De-Polarizing Politics: Faith Meets Science in the French Third Republic*

Gianandrea Lanzara (University of Bologna)
Mara P. Squicciarini (Bocconi University and CEPR)
Nico Voigtländer (UCLA and NBER)

This version: April 2026

Abstract

We investigate whether individual politicians can help reduce emotionality and polarization in parliamentary debates. Focusing on the French Third Republic (1880–1914), we study Members of Parliament (MPs) educated in Jesuit schools (collèges) – institutions combining religious formation with rigorous scientific, philosophical, and rhetorical training. Using newly digitized data on Jesuit school enrollments and over one million parliamentary speeches, we identify Jesuit-educated MPs (JeMPs) and analyze their linguistic and behavioral patterns through text embeddings and large language models. We find that JeMPs systematically employed less emotional and less polarized rhetoric, emphasizing reason and science in their arguments. They also engaged more directly with preceding speeches, as reflected in greater lexical overlap. We then examine the dynamics within parliamentary debates, where the order of speakers was by and large pre-assigned. Even after controlling for speaker-discussion fixed effects, we find that JeMP interventions reduced subsequent speakers' emotionality and polarization. These effects were particularly pronounced during the contentious 1905 Separation of Church and State Law.

JEL: K42, N43, O10, Z10

Keywords: Schooling, Political Polarization, Mediation, Speech Analysis

*We thank Jérôme Adda, Caterina Alfonzo, Francesco Barilari, Gemma Dipoppa (discussant), Joel Mokyr, Tommaso Sonno, Juliane Uhl, and audiences at 1st CEPR Applied Micro-Economic History Workshop, 2nd Bari Workshop in Applied Economics, ASREC 2024, Bologna University, CUFE, EIEF, IIES, LISER, NYU Stern, Southern Denmark University, Tilburg University, Tor Vergata University, UC Riverside, University Halle-Wittenberg, University of Bergamo, Utrecht, and Warwick for insightful comments and suggestions. Filippo Giorgis, Charlotte Giraud, Camilla Mastroberardino, Lorenzo Pedretti, and Daniel Pérez-Huerta provided excellent research assistance. This research was funded by the European Union (ERC) under grant agreement 101042055 - EDIPO. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council.

"...every good Christian is to be more ready to save his neighbor's proposition than to condemn it. If he cannot save it, let him inquire how he means it; and if he means it badly, let him correct him with charity."

— Ignatius of Loyola (1491–1556), *The Spiritual Exercises*, Annotation 22 (Loyola, 1914).

1 Introduction

Political polarization has become a defining feature of contemporary democracies. Beyond shaping the tone and animosity of political discourse, polarization can undermine legislative effectiveness, weaken institutional checks and balances, and increase policy uncertainty, with negative economic consequences (McCarty, Poole, and Rosenthal, 2006; Baker, Bloom, and Davis, 2016). A large literature documents how to measure polarization in political language and traces its evolution in legislative speech (Gentzkow, Shapiro, and Taddy, 2019), as well as its institutional and electoral underpinnings (Fiorina and Abrams, 2008). Yet much less is known about whether political actors themselves can mitigate polarization within legislative arenas. Can certain politicians – through their training, rhetoric, or mode of engagement – act as mediators in polarized environments, moderating debate and reducing the intensity of partisan conflict?

We examine this question in the context of the French Third Republic (1870–1914), a period marked by intense ideological conflict between a progressive republican coalition and its conservative opponents. We focus on members of parliament (MPs) who were educated in Jesuit schools (*collèges*), whom we refer to as Jesuit-Educated MPs (JeMPs). Jesuit education combined rigorous training in theology and humanities, but also in mathematics and the natural sciences, as well as an extensive instruction in rhetoric. Importantly, students in Jesuit schools practiced formal disputations designed to cultivate disciplined reasoning and engagement with opposing viewpoints.[1] We digitize the complete enrollment registries of Jesuit schools in France from the Archives Jésuites, which record student names, surnames, and grade-level information. These records cover approximately 70,000 individual students enrolled between 1850 and 1901. We match these student rosters to the universe of French MPs using detailed biographical information. We identify 284 parliamentarians with a Jesuit education, representing approximately 8% of the total number of MPs in our sample period. We then combine this matched dataset with a newly assembled corpus of all parliamentary

[1]The standardized Jesuit curriculum was codified in the *Ratio Studiorum*, rooted in the pedagogical principles of Ignatius of Loyola (1491–1556), who had founded the Society of Jesus in 1540. From the 16th century onward, Jesuits became one of the most influential teaching orders in Europe, operating a vast network of schools and universities. Historians of education emphasize the central role of Jesuit institutions in shaping early modern European schooling and their reputation for intellectual rigor and disciplined pedagogy (see for example O'Malley (1993) and Grendler (1989)).

debates in France between 1880 and 1914, covering more than one million speeches as well as the full set of roll-call votes cast during these decades.

We document three main findings. First, using Natural Language Processing (NLP) based measures of emotionality and polarization, we show that JeMPs' speeches were systematically less emotional and less polarized than those of other MPs, even within political area and after controlling for topic as well as speaker and speech characteristics. JeMPs also placed greater emphasis on scientific reasoning in their interventions. Second, we document evidence that JeMPs affected the debating style of other legislators. Exploiting institutional features of the Chamber that largely predetermined the order of speakers within a discussion, we estimate event-study specifications around the first JeMP intervention. Following a JeMP speech, subsequent interventions became significantly less emotional and less polarized. Our specifications account for speaker-by-discussion fixed effects, ruling out endogenous sorting of less emotional or less polarized MPs into the post-JeMP speaking order. To assess whether the estimated effects reflect distinctive features of Jesuit-educated MPs rather than generic characteristics of certain legislators or districts, we implement placebo exercises that assign "placebo JeMP status" in three ways: i) randomly selected non-Jesuit MPs; ii) the 5% least emotional (or least polarized) non-Jesuit MPs; iii) MPs elected in the 10% least emotional or polarized départements. Our results indicate that the moderating influence we document is not replicated by randomly chosen legislators, by MPs with similarly moderate rhetorical profiles, or by representatives from less polarized electoral environments.

Finally, we examine the 1905 *Law Separating Churches and State*, one of the most polarizing reforms of the Third Republic that made secularism the guiding principle of the country. The 1905 Law ended state recognition and funding for any religious affiliation, required worship to be organized through tightly defined associations, and reallocated many assets of the former worship establishments to public bodies and local authorities. Despite strong material and ideological reasons to oppose the reform, JeMPs employed comparatively neutral and less emotional language in this debate. Moreover, their interventions reduced the incidence of extreme pro-secularization rhetoric among subsequent speakers without symmetrically amplifying extreme opposition. We further provide exploratory evidence that JeMP participation in debates reduced pro-secularization votes among other MPs, consistent with a moderating influence that may extend beyond rhetoric.

Taken together, our findings suggest that depolarization need not rely exclusively on institutional redesign or shifts in mass preferences. It can emerge endogenously through the composition of the political elite: when legislators possess training that emphasizes disciplined reasoning, engagement with opposing viewpoints, and restraint in rhetoric, their participation can moderate deliberation and attenuate conflict. This highlights a complementary channel through which polarization may be shaped—not only by institutions or voters, but by the formative experiences of those who hold

office.

We contribute to several strands of literature. First, we add to the growing body of work measuring political polarization in legislative speech and behavior (McCarty et al., 2006; Bonica, 2014; Gentzkow et al., 2019). A large literature has developed tools to quantify ideological divergence in roll-call voting and high-dimensional text data. Our contribution shifts the focus from aggregate trends to individual-level heterogeneity in rhetorical style and, crucially, to its spillover effects within debates. Second, our findings speak to the political economy literature on the causes and consequences of polarization. Prior work emphasizes institutional and party-level drivers of polarization (Fiorina and Abrams, 2008; Moskowitz, Rogowski, and Jr., 2024; Canen, Kendall, and Trebbi, 2020, 2026). This literature primarily analyzes polarization through shifts in party competition and strategic positioning, rather than through differences in how individual legislators engage in deliberation. We complement this perspective by showing that even in highly polarized environments, the presence of certain legislators can measurably affect the rhetoric of others and alter the dynamics of debate.

Third, our paper contributes to research on political communication and elite behavior in polarized environments. Bäck, Carroll, Renström, and Ryan (2023) show that partisan source cues (that is, attributing an otherwise neutral message to either a voter's most liked or most disliked party) reinforce partisan social identities among voters and heighten intergroup differentiation in the electorate; by contrast, we examine how heterogeneity among elites themselves shapes the tone and dynamics of deliberation within legislative institutions. Henríquez, Sabat, and Sullivan (2023) show that cross-party "likes" among Chilean deputies are positively associated with subsequent cross-coalition voting, interpreting such interactions as costly signals of ideological agreement. Whereas they study strategic signaling between legislators in social media and its relationship to roll-call behavior, we examine how exposure to particular types of legislators shapes the tone and dynamics of deliberation within parliamentary debate itself. Fishkin, Siu, Diamond, and Bradburn (2021) document substantial reductions in both issue-based and affective polarization following structured, balanced cross-partisan discussion in a national deliberative field experiment ("America in One Room"). Their findings suggest that exposure to institutionalized, balanced deliberation can moderate political attitudes. Similarly, Doornbosch, van Vuuren, and de Jong (2025) examine three facilitated bipartisan face-to-face dialogues among U.S. members of Congress and identify communicative strategies – such as embracing multiple frames and foregrounding superordinate identities – that foster convergence across ideological and identity divides. Relatedly, Downey, Kuk, and Nanes (2024) provide experimental evidence that affective polarization is sustained in part by voters' misperceptions about how well their issue positions align with those of their co-partisans. The authors show that correcting these inaccurate beliefs reduces partisan affect, suggesting that polarization can be mitigated when individuals recognize cross-cutting issue misalignment. The results from these three papers are consistent with

the idea that systematic training in rhetoric and awareness/engagement with opposing views – such as that received by Jesuit-educated MPs – may foster moderation. We investigate this effect in a natural legislative setting, where deliberation unfolds within an actual parliament rather than in structured deliberative settings. We identify a specific group of politicians whose background predicts a systematic moderating influence on debate – even during one of the most contentious reforms of the period.

Fourth, a growing literature examines the formation and selection of political elites, highlighting how educational and career pathways shape who enters politics and how they govern once in office. Bó, Finan, Folke, Persson, and Rickne (2017) show that individuals with higher ability and stronger earnings potential are more likely to enter politics when the returns to office increase. At the same time, education shapes civic engagement and political behavior more broadly (Milligan, Moretti, and Oreopoulos, 2004), and the content of schooling influences long-run ideological beliefs (Cantoni, Chen, Yang, Yuchtman, and Zhang, 2017). State leaders' educational backgrounds affect governance outcomes and reform behavior (Besley, Montalvo, and Reynal-Querol, 2011; Dreher, Lamla, Lein, and Somogyi, 2009). This literature suggests that education shapes how politicians behave in office, suggesting that distinctive training environments (in our case, Jesuit schools) can influence the style and tone of legislative deliberation.

Finally, our historical setting contributes to the literature on the development of democratic institutions and legislative deliberation. The French Third Republic provides a highly polarized environment characterized by deep conflicts over secularization, church–state relations, and republican reform. By assembling a comprehensive corpus of parliamentary speeches and roll-call votes and linking them to individual-level educational backgrounds, we provide new large-scale evidence on deliberation in a foundational democratic period. Our analysis complements work emphasizing the importance of elite norms and informal institutions in sustaining democratic stability (Levitsky and Ziblatt, 2018) and the role of communicative practices in shaping democratic legitimacy (Habermas, 1996). It also connects to empirical studies of legislative speech showing that linguistic patterns predict legislators' ideological positions, highlighting the informational content of parliamentary rhetoric (Diermeier, Godbout, Yu, and Kaufmann, 2012). More broadly, the paper suggests that the structure of democratic debate is shaped not only by formal rules and electoral incentives, but also by the intellectual traditions and formative training of those who hold office.[2]

[2]More broadly, our analysis is also related to the literature on religion and political development. Many of these studies have focused on the Islamic world (e.g. Rubin, Jared, 2017), others have examined the relationship between politics and religion in Catholic Europe (e.g. Belloc, Drago, and Galbiati, 2016) and still others have looked at the period of the Protestant Reformation (e.g. Cantoni, Dittmar, and Yuchtman, 2018). For an overview of the relationship between religion and politics and, more broadly, of the economics of religion see Becker, Rubin, and Woessmann (2024), and for a theoretical framework on the relationship between political and religious authorities, see Bisin, Rubin, Seror, and Verdier (2024). To the best of our knowledge, we are the first to show how political elites, trained in religious schools, may contribute to de-polarizing the political debate.

The remainder of this paper is organized as follows: Section 2 discusses the historical background of Jesuit education and the political system during the French Third Republic. Section 3 presents our data and Section 4, our empirical strategy and results. Section 5 concludes.

2 Historical Background

This section provides the historical background for our analysis by tracing the development of Jesuit collèges in France and their conflict with the French state. It then situates these schools within the institutional setting of the Third Republic, which shaped the reforms at the center of our study.

2.1 Jesuits in France

In the early 18th century, the Society of Jesus was highly influential in France (van Kley, 1975). Jesuits ran a large network of schools (*collèges*) across the country and provided high-quality education for "future leaders of French society" (Thompson, 2018). This was based on rigorous training in humanities, rhetoric, theology, but also in sciences and mathematics. The Jesuits' contribution to sciences is well-known, especially during the Enlightenment era. A prominent example is the Jesuit mathematician and physicist Louis-Bertrand Castel (1688–1757). Castel, who entered the Jesuit order in 1703, was a prominent Paris-based intellectual and served as science editor of the Jesuit *Journal de Trévoux*.[3] He wrote on physics and universal gravity, proposed the famous "ocular harpsichord," while engaging in a critique of Newton's spectral description of prismatic color.[4]

The Jesuit school network was influential not only because it shaped elite formation, providing its students with humanistic and scientific training, but also because it created a nationally-connected community that many contemporaries often perceived as a rival source of authority to the crown, or later to the republican state. This perception partly contributed to a strong anti-Jesuit campaign in the early 1760s and to the later ban of the order in 1764.[5] The expulsion of Jesuits from the country lasted until 1814, when the Society of Jesus was permitted to return to France.

After a tentative rebuilding of the Jesuit school network, which suffered an additional closure in 1828, it was the 1850 Falloux Law that legitimated and expanded the space for non-state schooling and

[3]The *Journal de Trévoux (Mémoires de Trévoux)* was a scientific journal, founded in 1701 and run largely by Jesuit editors. It published critical reviews of contemporary books across "sciences & beaux-arts" (and broader learned culture). See https://artfl-project.uchicago.edu/journal-de-trevoux for an overview.

[4]Another example of a prominent Jesuit scientist is Antoine Gaubil (1689–1759), a leading astronomer who bridged between the Chinese and European scientific traditions, representing an important example of the Jesuit role in transcontinental scientific communication.

[5]The precipitating event was a scandal involving a Jesuit missionary in the French Caribbean, Father Lavalette. In 1761–62, the magistrates, under the strong influence of the Jansenists, claimed that the Jesuit principles were incompatible with the laws and traditions of the kingdom, thus suppressing the Jesuit collèges and requiring members of the order either to abjure their vows or to depart from French territory. Unable to effectively counter the decisions of the magistrates, Louis XV ultimately ratified the ban in 1764 (Vogel, 2010).

triggered the emergence of a longer-term and flourishing community of Jesuit educational institutions. Among these were the famous *Collège de l'Immaculée-Conception* (commonly called Vaugirard) and the *Ecole Sainte-Geneviève de la rue des Postes* (Langdon, 1978).[6]

A renewed wave of anti-Jesuit policies emerged under the Republican government. After the Franco-Prussian War and a short period dominated by conservative and monarchist forces (1870–1876), from 1877 onward, the political equilibrium shifted toward the progressive reforms implemented by the Republicans. One of the main objectives was the redefinition of the role of religion in public life toward building a secular state.[7] Secularization laws were poised to grant the State full control over education. This was intended as a means to ensure cohesion and unity of citizens beyond any past ideological divisions (Mayeur, 1981). However, to achieve this goal, Republicans had to weaken and dismantle their main competitors in the training of the French youth, i.e., the entire network of Catholic schools. Most secularization reforms addressed all Catholic institutions – such as the exemplary 1882 Jules Ferry Education Act.[8] In other cases, laws were specifically directed against the Jesuits. For instance, with the "Ferry Decrees" of March 1880, the Ferry administration orchestrated an attempt to finally expel all Jesuits from France and to close their *collèges* (Bush, 1975).[9] What drove this anti-Jesuit sentiment, despite their large contribution to the training of the country's elite? Not too differently from the anti-Jesuit spirit at the beginning of the 18th century, Jesuits were criticized "for the tendency to mobilize youth to their goal of state domination" (Doyle, 2017) and for training the Catholic elite of the time to cover important governmental positions, so that "the whole state bureaucracy, [...] was under siege by Jesuit affiliates" (Bush, 1975, p. 127).

The attempt to close all Jesuit collèges was not as effective as initially planned by the Republicans. Many schools were reorganized or reconstituted through alumni support and lay intermediaries.[10] By the early 1890s, the vast majority of Jesuit collèges became again fully operational, managing to

[6]The *Collège de l'Immaculée-Conception* was founded in Paris in 1852. Its nickname comes from its location on the rue de Vaugirard. It soon represented the flagship of the Jesuit collèges as the first school being established in Paris after 1850. The *Ecole Sainte-Geneviève de la rue des Postes* was particularly well-known for providing a post-baccalauréat instruction and its students had an extremely high admission rate at the *grandes écoles*.

[7]This was accompanied by a broader package of "Republican laws," which covered, for instance, freedom of the press, and freedom to form trade unions.

[8]The 1882 Jules Ferry Education Act is considered a milestone of the secularization program promoted by the Republican government that made primary education mandatory, free, and secular (see Squicciarini, 2020).

[9]As Jules Ferry himself puts it, in one of his speeches at the Chamber of Deputies, in 1879: "In education it's these [Jesuits] who give the tone; it's these who furnish the example for others; these who enjoy the success...who serve as model for all the ecclesiastical establishments," justifying the particularly strong opposition against the order (Bush, 1975).

[10]For the Vaugirard school, "under the supervision of *Monseigneur* D. Forges, Bishop of Ténarie and former auxiliary to Cardinal Brossais-Saint-Marc, Archbishop of Rennes, lay teachers provided the instruction; and a board of administration made up of former pupils and chaired by *Monsieur* Édouard Lefébure managed and oversaw everything relating to financial upkeep. As political passions gradually seem to subside, priests and scholastics of the Society reappear in the teaching posts of the classes and in those of advanced studies" (our translation from Delattre, 1949, pp. 1384–1385).

recruit a large set of students (Bush, 1975). The definitive closure took place with the 1901-1904 *Lois Anti-congréganistes*, when thousands of religious schools were closed – and the Jesuit collèges shared the same fate.[11] The secularization policies culminated in the 1905 Law "Concerning the Separation of Churches and State," which firmly enshrined the principle of laïcité in France.[12]

2.2 Jesuit *Collèges* in the 19th Century: Organization and Pedagogy

Figure 1 shows the evolution of Jesuit collèges and annual enrollment in Jesuit collèges from 1850 to 1910. As shown in the top panel, starting with the 1850 Falloux Law, the number of Jesuit collèges increased significantly over time, reaching a total of 30 institutions in operation (including three in Paris). The bottom panel in Figure 1 plots total enrollment in Jesuit schools, showing a steady increase until 1880, except for a temporary drop during the Franco-Prussian War in 1870. The negative impact of the Ferry decrees in 1880, as well as the partial recovery in subsequent years, which we discussed above, are also visible in this figure.

