Paulo* and *O Estado de São Paulo*, two of the biggest journals during that time, from 1980 and 1981. News related to activities in the *Câmara dos Deputados* is highlighted in red.

Figure 3: Time series of ideal point estimates

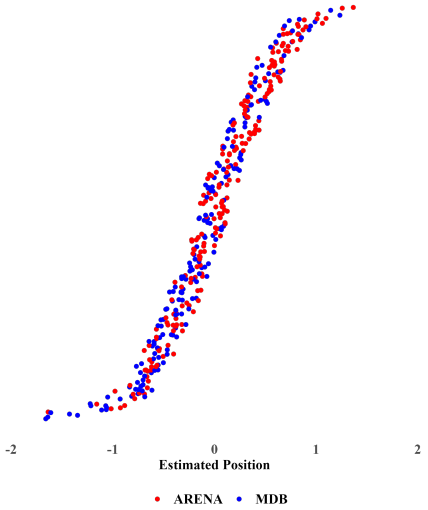


Notes: The figure displays the average θ (ideal point estimate) obtained from the *Wordshoal* model for each party from 1975 to 1982. We run the model annually at the politician-speech level, then calculate the average for each party and plot these averages. We considered congressional sessions that had at least two different speakers and five speeches. *New parties* include PP, PT, PTB, and PDT.

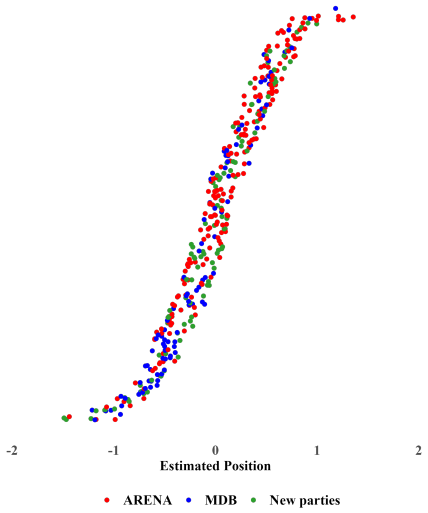
Figure 4: Individual ideal point estimates (θ)



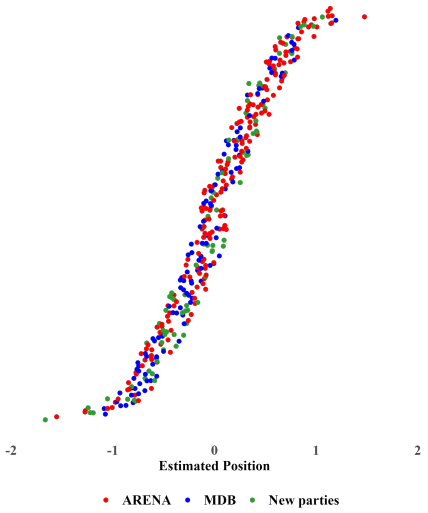
(a) 1978



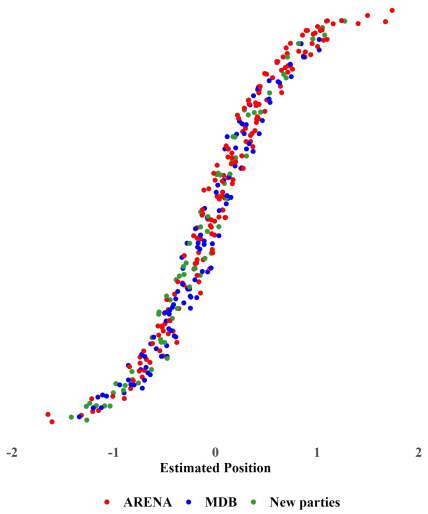
(b) 1979



(c) 1980



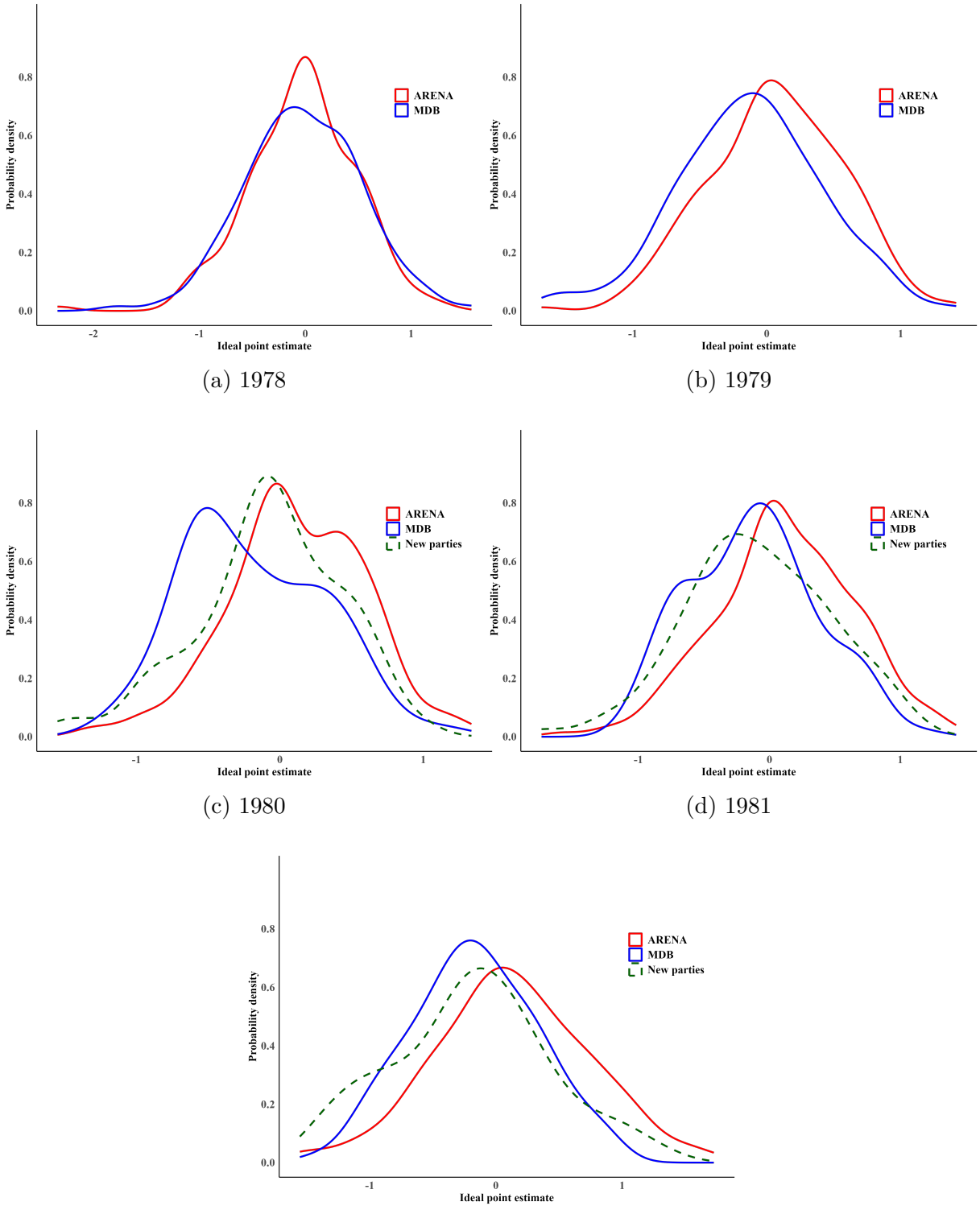
(d) 1981



(e) 1982

Notes: The figures display the mean *ideal point estimate* (θ) for each politician from 1978 to 1982. Each point represents the estimated position of a deputy in a given year. *New parties* include PP, PT, PTB, and PDT.

Figure 5: PDF of ideal point estimates



(a) 1978

(b) 1979

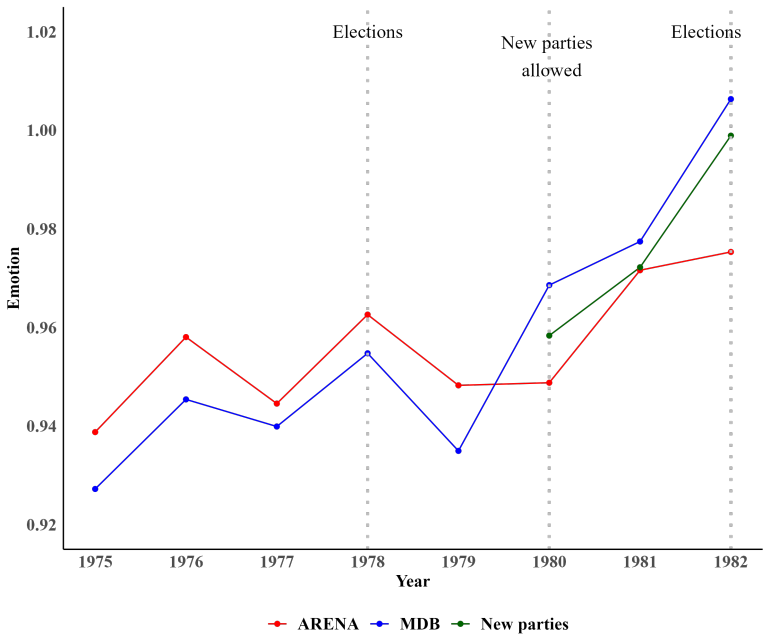
(c) 1980

(d) 1981

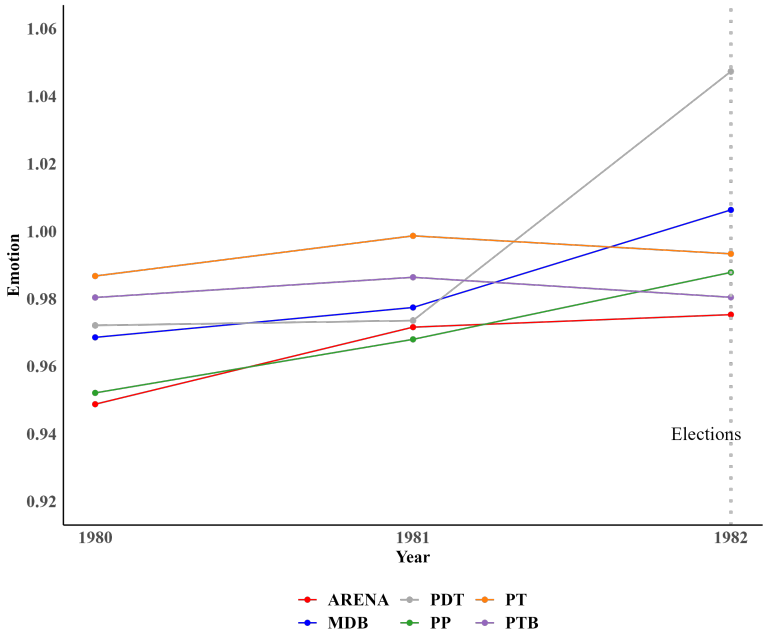
(e) 1982

Notes: The figures depict the probability density function of the *ideal point estimates* (θ), representing the weighted averages for each political party from 1978 to 1982. These estimates were derived at the annual-politician level using the *Wordshoal* model and then aggregated by taking a weighted average of all members within each party. We considered congressional sessions that had at least two different speakers and five speeches. *New parties* include PP, PT, PTB, and PDT.

Figure 6: Time series of emotion score



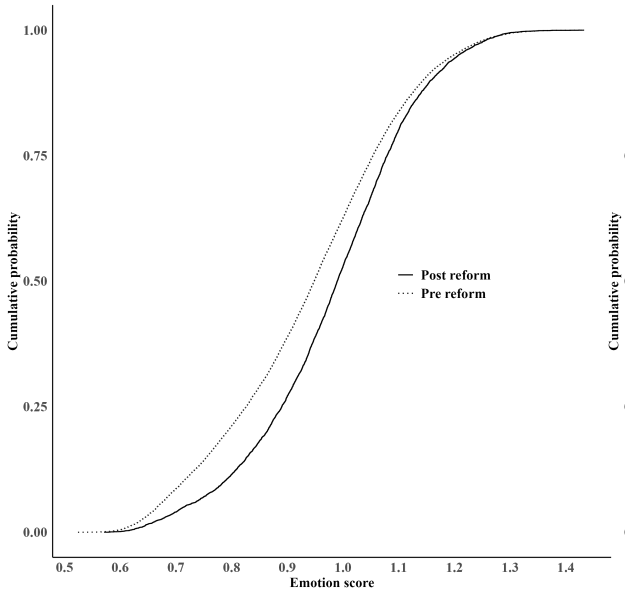
(a) 1975-1982



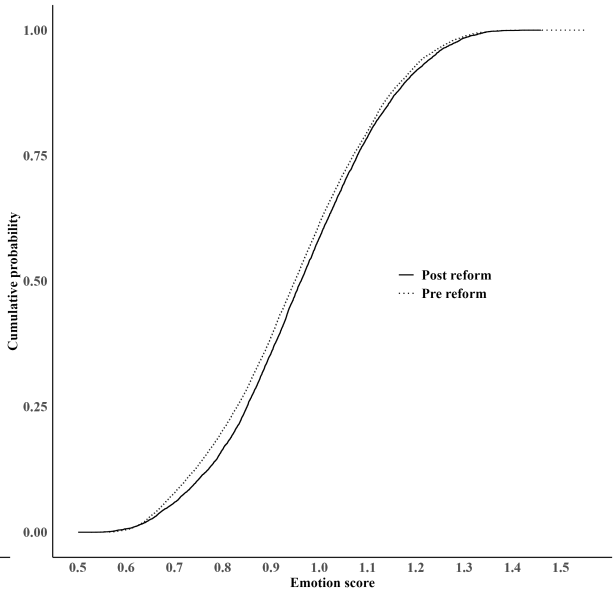
(b) 1980-1982

Notes: The figure displays the mean *emotion score* for each party from 1975 to 1982, and from 1980 to 1982 when more parties were allowed in the political system. *New parties* include PP, PT, PTB, and PDT. Emotion scores were obtained at the politician-speech level using the methodology described in Gennaro and Ash (2022) and then aggregated by taking a weighted average of all speeches within each party.