The increase in enrollment in Jesuit (and, more broadly, religious) schools after the 1870s is counterbalanced by a sharp decline in lay private schools, as both competed for the same clientele (Bush, 1975). This shift reflects the institutional advantage of Catholic education after the Falloux Law (1850), which enabled religious congregations to build stable, well-organized school networks, unlike the more fragmented lay private sector. At the same time, ecclesiastical schools better matched bourgeois expectations, combining moral discipline, social prestige, and preparation for the baccalauréat, which was central to status and careers (Bush, 1975). In addition, the political crisis of 1870–71 likely reinforced parental preference for religious institutions associated with order and stability (Harrigan, 1974). Overall, ecclesiastical schools emerged as the dominant private alternative, largely at the expense of weaker lay private schools.

Figure 2 depicts the distribution of Jesuit collèges across French departments, showing that they were distributed relatively evenly across the country.

Jesuit collèges typically enrolled students aged 8 to 18. Recruitment was addressed to relatively high-social-status families, but tuition cuts or scholarships were often granted to lower-income students. Tuition was comparable to that of secular lycées but generally higher than in non-Jesuit ecclesiastical schools (Bush, 1975).[13] Jesuit students were disproportionately successful in terms of educational achievement. For instance, their success rate in the baccalauréat examination was significantly higher than the national average. Even more strikingly, although they comprised only 5%

[11] The Law of July 1901 dissolved religious congregations, and the law of July 1904 prohibited all religious congregations from teaching, requiring congregational schools to close within ten years, and preventing the creation of any new congregational teaching institutions (Harrigan, 2001).

[12] We provide further details on the 1905 Law in Section 4.3.

[13] Tuition ranged from 1000 to 1600 francs for boarding students in Paris and from 800 to 1100 francs in the provinces (Bush, 1975). This corresponds roughly to the average annual income at the time (Coste, 1890).

of total secondary school enrollment, they accounted for 14.4% of students admitted to the grandes écoles between 1892 and 1901 (Bush, 1975).

Pedagogically, the curriculum, principles, and teaching methods of the Jesuit collèges were codified in the 1599 Ratio Studiorum, the comprehensive educational code of the Jesuit order. The Jesuit school curriculum blended intensive study of theology and the humanities with strong preparation in mathematics and the natural sciences, along with thorough training in rhetoric. A key feature of Jesuit education was the practice of formal disputations, through which students developed disciplined reasoning and engagement with opposing viewpoints.

In particular, the first years of Jesuit education were devoted to an immersion in classical letters and humanities. Then, students entered a three-year philosophy course, which, following the 1599 Ratio, included logic, physics, and astronomy together with methaphysics and philosphy. Additionally, mathematics was a core subject, to be studied daily by students in the second year of the philosophy course (Farrell, 1970). This structure and emphasis on sciences is evident also in the school curricula of French collèges in the 19th century. As an example, we provide the curriculum of the Jesuit school in Vannes in Figure 3. A second distinctive feature of Jesuit education was the emphasis on rhetoric. Its study was based on canonical authors – especially Cicero and Aristotle – combined with structured reading lists and stylistic training, including tropes, figures, and rules of style (Schwickerath, 1904).

2.3 Jesuit Education as Preparation for Mediation

The strong emphasis placed on disputation and rhetorical training in Jesuit education can be traced back to Ignatius of Loyola (1491–1556), the founder of the Society of Jesus. In the "Spiritual Exercises," Ignatius articulated a pedagogical and spiritual method grounded in disciplined argumentation. Our opening quote of Ignatius' Paragraph 22 (on top of Section 1) captures the key characteristics of Jesuit pedagogy and rhetoric: the defense of truth and the insistence that such defense be conducted with intellectual honesty and respect. This approach was formally institutionalized in the Ratio, which considers disputation among students as a central pedagogical practice, especially for the classes of theology and philosophy. The Ratio prescribed a specific ethos for argumentative engagement. Another illustrative quote is: "In arguing debatable questions, he should defend his view with such modesty and courtesy as to show respect for the contrary view...When it is possible to reconcile diverse views, an attempt should be made to do so" (Farrell, 1970, p. 26). Disputation was, thus, conceived as a disciplined exercise in reasoning, persuasion, and in the reconciliation of differing perspectives. The impact of this disputation tradition is evident in other Jesuit initiatives, such as the Conférence Olivaint. Founded in Paris in 1874, the Conférence Olivaint was a Jesuit student association, dedicated to political discussion and debates with public figures. Meetings were held weekly, and it represented a natural extension of the pedagogical model into the political sphere. Interestingly, the

association continues to exist today (www.conferenceolivaint.fr) and, despite not being run by Jesuits, it continues to engage with contemporary political and intellectual life.

More broadly, this pedagogical model arguably prepared Jesuit-educated students for a political role that went beyond advocacy. By training students to engage opposing arguments in a disciplined and respectful manner, and by emphasizing the reconciliation of divergent views, Jesuit education plausibly fostered skills that were especially valuable in parliamentary settings characterized by bargaining across factions. In this sense, Jesuit collèges may have produced not only eloquent political actors, but also individuals particularly well equipped to act as mediators.

2.4 The French Political System during the Third Republic

After France's defeat in the Franco-Prussian War (1870–1871) and the fall of Napoleon III's Second Empire, a National Assembly was elected by universal male suffrage in February 1871 to ratify the peace treaty with Prussia and determine the institutional form of the new regime. The Constitutional Laws of 1875 resulted from a compromise between the progressive republicans and the more moderate wing of the monarchist faction, which had come (at least temporarily) to accept the establishment of a republican regime. These laws established a bicameral parliamentary system composed of a lower chamber, the Chamber of Deputies, and an upper chamber, the Senate (Mayeur, 1984). [14]

While the two chambers possessed largely equal legislative powers, they differed in their electoral rules and eligibility requirements. As a concession to the monarchists, the Senate was designed as a more stable and conservative chamber, as reflected in four institutional features. First, while deputies were elected for four-year terms, elected senators served nine-year terms, alongside 75 senators who were appointed for life (after 1884, vacancies among these seats were filled through election, and the number of life senators gradually diminished). Second, deputies were elected by direct universal male suffrage, whereas senators were chosen by electoral colleges in each department composed of delegates from municipal councils, deputies, and members of other local bodies. Third, candidates for the Senate had to be at least 40 years old, while candidates for the Chamber of Deputies had to be at least 25. Fourth, and finally, the Senate was only partially renewed: one third of the elected senators were replaced every three years.[15]

With regard to the Chamber of Deputies, the electoral system established by the Constitutional Laws of 1875 and the electoral laws of 1875 and 1876 provided for elections in single-member districts under a two-round majority system. In the first round, candidates were required to obtain an absolute majority of votes and at least 25% of registered voters; in the second round, a relative major-

[14] The Senate comprised 300 senators, while the number of deputies in the Chamber fluctuated from approximately 530 to 600 depending on the electoral laws implemented over time.

[15] For details on the constitutional laws see https://www.elysee.fr/la-presidence/les-lois-constitutionnelles-de-1875.

ity sufficed.[16] With the exception of a brief interlude between 1885 and 1889[17], this system remained in place until 1936.

In the early years of the Third Republic, political organization was only weakly structured by formal parties. This was even reflected by the fact that deputies were seated alphabetically in the Chamber, embodying the theory of the representative mandate rather than partisan alignment. Once in the Chamber, they gradually affiliated with parliamentary groups, which became the main units of political coordination. The institutional recognition of these groups emerged only progressively, particularly under pressure from left-wing deputies: a reform of the Chamber's rules in November 1911 allowed groups to designate their representatives to committees, and by 1914 seating in the Chamber itself was organized by group from the beginning of each legislature (Davin, 1985).

Speaker Order in Parliamentary Debates. The rules of procedure of the French Chamber of Deputies during the Third Republic (Bonnard, 1926) suggest that parliamentary debate followed a structured and largely predetermined order. This is important for our event-study design in Section 4.2. The rules established that the President of the Chamber opened sessions, directed deliberations, and enforced order (Article 37). Deputies were required to request the floor and obtain permission before speaking (Articles 41–42), and their requests were recorded by secretaries in a formal speaking list (*ordre des demandes*). At the same time, certain actors – such as ministers, committee presidents, rapporteurs, and government commissioners – were not bound by this list and could intervene more flexibly (Article 43). We control for those speaker roles in our regressions. In addition, while speeches in general sessions were formally limited to 15 minutes, the President retained discretion to interrupt speakers or allow deviations from the prescribed order. In practice, however, debates were not entirely rigid. Frequent interruptions, short interjections, and spontaneous interventions were common features of parliamentary exchanges. We include speech length fixed effects to filter out such short interventions.

3 Data

Our analysis relies on three main sources of the data. First, we digitize the enrollment registries of Jesuit schools in France from 1850 to 1901. Second, we digitize the universe of parliamentary speeches for the French Third Republic from 1881 to 1914. Third, we use individual level data on all MPs during the Third Republic with detailed biographical information. In the remainder of this section, we describe each of these sources, and how we link them together. The final output is a dataset of parliamentary speeches where each speaker, whenever elected as a member of parliament is associated with key personal and political characteristics, among which, importantly, whether they

[16] For details on the 1875 electoral law, see https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000521258.

[17] The 1885 elections were held under a two-round list system at the department level. Each department elected multiple deputies and voters cast ballot for party lists.

were educated in a Jesuit school. More details on the data construction are provided in Appendix A.

3.1 Jesuit Students

Data on students enrolled in Jesuit schools in France are available at the *Archives Jésuites* in Vanves, near Paris,[18] which preserve enrollment registries for 30 of the 32 Jesuit schools that operated in France between 1850 and 1901.[19] The student registers are preserved in 142 separate documents dating from 1850 to 1901. For 37 of these documents, only the handwritten (hereafter HW) version has survived. Figure A.1 in Appendix A.1.1 illustrates the student registry for school year 1885-86 at the *Collège de l'Immaculée-Conception*, located on *Rue de Vaugirard* in Paris, while Figure A.3 shows a handwritten document for the same school for school year 1853–54 . As in these examples, registers from Jesuit schools typically report the list of students enrolled by school year along with information on each student's grade.[20] As an exception to the standard format, eight documents report only the name and year of entry for students.

For printed school registries, our digitization process consists of four main steps: (i) we manually scan all front pages, legends, and pages reporting the lists of students in alphabetical order; (ii) we perform Optical Character Recognition (OCR) on each document using Google Document AI;[21] (iii) we parse and standardize the OCR output to harmonize column types (e.g., name, student number, class, year of entry) despite substantial variation in formatting; (iv) with the help of research assistants, we manually reviewed all final Excel files to identify and correct potential OCR errors. For handwritten (HW) documents, we enter the data manually for the entire document. At the end of this process, we obtain a dataset with 268,216 student records. Some of these records belong to the same student observed in multiple school years, while others refer to students who appear only once in the dataset along with their year of entry at their school.

We then reconstruct the school histories of Jesuit-school students based on their names and the years during which they appear in the records. The main objective is to determine the earliest and latest years during which each student attended a Jesuit school. This information, in turn, helps us estimate the student's age, which is ultimately needed to match student records to parliamentarians. Concretely, we begin by constructing the variables $y_r^{\min}$ and $y_r^{\max}$ for each *record* r as follows: (i) if the document reports the school year in standard format, e.g., 18XX–18(XX+1), we assign 18XX to $y_r^{\min}$ and 18(XX+1) to $y_r^{\max}$; (ii) if the document only reports the year of entry, we assign this value to both $y_r^{\min}$ and $y_r^{\max}$. We then implement an iterative procedure that links student records belonging

[18] See https://www.jesuites.com/contact/bureaux-archives-jesuites for an overview.

[19] The archives lack documentation for the schools in Montauban and Mende, part of the Province of Toulouse, as well as for those in Algiers and Oran, located in colonial Algeria and associated with the Province of Lyon.

[20] The 1885-1886 register in Figure A.1 also reports information on section, division (a group of students spanning multiple grades who shared the same boarding facilities), and boarding status.

[21] A key advantage of Google Document AI is its ability to detect and preserve the tabular structure of the pages.

to the same student and updates enrollment windows accordingly. In each round, we link records whose current enrollment intervals overlap within a four-year window and whose names match under a conservative fuzzy criterion. The procedure iterates until no further links are formed.[22] We obtain 70,452 linked students.

As a final step, we use class information reported in the original documents, whenever available, to estimate the age of each student. First, we harmonize class codes across schools and years using the legends provided in the school registries (see Figure A.2). Jesuit schools followed the French educational system, where grades ran in reverse order. *Rhétorique*, the final year of the standard curriculum, was typically attended at ages 16-17, corresponding to "grade 1." Most schools offered additional years in Philosophy and Science, attended at around age 18. Entry-level grades varied across Jesuit schools: while some schools began at grade 7, approximately at age 10, others offered preparatory courses starting at grade 9, approximately at age 8. Accordingly, we estimate the implied age for each record r using the formula: $Age_r = 8 + (9 - Grade_r)$.

3.2 Parliamentary Speeches

The transcripts of parliamentary debates of the French Third Republic were published in the official gazette (*Journal officiel de la République française*) for both the Chamber of Deputies and the Senate. These publications are available online in PDF format on Gallica, the digital repository of the French National Library.[23] While the *Journal officiel* issues had a three-column structure, some pages embed numerical tables that disrupt the three-column layout. Because OCR software may fail to reliably detect column breaks, we use Detectron2 image detection algorithms to segment each page into text columns/blocks and tables, and then apply Google Cloud Vision for optical character recognition on each detected text region.[24]

Next, we split the digitized transcripts into parliamentary sittings, agenda items, and individual speeches. First, we separate distinct sittings within each transcript document (when Parliament convened more than one time on the same date, the corresponding record typically reports all sittings). Second, we partition each sitting into consecutive sections of the debate – hereafter, *discussions* – each centered on a specific issue or bill that Parliament was called upon to examine. Finally, we split each discussion into individual speeches. To illustrate, on April 5, 1905, the Chamber of Deputies convened for a single sitting. After discussing the construction of a railway between the towns of Candé

[22] Further details on the linkage procedure are provided in Appendix A.1.2.

[23] See https://gallica.bnf.fr. Gallica also provides an OCR-generated text version of these documents. However, we digitize the documents directly from the scanned pages in order to obtain a higher-quality text extraction. Our approach to parsing individual speeches from the transcripts and linking them to MPs builds on Lanzara, Lazzaroni, Masella, and Squicciarini (2024), who rely on Gallica's OCR text.

[24] See Appendix Section A.2 for more details. For information on Detectron2, see https://github.com/facebookresearch/detectron2; for information on Google Cloud Vision, see https://docs.cloud.google.com/vision/docs/ocr.

and Angers, it debated the law on the Separation of Churches and the State. In this discussion, 42 distinct speakers took the floor and delivered 237 interventions (see Figure A.5 in the Appendix). We leverage procedural and stenographic conventions that recur throughout the study period to construct regular-expression patterns that identify sitting boundaries and discussion headers, and extract individual speeches along with the reported name of the speaker.[25] Our final sample comprises 1,660,083 speeches, distributed across 82,364 discussions and 8,191 parliamentary sittings; 1,173,687 of these speeches -— 70% of the total —- were delivered in the Chamber of Deputies.

3.3 Members of Parliament (MPs)

Data on the personal and political biographies of all parliamentarians of the French Third Republic are available on the institutional websites of the *Assemblée nationale* (today's lower chamber) and the *Sénat*, respectively.[26] We focus on 2,464 deputies and 865 senators who served in the French Parliament between January 1881 and March 1914. Each MP has a dedicated page reporting his name, birth and death details, as well as his political biography.[27]

For deputies, these pages also report political affiliation for each individual mandate. These "political groups" comprise around 200 distinct categories that sometimes correspond to political parties, sometimes to broader coalitions, and sometimes to narrower subgroups or internal factions. We map the parliamentary groups reported for the Chamber of Deputies into five broad political areas: left, center-left, center, center-right, and right. This mapping relies on the classifications in Cagé and Piketty (2025) and is supplemented with information from https://www.france-politique.fr/ for groups that appear in our data but are not covered in Cagé and Piketty.[28] By contrast, political affiliation is not available for senators.

Finally, each page includes one or two biographical excerpts drawn from the following references: the *Dictionnaire des parlementaires français de 1789 à 1889* by Adolphe Robert and Gaston Cougny, and the *Dictionnaire des parlementaires français de 1889 à 1940* by Jean Jolly. We use ChatGPT to infer each MP's occupation from these excerpts, focusing on the following pre-defined list of broad professional groups: other government officials, other skilled professionals, legal profession-

[25] For example, the President of the Chamber announces the start of proceedings with the formula *la séance est ouverte*. Each agenda item is introduced by an uppercase header. Each intervention then starts on a new line – typically after a period or an uppercase header – and the speaker is introduced by "M." followed by a surname and a period. See Figure A.5 and Appendix Section A.2 for more details.

[26] See www2.assemblee-nationale.fr/sycomore/recherche for the *Assemblée nationale* and www.senat.fr/senateurs-3eme-republique/senatl.html for the Senate.

[27] Lanzara et al. (2024) compiled this dataset and extensively reviewed and standardized these biographical fields, correcting minor inconsistencies in birth information and in mandate start and end dates.

[28] Importantly, the assignment of political groups to broad areas is time-varying. For instance, the Republican Radicals are classified as left in 1881 but as center-left in 1885, reflecting a shift in the political spectrum following the rise of the Socialist Party.

als, skilled manual workers, industrial owners, education professionals, senior government officials, military personnel, clergy, agricultural owners, unskilled workers, and traders.

3.4 Linking Datasets

Matching Students to MPs. We match students from Jesuit schools and parliamentarians based on their names and year of birth. For students, we approximate their year of birth using either grade-based age information (available for about 90% of records) or, when age is missing, the years during which the student appears in Jesuit school records.[29] For each parliamentarian, we then search for potential matches among students using a four-step procedure: (i) we restrict attention to students observed in the data when the parliamentarian would have been between ages 4 and 22; (ii) we require the student's birth-year interval to overlap with the MP's birth year within a ±3-year margin; (iii) we identify candidate matches based on similarity between students' and MPs' surnames using the Jaro string-distance metric; (iv) we apply a similar fuzzy matching procedure to first names; (v) all candidate matches are manually reviewed to eliminate false positives arising from spelling variations or transcription errors. This procedure identifies 133 parliamentarians who attended Jesuit schools before serving in the French Parliament between 1881 and 1914. Of these, 117 served in the Chamber of Deputies, while 16 served only in the Senate.

Matching parliamentary speakers to MPs. We link the extracted speeches to our biographical dataset of MPs using the speaker name and the date of the parliamentary sitting. Our matching procedure consists of six steps: (i) for each branch–date combination, we extract the list of unique speakers appearing in the transcript and retain a single observation per speaker; (ii) we retrieve from the biographical dataset the MPs who were in office in that branch on that date, using the start and end dates of their mandates;[30] (iii) for speakers whose cleaned surname consists of a single component, we attempt to match directly to MPs' cleaned surnames, retaining candidate matches whenever the Levenshtein distance is small (e.g., distance ≤ 1); (iv) for remaining speakers with one name component, we attempt matches against MPs' multi-part surnames, as in these cases the speaker's surname might not coincide with the one that we identified from the biographical dataset; (v) for speakers with multi-part names, we match the full lists of name components across the two datasets using Levenshtein distance. For each potential speaker–MP pair, we first compute the matrix of pairwise

[29] Jesuit students typically attended school between the ages of 8 and 18, with a full trajectory lasting 11 years. Based on this, we compute a conservative interval for the year of birth:

$$ybirth_i^{\min} = \min(y_i^{\min} - 19 + n,\ y_i^{\max} - 7 - n),$$
$$ybirth_i^{\max} = \max(y_i^{\min} - 19 + n,\ y_i^{\max} - 7 - n),$$

where n is the number of years the student appears in the data, capped at 11. This formula incorporates the idea that the more years a student is observed, the more accurately we can infer their age.