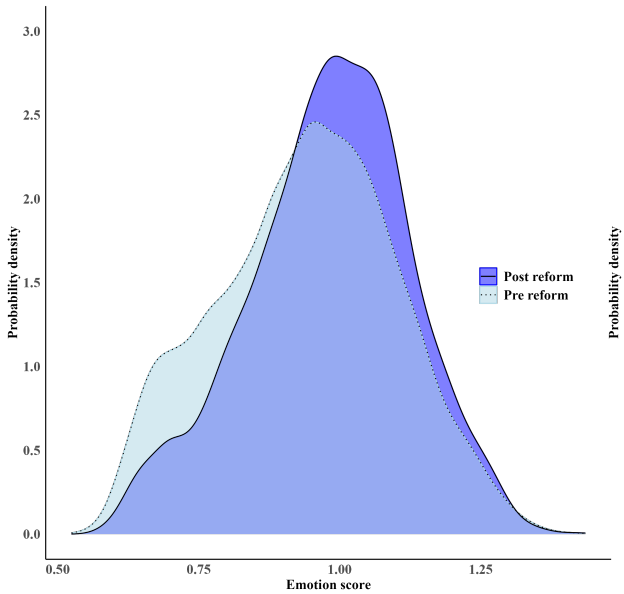
Figure 7: Shifts in the distribution of emotional speech after the reform



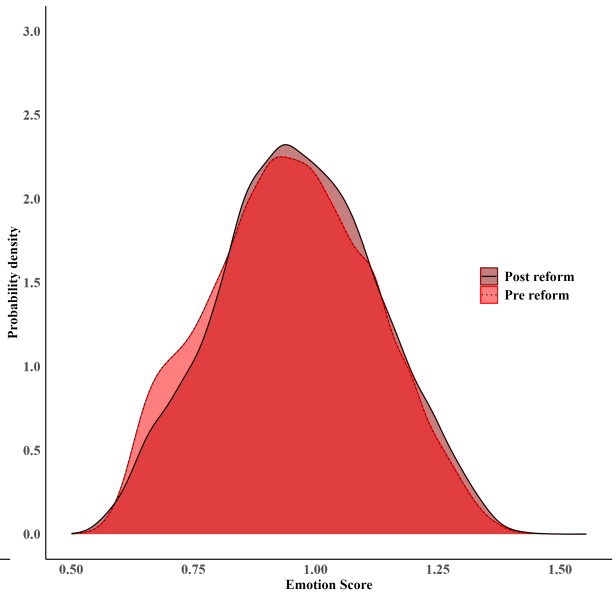
(a) Cumulative distribution: MDB



(b) Cumulative distribution: ARENA



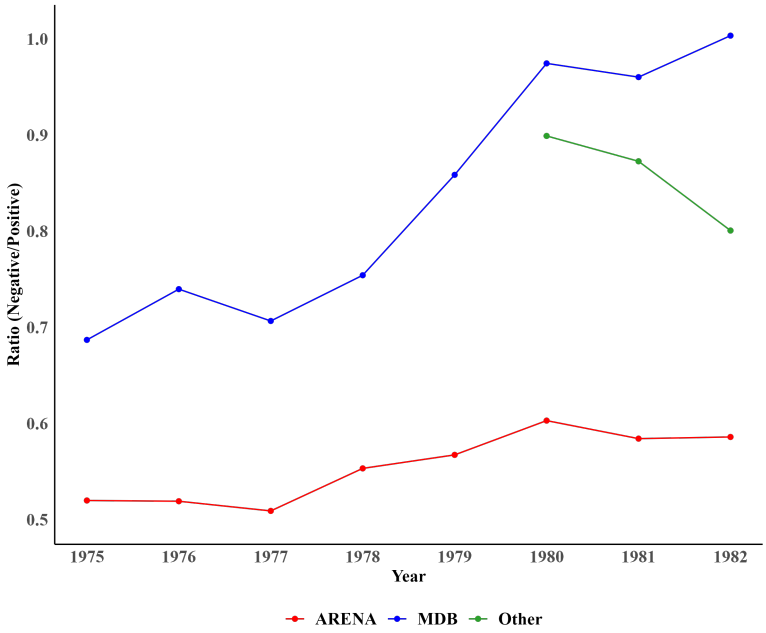
(c) Probability density: MDB



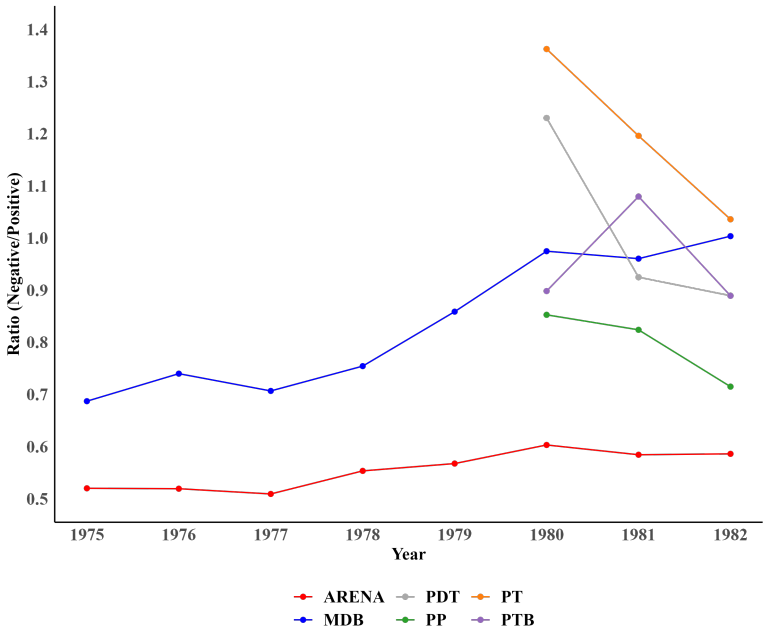
(d) Probability density: ARENA

Notes: The figures display the cumulative distribution function (CDF) and the probability density function (PDF) of the emotion score for (a) MDB members and (b) ARENA members, covering the period from 1975 to 1982. The *Pre reform* period encompasses 1975 to 1979, while the *Post reform* period extends from 1980 to 1982. The data is at the politician-speech level.