[30] To account for small recording errors, we apply a seven-day buffer on each side of the mandate interval.

Levenshtein distances between all speaker name components and all MP name components. We then construct a one-to-one assignment by iteratively selecting the smallest-distance entry in this matrix and removing the corresponding row and column, so that each component can be used at most once. The overall match distance for the pair is defined as the average Levenshtein distance across these matched component pairs; (vi) we classify a speaker–MP link as a successful match when it is both sufficiently close in distance and unambiguous, i.e., no ties among best matches.

3.5 Text Analysis of Parliamentary Speeches

This section describes how we derive emotionality and polarization scores from speeches in the French Parliament. First, we preprocess the text to harmonize case and accents and retain only informative word stems. Second, we train a Word2Vec model and perform a Latent Dirichlet Allocation (LDA) topic analysis. Finally, we use these techniques to assign emotionality and polarization scores to each speech. Further details on the text analysis are provided in Appendix B.

Preprocessing. We preprocess parliamentary speeches to reduce dimensionality (Gentzkow, Kelly, and Taddy, 2019). First, we assign part-of-speech tags and retain only adjectives, adverbs, nouns, and verbs. We then normalize tokens by removing accents, converting text to lowercase, stripping punctuation and digits, and keeping only alphabetic strings. Finally, we apply the Porter2 stemming algorithm[31] to reduce words to their morphological roots (Porter, 2009). We further remove uninformative language by dropping short tokens (fewer than three characters), filtering out rare words (appearing in five or fewer documents or ten or fewer times in the corpus), and removing French stopwords using a custom list.[32] Because parliamentary debates contain a large amount of procedural discourse, we also remove procedural words following Gentzkow et al. (2019). To do so, we digitize the Rules of Procedure (*Règlements*) of the Chamber and Senate during the Third Republic (Bonnard, 1926), use an auxiliary Word2Vec model to embed these terms, construct a frequency-weighted "procedural" centroid, and drop the bottom 1% of words in the speech corpus ranked by cosine distance to this centroid.[33] After these steps, our sample comprises 1,468,790 speeches with nonempty text.

Word Embeddings and Topic Analysis. We train a Word2Vec model on the resulting corpus using `gensim` (Mikolov, Sutskever, Chen, Corrado, and Dean, 2013), with 300 dimensions and a context window of 8 words, yielding vector representations for all tokens. We then compute a document embedding for each speech as an inverse-frequency-weighted average of its word vectors. For speech

[31] We use the French version of this algorithm, implemented in the R package `SnowballC`.

[32] We start from the French stopword list in the `stopwords` package and supplement it with days of the week, months, names of French *départements*, and additional low-information terms (e.g., *faire*, *parl*, *temps*, *nombre*).

[33] Operationally: (i) we preprocess the *Règlements* using the same pipeline (except POS tagging); (ii) we train an auxiliary Word2Vec model on the full speech corpus; (iii) we compute a frequency-weighted centroid of procedural terms; (iv) we compute cosine distances between each vocabulary word and this centroid; and (v) we remove the 1% most similar words. Appendix Figure A.8 illustrates the resulting procedural lexicon.

$\mathcal{D}_i$, the document vector is

$$\vec{d_i} \;=\; \frac{1}{|\mathcal{D}_i|} \sum_{w \in \mathcal{D}_i} \frac{\alpha}{\alpha + p(w)} \vec{w}, \tag{1}$$

where $\vec{w}$ is the vector representation of word w, $p(w)$ is its relative frequency in the corpus, and $\alpha = 0.001$ (Gennaro and Ash, 2022). The inverse-frequency weights ensure that words that appear frequently in the corpus are down-weighted to reduce their influence, while rarer and more distinctive words are up-weighted. The summation in equation (1) runs over all tokens in $\mathcal{D}_i$, so that repeated words are counted multiple times.

We extract latent policy domains via unsupervised topic modeling. Specifically, we apply Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan, 2003) and assign each speech to its highest-probability topic. Our preferred specification uses 20 topics, which map closely to major policy domains (e.g., schooling, labor, military affairs; Appendix Figure A.9).[34] Second, we construct two speech-level measures: emotionality and polarization.

Emotionality and Polarization (Escore and Pscore). To measure emotionality, we follow Gennaro and Ash (2022) and use the French LIWC dictionaries (2007 release) to define lists of *affective* and *cognitive* terms (Tausczik and Pennebaker, 2010).[35] We then compute inverse-frequency-weighted centroid vectors for the emotional and cognitive dictionaries, denoted $\mathbf{E}$ and $\mathbf{C}$, and define the emotionality score of speech i as

$$Escore_i \;=\; \frac{\text{sim}(\vec{d_i}, \mathbf{E}) + 1}{\text{sim}(\vec{d_i}, \mathbf{C}) + 1}, \tag{2}$$

where $\text{sim}(\cdot, \cdot)$ denotes cosine similarity. Higher values indicate language that is relatively closer to affective content than to cognitive content (Appendix Figure A.10).

To measure polarization, we adapt the methodology of Gennaro and Ash (2022) to capture partisanship in political speech during the French Third Republic. Because no exogenous dictionaries of ideologically charged language exist for this context, we construct left- and right-wing lexicons from our own data using speeches by MPs classified as left- or right-aligned. Moreover, since polarized language likely varies across policy domains and over time, we allow these lexicons to be year- and topic-specific. For each year-topic group g, we compute inverse-frequency-weighted centroid vectors of left- and right-wing discourse, $\mathbf{L}_g$ and $\mathbf{R}_g$, excluding the president of the chamber and members of

[34] We estimate the LDA using the `topicmodels` package with Gibbs sampling, following standard choices for priors and convergence diagnostics; see Appendix Section B.4 for full details.

[35] See Appendix B.5 for additional details on the LIWC dictionaries. We refine these dictionaries by expanding wildcards using French lemmas in the WordNet lexical database, removing overlaps across categories, trimming semantically distant terms, applying the same normalization and stemming as for the corpus, and removing stopwords.

government. For a speech i in group g, we compute a right-wing alignment score

$$Rscore_i \;=\; \frac{\text{sim}(\vec{d_i}, \mathbf{R}_g) + 1}{\text{sim}(\vec{d_i}, \mathbf{L}_g) + 1}, \tag{3}$$

and define the polarization score as the absolute deviation of this alignment from the median:

$$Pscore_i \;=\; \big|\, rscore_i - \text{median}(rscore) \,\big|. \tag{4}$$

This measure captures how far a speech departs from the median rhetorical position in that year-topic debate, whether toward the left or the right (Appendix Figure A.11).

Support for Secularization Policies (Proscore). In our empirical analysis, we provide a case study on the role of JeMPs in the parliamentary debates on the Law of Separation of churches and the State, widely regarded as one of the most contentious reforms of the French Third Republic and an ideal laboratory to investigate the role of moderators. These debates comprise 12,010 speeches delivered from February to July 1905 in around 60 distinct parliamentary discussions.

One key advantage of focusing on a single piece of legislation is that roll-call votes allow us to observe each MP's stance on the bill. We use the digitized roll-call votes on the 1905 Law of Separation.[36] These votes include decisions on specific paragraphs, on the full text of individual articles, and on the final version of the bill. For each of these acts, we coded the exact point of the parliamentary discussion where a vote on the bill occurred. This allows us to partition each discussion into segments defined by consecutive roll-call votes. We then assign to each roll call all speeches delivered in the corresponding segment.

We leverage this information to construct novel text-based measures of support for, and opposition to, secularization. First, we identify MPs who voted consistently in favor of the Secularization Law and MPs who voted consistently against it in both the first and final roll-call votes, thereby excluding MPs who changed position (and may have shifted their language) during the debate. We then build two dictionaries from the speeches of these pro-secularization and anti-secularization MPs and represent each dictionary by its centroid in the word-embedding space, denoted $\mathbf{P}$ (pro-law) and $\mathbf{A}$ (anti-law). For each speech $\mathcal{D}_i$ with embedding $\mathbf{d}_i$, we compute our *pro-secularization score* as

$$\text{Proscore}_i \;=\; \frac{\text{sim}(\vec{d_i}, \mathbf{P}) + 1}{\text{sim}(\vec{d_i}, \mathbf{A}) + 1}, \tag{5}$$

[36] Roll-call votes on the provisions that would define the Law of Separation began on April 16, 1905, when the Chamber approved Article 2, and concluded with the final vote on the full text of the bill on July 3, 1905. See Appendix A.4 for details on the voting data, and B.7 for details on the construction of the Proscore.

where sim denotes cosine similarity. In analogy with the emotionality and polarizationscores, this score captures how closely a speech aligns with the rhetoric of MPs supporting the law relative to the rhetoric of MPs opposing it.

In the context of the Separation Law case study, the smaller sample size also allows us to complement this measure with a large language model (LLM) assessment of each speech's pro-secularization.[37] Specifically, we asked the GPT4o-mini model (OpenAI) to score the intensity of support for the Separation Law on a five-point scale, with values -2 (strongly against), -1 (moderately against), 0 (neutral), 1 (moderately in favor), and 2 (strongly in favor), allowing for missing values in case of purely procedural or irrelevant interventions. To help the model interpret nuances and implicit references, we include in the prompt, alongside the target speech, up to four immediately preceding speeches from the same discussion, as well as a small set of illustrative examples. Figure A.12 in the appendix shows that the LLM coding is highly correlated with our Word2Vec-based Proscore defined in equation (5).

3.6 Aggregate Patterns

Figure 4 shows the share of JeMPs among members of the Chamber of Deputies. After the first JeMP in our dataset enters Parliament in 1876, the share of JeMPs increases steadily until 1902 and continues to rise until 1920, albeit at a more moderate pace. Following a peak in 1919, corresponding to the legislative elections, the share of JeMPs in Parliament declines, although it remains positive until the dissolution of the Third Republic in 1940. The growing numerical importance of JeMPs reflects the expansion of Jesuit schools in France and the evolution of total enrollments shown in Figure 1, with a lag due to the time between entry into the school system and eventual election to Parliament.

How did future JeMPs perform in school compared to their fellow Jesuit students? Our data allows us to construct two measures of student achievement: the probability of repeating a grade and the highest grade attained during their school career. Figure A.13 shows that Jesuit students who later became MPs have better school performance than other Jesuit students according to both metrics. Thus, JeMPs represent the brightest students from already highly selective educational institutions.

Figure A.14 shows the distribution of deputies across political areas (recall that this information is not systematically available for senators). Consistent with their religious-conservative educational background, JeMPs are overrepresented among center-right and independent MPs, and they are underrepresented among left-wing and center-left MPs. Figure A.15 shows that the distribution of deputies across occupational categories is remarkably similar for JeMPs and other MPs. Legal professions are the most common among parliamentarians, accounting for about 30% of both JeMPs and other MPs.

[37] Extending this exercise to the full sample of more than one million speeches would be computationally costly and time-intensive. We therefore leave the application of the LLM-based measure to the full corpus to future revisions.

Figure 5 compares the average emotionality score of JeMPs and other MPs across political areas. Two main facts stand out. First, across parties, MPs from the center and center-left employed a less emotional language, while those on the left wing used by far the most emotional language. Second, JeMPs used significantly less emotional language across all political parties, with the exception of left-wing parties: among those – who were already considerably emotional – the relatively few JeMPs on the left (11 in total) were even *more* emotional.[38] Given their limited number, their unusual political alignment relative to their educational background, and their atypical levels of emotionality, we treat left-wing JeMPs separately and include a dummy for them in our baseline controls.

Figure A.16 plots the average emotionality and polarization measures over time, showing that there are no major trends between 1881 and 1914. Figure A.17 shows that the lion's share of the variation in the emotionality and polarization arises within discussions rather than across discussions. To illustrate this, the figure sorts speeches into four quartiles, with the highest quartiles reflecting the most emotional/polarized discussions. As illustrated in the figure, the variation within each discussion quartile is significantly larger than the differences between the quartiles, suggesting that emotionality and polarization can fluctuate widely within discussions. We find a very similar pattern when we instead classify speeches according to the average emotionality/polarization of each speaker: Figure A.18 shows that the majority of variation occurs within, rather than accross speakers. These findings motivate our focus on within-speaker and within-discussion variation in the event study below.

4 Empirical Results

In this section, we examine how Jesuit-educated MPs differed in their speech patterns and how they affected other speakers during parliamentary debates. Our sample consists of all speeches that took place in France's Chamber of Deputies between January 11th, 1881 and March 4th, 1914.[39] More than 85% of these speeches are delivered by elected deputies, some of whom may simultaneously

[38]For a few left-wing JeMPs, we have access to detailed biographical information. Their cases suggest political trajectories that differed sharply from the broader JeMP pattern and from the religious and conservative orientation of the schools in which they were educated. The biographies of left-wing JeMPs Arthur-Vincent Delmas, Gabriel Baron, and Alfred Massé place them firmly in the republican, socialist, and anticlerical camp. Delmas, affiliated with the radical left, campaigned as a républicain radical for a "politique de progrès et de liberté" and explicitly endorsed both the denunciation of the Concordat and the rupture of the link between Church and State. Gabriel Baron was described as a "militant socialiste" fighting for the refonte of the social order and the suppression of financial privileges; in Parliament he supported the Combes ministry. Alfred Massé, for his part, ran on a radical-socialist platform that included progressive taxation, workers' pensions, and the separation of Church and State. Taken together, these biographies show that some Jesuit-educated deputies became prominent actors of the republican left and of the secularizing politics of the early twentieth century.

[39]There are two main reasons for focusing on the Chamber of Deputies. First, the Chamber of Deputies constituted the central arena of democratic political life under the Third Republic (Mayeur, 1984). Second, and most importantly, we observe only 29 JeMPs in the Senate. We begin our analysis in 1881 because the parliamentary transcripts available on Gallica from that year onward provide a consistent and structured record of debates. We end our analysis in 1914, before the outbreak of World War I, a major discontinuity in both parliamentary debates and political life more broadly.

hold institutional roles such as President of the Chamber, government ministers (including the Prime Minister), or rapporteurs on specific laws.[40] In the analysis below, we standardize all continuous measures, so that all coefficients can be interpreted in units of standard deviations.[41]

4.1 Cross-Sectional Results: Jesuit-Educated MPs Debate Differently

We begin by examining differences in rhetorical style between JeMPs and other legislators. We use the following specification:

$$y_{ist} = \beta_0 + \beta_1 \text{ JeMP}_s + \gamma X_{ist} + \delta_t + \varepsilon_{ist}, \tag{6}$$

where y_{ist} is our measure of emotionality or polarization for speech i delivered by speaker s at date t, and JeMP_s indicates whether the speaker was educated in a Jesuit school. The vector X_{ist} includes numerous controls that we group as follows: *Baseline Controls* include dummies for the President of the Chamber, members of government, rapporteurs, for procedural speeches, for speeches that remain empty after pre-processing, for left-wing JeMPs (as explained in Section 3.6), and for discussions with at least one JeMP speaker; *Speech Length* includes dummies for quartiles of speech length (measured by total word count); *Topic FE* assign a dummy from 20 topics to each speech, where topics are identified using a Latent Dirichlet Allocation algorithm (see Appendix B.4 for details); *Speaker Characteristics* comprise dummies for the speaker's age (in 5-year intervals) and for 12 occupation categories (as listed in Appendix A.3); and *Speaker Politics* includes dummies for the department in which the speaker was elected, as well as dummies for the speaker's political area (left, center-left, center, center-right, right, independent). In addition, we always include year fixed effects δ_t to absorb common shifts in rhetoric over time. Finally, ε_{ist} is the error term. Our coefficient of interest, β_1, reflects the average gap in emotionality or polarization between JeMPs and other legislators, conditional on the rich set of observable characteristics.

Emotionality. Table 1 presents our results for the (standardized) emotionality score of speech i, showing that JeMPs use systematically less emotional language than other MPs. Across all specifications, the coefficient β_1 is negative and statistically highly significant. In the baseline specification (col 1), JeMP speeches are about 5% of a standard deviation less emotional. The estimated gap is essentially unchanged when we additionally control for speech length (col 2), and it persists, although slightly

[40] The remaining speeches are delivered either by members of the government elected to the Senate, by unelected members of the government (as is often the case, for instance, for the Ministers of War or the Navy), or by speakers whom we are unable to match to any of these categories. Members of the government – whether elected or not – could speak in both chambers in their ministerial capacity.

[41] As noted above, some speeches – around 12% – remain empty after our pre-processing routine (usually because they were very short to start with). To preserve the original sequence of speeches without gaps, we assign these observations a "neutral" standardized value of zero and include a dummy variable for these speeches in all regressions.

attenuated, even after adding topic fixed effects (col 3), fixed effects for the speaker's age and occupation (col 4), and fixed effects for political affiliation and department of election (col 5).[42] The result in col 5 implies that even within the same occupation, age group, political party, and within the same topic of a speech, JeMPs were still significantly less emotional – by about 2% of a standard deviation.

Polarization. Table 2 reports the analogous specifications for the polarization score and shows that JeMPs use systematically less polarized language than other MPs. The JeMP coefficient is negative and statistically significant across all specifications. The most demanding specification in col 5 shows that the polarization score of JeMPs is 1.2% of a standard deviation lower as compared to other MPS within the same occupation category, age group, political party, and within the same topic of a speech.

Overall, Tables 1 and 2 establish a robust pattern: JeMPs adopt a systematically less emotional and less polarized rhetorical register, even after controlling for a host of observable characteristics. This evidence aligns with the distinctive environment to which JeMPs were exposed during their formative years in Jesuit schools. As discussed in the historical background (Section 2), Jesuit education combined rigorous training in philosophy and rhetoric with structured disputation – explicitly training students to engage opposing viewpoints constructively – and emphasized the integration of theology and scientific inquiry. To further corroborate this mechanism, we examine whether JeMP speeches are differentially framed around religion and science.

Proximity to Religion and Science. We use our word embeddings to measure the proximity of each speech to the semantic fields associated with the concepts "religion and "science."[43] Figure 6 displays word clouds of the terms most closely associated with these concepts in our corpus. Terms closely related to religion in parliamentary speeches include "christian," "catholic," and "belief," whereas those related to science include difference scientific fields such as "chemistry," "biology," and "sociology." We then employ the proximity measures of MP speeches as dependent variables in regression (6). Table 3 examines whether JeMPs differ from other MPs in referring to religion (cols 1-3) and science (cols 4-6) in their speeches. For religion, we find ambiguous results: JeMPs refer slightly more to religion when we only include our standard set of controls (col 1). However, this appears to

[42]The number of observations drops from col 3 to col 4 in Tables 1, 2, and 3 because some speeches could not be matched to an elected member of parliament and therefore have missing biographical information. The sample shrinks further in col 5 because some government members who spoke in the Chamber were in fact elected to the Senate. As ministers or heads of government, these individuals could speak in both chambers, and we therefore observe some of their speeches in the Chamber of Deputies. However, because they were serving a term in the Senate, their personal and political information comes from the Senate biographical records. In these records, age and occupation are available, whereas detailed political affiliation is coded only for members of the Chamber of Deputies. We did not code this variable for senators as information on the political party is not systematically available for senators.