Figure 8: Ratio of negative to positive emotional words



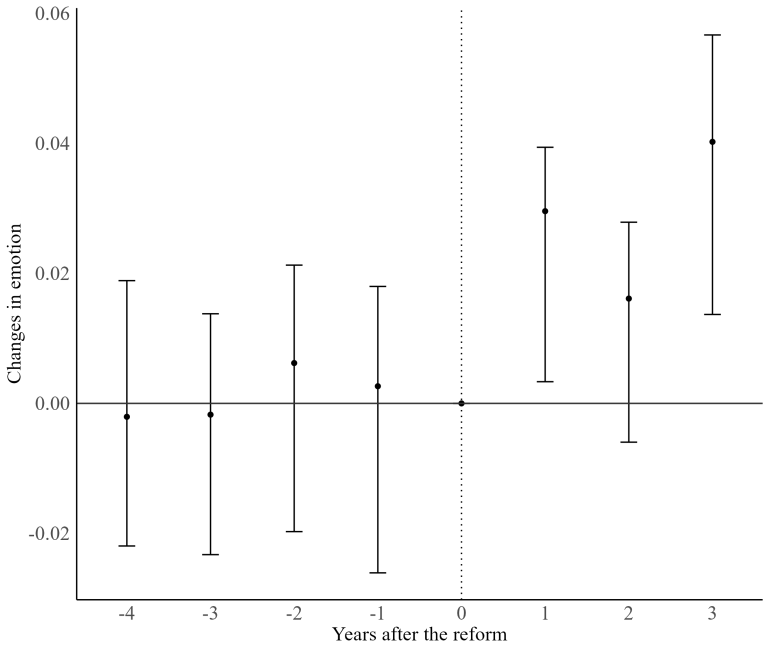
(a) 1975-1982



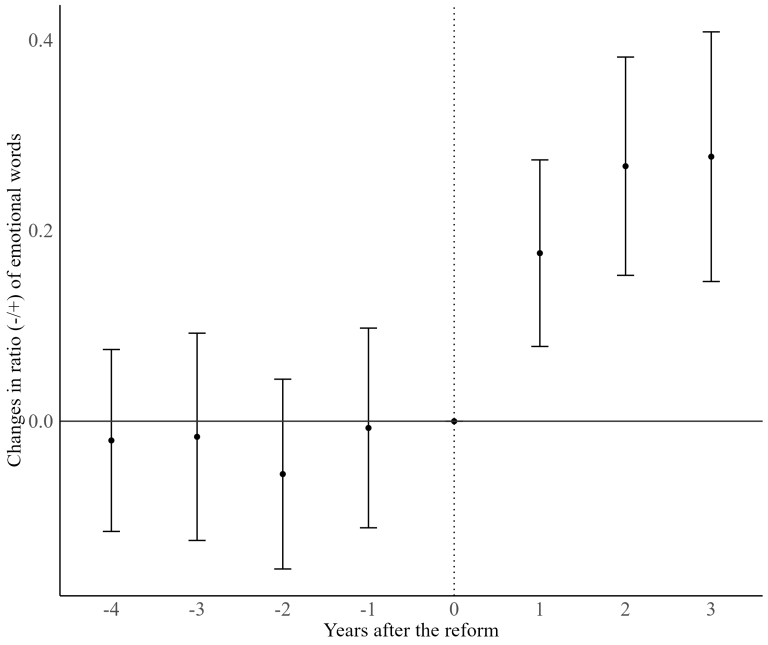
(b) 1975-1982

Notes: The figure displays the ratio of *negative* to *positive* words in the emotion dictionary used as a seed for the *emotion score*, for each party from 1975 to 1982, and from 1980 to 1982, when more parties were allowed in the political system. New parties include PP, PT, PTB, and PDT. The data is at the politician-speech level.

Figure 9: Trends in emotion during 1975-1982 - MDB



(a) Emotion score



(b) Ratio of negative to positive words

Notes: The figure displays the γ_τ and δ_τ coefficients for the regression $Y_{its} = \lambda_t + \sum_{\tau=-4}^{-1} \gamma_\tau D_{i\tau} + \sum_{\tau=1}^{3} \delta_\tau D_{i\tau} + x_{ist} + \theta_s + \varepsilon_{ist}$, with 1979 as the reference year ($\tau = 0$). The regressions are the same as in Columns (3) and (6) from Table 5 and include the full set of controls.

Tables

Table 1: Descriptive statistics of deputies and speeches

Year	Total elected	Total in sample	ARENA	MDB	Other
Panel A: Number of Deputies					
1975	364	354	197	157	-
1976	364	333	179	154	-
1977	364	334	185	149	-
1978	364	317	167	150	-
1979	420	398	214	184	-
1980	420	396	194	99	103
1981	420	402	200	115	87
1982	420	348	172	107	69
Panel B: Number of Speeches					
1975		6643	2800	3843	-
1976		6750	3017	3733	-
1977		9075	3939	5136	-
1978		5756	2215	3541	-
1979		7700	3359	4341	-
1980		7253	3093	2260	1900
1981		7863	3349	2862	1652
1982		4123	1900	1478	745

Notes: This table displays the number of deputies and their annual speeches. *Total Elected* indicates the number of deputies elected for the current legislature. *Total in sample* is smaller than *Total Elected* because not all elected deputies chose to give speeches in a given year. *Other* includes PP, PDT, PT, and PTB.

Table 2: Switches between parties during 1980-1982

		To			
		ARENA	MDB	PP	PT/PTB/PDT
From	ARENA	-	9	26	0
	MDB	23	-	41	27
	PP	3	3	-	0
	PT/PTB/PDT	0	0	0	-

Notes: This table shows the transitions of party members between 1980 and 1982, categorized by their departure from the source party (**From**) and their arrival at the destination party (**To**). For example, the second row of the first column indicates that 22 deputies left MDB and went to ARENA from 1980 to 1982.

Table 3: Descriptive statistics of main variables

		MDB		ARENA		Other parties
		Pre-reform	Post-reform	Pre-reform	Post-reform	Post-reform
Panel A: Individual characteristics						
Gender	Shares					
Female		0.011	0.015	0.003	0.009	0.017
Male		0.989	0.985	0.997	0.991	0.983
Age						
23-35		0.139	0.041	0.064	0.043	0.067
36-45		0.286	0.411	0.226	0.229	0.296
46-55		0.297	0.371	0.429	0.413	0.361
55+		0.278	0.177	0.281	0.315	0.276
Profession						
Agriculture and Natural Resources		0.044	0.066	0.087	0.085	0.009
Arts and Media		0.052	0.058	0.042	0.034	0.051
Business and Administration		0.093	0.080	0.127	0.124	0.077
Education		0.089	0.095	0.102	0.085	0.120
Engineering and Technical		0.074	0.066	0.096	0.090	0.094
Healthcare		0.059	0.036	0.075	0.085	0.085
Legal and Public Service		0.504	0.526	0.401	0.397	0.530
Manual and Skilled Trades		0.056	0.044	0.060	0.077	0.026
Religious and Community Services		0.004		0.003		
Unknown		0.026	0.029	0.006	0.021	0.009
Education						
High school degree (complete or incomplete)		0.022	0.022	0.057	0.064	0.009
College degree (complete or incomplete)		0.874	0.869	0.810	0.791	0.949
Postgraduate degree		0.004	0.007	0.012	0.009	0.009
Unknown		0.100	0.102	0.120	0.137	0.034
Panel B: Speeches						
Numbers of speeches per politician	Mean	76	48	46	36	37
	SD	106	46	98	47	44
	p10	7.9	3.6	3.0	3.0	2.0
	p50	38	35	18	21	23
	p90	180	103	102	77	91
	Observations	270	137	332	234	117
Speech length (words)	Mean	380	453	375	417	405
	SD	323	324	314	303	313
	p10	20	78	23	81	33
	p50	321	385	315	352	343
	p90	818	894	785	834	813
	Observations	20,594	6,600	15,330	8,342	4,297
Panel C: Outcome variables						
Ideal point estimate (θ)	Mean	-0.026	-0.031	0.010	0.003	0.006
	SD	0.547	0.532	0.540	0.558	0.525
	p10	-0.733	-0.719	-0.697	-0.710	-0.704
	p50	-0.020	-0.019	0.000	0.000	0.000
	p90	0.677	0.681	0.717	0.726	0.671
	Observations	1,698	1,042	2,029	1,847	838
Emotion score	Mean	0.940	0.981	0.950	0.964	0.971
	SD	0.160	0.143	0.166	0.163	0.153
	p10	0.712	0.786	0.721	0.745	0.752
	p50	0.948	0.990	0.949	0.962	0.981
	p90	1.144	1.158	1.172	1.181	1.160
	Observations	20,594	6,600	15,330	8,342	4,297
Ratio of negative to positive emotional words	Mean	0.964	1.35	0.671	0.779	1.175
	SD	1.157	1.522	0.906	0.957	1.371
	p10	0	0.2	0	0	0.124
	p50	0.667	1	0.429	0.5	0.8
	p90	2	3	1.5	1.8	2.5
	Observations	16,420	5,744	12,737	7,392	3,659