[43]This measure of proximity is the cosine similarity between each vectorized document (see equation 1) and the word vectors corresponding to the "religion" and "science" in the word embedding space. Almelhem, Iyigun, Kennedy, and Rubin (2026) use topic analysis to detect religious and scientific language in a corpus of English works published between 1500 and 1900. In our corpus of parliamentary speeches, topic analysis reveals broad policy domains, see Appendix B.4.

be driven by their greater participation in debates in which religious issues were particularly salient, as the coefficient becomes zero when we control for topic fixed effects in col 2. When we add further speaker-specific characteristics and party affiliation (col 3), the JeMP coefficient becomes significantly *negative*. This suggests that within their political party, JeMPs referred less to religion than their fellow party members of similar background and age. In contrast, we find a coherent pattern of JeMPs using more scientific language across all specifications. One interpretation of these patterns is that Jesuit-educated MPs deliberately framed arguments in scientific and rational terms that carried broader legitimacy in an increasingly secular and polarized political environment. Compared to their closest peers, they also avoided explicit religious appeals in parliamentary debate. Rather than mobilizing confessional rhetoric, they may have translated religious or moral commitments into universalistic, evidence-based arguments. This rhetorical strategy would be consistent with their moderating influence in debate, as secular and scientifically framed language may have lowered the risk of partisan escalation and facilitated engagement across ideological lines.

Lexical Engagement with Preceding Speakers. Finally, Table 4 examines whether JeMPs differed from other MPs in the extent to which they engaged with preceding speakers. The dependent variable is the Jaccard Index, a measure of lexical overlap, between each speech and the preceding speech.[44] Across all three specifications – baseline controls, topic fixed effects, and additional speaker-level controls – we find that JeMPs were significantly more likely to lexically engage with the immediately preceding intervention. This pattern holds within parties and conditional on debate topics, indicating that JeMPs were systematically more engaged in the language of ongoing exchanges than their fellow legislators. By explicitly acknowledging and addressing prior arguments, JeMPs may have helped transform debates from parallel monologues into more structured interaction, thereby limiting rhetorical escalation. We turn next to examine whether these cross-sectional differences in JeMPs' own interventions also led to measurable reductions in emotionality and polarization among *subsequent* speakers.

4.2 Event Studies: Jesuit-Educated MPs Influenced Political Debates

Having established that JeMPs adopt a more rational, neutral, and science-oriented rhetorical style, we next ask whether this distinctive register can shape the dynamics of parliamentary debates, reducing their emotional tone and intensity of partisan conflict. To address this question, we implement an event-study design that exploits the timing of the first JeMP intervention within a discussion.

[44] More generally, the Jaccard Index measures the similarity between two sets and is defined as the size of their intersection divided by the size of their union. In our context, it equals the number of unique words shared by two speeches divided by the total number of unique words across both speeches. Because the measure is mechanically normalized by the combined vocabulary size of the two speeches, it is independent of speech length; we therefore omit speech length as a control in Table 4. The Jaccard Index should be interpreted as capturing lexical overlap with the immediately preceding speech, rather than as a direct measure of deeper argumentative engagement, persuasion, or substantive agreement.

Our main specification includes speaker $\times$ discussion fixed effects, which absorb all determinants of rhetorical style that are fixed for a given speaker within a given discussion. As a result, identification of the event-time coefficients comes from comparing whether a given speaker becomes less emotional or less polarized after a JeMP intervention than before it, within the same discussion and relative to that speaker's own fixed effect in that discussion. We use the following specification, examining the event window of $\pm$ 5 speeches around the first JeMP intervention in a discussion:

$$y_{isd} = \sum_{\substack{k \in \{-5 \ldots +5\} \\ k \neq -1}} \beta_k D_{i,t-k} + \gamma X_{isd} + \alpha_{sd} + \varepsilon_{isd} \tag{7}$$

where y_{isd} is the emotionality or polarization score for speech i delivered by speaker s in discussion d, and $D_{i,t-k}$ is an indicator that speech i occurs k speech turns relative to t, the first JeMP intervention in discussion d. The control vector X_{isd} includes quartiles of speech length, an indicator for procedural speeches, an indicator for speeches that remain empty after pre-processing, event-time indicators relative to interventions by left-wing JeMPs (as explained in Section 2.4), as well as fixed effects for the speaker's speech count within a discussion (i.e., whether the speech is the speaker's 1st, 2nd, 3rd,...intervention) to flexibly absorb possible patterns in rhetoric over the course of a debate. Finally, the term α_{sd} denotes speaker $\times$ discussion fixed effects, and ε_{isd} is an error term. The key identification assumption is that the error term – i.e., the unobserved factors that affect the speech tone – are not systematically correlated with the timing of the JeMP entry. We discuss the plausibility of this assumption below, focusing in particular on pre-trends and the scope for strategic timing of JeMP interventions.

Our main objects of interest are the event-time coefficients β_k. The post-JeMP coefficients ($k \geq 0$) trace out the dynamic response of emotionality or polarization following JeMP intervention. The pre-JeMP coefficients ($k < 0$) serve as a falsification test, allowing us to assess whether rhetorical style was already changing prior to JeMP entry. We omit the dummy for $k = -1$ (the speech immediately preceding JeMP entry), so that all β_k are interpreted relative to that baseline.

Figure 7 reports the event-study estimates for the standardized emotionality score. The x-axis indexes speech order relative to the first JeMP intervention ($t = 0$), and the y-axis plots the estimated β_k coefficients from equation (7), along with 90% confidence intervals, relative to the omitted category $k = -1$. Three main patterns emerge from the figure. First, the JeMP intervention itself at t shows a particularly low emotionality score. Note that this is relative to the JeMP's own speaker-discussion fixed effect, suggesting that on average, the first JeMP intervention in a discussion is delivered in a particularly moderating tone.[45] Second, following the first JeMP intervention the emotional tone of

[45]Below, we extend the analysis by examining the effects of later JeMP interventions within the same discussion.

the parliamentary debate declines markedly over the next five speaking turns, reaching a trough of about 7 percent of a standard deviation at $t+4$. While emotionality begins to decline already at $t+1$, the post-JeMP coefficients are statistically significant at the 90% confidence level from $t+2$ through $t+4$. Third, all pre-JeMP coefficients are close to zero and do not exhibit any trend, indicating that JeMPs do not systematically intervene at moments when the emotional intensity of the debate is already trending upward or downward.

Figure 8 reports the analogous results for the polarization score. We again find a marked decline immediately after the JeMP intervention. The depolarizing effect is sizeable, with estimates ranging from 5 percent of a standard deviation at $t+1$ to 9 percent of a standard deviation at $t+5$. As in Figure 7, the pre-JeMP coefficients are close to zero and do not exhibit any discernible trend.

A natural concern with this design is that JeMPs may have entered debates strategically, rather than at quasi-random moments in the evolution of rhetorical tone. Reassuringly, there are no pre-trends in Figures 7 and 8. This helps alleviate the concern that JeMPs systematically intervened at moments when rhetorical temperature was already about to decline. We can address remaining concerns about strategic JeMP interventions with three additional arguments: First, speaking order was to a large extent predetermined by parliamentary procedure, limiting the scope for fine-grained strategic timing (see the discussion on speaker order in Section 2.4). Second, even if speakers had substantial discretion, it is not clear how JeMPs would have systematically forecast sudden changes in rhetorical tone at such short horizons, especially given the flat pre-trends and the rich set of controls included in the specification. In particular, because the specification includes speaker-discussion fixed effects, it is not possible for the estimated effects to be driven by JeMPs systematically timing their interventions to precede or follow speakers who are, on average, more or less emotional or polarized within a given debate. Such speaker-specific differences are absorbed by the fixed effects. Finally, even if JeMPs could time their interventions around sudden declines in rhetorical heat and had incentives to do so, rationalizing our results would still require that such timing ability or incentives be systematically stronger for JeMPs than for other MPs. Under this interpretation (which is itself unlikely, given the absence of pre-trends), the estimates would speak less to the causal effect of JeMP interventions on subsequent speeches, but they would still imply that JeMPs were distinctive in how they entered and shaped parliamentary debates.

Placebo Exercises. Another concern is that the estimated decline in emotionality and polarization following JeMPs interventions may not reflect a uniquely JeMP effect, but could rather arise mechanically from generic properties of low-emotional speakers. To address this concern, we implement three placebo exercises. The first asks whether a similar pattern would emerge for randomly selected MPs;

We find a very similar post-JeMP pattern, although the drop in emotionality at t is somewhat smaller for later JeMP interventions.

the second examines whether the results can be explained simply by interventions from systematically calm or unpolarized speakers; and the third assesses whether they may reflect constituency-level moderation or local political culture. In our first placebo test, we randomly sample 5% of MPs (excluding JeMPs) and use them as a treatment group in the event-study. We repeat this procedure 1,000 times and show the distribution of the event-time coefficients, for periods 0 to 5, in Figure 9 for emotionality and in Figure 10 for polarization. These figures show that the observed pattern of JeMP interventions is indeed unique, as our event study coefficients typically lie outside the distributions of coefficients obtained from the placebo exercises. Second, we estimate year-specific MP fixed effects from regressions of emotionality and polarization (excluding JeMPs). We then identify the 5% of MPs with the lowest fixed effects (i.e., those who are systematically the least emotional or least polarized speakers within a given year) and treat them as placebo JeMPs in the event-study framework. We show these results in Figure A.19 in the appendix. While these "calmest" MPs are themselves less emotional and polarized in their interventions, they do not trigger a decrease in emotionality or polarization of subsequent speakers. In other words, being individually less emotional or less polarized was not sufficient to shift the tone of subsequent speakers. This exercise suggests that JeMP interventions had a distinctive effect on the subsequent tone of debate, beyond what one would expect from speakers who were simply calmer or less polarized on average. Third, we examine whether features of the locations that elected individual MPs may be driving the results. We estimate department-of-election fixed effects for emotionality and polarization (excluding JeMPs). From those, we identify the 5% of departments with the lowest fixed effects and use the MPs from these departments as placebo JeMPs. Again, we do not find comparable post-intervention declines in this placebo (see Figure A.20), suggesting that the JeMP effect is not simply attributable to constituency-level moderation or local political culture.

Event Studies with Multiple JeMP Interventions. Our main event study focused on the first JeMP intervention. This is the cleanest exercise, as the subsequent evolution of the discussion – including later JeMP interventions – may itself be affected by the initial treatment. However, our results are robust to allowing for multiple JeMP treatment events within a discussion. To show this, we define treatment events in regression (7) as not only the first JeMP intervention in a discussion, but also any subsequent JeMP interventions that occur at least 30 speeches after the previous treatment.[46] To maintain the same identifying variation as in the main event-study (which had only one event window per discussion), we add speaker $\times$ event-window fixed effects to equation (7). Thus, we effectively compare the same speaker within the same 10-speech event discussion window, before and after a

[46]To illustrate, suppose that in a given discussion, JeMPs take the floor at speaking turns t_0, $t_0 + 10$, and $t_0 + 40$: we focus on the impact of JeMP speeches at t_0, as in the main event-study, and $t_0 + 40$; by contrast, we do not consider $t_0 + 10$ as a treatment event because it is too close to the previous treatment. Our results are similar when using alternative cutoffs such as 20 or 50.

JeMP takes the floor. We report our main result – the estimated β_k coefficients for emotionality – in Figure A.21, where the x-axis indexes speech order relative to a JeMP treatment event. The estimates confirm the main findings of our previous analysis: while no discernible trends are visible for the speaking turns leading up to the JeMP intervention, emotionality of the debate declines markedly. For polarization, the pattern is less pronounced than for the first JeMP intervention in our main results. This suggests that mitigating polarized partisan conflict may become more difficult further into the discussion.

4.3 Case Study: The 1905 Separation Law

The 1905 Separation Law is widely regarded as one of the most significant and polarizing pieces of legislation in modern French history. Building on earlier republican efforts to reorganize church-state relations, it made secularism the guiding constitutional principle of public life. The law guaranteed freedom of conscience and the free exercise of religion, while also establishing that the state would no longer recognize, fund, or support any religion. It further imposed strict rules on religious activity through the framework of the 1901 Law on Associations, requiring religious groups to operate as regulated associations. Finally, implementation involved a mandatory inventory of church property to distinguish assets deemed necessary for worship from those that could be confiscated. Importantly, the 1905 Separation Law was the final building block in an anti-clerical agenda that had already contributed to the closure of Jesuit schools through the 1901 Law on Associations and the 1904 "Loi Combes," which prohibited teaching by members of religious congregations and organized the liquidation of their assets. These features make the Separation Law an ideal setting to study the role of Jesuit-educated MPs in a highly polarized political environment.

To conduct this analysis, we focus on the discussions in the Chamber of Deputies specifically devoted to the Separation Law, which we identify using section titles. This sample comprises 12,010 speeches delivered in the Chamber of Deputies in 60 distinct discussions between February 2 and July 3, 1905, the date of the final vote on the full text of the bill. Around 60% of JeMPs voted against the law, as compared to 40% of other MPs. Furthermore, 80% of the JeMP speeches in this sample come from JeMPs who voted against the law, while this percentage is close to 50% for other MPs. This is not surprising: JeMPs had been educated in institutions that the incumbent left coalition had recently dismantled, giving them strong personal and political reasons to oppose the reform.

There are at least three advantages for focusing on France's Separation law. First, the debates on the Law of Separation provide a particularly demanding test of the moderation hypothesis. If any setting should have led JeMPs to take a stronger position, to speak in a more emotional tone, or to contribute to more polarized exchanges, it is this one. France's secularization laws directly targeted institutions tied to JeMPs' educational background, and they did so in the midst of a highly

salient ideological conflict. Second, by zooming into a single piece of legislation, we are able to hand-match roll-call votes to the individual debates, thereby linking each MP's recorded position to his parliamentary speech (see Section 3.5 for detail). Finally, this focus on a single law allows us to construct a *Proscore* index that measures the extent to which MPs supported (or opposed) this law (see Section 3.5). As the Proscore is constructed using exclusively roll-call votes and debates on the Separation Law, it provides a more precise measure of MPs' stance on secularization than the broader polarization index (which is based on all debates).

In our empirical analyses below we focus on JeMPs who opposed the Secularization Law, voting consistently against it in both the first and final roll-call votes.[47] We denote these by JeMP^{Opp}. This restriction further stacks the odds against finding moderating behavior, since these JeMPs not only had educational ties to institutions targeted by secularization, but also revealed a clear and consistent political opposition to the law. If moderation were to break down anywhere, one would expect it to do so for this group.[48]

Cross-Sectional Results. Table 5 presents our first set of results on the Separation Law. We report results with our baseline controls in Panel A and then add fixed effects for political area in Panel B. This allows us to assess the extent to which the results are driven by the fact that JeMP^{Opp} represented a more conservative political bloc – 50% of them belonged to the center-right, and another 49% to parties on the right, as compared to 24% and 7%, respectively, among all other MPs.[49] Column 1 uses the *Proscore* as the outcome variable in our baseline cross-sectional specification with JeMP^{Opp} as main explanatory variable. We find that even in this highly sensitive policy setting, and even when focusing on JeMPs who opposed secularization, they use relatively neutral language: the *Proscore* of their speeches does not differ significantly from that of the average MP in Panel A, with a relatively precisely estimated close-to-zero coefficient. The result is even more striking in Panel B: When we compare JeMP^{Opp} to MPs from the same political area (center-right and right), we find that they used a language that supported separation relatively *more*. In other words, within the subset of politicians from the right and center-right who overwhelmingly opposed the Separation Law, those with a Jesuit education used a more moderate language, with a relatively more pro-law orientation. Next, column 2 shows that JeMP^{Opp} do not differ from other MPs in terms of the Polarization Score (in both Panels A and B). Remarkably, column 3 shows that JeMP^{Opp} maintained a lower Emotionality Score than their colleagues even when debating a Law that they opposed and that would effectively dismantle

[47] This is the same criterion that we adopted to construct the reference groups for the *Proscore*. All JeMPs who voted against the full text of the law also voted against it in the first roll-call.

[48] Using all JeMPs instead yields similar results, since JeMP^{Opp} account for more than four-fifths of all JeMP speeches in this sample.

[49] All regressions include a dummy for the remaining JeMPs who did not oppose the Separation Law. Thus, coefficients compare JeMP^{Opp} to the average non-Jesuit educated MP.

their former schools. This emotionality gap is even larger when we include political area fixed effects in Panel B. Finally, JeMP^{Opp} who opposed the secularization law used a language that was closer to Religion and, especially, to Science (cols 4 and 5, in both Panels A and B). This pattern is especially interesting because, even in a setting where JeMPs naturally referred more to religion as opponents of secularization, their rhetoric still revolved strongly around science and rationality rather than around an exclusively religious mode of argument. Taken together, these findings suggest that the debates on the Law of Separation do not overturn the broader moderation pattern documented above. If anything, they strengthen it: Even when focusing on JeMPs who opposed the reform, and in a setting where one might have expected them to show stronger opposition, greater emotionality, and more polarized rhetoric, their speeches remain relatively neutral, emotionally restrained, and no more polarized than those of other MPs.

Within-Discussion Results. Next, we examine whether the comparatively moderate rhetorical style of JeMPs in the Separation Law debates influenced the interventions of other MPs within the same discussion. We perform a similar empirical exercise to the event study described in Section 4.2. Given the limited number of discussions, we introduce two departures from our main specification in equation (7). First, we allow for multiple JeMP^{Opp} treatment events within the same discussion, provided that these 'event windows' are at least 30 speaking turns apart. We thus use a similar specification as in the robustness event study for the full sample reported in Figure A.21. Second, in a difference-in-differences spirit, we compare the five speeches following a JeMP^{Opp} intervention to the five preceding speeches. Our specification includes speaker $\times$ discussion fixed effects, so that identification comes from within-speaker changes in language within the same discussion, before and after a JeMP^{Opp} intervention.[50]

As described in Section 3.5, we use two complementary methodologies to measure a speech's stance toward secularization policies. First, we use our embedding-based Proscore, where higher values indicate that a speech is semantically closer to those of MPs who support the law. Second, in the context of the debates on the Law of Separation, the more limited sample size allows us to use a large language model (ChatGPT4o-mini) to classify the pro-secularization stance of each speech. This classification is based on a five-point scale ranging from -2 (strong opposition) to +2 (strong support), and takes into account the four preceding speeches within the same discussion. Thus, the Proscore captures semantic similarity to fixed reference points across the full vocabulary, while the LLM-based measure provides a context-aware classification of each speech's stance, capturing richer semantic and contextual nuances.

[50]The JeMP intervention itself has its own dummy (as in our main event study in Section 4.2). This dummy is not reported here. Also, we obtain similar results when excluding the 20% of JeMP speeches from those JeMPs who did not consistently oppose secularization.

Column 1 of Table 6 shows that JeMP^{Opp} interventions reduce average support for the Law of Separation. The effect is sizable: the average Proscore of the five subsequent speeches declines by about 13% of a standard deviation. Columns 2–5 further examine whether the effect operates at different points of the distribution of pro-secularization stance. In particular, we ask whether JeMPs^{Opp} make subsequent speakers even more opposed to the law than they already were, or whether they attenuate the most fervent expressions of support, thereby contributing to a depolarization of the debate. To address this question, we construct two indicator variables for whether a speech falls in the top or bottom decile of the Proscore distribution, as well as two indicators for whether the LLM classifies the speech as strongly opposed to or strongly in favor of the Separation Law.[51] Columns 2 and 3 report the results based on the embedding measure, while columns 4 and 5 report the results based on the LLM classification. Both approaches yield a consistent picture. Following JeMP^{Opp} interventions, speeches are less likely to express strong support for the Separation Law (columns 2 and 4), whereas they are not significantly more likely to express strong opposition (columns 3 and 5).

These findings have broad implications. First, they show that the JeMPs' distinctive rhetorical style was able to influence political debate on a major and highly contentious piece of legislation, shifting other speakers' language closer to their own positions. Second, they suggest that JeMP^{Opp} arguments were primarily directed at their adversaries, echoing the epigraph from Ignacio de Loyola. They were able to moderate the strongest supporters of secularization without fueling fervent opposition. Recall that our cross-sectional analysis in Table 5 showed that JeMPs^{Opp} tend to use less emotional language, and rely more on rational and scientific argumentation; furthermore, in the specific context of the Law of Separation, their language was also comparatively neutral, despite their opposition to the reform. One interpretation of these results is therefore that JeMPs^{Opp} were able to frame their arguments in a way that was perceived as legitimate by the opponents of the reform. This is particularly striking in a political context characterized by intense conflict between the secular and progressive camp and the more conservative factions of Parliament. Now, if JeMPs^{Opp} interventions were able to shape the language used by other speakers, a natural question arises: does this impact remain confined to rhetoric, or does it also translate into voting behavior?

Voting Behavior. Finally, we examine whether the effect of JeMP interventions on the rhetorical style of subsequent speakers is accompanied by a shift in voting behavior, again focusing on JeMPs who consistently voted against the secularization reform (JeMP^{Opp}). We use the digitized roll-call votes on the 1905 Law of Separation. Roll-call votes on the provisions that would define the Law of Separation began on April 16, 1905, and concluded with the final vote on the full text of the bill on July 3, 1905. We focus on the subset of 18 acts for which left-wing parliamentarians were, on average, at least 25%

[51] The LLM classifies approximately 10% of speeches as strongly in favor of the law and 10% as strongly opposed to it. Thus, our LLM-based analysis is coherent with our focus on the top and bottom deciles of the Proscore distribution.

more likely than right-wing parliamentarians to vote in favor. This restriction helps us ensure that a recorded "for" vote reflects *support* for the overall agenda of the Separation Law.[52]

In attempting to identify the impact of JeMP$^{\text{Opp}}$ speeches on the voting behavior of other parliamentarians, we face an empirical challenge. While speeches occur sequentially, allowing us to precisely distinguish those that precede and follow a treatment intervention, votes on a given legislative act are cast simultaneously at the end of a segment of the debate. As a result, our sharp event study design from Section 4.2 is not viable when the dependent variable is the MP's vote. Instead, we focus on three measures of JeMP$^{\text{Opp}}$ speech engagement ($JeMP^{\text{Opp}}\ Speech_{\ell}$) during the discussion of parliamentary acts ℓ. First, we use the share of JeMP$^{\text{Opp}}$ speeches relative to the overall number of speeches in the debate of an act. Second, in the spirit of exploiting variation in the *timing* of the JeMP$^{\text{Opp}}$ interventions across roll-call votes, we construct an index of "JeMP$^{\text{Opp}}$ closeness to the vote," ranging from 0 to 1, which captures how close to the final vote JeMPs$^{\text{Opp}}$ engage in the debate.[53] Third, we also compute an alternative measure: rather than using the average order of all JeMP$^{\text{Opp}}$ contributions, "Last JeMP$^{\text{Opp}}$ Closeness to Vote" uses only the last JeMP$^{\text{Opp}}$'s intervention before the roll-call vote, computing its proximity to the vote (again, between 0 and 1).

We then estimate the following specification, focusing only on the voting behavior of non-Jesuit educated MPs:

$$Vote_{s,\ell} = \beta_0 + \beta_1 JeMP^{\text{Opp}}\ Speech_{\ell} + \alpha_s + \varepsilon_{s,\ell} \tag{8}$$

where $Vote_{s,\ell}$ is a dummy that takes value one if non-JeMP parliamentarian s voted in favor of legislative act ℓ, and $JeMP^{\text{Opp}}\ Speech_{s\ell}$ represents our three measures of JeMP$^{\text{Opp}}$ engagement during the parliamentary session that debated act ℓ. Importantly, we include MP fixed effects α_s, so that the identifying variation is within-MP voting behavior across bills characterized by different levels of JeMP$^{\text{Opp}}$ engagement. Finally, $\varepsilon_{s,\ell}$ is an error term.

Column 1 of Table 7 shows that a higher share of JeMP$^{\text{Opp}}$ speeches in a discussion segment is associated with a lower share of MPs voting in favor of a pro-secularization act. The estimated coefficient implies that a one percentage point increase in the share of JeMP$^{\text{Opp}}$ speeches reduces the probability of a favorable vote by 0.025 percentage points. This result is confirmed – and, if anything, stronger – when we restrict attention to discussion segments with at least one JeMP$^{\text{Opp}}$ speaker (column 2). A natural concern is that this relationship may partially reflect reverse causality: opposing JeMPs may have been more active prior to roll-call votes where support for the bill was

[52] We adopt this procedure because the political objective of a roll call is often difficult to infer from its title alone (e.g., "On the additional article by M. Allard"), and the full text of the proposed act is not consistently available in the parliamentary records. See Appendix A.4 for details.

[53] Specifically, we compute the average speaking-turn position of JeMP$^{\text{Opp}}$ within each segment and normalize it by the total number of speeches. Hence, a low value of "JeMP$^{\text{Opp}}$ closeness to the vote" corresponds to JeMP$^{\text{Opp}}$ interventions occurring mostly at the beginning of the discussion.

expected to be lower, for instance because such votes were perceived as more contested (recall that on average, about 60% of MPs voted in favor of secularization, so that acts with lower support were closer to the 50% threshold). Reassuringly, however, our second measure of JeMP^{Opp} engagement – capturing the timing of JeMP^{Opp} interventions – is also negative and statistically significant, both when using the average timing (column 3) and the timing of the last intervention (column 4). These patterns suggest that, even conditional on the overall share of JeMP^{Opp} speeches, JeMPs^{Opp} were more effective at influencing voting behavior when they spoke closer to the final vote.

Taken together, these results suggest that the moderating role of JeMP^{Opp} was not confined to the rhetorical tone of parliamentary debates, but may also have extended to voting behavior. In particular, JeMP^{Opp} interventions are associated with lower support for pro-secularization acts among other MPs, especially when those interventions occurred closer to the vote. While this evidence is necessarily more indirect than our speech-level results, it is consistent with the view that JeMPs helped shape not only the tone of the debate, but potentially also its legislative consequences.

5 Conclusions

We examined whether certain types of political elites can moderate polarization within legislative arenas. Using newly digitized archival records on Jesuit student enrollment – linking them to the universe of French parliamentary speeches and roll-call votes during the Third Republic – we showed that Jesuit-educated MPs employed systematically less emotional and less polarized rhetoric and influenced subsequent speakers to adopt more moderate language. These results persisted under stringent fixed effects specifications, while they were absent in a range of placebo exercises, suggesting that our findings reflect distinctive characteristics of JeMPs rather than generic features of particular legislators or electoral environments. Even during the highly contentious debates surrounding the 1905 separation of Church and State, JeMPs used comparatively neutral rhetoric and dampened extreme discourse among subsequent speakers.

Our findings contribute to the literature on political polarization by shifting attention from institutional or party-level drivers to heterogeneity within the political elite. Much of the existing literature explains polarization through electoral competition, party sorting, or institutional incentives. While these forces undoubtedly shaped aggregate trends, our results suggested that polarization within legislatures was not solely an equilibrium outcome of institutional constraints. Instead, it could be moderated from within, through the presence and participation of individuals with distinctive backgrounds and interaction styles.

Beyond these contributions, our findings point to two broader implications. First, they suggest that polarization within legislatures may be a dynamic and path-dependent process. The spillover effects that we document indicate that the tone of debate can shift depending on who intervenes and

when. If certain legislators systematically de-escalate rhetoric, early interventions by such actors may move discussions onto lower-temperature trajectories, whereas the absence of moderating voices may allow emotional escalation to persist. This perspective highlights the importance of agenda control, speaking order, and procedural rules in shaping polarization dynamics, and suggests that relatively small changes in elite composition or participation could generate disproportionate effects on deliberative outcomes.

Second, our results raise the possibility that rhetorical moderation functions as a form of elite self-restraint that generates positive externalities within democratic institutions. By tempering the tone of debate, moderating legislators may improve the quality of deliberation not only through their own speeches but also by influencing others. This perspective resonates with arguments emphasizing the importance of informal "norms of forbearance" and mutual toleration for democratic stability (Levitsky and Ziblatt, 2018), as well as with theories that view democracy as sustained by self-enforcing limits among political actors (Weingast, 1997). Yet, the incentives to supply such restraint may be limited, particularly in polarized environments where partisan signaling and mobilization are electorally rewarded. If rhetorical moderation constitutes a form of institutional forbearance, then its provision may depend critically on elite composition and formation, suggesting that the selection and training of political leaders have first-order implications for the resilience of democratic institutions.

References

Almelhem, A., M. Iyigun, A. Kennedy, and J. Rubin (2026, 02). Enlightenment Ideals and Belief in Progress in the Run-up to the Industrial Revolution: A Textual Analysis. *Quarterly Journal of Economics 141*(1), 263–314.

Bäck, H., R. Carroll, E. Renström, and A. Ryan (2023). Elite Communication and Affective Polarization Among Voters. *Electoral Studies 84*, 102639.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics 131*(4), 1593–1636.

Becker, S. O., J. Rubin, and L. Woessmann (2024). Religion and growth. *Journal of Economic Literature*.

Belloc, M., F. Drago, and R. Galbiati (2016). Earthquakes, religion, and transition to self-government in italian cities. *Quarterly Journal of Economics 131*(4), 1875–1926.

Besley, T., J. G. Montalvo, and M. Reynal-Querol (2011). Do Educated Leaders Matter? *The Economic Journal 121*(554), F205–F227.

Bisin, A., J. Rubin, A. Seror, and T. Verdier (2024). Culture, institutions and the long divergence. *Journal of Economic Growth 29*(1), 1–40.

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

Bó, E. D., F. Finan, O. Folke, T. Persson, and J. Rickne (2017). Who Becomes a Politician? *Quarterly Journal of Economics 132*(4), 1877–1914.

Bonica, A. (2014). Mapping the Ideological Marketplace. *American Journal of Political Science 58*(2), 367–386.

Bonnard, R. (1926). *Les règlements des assemblées législatives de la France depuis 1789: notices historiques et textes*. Paris: Société anonyme du Recueil Sirey.

Bush, J. W. (1975). Education and Social Status: The Jesuit College in the Early Third Republic. *French Historical Studies 9*, 125–140.

Cagé, J. and T. Piketty (2025). *A History of Political Conflict*. Cambridge, MA: Harvard University Press.

Canen, N., C. Kendall, and F. Trebbi (2020). Unbundling Polarization. *Econometrica 88*(3), 1197–1233.

Canen, N., C. Kendall, and F. Trebbi (2026). Political Parties as Drivers of U.S. Polarization: 1927–2018. *Journal of Political Economy*. Forthcoming.

Cantoni, D., Y. Chen, D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2017). Curriculum and Ideology. *Journal of Political Economy 125*(2), 338–392.

Cantoni, D., J. Dittmar, and N. Yuchtman (2018). Religious competition and reallocation: The political economy of secularization in the protestant reformation. *Quarterly Journal of Economics 133*(4), 2037–2096.

Coste, A. (1890). étude statistique sur les salaires des travailleurs et le revenu de la France. *Journal de la Société de Statistique de Paris 31*, 225–240.

Davin, J.-P. (1985). Les Groupes Parlementaires. *Pouvoirs 34*, 23–29.

Delattre, P. (Ed.) (1949). *Les Établissements des Jésuites en France depuis quatre siècles: Répertoire topo-bibliographique publié à l'occasion du quatrième centenaire de la fondation de la Compagnie de Jésus, 1540–1940*, Volume 1. Enghien and Wetteren, Belgium: Institut supérieur de Théologie and Imprimerie De

Meester Frères.

Diermeier, D., J.-F. Godbout, B. Yu, and S. Kaufmann (2012). Language and Ideology in Congress. *British Journal of Political Science 42*(1), 31–55.

Doornbosch, L. M., M. van Vuuren, and M. D. de Jong (2025). Moving Beyond Us-Versus-Them Polarization Towards Constructive Conversations. *Democratization 32*(1), 96–122.

Downey, M., J. Kuk, and M. Nanes (2024). (Not) Feelin' the (Cross) Pressure: Experimental Evidence on Issue Misalignment and Polarization. Working paper.

Doyle, A. (2017). Catholic Church and State Relations in French Education in the Nineteenth Century: The Struggle between Laïcité and Religion. *International Studies in Catholic Education 9*(1), 108–122.

Dreher, A., M. J. Lamla, S. M. Lein, and F. Somogyi (2009). The Impact of Political Leaders' Profession and Education on Reforms. *Journal of Comparative Economics 37*(1), 169–193.

Farrell, A. P. (1970). *The Jesuit Ratio Studiorum of 1599: Translated into English, with an Introduction and Explanatory Notes*. Washington, D.C.: Conference of Major Superiors of Jesuits.

Fiorina, M. P. and S. J. Abrams (2008). Political Polarization in the American Public. *Annual Review of Political Science 11*, 563–588.

Fishkin, J. S., A. Siu, L. Diamond, and N. Bradburn (2021). Is Deliberation an Antidote to Extreme Partisan Polarization? Reflections on "America in One Room". *American Political Science Review 115*(4), 1464–1481.

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

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as Data. *Journal of Economic Literature 57*(3), 535–574.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica 87*(4), 1307–1340.

Grendler, P. F. (1989). *Schooling in Renaissance Italy: Literacy and Learning, 1300–1600*. Baltimore, MD: Johns Hopkins University Press.

Habermas, J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge, MA: MIT Press.

Harrigan, P. J. (1974). The Social Appeals of Catholic Secondary Education in France in the 1870s. *French Historical Studies 8*, 443–468.

Harrigan, P. J. (2001). Church, State, and Education in France from the Falloux to the Ferry Laws: A Reassessment. *Canadian Journal of History 36*, 51–83.

Henríquez, P. A., J. Sabat, and J. P. Sullivan (2023). Politicians' Willingness to Agree: Evidence from the Interactions in Twitter of Chilean Deputies. *Journal of Information Technology & Politics 20*(1), 92–111.

Langdon, J. W. (1978). The Jesuits and French Education: A Comparative Study of Two Schools, 1852–1913. *History of Education Quarterly 18*(1), 49–60.

Lanzara, G., S. Lazzaroni, P. Masella, and M. P. Squicciarini (2024). Economic Shocks and Assimilation Policies: Phylloxera and Educational Expansion in French Algeria. BAFFI CAREFIN Centre Research Paper 221.

Levitsky, S. and D. Ziblatt (2018). *How Democracies Die*. New York: Crown.

Loyola, I. o. (1914). *The Spiritual Exercises*. Christian Classics Ethereal Library. Accessed 27 Feb. 2026.

Mayeur, F. (1981). *Histoire Générale de l'Enseignement et de l'Éducation en France: Tome III. De la Révolution à l'École Républicaine (1789–1930)*. Paris: Nouvelle Librairie de France.

Mayeur, J.-M. (1984). *La vie politique sous la Troisième République*. Points Histoire. Paris: Éditions du Seuil.

McCarty, N., K. T. Poole, and H. Rosenthal (2006). *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, MA: MIT Press.

Mikolov, T., I. Sutskever, K. Chen, G. Corrado, and J. Dean (2013). Distributed Representations of Words and Phrases and Their Compositionality. *arXiv preprint arXiv:1310.4546*.

Milligan, K., E. Moretti, and P. Oreopoulos (2004). Does Education Improve Citizenship? Evidence from the United States and the United Kingdom. *Journal of Public Economics 88*(9–10), 1667–1695.

Moskowitz, D. J., J. C. Rogowski, and J. M. S. Jr. (2024). Parsing Party Polarization in Congress. *Quarterly Journal of Political Science 19*(4), 357–385.

O'Malley, J. W. (1993). *The First Jesuits*. Cambridge, MA: Harvard University Press.

Porter, M. (2009). The English (Porter2) Stemming Algorithm. http://snowball.tartarus.org/algorithms/english/stemmer.html.

Rubin, Jared (2017). *Rulers, Religion, and Riches: Why the West Got Rich and the Middle East Did Not*. Cambridge University Press.

Schwickerath, R. (1904). *Jesuit Education: Its History and Principles Viewed in the Light of Modern Educational Problems* (Second ed.). St. Louis, MO: B. Herder.

Squicciarini, M. P. (2020). Devotion and Development: Religiosity, Education, and Economic Progress in Nineteenth-Century France. *American Economic Review 110*(11), 3454–3491.

Tausczik, Y. R. and J. W. Pennebaker (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology 29*(1), 24–54.

Thompson, D. G. (2018). The Jesuit Province of France on the Eve of its Destruction in 1762. *Archivum Historicum Societatis Iesu 87*(173).

van Kley, D. K. (1975). *The Jansenists and the Expulsion of the Jesuits from France, 1757–1765*. Yale Historical Publications: Miscellany. New Haven: Yale University Press.

Vogel, C. (2010). The Suppression of the Society of Jesus, 1758–1773. European History Online (EGO).

Weingast, B. R. (1997). The Political Foundations of Democracy and the Rule of Law. *American Political Science Review 91*(2), 245–263.