Notes: The table presents descriptive statistics of the main variables: share, mean, standard deviation (SD), 10th percentile (p10), median (p50), and 90th percentile (p90). For the emotion score, the data is at the politician-date-speech level. The data is at the politician-year level for the ideal point estimate (θ). The period for *Pre-reform* is 1975-1979, and the period for *Post-reform* is 1980-1982. *Other parties* include PP, PT, PTB, and PDT. *Healthcare:* Physicians, Dentists, Pharmacists, Veterinary Physicians, and Radiotelegraphists. *Arts and Media:* Visual Artists, Musicians, Writers, Actresses, Journalists, Advertising Professionals, Communicators, Poets, and Sociologists. *Engineering and Technical:* Engineers, Chemists, Architects, Technical Professionals, Mechanics, Electricians, Surveyors, Toolmakers, and Accountants. *Business and Administration:* Economists, Businessmen, Administrators, Bankers, Bank Employees, Consultants, and Tax Agents. *Legal and Public Service:* Lawyers, Public Prosecutors, Justice Prosecutors, Justice Clerks, Notaries, Court Clerks, Public Servants, Inspectors, Police Officers, and Military Personnel. *Education:* Teachers, Pedagogues, and Educational Advisors. *Agriculture and Natural Resources:* Agronomists, Farmers, Agriculturalists, Cultivators, Livestock Breeders, and Agricultural Extension Workers. *Manual and Skilled Trades:* Factory Workers, Civil Constructors, Dispatchers, Mechanics, Electricians, Professional Drivers, Stevedores, Retail Workers, and Brokers. *Religious and Community Services:* Pastors, Priests, and Community Leaders.

Table 4: Variations in emotionality across demographics

	(1)	(2)	(3)	(4)
	Emotion		Negative to positive words	
Female	0.038**	0.008**	0.110*	0.111**
	(0.017)	(0.004)	(0.057)	(0.050)
36-45	0.018*	0.008*	0.036	0.042
	(0.010)	(0.005)	(0.038)	(0.033)
46-55	0.015	0.009*	-0.029	0.009
	(0.011)	(0.005)	(0.038)	(0.032)
55+	0.023*	0.012*	-0.090**	-0.021
	(0.013)	(0.006)	(0.045)	(0.036)
MDB	0.001	0.009**	0.360***	0.247***
	(0.007)	(0.004)	(0.026)	(0.020)
Other parties	0.004	0.010*	0.298***	0.205***
	(0.008)	(0.005)	(0.053)	(0.041)
College or Postgraduated	-0.033***	-0.021***	0.022	0.039
	(0.010)	(0.006)	(0.042)	(0.033)
Unknown education	-0.026**	-0.022***	-0.045	-0.010
	(0.011)	(0.007)	(0.048)	(0.038)
Agric. and Natural Resources	0.011	0.015**	0.003	0.039
	(0.011)	(0.007)	(0.043)	(0.034)
Arts and Media	0.028**	0.010	-0.059	-0.035
	(0.013)	(0.007)	(0.073)	(0.055)
Business and Management	0.006	0.002	-0.097**	-0.054*
	(0.008)	(0.005)	(0.039)	(0.033)
Education	-0.001	-0.002	-0.041	-0.020
	(0.010)	(0.006)	(0.048)	(0.035)
Engineering and Technical	-0.006	-0.0002	-0.021	-0.008
	(0.014)	(0.006)	(0.041)	(0.035)
Healthcare	-0.002	0.005	-0.115**	-0.069*
	(0.012)	(0.007)	(0.048)	(0.039)
Manual and Skilled Trades	0.018*	0.008	-0.002	0.034
	(0.010)	(0.006)	(0.056)	(0.034)
Religious and Community Services	0.043***	-0.004	-0.285***	-0.146***
	(0.007)	(0.004)	(0.028)	(0.022)
Observations	54,672	54,473	45,563	45,423
R^2	0.019	0.510	0.039	0.105
Dependent variable mean	0.954	0.954	0.917	0.917
Year FE	✓	✓	✓	✓
Topic FE		✓		✓

Notes: Each column shows the OLS regression of the emotionality score and the ratio of negative to positive words in a given speech on individual politician characteristics and year. Regressions take the form $Y_{ist} = \gamma_1 * Gender + \gamma_2 * Age + \gamma_3 * Party + \gamma_4 * Education + \gamma_5 * Profession + \lambda_t + \varepsilon_{ist}$. The sample consists of all speeches pronounced by the members of Congress between 1975 and 1982. All columns include year-fixed effects. Columns (2) and (4) also include topic-fixed effects. Standard errors are clustered at the politician level. $*$, $**$, $***$ denote significance at the 10%, 5% and 1% levels, respectively. The reference values for gender, age group, party, educational level, and profession are, respectively, Male, 23-35, ARENA, High school (completed or not), and Public service. The sample is at the politician-speech level for columns (1)-(4) and politician-year for columns (5)-(6).