TABLES

Table 1: Emotionality of JeMP Speeches

Dependent Variable: *Emotionality Score of Parliamentary Speeches*

	(1)	(2)	(3)	(4)	(5)
JeMP	-0.051***	-0.047***	-0.032***	-0.028***	-0.020***
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Baseline Controls	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Speech Length		✓	✓	✓	✓
Topic FE			✓	✓	✓
Speaker Characteristics				✓	✓
Speaker Politics					✓
R^2	0.05	0.06	0.29	0.29	0.30
Observations	1,128,866	1,128,866	1,128,866	1,007,382	983,727

Notes: The table shows that Jesuit-educated MPs (JeMPs) had significantly lower levels of emotionality in their parliamentary speeches. Regressions are run at the speech level. *Baseline Controls* include dummies for the President of the Chamber, members of government, rapporteurs, for procedural speeches, for left-wing JeMPs, and for discussions with at least one JeMP speaker. *Year FE* are fixed effects for the year during which a speech occurred. *Speech Length* are dummies for quartiles of speech length (measured by total word count); *Topic FE* are dummies for 20 topics, where topics are identified using a Latent Dirichlet Allocation algorithm (see Appendix B.4 for details); *Speaker Characteristics* are dummies for the speaker's age (in 5-year intervals) and for 12 occupation categories (as listed in Appendix A.3); and *Speaker Politics* includes dummies for the department in which the speaker was elected, as well as dummies for the speaker's political area (left, center-left, center, center-right, right, independent). Robust standard errors are reported in parentheses. Notation for statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Partisan Polarization of JeMP's Speeches

Dependent Variable: Polarization Score

	(1)	(2)	(3)	(4)	(5)
JeMP	-0.016***	-0.019***	-0.015***	-0.019***	-0.012**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Baseline Controls	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Speech Length		✓	✓	✓	✓
Topic FE			✓	✓	✓
Speaker Characteristics				✓	✓
Speaker Politics					✓
R^2	0.02	0.03	0.13	0.13	0.13
Observations	1,128,866	1,128,866	1,128,866	1,007,382	983,727

Notes: The table shows that Jesuit-educated MPs (JeMPs) debated in a less polarized fashion in parliament. Regressions are run at the speech level. See the notes to Table 1 for a description of the control variables. Robust standard errors are reported in parentheses. Notation for statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Proximity to Religion and Science in MP Speeches

Dependent Variable: Proximity to Religion / Science

	Religion			Science		
	(1)	(2)	(3)	(4)	(5)	(6)
JeMP	0.011*	-0.001	-0.017***	0.044***	0.056***	0.054***
	(0.006)	(0.004)	(0.005)	(0.006)	(0.005)	(0.005)
Baseline Controls	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Speech Length	✓	✓	✓	✓	✓	✓
Topic FE		✓	✓		✓	✓
Speaker Characteristics			✓			✓
Speaker Politics			✓			✓
R^2	0.06	0.48	0.48	0.07	0.35	0.35
Observations	1,128,866	1,128,866	983,727	1,128,866	1,128,866	983,727

Notes: The table shows that the parliamentary speeches of JeMPs were closer to science as compared to other MPs, but not necessarily closer to religion. Regressions are run at the speech level. Proximity to religious and scientific language measures the proximity, in a word-embedding space, between each speech and, respectively, the words "religion" and "science." Cols 1-3 reports the results for religion, while cols 4-6 report the results for science. See the notes to Table 1 for a description of the control variables. Robust standard errors are reported in parentheses. Notation for statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Vocabulary Overlap with the Preceding Speaker

Dep. Var.: Jaccard Similarity Index with the Previous Speech

	(1)	(2)	(3)
JeMP	0.013**	0.014**	0.011*
	(0.006)	(0.006)	(0.006)
Baseline Controls	✓	✓	✓
Year FE	✓	✓	✓
Topic FE		✓	✓
Speaker Characteristics			✓
Speaker Politics			✓
R^2	0.01	0.02	0.02
Observations	1,128,866	1,128,866	983,727

Notes: The table shows that JeMPs engaged more with the arguments of preceding speakers. The dependent variable, the Jaccard Index, measures the vocabulary overlap between each speech and the preceding speech. See footnote 44 for detail. Regressions are run at the speech level. See the notes to Table 1 for a description of the control variables. Robust standard errors are reported in parentheses. Notation for statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: The Law of Separation of Church and State: JeMPs' Rhetorical Style

Dependent Variable: As indicated in the table header

	(1) Pro Score	(2) Polarization Score	(3) Emotionality Score	(4) Proximity to Religion	(5) Proximity to Science
Panel A: Baseline Controls					
JeMP^{Opp}	0.004	0.031	-0.076**	0.089*	0.112***
	(0.036)	(0.024)	(0.037)	(0.051)	(0.039)
R^2	0.05	0.02	0.06	0.11	0.09
Observations	12,010	12,010	12,010	12,010	12,010
Panel B: With Political Area Fixed Effects					
JeMP^{Opp}	0.133***	0.027	-0.133***	0.148***	0.163***
	(0.038)	(0.025)	(0.041)	(0.055)	(0.042)
R^2	0.07	0.02	0.07	0.11	0.09
Observations	11,552	11,552	11,552	11,552	11,552

Notes: The table shows that, when debating the 1905 Law on the Separation of Churches and the State, JeMPs who opposed the law of Separation (JeMP^{Opp}) by voting against it nevertheless used a moderating language that was less emotional, and referring more to both religion and science in their speeches. Panel A reports the results with our *Baseline Controls* (see the note to Table 1), while Panel B additionally includes political area fixed effects. See Section 3.5 for detail on the construction of the *Proscore*, which measures support for the 1905 Law. All regressions include a dummy for the remaining JeMPs who did not oppose the Separation Law. Thus, coefficients compare JeMP^{Opp} to the average non-Jesuit educated MP. Robust standard errors are reported in parentheses. Notation for statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: The Law of Separation of Church and State: Speakers following a JeMP^{Opp} Speech

Dependent Variable: As indicated in the column headers

	Word2Vec			LLM	
	(1)	(2)	(3)	(4)	(5)
	Proscore (Continuous)	Very Pro (Top Decile)	Very Against (Bottom Decile)	Very Pro	Very Against
Post JeMP^{Opp}	-0.121**	-0.035**	0.019	-0.021**	0.026
	(0.047)	(0.017)	(0.018)	(0.010)	(0.017)
Baseline Controls	✓	✓	✓	✓	✓
Speech Length	✓	✓	✓	✓	✓
MP × Discussion FE	✓	✓	✓	✓	✓
R^2	0.480	0.331	0.338	0.462	0.382
Observations	1,729	1,729	1,729	1,729	1,729

Notes: The table provides evidence that, during debates on the 1905 Law on the Separation of Churches and the State, JeMP interventions reduced support for the bill. Regressions are run at the speech level. In Colum 1, the dependent variable is the Proscore, a continuous measure of pro-secularization stance; in cols 2 and 3, the dependent variable is an indicator equal to one if the speech falls, respectively, in the top decile or in the bottom decile of the Proscore distribution; in columns 4 and 5, the dependent variable is an indicator equal to one if the large language model GTP40-mini classifies the speech as, respectively, strongly in favor of or strongly against secularization. See Section 4.3 for details. JeMP^{Opp} denotes JeMPS who consistently voted against the Secularization Law. *Post JeMP^{Opp}* is an indicator equal to one for the five speeches following a JeMP^{Opp} treatment event and zero for the five preceding speeches. The JeMP^{Opp} treatment events include the first JeMP^{Opp} intervention within a discussion, as well as subsequent JeMP^{Opp} interventions that occur more than 30 speaking turns after the previous treatment event. *Baseline Controls* include an indicator for procedural speeches, an indicator for speeches that remain empty after pre-processing, an indicator for the five speeches that follow a left-wing JeMP treatment intervention, as well as an indicator for the actual JeMP treatment events, which are included in the post-period. *Speech Length* represent quartiles of speech length (as measured by total word count). See the note to Table 1 for a description of *Baseline Controls* and *Speech Length* controls. Robust standard errors are reported in parentheses. Notation for statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

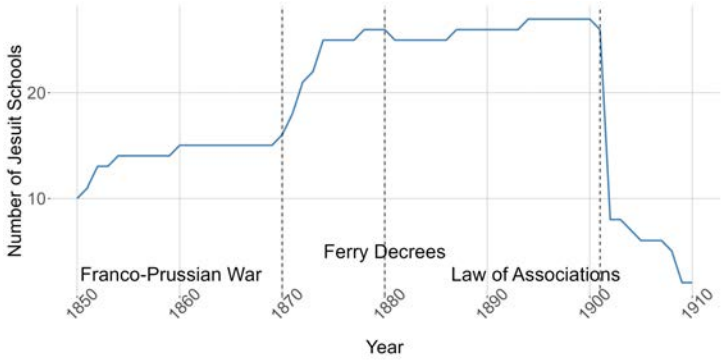
Table 7: The Law of Separation of Church and State: JeMPs^{Opp} Interventions and Voting

Dependent Variable: *MP Voted in Favor of a Separation Act*

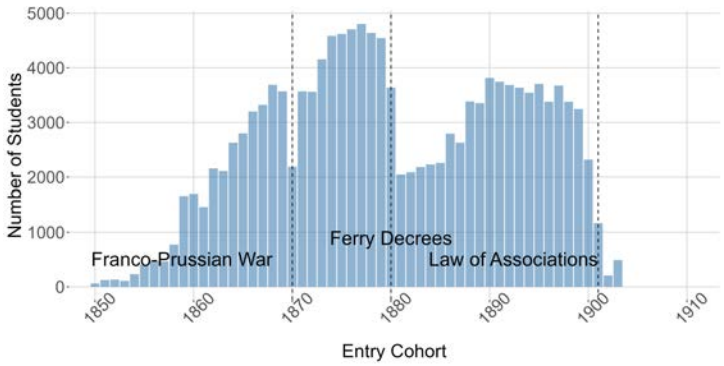
	All Voting Rounds	With JeMP^{Opp} speeches		
	(1)	(2)	(3)	(4)
Share JeMP$^{\text{Opp}}$ Speeches	-0.025***	-0.037***	-0.038***	-0.038***
	(0.001)	(0.001)	(0.001)	(0.001)
Avg. JeMP$^{\text{Opp}}$ Closeness to Vote			-0.055***	
			(0.015)	
Last JeMP$^{\text{Opp}}$ Closeness to Vote				-0.052***
				(0.011)
MP FE	✓	✓	✓	✓
R^2	0.72	0.69	0.69	0.69
Observations	9,862	7,125	7,125	7,125

Notes: The table suggests that, during debates on the 1905 Law on the Separation of Churches and the State, JeMPs were able to influence the voting behavior of other MPs, reducing support for pro-secularization acts. Regressions are run at the MP-discussion-segment level, using all discussion segments in parliament about the Separation Law that were followed by a roll-call vote on either single articles or the full text of the law. The dependent variable is an indicator equal to one if, at the end of the discussion segment, a non-JeMP MP voted in favor of the corresponding legislative act supporting the law. JeMP$^{\text{Opp}}$ denotes JeMPs who consistently voted against the secularization law. *Share JeMP^{Opp} Speeches* measures the share of JeMP$^{\text{Opp}}$ speeches (scaled between 0 and 100) in a discussion segment leading up to a roll-call vote. *Avg. JeMP^{Opp} Closeness to the Vote* represents the average index of JeMP$^{\text{Opp}}$ speeches within the discussion segment, where the speech index is normalized by the total number of speeches in the segment (so that 0 and 1 represent, respectively, the start and end of the segment). *Last JeMP^{Opp} Closeness to the Vote* is the normalized speech index of the last JeMP$^{\text{Opp}}$ in the segment. Robust standard errors are reported in parentheses. Notation for statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

FIGURES



(a) Number of Jesuit Schools in France, 1850–1910



(b) Number of Students in Jesuit Schools by Entry Cohort

Figure 1: Number of Jesuits Schools and Students in France, 1850-1910

Notes: The top panel shows the number of Jesuit schools operating in France between 1850 and 1910. In 1850, the Falloux Law legitimated non-State schooling and triggered the emergence of Jesuit educational institutions. The increase in 1870 can be explained as a transfer from weaker lay private schools to better organized and socially attractive Catholic institutions competing for the same social clientele. In 1901, the Law of Association dissolved religious congregations. The bottom panel show total enrollment in Jesuit schools in France.
Source: Authors' computation from data on school and student registries, collected at the *Archives Jésuites* in Vanves. See Section 3.1 for details.

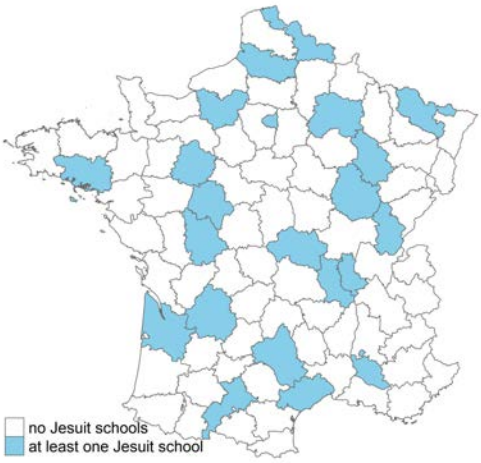
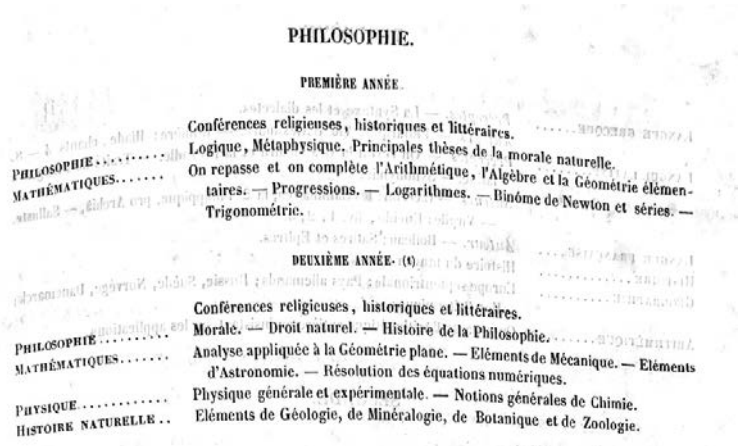


Figure 2: Distribution of Jesuit Schools across French Departments

Notes: The figure shows French departments hosting at least one Jesuit School from 1850 to 1910 (highlighted in light blue).
Source: Data on school registries, collected at the *Archives Jésuites* in Vanves. See Section 3.1 for details.



PHILOSOPHIE.

PREMIÈRE ANNÉE.

PHILOSOPHIE..... Conférences religieuses, historiques et littéraires.
Logique, Métaphysique. Principales thèses de la morale naturelle.
MATHÉMATIQUES..... On repasse et on complète l'Arithmétique, l'Algèbre et la Géométrie élémentaires. — Progressions. — Logarithmes. — Binôme de Newton et séries. — Trigonométrie.

DEUXIÈME ANNÉE. (1)

PHILOSOPHIE..... Conférences religieuses, historiques et littéraires.
Morale. — Droit naturel. — Histoire de la Philosophie.
MATHÉMATIQUES..... Analyse appliquée à la Géométrie plane. — Eléments de Mécanique. — Eléments d'Astronomie. — Résolution des équations numériques.
PHYSIQUE..... Physique générale et expérimentale. — Notions générales de Chimie.
HISTOIRE NATURELLE.. Eléments de Géologie, de Minéralogie, de Botanique et de Zoologie.

Figure 3: Curriculum in the Final Two Years of Instruction at a Jesuit School

Notes: The figure reproduces the curriculum for the final two years of instruction, referred to as "Philosophy" (*Philosophie*), at the Jesuit school in Vannes. Following the 1599 *Ratio Studiorum*, the Curriculum included logic, physics, and astronomy together with methaphysics and philosphy.
Source: Authors' photo, taken at the *Archives Jésuites* in Vanves.

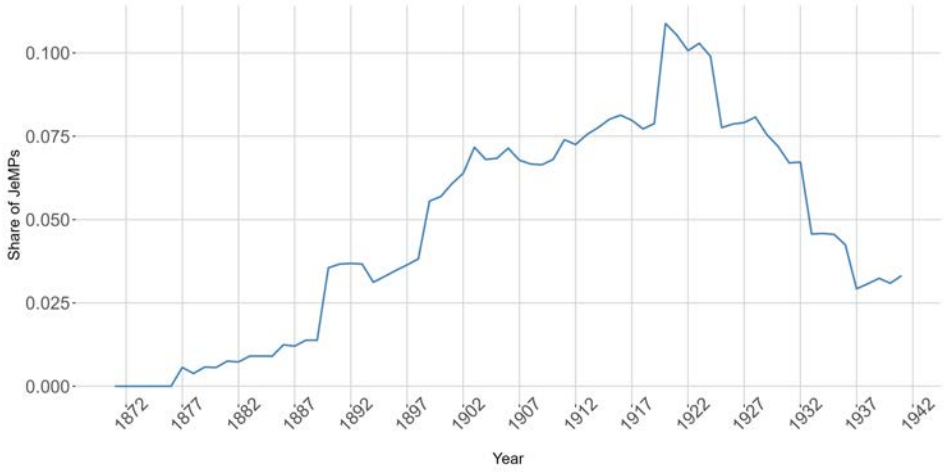


Figure 4: Share of Jesuit-educated MPs in the Chamber of Deputies

Notes: The figure shows the share of deputies who received a Jesuit education among members of the French Chamber of Deputies from 1871 to 1941. The share of Jesuit-educated MPs (JeMPs) rises until 1921 with the expansion of Jesuit schooling in France in the late 19th century and reaches around 7–10% between 1901 and 1921. The subsequent decline is due to the dissolution of religious congregations following the Law of Associations (1901).
Source: Authors' computation, obtained matching student registries from the *Archives Jésuites* in Vanves with MPs biographies from the official website of the *Assemblée Nationale*. For details, see Sections 3.1 and 3.3.

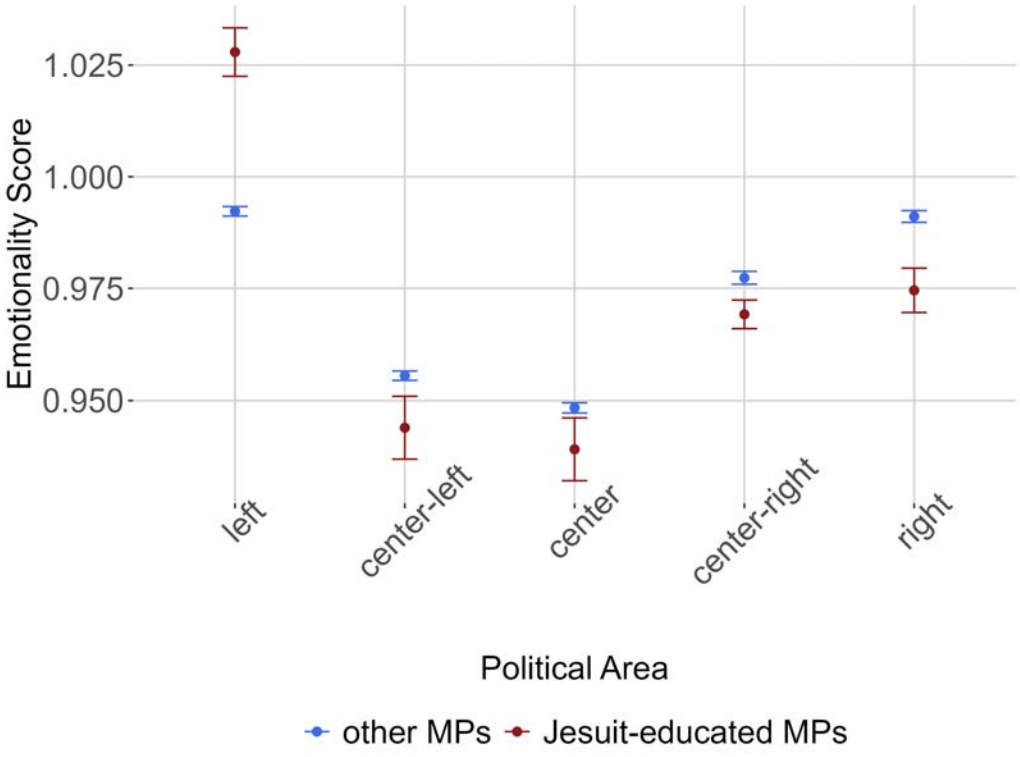


Figure 5: Emotionality of MP's Language Across Political Areas

Notes: The figure shows that Jesuit-educated MPs were less emotional in their speeches than other MPs across political areas except for the left. The dots represent the average emotionality score within each political area, while the error bars represent 90% confidence intervals.



(a) Religion

(b) Science

Figure 6: Word Clouds: Religion and Science in Parliamentary Speeches

Notes: The figure shows the word clouds of the 100 words in the corpus with the smallest cosine distance to the words "religion" and "science," computed with our main word-embedding model.

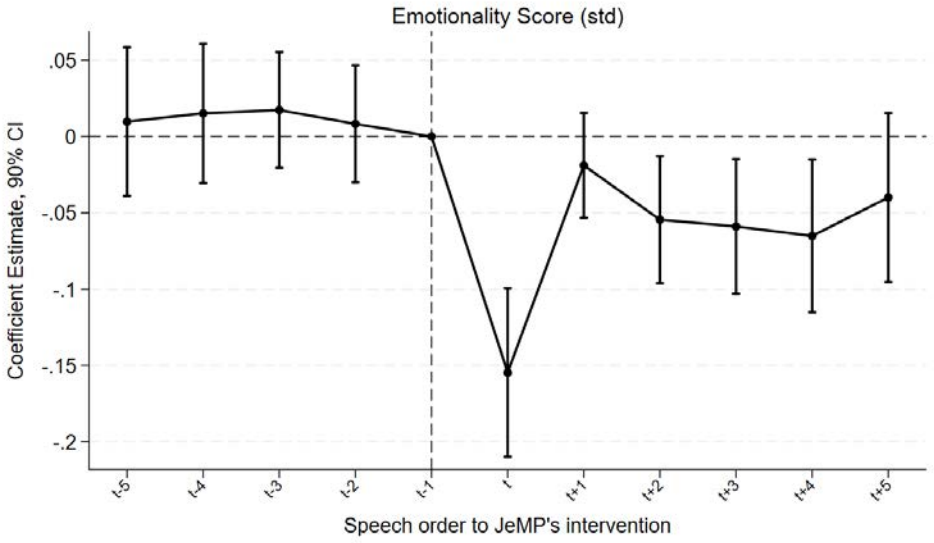


Figure 7: Event Study: JeMPs Reduce the Emotionality of Subsequent Speakers

Notes: The figure shows that following a JeMP speech, the emotional tone of the parliamentary debate declines markedly. The emotionality score measures the relative distance, in a word-embedding space, between each speech and pre-defined dictionaries of affection and cognitive terms. The x-axis indexes speech order relative to the first JeMP intervention ($t = 0$), and the y-axis plots the estimated β_k coefficients from regression (7), along with 90% confidence intervals, relative to the omitted category $k = -1$. The event-study specification includes speaker $\times$ discussion fixed effects, as well as controls for quartiles of speech length, an indicator for procedural speeches, an indicator for speeches that remain empty after pre-processing, event-time indicators relative to interventions by left-wing JeMPs, and fixed effects for the speaker's speech order within a discussion. Standard errors are clustered at the discussion level.

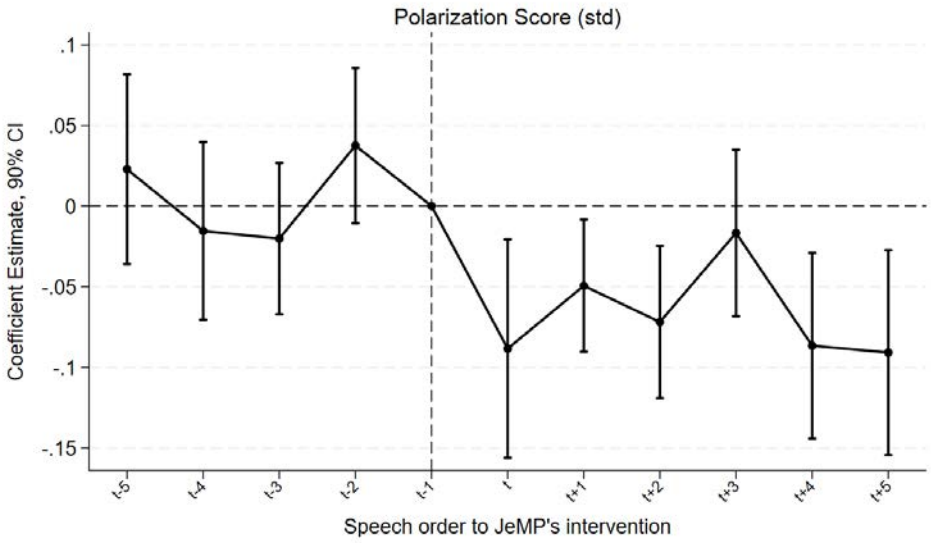
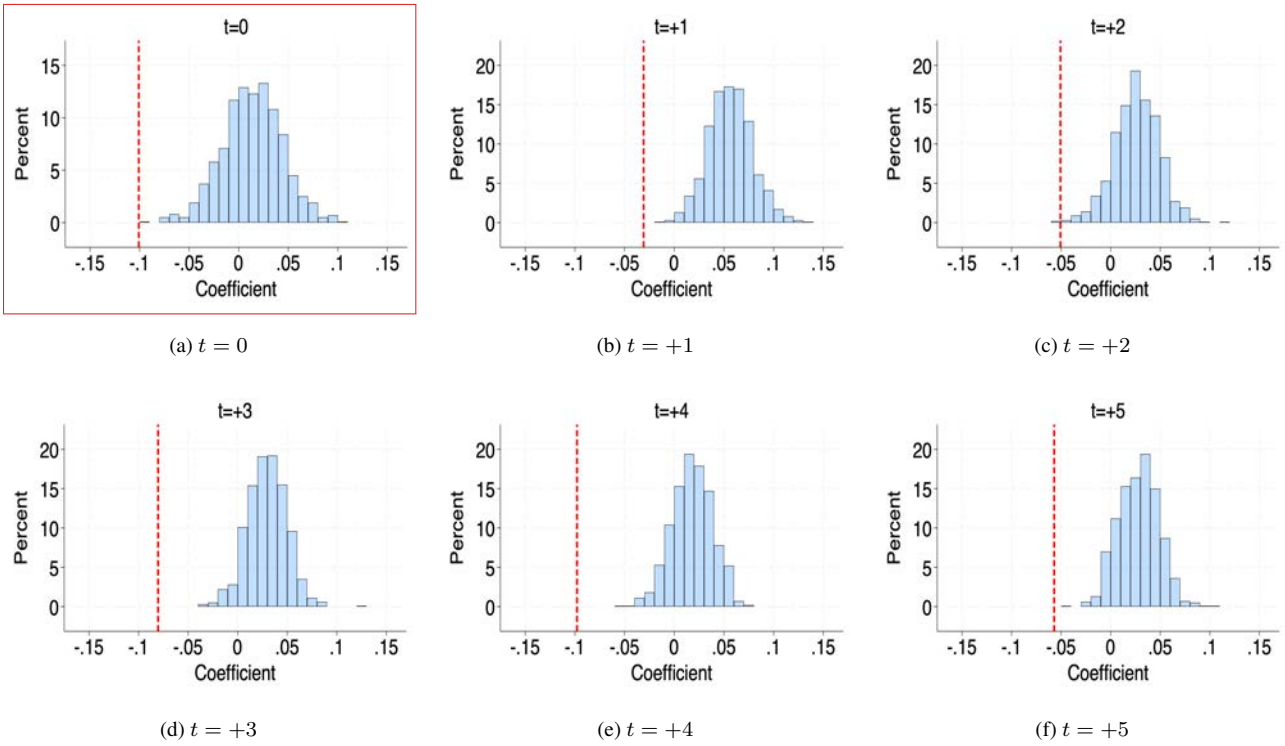


Figure 8: Event Study: JeMPs Reduce the Polarization of Subsequent Speakers

Notes: The figure shows that the partisan polarization of the parliamentary debate declines following a JeMP speech. The polarization score measures the relative distance, in a word-embedding space, between each speech and year–topic-specific dictionaries constructed from left-wing and right-wing parliamentarians' speeches, where topics are identified using a Latent Dirichlet Allocation algorithm. The x-axis indexes speech order relative to the first JeMP intervention ($t = 0$), and the y-axis plots the estimated β_k coefficients from equation (7), along with 90% confidence intervals, relative to the omitted category $k = -1$. The event-study specification includes speaker $\times$ discussion fixed effects, as well as controls for quartiles of speech length, an indicator for procedural speeches, an indicator for speeches that remain empty after pre-processing, event-time indicators relative to interventions by left-wing JeMPs, and fixed effects for the speaker's speech count within a discussion. Standard errors are clustered at the discussion level.

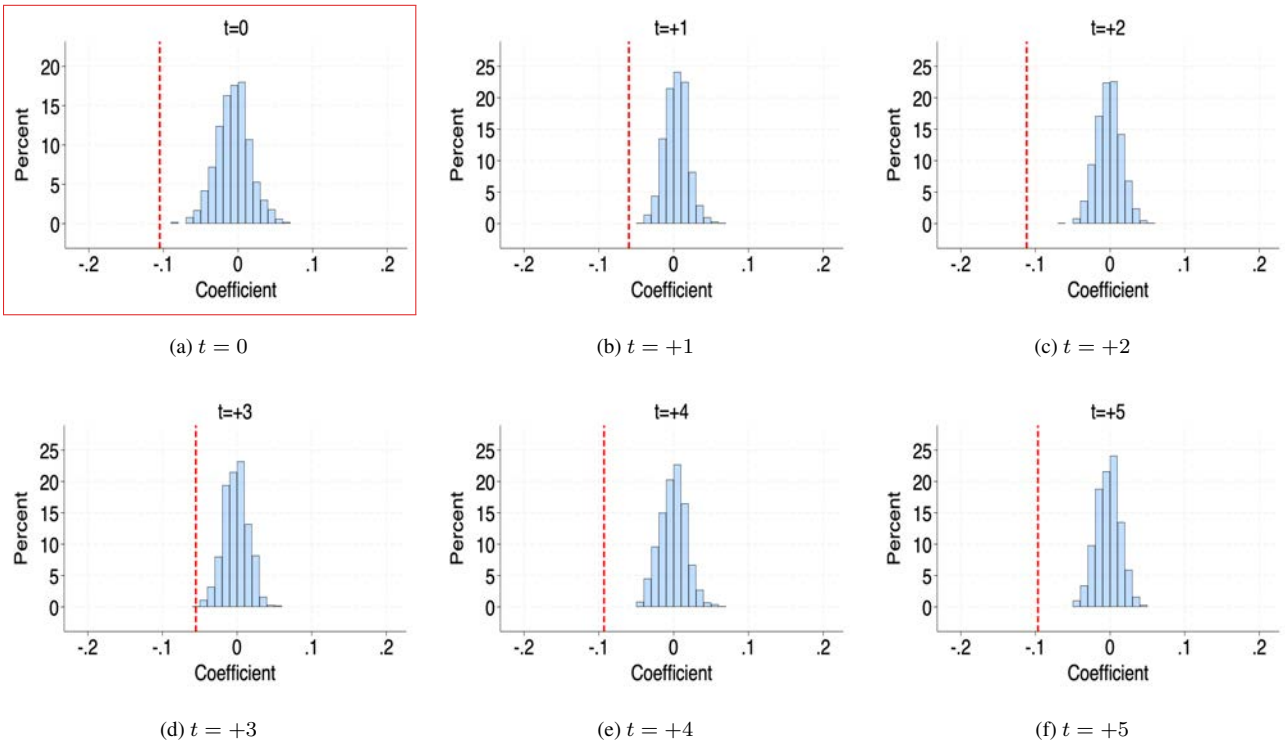


(a) $t = 0$ (b) $t = +1$ (c) $t = +2$

(d) $t = +3$ (e) $t = +4$ (f) $t = +5$

Figure 9: Placebo Event-Study for Emotionality: Random Samples of (non-JeMP) MPs

Notes: The figure repeats the event study from Figure 7 using 1,000 random samples of "Placebo JeMPs" as treatment. To construct those: (i) we exclude actual JeMPs, the President of the Chamber, and members of government; (ii) we randomly draw the 5% MPs and treat those as "Placebo JeMPs." Panel (a) corresponds to the treatment period, and Panels (b)-(f) correspond to each of the 5 post-treatment periods. In each panel, the histogram represents the distribution of the estimated coefficients on the corresponding event-time dummy across the 1,000 simulations. The dashed red line represents the coefficient on the actual JeMPs from the event-study in Figure 7. See the note to Figure 7 for detail on the event study specification.



(a) $t = 0$ (b) $t = +1$ (c) $t = +2$

(d) $t = +3$ (e) $t = +4$ (f) $t = +5$

Figure 10: Placebo Event-Study for Polarization: Random Samples of (non-JeMP) MPs

Notes: The figure repeats the event study from Figure 8 using 1,000 random samples of "Placebo JeMPs" as treatment. Those are constructed in the same way as described in the note to Figure 9. In each panel, the histogram represents the distribution of the estimated coefficients on the corresponding event-time dummy across the 1,000 simulations, and the dashed red line represents the coefficient on the actual JeMPs from the event-study in Figure 8. See the note to Figure 7 for detail on the event study specification.

Online Appendix

De-Polarizing Politics: Faith Meets Science in the French Third Republic

Gianandrea Lanzara Mara P. Squicciarini Nico Voigtländer

A Data

A.1 Students Enrolled in Jesuit Schools in France, 1850-1901

A.1.1 Digitizing Student Registries

Data on students enrolled in Jesuit schools in France are available at the *Archives Jésuites* in Vanves near Paris.[1] The Jesuit archives are organized into four collections, each corresponding to one of the four Jesuit Provinces established in France: Paris, Lyon, Toulouse and Champagne. They contain the student registers for 30 of the 32 Jesuit schools that operated in France between 1850 and 1901.[2] The student registers are preserved in 142 separate documents dating from 1850 to 1901. For 37 of these documents, only the handwritten (hereafter HW) version has survived.[3]

Registers from each Jesuit school typically report the list of students enrolled by school year, with each document often spanning several school years. In most cases, these documents also report additional information on each student's grade, section, division (a group of students spanning multiple grades who shared the same boarding facilities), and boarding status. As an exception to the standard format, eight documents report only the name and year of entry for students.

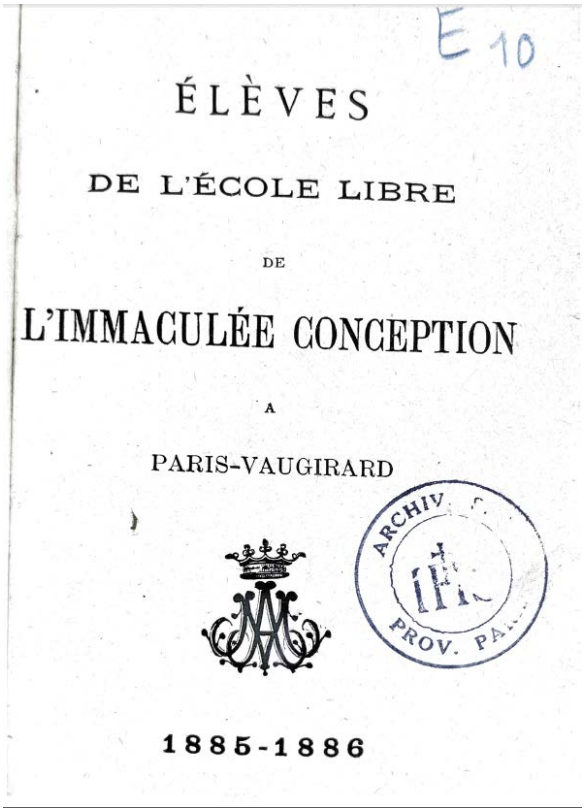
Figure A.1 illustrates the student registry for school year 1885-86 at the *Collège de l'Immaculée-Conception*, located in *Rue de Vaugirard* in Paris. The left panel shows the front page of the document, while the right panel displays the list of students in alphabetical order. The columns on this page, from left to right, indicate: (i) student number (a unique numeric identifier); (ii) student name; (iii) class (i.e., grade and section); and (iv) the division. The codes used for class, section, and division vary across schools and years, and each document typically includes the legends to explain them. Figure A.2 presents the relevant legend for this specific school and school year. To illustrate, Fernand

[1] See https://www.jesuites.com/contact/bureaux-archives-jesuites for an overview.

[2] The Provinces were, and still are, the main administrative divisions of the Society of Jesus. The archives lack documentation for the schools in Montauban and Mende, part of the Province of Toulouse, as well as for those in Algiers and Oran, located in colonial Algeria and associated with the Province of Lyon.

[3] According to private conversations with the archive staff, Jesuit school officials typically compiled the handwritten versions and sent them to a printing house.

Adhémar, Josse Allard, and Henri d'Anchald were all enrolled in the "Rhetoric" grade ("Rh" in the legend, short for *Rhétorique*), although each of them attended one of the three different sections that the school offered for this grade.[4] All three belonged to the first division. However, while Fernand and Josse were boarders, the asterisk indicates that Henri was a half-boarder. In our main analysis, we will only use the grade to infer the student's age. Below, we provide further details on the grade levels and curricular structure of Jesuit schools.