Table 5: Main results - trends in emotion

	(1)	(2)	(3)	(4)	(5)	(6)
	Emotion Score			Negative to positive words		
MDB × 1975 (γ_{-4})	0.002	-0.002	-0.002	-0.057	-0.038	-0.020
	(0.012)	(0.011)	(0.010)	(0.050)	(0.047)	(0.049)
MDB × 1976 (γ_{-3})	0.0007	-0.002	-0.005	-0.025	-0.021	-0.016
	(0.011)	(0.011)	(0.009)	(0.058)	(0.055)	(0.055)
MDB × 1977 (γ_{-2})	0.009	0.006	0.0008	-0.070	-0.061	-0.056
	(0.011)	(0.011)	(0.010)	(0.050)	(0.050)	(0.051)
MDB × 1978 (γ_{-1})	0.005	0.003	-0.004	-0.021	-0.002	-0.007
	(0.013)	(0.012)	(0.011)	(0.055)	(0.055)	(0.053)
MDB × 1980 (δ_1)	0.033***	0.030***	0.021**	0.186***	0.175***	0.176***
	(0.010)	(0.010)	(0.009)	(0.051)	(0.051)	(0.050)
MDB × 1981 (δ_2)	0.019**	0.016*	0.011	0.264***	0.255***	0.268***
	(0.009)	(0.009)	(0.009)	(0.058)	(0.058)	(0.058)
MDB × 1982 (δ_3)	0.044***	0.040***	0.035***	0.281***	0.274***	0.278***
	(0.011)	(0.011)	(0.011)	(0.066)	(0.067)	(0.067)
Observations	50,866	50,408	50,296	42,293	41,936	41,825
R^2	0.011	0.019	0.136	0.039	0.041	0.086
Dep. var. mean	0.952	0.952	0.952	0.896	0.895	0.895
Controls						
Year	✓	✓	✓	✓	✓	✓
Gender		✓	✓		✓	✓
Age		✓	✓		✓	✓
Profession		✓	✓		✓	✓
Education		✓	✓		✓	✓
Speech length			✓			✓

Notes: The data are at the politician-speech level. The sample consists of speeches by politicians who were members of MDB and ARENA during the 1975-1982 period. Regressions take the form $Y_{its} = \lambda_t + \sum_{\tau=-4}^{-1} \gamma_\tau D_{i\tau} + \sum_{\tau=1}^{3} \delta_\tau D_{i\tau} + x_{ist} + \theta_s + \varepsilon_{ist}$ where Y_{its} is the outcome variable for individual i at year t and speech s. The term λ_t denotes year-fixed effects, capturing common shocks or trends at time t. $D_{i\tau}$ is an indicator variable equal to 1 if individual i is affiliated with MDB τ years after the reform, and 0 otherwise. γ_τ captures the effect of being affiliated with MDB at different times pre-reform and δ_τ measures the impact of MDB affiliation at different times post-reform. x_{ist} represents a vector of individual characteristics (gender, age group, profession, education status) and θ_s controls for speech length. ε_{ist} is the error term, capturing unobserved factors affecting the outcome variable. Standard errors are clustered at the politician level. $*$, $**$, $***$ denote significance at the 10%, 5% and 1% levels, respectively.

Table 6: Main results - emotion after the reform

	(1)	(2)	(3)	(4)	(5)	(6)
	Emotion Score			Negative to positive words		
MDB	-0.010	-0.006	-0.003	0.293***	0.282***	0.279***
	(0.008)	(0.008)	(0.007)	(0.030)	(0.028)	(0.028)
Post-reform	0.014**	0.015**	0.007	0.108***	0.112***	0.085***
	(0.007)	(0.006)	(0.006)	(0.023)	(0.023)	(0.024)
MDB × Post-reform	0.026***	0.025***	0.022**	0.278***	0.257***	0.259***
	(0.010)	(0.009)	(0.009)	(0.044)	(0.043)	(0.043)
Observations	50,866	50,408	50,296	42,293	41,936	41,825
R^2	0.007	0.016	0.134	0.037	0.039	0.084
Dep. var. mean	0.952	0.952	0.952	0.896	0.895	0.895
Controls						
Age		✓	✓		✓	✓
Gender		✓	✓		✓	✓
Profession		✓	✓		✓	✓
Education		✓	✓		✓	✓
Speech length			✓			✓

Notes: The data are at the politician-speech level. The sample consists of speeches by politicians who were members of MDB and ARENA during the 1975-1982 period. Regressions take the form $Y_{ist} = \delta(\text{MDB}_i \times \text{Post}_t) + \beta\text{MDB}_i + \alpha\text{Post}_t + \gamma'\mathbf{X}_{ist} + \phi\text{Length}_s + \varepsilon_{ist}$, where Y_{ist} is the outcome variable for speech s by politician i in period t, MDB_i is a dummy variable indicating membership in MDB, Post_t is a dummy variable indicating the post-treatment period (1980 onwards), $\mathbf{X}_{ist}$ is a vector of control variables including gender, age group, profession, and education status, and Length_s is the length of the speech in words. Standard errors are clustered at the politician level. $*$, $**$, $***$ denote significance at the 10%, 5% and 1% levels, respectively.

Table 7: Robustness checks - No switchers only

	(1)	(2)	(3)	(4)	(5)	(6)
	Emotion Score			Negative to positive words		
MDB × 1975 (γ_{-4})	0.002	-0.0010	-0.003	-0.094*	-0.068	-0.050
	(0.012)	(0.012)	(0.011)	(0.052)	(0.049)	(0.050)
MDB × 1976 (γ_{-3})	0.0001	-0.002	-0.006	-0.048	-0.039	-0.030
	(0.012)	(0.011)	(0.010)	(0.061)	(0.058)	(0.058)
MDB × 1977 (γ_{-2})	0.012	0.010	0.004	-0.101*	-0.086	-0.077
	(0.012)	(0.011)	(0.011)	(0.053)	(0.053)	(0.053)
MDB × 1978 (γ_{-1})	0.007	0.004	-0.003	-0.048	-0.025	-0.028
	(0.014)	(0.013)	(0.012)	(0.058)	(0.057)	(0.056)
MDB × 1980 (δ_1)	0.027***	0.025**	0.018*	0.149***	0.148***	0.153***
	(0.010)	(0.010)	(0.009)	(0.053)	(0.052)	(0.052)
MDB × 1981 (δ_2)	0.016*	0.015	0.010	0.260***	0.255***	0.270***
	(0.009)	(0.009)	(0.009)	(0.059)	(0.059)	(0.059)
MDB × 1982 (δ_3)	0.045***	0.041***	0.036***	0.278***	0.273***	0.281***
	(0.011)	(0.011)	(0.011)	(0.070)	(0.070)	(0.070)
Observations	47,298	46,946	46,832	39,428	39,133	39,020
R^2	0.011	0.020	0.133	0.042	0.044	0.091
Dep. var. mean	0.953	0.953	0.953	0.898	0.897	0.897
Controls						
Year	✓	✓	✓	✓	✓	✓
Gender		✓	✓		✓	✓
Age		✓	✓		✓	✓
Profession		✓	✓		✓	✓
Education		✓	✓		✓	✓
Speech length			✓			✓

Notes: The data are at the politician-speech level. The sample consists of speeches from politicians who were consistent members of either MDB or ARENA throughout the 1975-1982 period, with no switches between these parties during that time. Regressions take the form $Y_{its} = \lambda_t + \sum_{\tau=-4}^{-1} \gamma_\tau D_{i\tau} + \sum_{\tau=1}^{3} \delta_\tau D_{i\tau} + x_{ist} + \theta_s + \varepsilon_{ist}$ where Y_{its} is the outcome variable for individual i at year t and speech s. The term λ_t denotes year-fixed effects, capturing common shocks or trends at time t. $D_{i\tau}$ is an indicator variable equal to 1 if individual i is affiliated with MDB τ years after the reform, and 0 otherwise. γ_τ captures the effect of being affiliated with MDB at different times pre-reform and δ_τ measures the impact of MDB affiliation at different times post-reform. x_{ist} represents a vector of individual characteristics (gender, age group, profession, education status) and θ_s controls for speech length. ε_{ist} is the error term, capturing unobserved factors affecting the outcome variable. Standard errors are clustered at the politician level. $*$, $**$, $***$ denote significance at the 10%, 5% and 1% levels, respectively.

A Data, variables and institutional setting

A.1 Data sources and variables

Table A.1: Scraping results

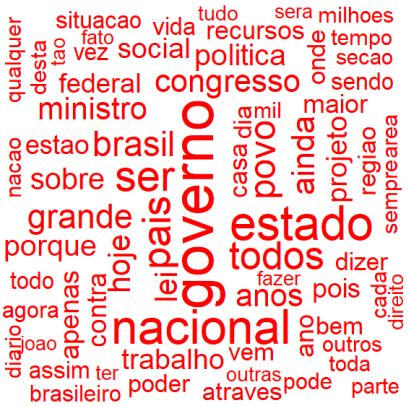
Year	Total	Regular (%)	Fuzzy (%)	PDF Problem (%)	Not Found (%)
1975	9,542	79.03	18.54	0.34	2.10
1976	8,856	84.18	13.02	0.78	2.02
1977	10,994	90.69	8.00	0.39	0.91
1978	7,211	89.04	8.36	0.76	1.83
1979	10,601	82.87	12.60	1.77	2.75
1980	9,844	85.89	11.48	0.54	2.09
1981	10,413	88.30	9.66	0.73	1.31
1982	5,121	92.05	6.19	0.43	1.33
1983	8,702	91.52	5.95	0.94	1.59
1984	9,407	92.95	5.05	0.96	1.04
1985	8,748	89.71	7.74	1.05	1.50
1986	5,225	91.06	6.60	1.03	1.30

Notes: This table presents the results of the speech scraping process. The *Total* column represents the number of speeches sourced from the *Câmara dos Deputados* website. The *Regular* method involves matching speeches based on a 100% name match of the politicians, while the *Fuzzy* method employs an 80% similarity threshold for matching.

Figure A.1: Wordclouds: 1975-1982

(a) ARENA: pre-reform

(b) MDB: pre-reform



(c) ARENA: post-reform

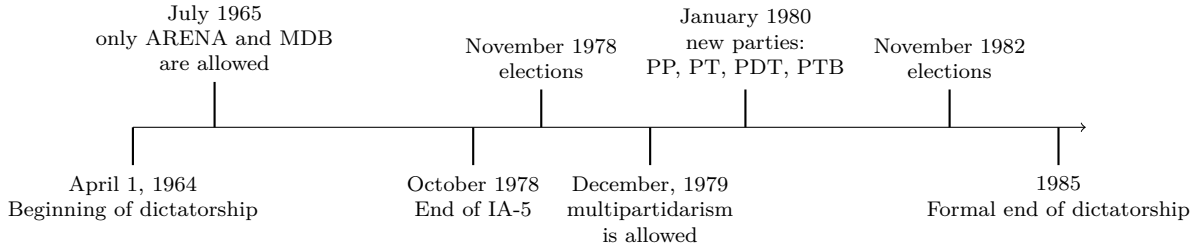
(d) MDB: post-reform

(e) New parties: post-reform

Notes: This figure displays word clouds with the 100 most frequent terms from congressional speeches for each party, both before and after the reform. The pre-reform period covers the years 1975-1979, while the post-reform period spans 1980-1982. Procedural words such as "deputies" (deputados) and "president" (presidente) have been filtered out, as well as names of politicians and localities. New parties emerging in the post-reform period include PP, PDT, PT, and PTB.

A.2 Timeline of events

Figure A.2: Timeline of political events



B Additional checks

B.1 Affective and cognitive scores

Table B.1: Affective and cognitive scores

	(1)	(2)	(3)	(4)	(5)	(6)
	Affective Score			Cognition Score		
MDB × 1975 (γ_{-4})	0.018	0.024	0.016	0.018	0.020	0.012
	(0.020)	(0.020)	(0.016)	(0.014)	(0.014)	(0.011)
MDB × 1976 (γ_{-3})	-0.012	-0.002	0.002	-0.014	-0.007	-0.005
	(0.019)	(0.019)	(0.015)	(0.014)	(0.013)	(0.011)
MDB × 1977 (γ_{-2})	-0.007	0.002	0.008	0.0002	0.006	0.006
	(0.018)	(0.019)	(0.016)	(0.013)	(0.013)	(0.012)
MDB × 1978 (γ_{-1})	-0.007	-0.001	0.018	-0.003	-0.0007	0.012
	(0.021)	(0.021)	(0.016)	(0.013)	(0.012)	(0.011)
MDB × 1980 (δ_1)	-0.063***	-0.056***	-0.036***	-0.029**	-0.026**	-0.014*
	(0.018)	(0.018)	(0.014)	(0.012)	(0.011)	(0.009)
MDB × 1981 (δ_2)	-0.045***	-0.038**	-0.027*	-0.027**	-0.023**	-0.017*
	(0.016)	(0.016)	(0.014)	(0.011)	(0.011)	(0.009)
MDB × 1982 (δ_3)	-0.093***	-0.085***	-0.069***	-0.047***	-0.044***	-0.033***
	(0.020)	(0.020)	(0.018)	(0.012)	(0.013)	(0.011)
Observations	50,866	50,408	50,296	50,866	50,408	50,296
R^2	0.028	0.031	0.308	0.038	0.046	0.331
Dep. var. mean	0.882	0.882	0.882	0.830	0.830	0.830
Controls						
Year	✓	✓	✓	✓	✓	✓
Gender		✓	✓		✓	✓
Age		✓	✓		✓	✓
Profession		✓	✓		✓	✓
Education		✓	✓		✓	✓
Speech length			✓			✓

Notes: The data are at the politician-speech level. The sample consists of speeches by politicians who were members of MDB and ARENA during the 1975-1982 period.Regressions take the form $Y_{its} = \lambda_t + \sum_{\tau=-4}^{-1} \gamma_\tau D_{i\tau} + \sum_{\tau=1}^{3} \delta_\tau D_{i\tau} + x_{ist} + \theta_s + \varepsilon_{ist}$ where Y_{its} is the outcome variable for individual i at year t and speech s. The term λ_t denotes year-fixed effects, capturing common shocks or trends at time t. $D_{i\tau}$ is an indicator variable equal to 1 if individual i is affiliated with MDB τ years after the reform, and 0 otherwise. γ_τ captures the effect of being affiliated with MDB at different times pre-reform and δ_τ measures the impact of MDB affiliation at different times post-reform. x_{ist} represents a vector of individual characteristics (gender, age group, profession, education status) and θ_s controls for speech length. ε_{ist} is the error term, capturing unobserved factors affecting the outcome variable. Standard errors are clustered at the politician level. $*$, $**$, $***$ denote significance at the 10%, 5% and 1% levels, respectively.

B.2 Extended sample

Table B.2: Extended sample until 1986

	(1)	(2)	(3)	(4)	(5)	(6)
	Emotion Score			Negative to positive words		
MDB × 1975 (γ_{-4})	0.002	-0.002	-0.002	-0.057	-0.036	-0.019
	(0.012)	(0.011)	(0.010)	(0.050)	(0.047)	(0.048)
MDB × 1976 (γ_{-3})	0.0007	-0.002	-0.006	-0.025	-0.021	-0.015
	(0.011)	(0.010)	(0.009)	(0.058)	(0.055)	(0.054)
MDB × 1977 (γ_{-2})	0.009	0.006	-0.0002	-0.070	-0.059	-0.051
	(0.011)	(0.011)	(0.010)	(0.050)	(0.050)	(0.051)
MDB × 1978 (γ_{-1})	0.005	0.002	-0.005	-0.021	-0.0009	-0.002
	(0.013)	(0.013)	(0.011)	(0.055)	(0.054)	(0.053)
MDB × 1980 (δ_1)	0.033***	0.030***	0.022**	0.186***	0.174***	0.181***
	(0.010)	(0.010)	(0.009)	(0.051)	(0.051)	(0.049)
MDB × 1981 (δ_2)	0.019**	0.017*	0.011	0.264***	0.254***	0.274***
	(0.009)	(0.009)	(0.009)	(0.058)	(0.058)	(0.058)
MDB × 1982 (δ_3)	0.044***	0.040***	0.036***	0.281***	0.275***	0.282***
	(0.011)	(0.011)	(0.011)	(0.066)	(0.066)	(0.066)
MDB × 1983 (δ_4)	0.028***	0.024**	0.019*	-0.018	-0.017	-0.012
	(0.010)	(0.010)	(0.010)	(0.060)	(0.059)	(0.059)
MDB × 1984 (δ_5)	0.024**	0.021*	0.015	-0.132**	-0.132**	-0.131**
	(0.012)	(0.011)	(0.011)	(0.060)	(0.060)	(0.060)
MDB × 1985 (δ_6)	0.018	0.011	0.005	-0.247***	-0.236***	-0.230***
	(0.011)	(0.011)	(0.011)	(0.061)	(0.061)	(0.060)
MDB × 1986 (δ_7)	0.016	0.017	0.011	-0.289***	-0.297***	-0.314***
	(0.015)	(0.014)	(0.013)	(0.068)	(0.069)	(0.069)
Observations	71,268	70,406	70,334	61,045	60,316	60,245
R^2	0.018	0.026	0.111	0.031	0.033	0.067
Dep. var. mean	0.961	0.961	0.961	0.943	0.942	0.942
Controls						
Year	✓	✓	✓	✓	✓	✓
Individual characteristics		✓	✓		✓	✓
Speech length			✓			✓

Notes: The data are at the politician-speech level. The sample consists of speeches by politicians who were members of MDB and ARENA during the 1975-1986 period. Regressions take the form $Y_{its} = \lambda_t + \sum_{\tau=-4}^{-1} \gamma_\tau D_{i\tau} + \sum_{\tau=1}^{7} \delta_\tau D_{i\tau} + x_{ist} + \theta_s + \varepsilon_{ist}$ where Y_{its} is the outcome variable for individual i at year t and speech s. The term λ_t denotes year-fixed effects, capturing common shocks or trends at time t. $D_{i\tau}$ is an indicator variable equal to 1 if individual i is affiliated with MDB τ years after the reform, and 0 otherwise. γ_τ captures the effect of being affiliated with MDB at different times pre-reform and δ_τ measures the impact of MDB affiliation at different times post-reform. x_{ist} represents a vector of individual characteristics (gender, age group, profession, education status) and θ_s controls for speech length. ε_{ist} is the error term, capturing unobserved factors affecting the outcome variable. Standard errors are clustered at the politician level. $*$, $**$, $***$ denote significance at the 10%, 5% and 1% levels, respectively.

B.3 Topics

Table B.3: Percentage distribution of topics by party and period

	ARENA		MDB		Other parties
	Pre-reform	Post-reform	Pre-reform	Post-reform	Post-reform
Agribusiness	8	10.2	6.6	7	5.3
Defense	1.7	2.7	4.7	10.3	7.2
Economic Development	4.9	6.1	1.9	1.5	2.2
Finance	3.9	5.9	5.1	5.4	6.2
Health and Education	5.3	5.3	6.5	6.6	7.6
Labor Rights	1	2.1	4	7.3	6.2
Land and Agriculture	3.1	4.6	3.7	5.9	3.2
Legal Measures	2.5	2.7	2.1	1.5	1.5
Legal System	12.9	6.6	14.2	6.5	8.2
Legislative Debate	11.1	6.4	6.9	3.8	5.3
Media	4.6	4.1	3.9	2.4	2.8
Oil Issues	3.3	2.8	3.2	3.4	2.9
Parliamentary Discourse	5.4	4.4	4	3.9	4.1
Pensions	5	6	7	5.8	7.5
Poverty	1.1	1.5	2.6	4.2	4.3
Politics and Democracy	4.2	5.9	7.2	11.6	9.6
Regional Development	6.4	8.6	3.3	4	4.2
Religion	7.2	7.6	5.6	4.4	5.7
Transportation	8.5	6.6	7.3	4.7	6

Notes: This table displays the percentage (%) of each topic within different parties and periods. Topics were identified using an LDA (Latent Dirichlet Allocation) model and subsequently classified manually. Each speech was assigned to a major topic based on the LDA model's output.

Table B.4: Average emotion of topics by party and period

	ARENA		MDB		Other parties
	Pre-reform	Post-reform	Pre-reform	Post-reform	Post-reform
Agribusiness	0.94	0.93	0.95	0.96	0.96
Defense	1.02	1.00	1.02	1.05	1.04
Economic Development	0.91	0.91	0.92	0.92	0.91
Finance	0.81	0.84	0.85	0.88	0.87
Health and Education	0.91	0.94	0.93	0.95	0.94
Labor Rights	0.93	0.95	0.98	0.99	1.02
Land and Agriculture	0.96	0.97	0.97	1.02	1.00
Legal Measures	0.92	0.94	0.94	0.91	0.93
Legal System	0.73	0.70	0.73	0.74	0.74
Legislative Debate	1.07	1.11	1.03	1.08	1.05
Media	1.10	1.10	1.07	1.05	1.07
Oil Issues	0.95	0.99	0.96	1.00	0.98
Parliamentary Discourse	1.00	1.04	0.98	1.01	1.02
Pensions	0.84	0.85	0.84	0.87	0.85
Poverty	1.05	1.06	1.07	1.08	1.09
Politics and Democracy	0.98	0.99	0.99	1.00	1.00
Regional Development	0.95	0.96	0.96	1.00	0.97
Religion	1.18	1.19	1.17	1.18	1.18
Transportation	0.97	0.97	0.97	0.98	0.98

Notes: This table displays the average emotion score of each topic within different parties and periods. Topics were identified using an LDA (Latent Dirichlet Allocation) model and subsequently classified manually. Each speech was assigned to a major topic based on the LDA model's output.