ÉLÈVES

DE L'ÉCOLE LIBRE

DE

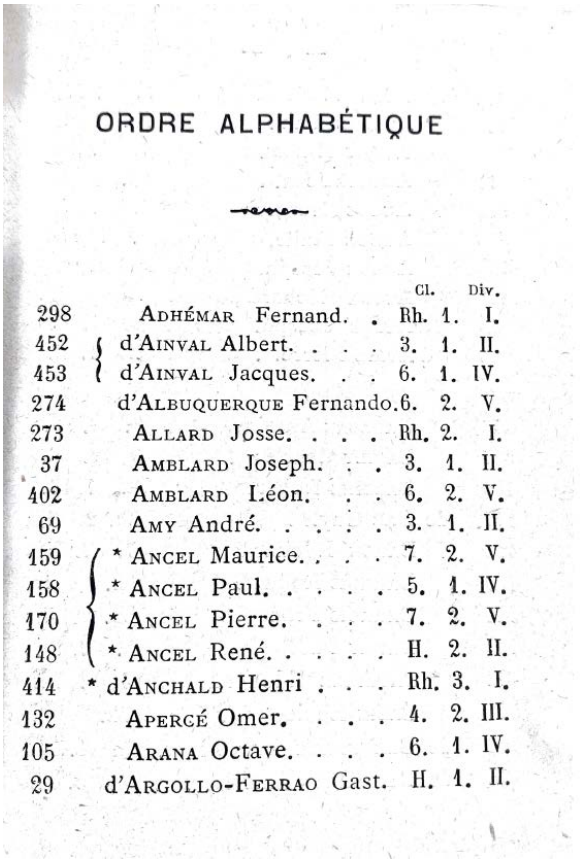
L'IMMACULÉE CONCEPTION

A

PARIS-VAUGIRARD

1885-1886

(a) Front Page



ORDRE ALPHABÉTIQUE

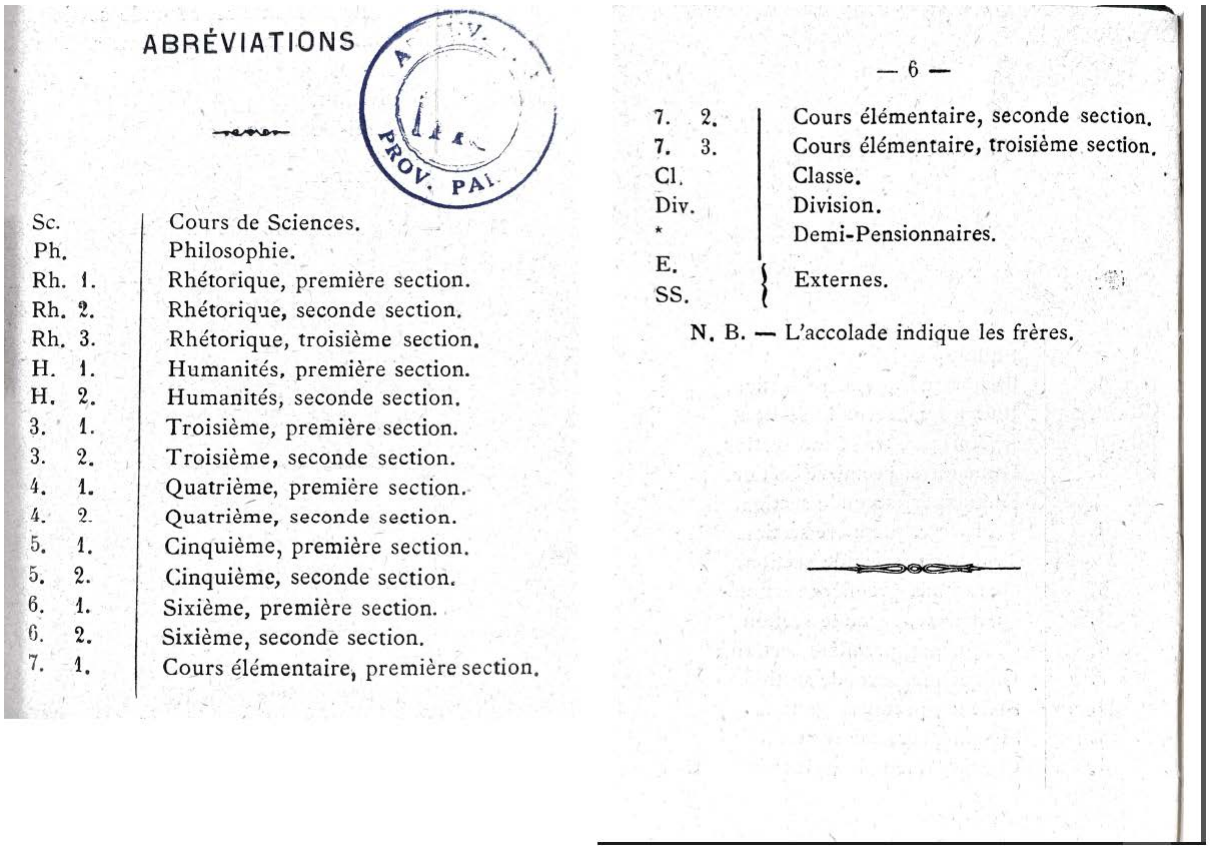
		Cl.	Div.
298	Adhémar Fernand. .	Rh.	1. I.
452	d'Ainval Albert. . . .	3.	1. II.
453	d'Ainval Jacques. . .	6.	1. IV.
274	d'Albuquerque Fernando.	6.	2. V.
273	Allard Josse. . . .	Rh.	2. I.
37	Amblard Joseph. . .	3.	1. II.
402	Amblard Léon. . .	6.	2. V.
69	Amy André.	3.	1. II.
159	* Ancel Maurice. . . .	7.	2. V.
158	* Ancel Paul.	5.	1. IV.
170	* Ancel Pierre. . . .	7.	2. V.
148	* Ancel René.	H.	2. II.
414	* d'Anchald Henri . . .	Rh.	3. I.
132	Apercé Omer. . . .	4.	2. III.
105	Arana Octave. . . .	6.	1. IV.
29	d'Argollo-Ferrao Gast.	H.	1. II.

(b) List of students in alphabetical order

Figure A.1: Student Registry – *Collège de l'Immaculée-Conception* (Paris), school year 1885-86

Note: The left panel shows the front page of the registry, while the right panel displays the list of students in alphabetical order. The columns from left to right indicate: student number, student name, class, and division (a group of students, spanning multiple grades, who shared the same boarding facilities). Asterisks indicate half-boarders, whereas curly brackets indicate brothers (see the legend in Figure A.2).

[4] At Jesuit schools, grades ran in reverse order, with *Rhétorique* corresponding to the final year of the standard curriculum. Some schools, including this one, offered additional post-secondary classes in Philosophy and Science.



ABRÉVIATIONS

Sc.	Cours de Sciences.
Ph.	Philosophie.
Rh. 1.	Rhétorique, première section.
Rh. 2.	Rhétorique, seconde section.
Rh. 3.	Rhétorique, troisième section.
H. 1.	Humanités, première section.
H. 2.	Humanités, seconde section.
3. 1.	Troisième, première section.
3. 2.	Troisième, seconde section.
4. 1.	Quatrième, première section.
4. 2.	Quatrième, seconde section.
5. 1.	Cinquième, première section.
5. 2.	Cinquième, seconde section.
6. 1.	Sixième, première section.
6. 2.	Sixième, seconde section.
7. 1.	Cours élémentaire, première section.

PROV. PAI

7. 2.	Cours élémentaire, seconde section.
7. 3.	Cours élémentaire, troisième section.
Cl.	Classe.
Div.	Division.
*	Demi-Pensionnaires.
E.	Externes.
SS.	Externes.

N. B. — L'accolade indique les frères.

Figure A.2: Legend for *Collège de l'Immaculée-Conception* (Paris), year 1885-86.

Notes: From "Sc" to "7.3", the legend maps the class codes to the corresponding grade-section. For example, "Rh. 1." indicates the Rhetoric grade, first section (*Rhétorique, première section*). "Cl" and "Div" explain the names of the additional columns in the registry, respectively Class and Division (see Figure A.1). From "*" to "SS", the legend defines boarding status: half-boarders (*Demi-Pensionnaires*), and day students (*Externes*), while boarders are listed without a symbol. The final note specifies that curly brackets indicate brothers.

For printed school registries (as shown above), our digitization process consists of four main steps:

1. We manually scan all front pages, legends, and pages reporting the lists of students in alphabetical order;
2. We perform Optical Character Recognition (OCR) on each document using Google Document AI. A key advantage of Google Document AI is its ability to detect and preserve the tabular structure of the pages;
3. We parse and standardize the OCR output using custom R code, with two main objectives:
    - Associate each student list with the corresponding school year, typically extracted from the front page;
    - Infer and harmonize column types (e.g., name, student number, class, year of entry) de-

spite substantial variation in formatting. At this stage, grade, section, division, and boarding status, when available, are grouped into a single "class information" field.

This step includes numerous document-specific exceptions to accommodate irregular layouts and atypical reporting practices;

4. With the help of research assistants, we manually reviewed all final Excel files to identify and correct potential OCR errors.

For handwritten (HW) documents, we enter the data manually for the entire document, relying on a French research assistant with a background in history and prior experience reading historical French calligraphy. Figure A.3 illustrates an example of a handwritten document: the list of students for the school year 1853-54 for the same school shown in Figure A.1.

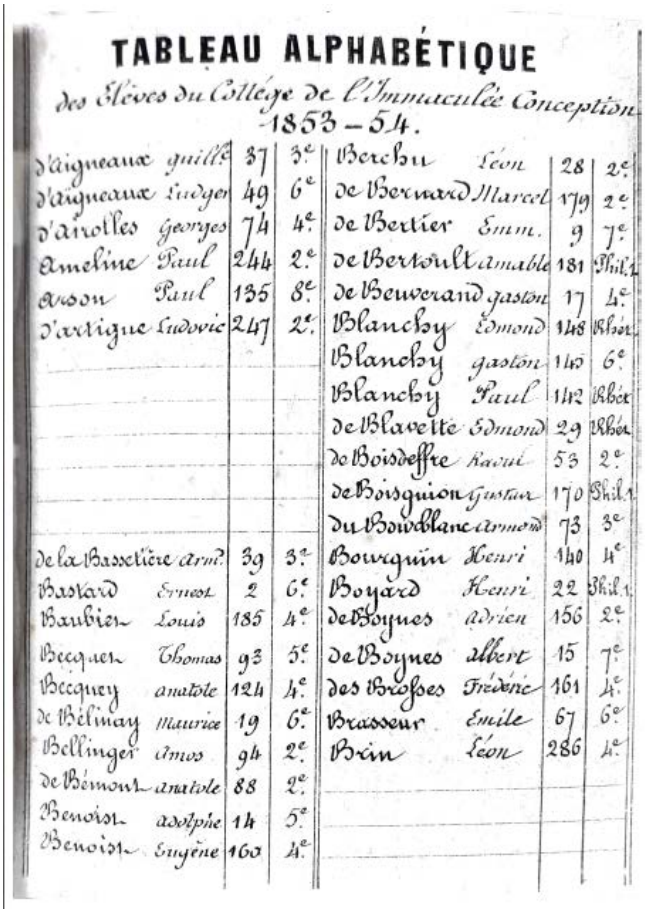


TABLEAU ALPHABÉTIQUE

des Élèves du Collége de l'Immaculée Conception.

1853 – 54.

d'Aigneaux	guillt	37	3e	Berchu	Léon	28	2e
d'Aigneaux	Ludger	49	6e	de Bernard	Marcel	179	2e
d'Arolles	georges	74	4e	de Bertier	Emm.	9	7e
Ameline	Paul	244	2e	de Bertoult	Amable	181	Phil.
Arson	Paul	135	8e	de Beuverand	gaston	17	4e
d'Artigue	Ludovic	247	2e	Blanchy	Edmond	148	Rhon
				Blanchy	gaston	145	6e
				Blanchy	Paul	142	Rhet
				de Blavette	Edmond	29	Rhet
				de Boisdeffre	Raoul	53	2e
				de Boisguion	Gustave	170	Phil.
				du Bouvblanc	Armand	73	3e
de la Bassetière	Arm.	39	3e	Bourguin	Henri	140	4e
Bastard	Ernest	2	6e	Boyard	Henri	22	Phil.
Baubier	Louis	185	4e	de Boynes	Adrien	156	2e
Becquer	Thomas	93	5e	de Boynes	Albert	15	7e
Becquey	Anatole	124	4e	des Brosses	Frédéric	161	4e
de Bélimay	Maurice	19	6e	Brasseur	Emile	67	6e
Bellinger	Amos	94	2e	Brin	Léon	286	4e
de Bémont	Anatole	88	2e				
Benoist	Adolphe	14	5e				
Benoist	Eugène	160	4e				

Figure A.3: Student Registry – *Collège de l'Immaculée-Conception* (Paris), school year 1853-54

Notes: List of students in alphabetical order for a handwritten document.

At the end of this process, we obtain a dataset with 268,216 student records. Some of these

records belong to the same student observed in multiple school years, while others refer to students who appear only once in the dataset along with their year of entry at their school.

A.1.2 Linking Student Records

We reconstruct the school histories of Jesuit-school students based on their names and the years during which they appear in the records. The main objective is to determine the earliest and latest years during which each student attended a Jesuit school. This information, in turn, helps us estimate the student's age, which is ultimately needed to match student records to parliamentarians (see Section A.1.3 below).

First, we clean the student names with the following steps: (i) we expand common abbreviations to full names (e.g., *Pre* $\rightarrow$ *Pierre*, *Lt* $\rightarrow$ *Laurent*); (ii) we remove all punctuation marks, which typically reflect OCR or transcription errors, with the exception of apostrophes and dashes (unless these appear at the end of the string); (iii) we remove all non-ASCII characters; (iv) we set the text to lower case.

Second, we extract the first name and the last name from the name field. To do so, we must determine whether names in a given document-year follow a "firstname lastname" or "lastname firstname" format. We check the position of common French first names (e.g., *Jean*, *Louis*, *Henri*) within the name string. Based on the predominant pattern, we extract the first name either from the beginning or end of the cleaned name string. For HW documents, we directly classify first and last names in transcribing the handwritten records.

Third, we remove all nobiliary particles such as *de*, *du*, *la*, and *le* from surnames, except when they are part of a hyphenated compound name (e.g., *d'aubigny* $\rightarrow$ *aubigny*, but *mery-le-beuve* $\rightarrow$ *mery-le-beuve*).

Fourth, we construct the variables $y_r^{\min}$ and $y_r^{\max}$ for each *record* r as follows: (i) if the document reports the school year in standard format, e.g., 18XX–18(XX+1), we assign 18XX to $y_r^{\min}$ and 18(XX+1) to $y_r^{\max}$; (ii) if the document only reports the year of entry, we assign this value to both $y_r^{\min}$ and $y_r^{\max}$.[5]

Once we have the cleaned first name, last name, and the narrowest possible time window for each record, we implement an iterative procedure to link student records across school years. As multiple records are linked, we update $y_i^{\min}$ and $y_i^{\max}$ for each *student* i accordingly. Let $\mathcal{R}_i^k$ be the set of records associated with student i at the end of iteration k; then,

$$y_i^{\min,k+1} = \min_{r \in \mathcal{R}_i^k} y_r^{\min,k}, \quad \text{and} \quad y_i^{\max,k+1} = \max_{r \in \mathcal{R}_i^k} y_r^{\max,k}.$$

[5]Two documents, one for the school in Mongre and one for the school in Saint-Etienne, also report the year of exit. In these cases, we assign the year of entry to $y_r^{\min}$ and the year of exit to $y_r^{\max}$. These records are included in the iterative procedure, as additional records for the same students may exist in other documents from the same school. If the entry and exit years produced by the iterative procedure differ from those reported in the original document, we retain the former.

Once all records for student i are linked, $y_i^{\min}$ and $y_i^{\max}$ represent the earliest and latest years during which student i is observed in a Jesuit school.

The iterative procedure works as follows. As an initial step, we link records with an exact name match within the same school, provided that the student numbers as reported in the original documents coincide (as there could be cases of multiple names within the same schools).[6] Next, we iterate through the following steps:

- We consider all possible pairs of student records within the same school, under the restriction that their enrollment intervals in the current iteration – defined by their respective $y_i^{\min,k}$ and $y_i^{\max,k}$ – overlap within a four-year window;[7]
- If both the first and last names are identical, we link the records;
- We then consider four additional, and strictly looser, criteria for linking records:
  - One surname is a subset of the other, and the first names are identical;
  - One first name is a prefix of the other, and the last names are identical;
  - The first names have a Jaro distance below 0.1, and the last names are identical[8];
  - The last names have a Jaro distance below 0.1, and the first names are identical.
- We update $y_i^{\min,k+1}$ and $y_i^{\max,k+1}$ for all linked records and repeat the procedure.

Because the enrollment window may expand after each round of matching, new records may be linked in subsequent iterations. The process continues until no further links are formed. At the end, we obtain 70,452 linked students.

A.1.3 Determining the age of students

We use class information reported in the original documents, whenever available, to estimate the age of each student. As a first step, we harmonize class codes across schools and years using the legends provided in the school registries (see Figure A.2). At Jesuit schools, following the French educational system, grades ran in reverse order, with *Rhétorique*, the final year of the standard curriculum, typically attended at ages 16-17, corresponding to grade 1. Most schools, such as the one in *Rue de Vaugirard* in Paris (illustrated in Figure A.2), offered additional classes in Philosophy and Science,

[6] Although these individual identifiers were intended to track students across school years, they exhibit too many inconsistencies to be reliably used on their own as linking variables. Moreover, they are not available for all entries. As a result, we use them only in this initial step to reduce the dimensionality of the name-matching problem.

[7] This restriction ensures that we do not link records with enrollment periods that are implausibly far apart, even if the student names match. For example, suppose we have a record for Jean Dubois in "1870-1871" and another one in the same school for "1880-1881," with no information on the student in between. These records are unlikely to belong to the same student, although the names match. Our four-year time-window restriction ensures that when comparing pairs of records At the same time, the four-year window is flexible enough to allow links between records that are reasonably close in time, allowing for temporary school absence, missing records, or OCR errors leading to gaps in student records.

[8] The Jaro distance measures the similarity between two sequences, taking into account the number of matching characters. The score is normalized to range from 0 (exact match) to 1 (completely dissimilar). A threshold of 0.1 is conservative: for example, "Smith" and "Smyth" have Jaro distance of 0.13.

attended at approximately age 18. We code these as grades 0 and -1, respectively. Entry-level grades at Jesuit schools varied: while some schools began at grade 7, approximately at age 10, others offered preparatory courses starting as early as grade 9, approximately at age 8.

As a validation for both our student-record-linkage procedure and our coding of class levels, we compute grade-to-grade transition matrices for all students observed in consecutive years. Reassuringly, 96% of students either advanced or repeated the grade (with 82% advancing and 14% repeating). As an example, Figure A.4 below shows the transition matrix for students enrolled at the *Collège de l'Immaculée-Conception* on *Rue de Vaugirard* in Paris.

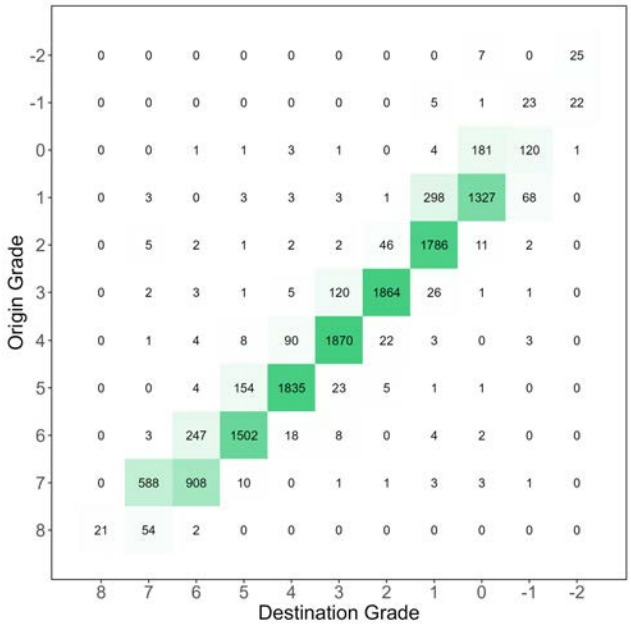


Figure A.4: Grade-to-grade transition matrix for the *Collège de l'Immaculée-Conception* (Paris)

Note: The figure displays the grade-to-grade transition matrix for students observed in at least two consecutive school years at the *Collège de l'Immaculée-Conception* on *Rue de Vaugirard* in Paris. Each row corresponds to the grade a student attended in year t, and each column to the grade attended in year $t+1$. The cells of the matrix report the number of observed transitions from a given grade to another grade in the subsequent year.

As mentioned above, in the French school system grades ran in reverse order, with grade 9 corresponding approximately to age 8. Accordingly, we estimate the implied age for each record r using the formula: $Age_r = 8 + (9 - Grade_r)$.

A.2 Parliamentary Speeches, 1881–1914

We digitized the universe of French parliamentary debates from 1881 to 1914 using the original transcripts published in the official gazette (*Journal officiel de la République française*). These pub-

lications are available online in PDF format and as OCR-generated text layers on Gallica, the digital repository of the French National Library.[9] While Gallica provides a text version, we digitize the documents from scratch to obtain higher-quality text extraction. Building on Lanzara et al. (2024), who rely on Gallica's OCR text, we then split the transcripts into individual speeches and link speaker names to biographical data on elected MPs. Section A.3 below describes the biographical data, while Section A.5.2 describes the matching procedure.

A.2.1 Digitizing Parliamentary Debates

Figure A.5 shows the front page of the parliamentary transcript for the sitting of April 5th, 1905 at the Chamber of Deputies. While this page illustrates the typical three-column structure of the *Journal officiel*, some pages embed numerical tables that disrupt the three-column layout. Because OCR software may fail to reliably detect column breaks, we use a Detectron2 image detection algorithm to segment each page into text columns/blocks and tables, and then apply Google Cloud Vision for optical character recognition on each detected text region.[10]

More specifically, our digitization process consists of six main steps:

1. We download all PDFs for both branches of Parliament from 1881 to 1942. We manually verify cases in which multiple PDFs are associated with the same date and discard duplicates when they contain transcripts of the same parliamentary sitting.[11]

2. We convert all PDF pages into high-resolution JPEG images.

3. We train an image-detection model to identify pages containing tables, using a manually annotated sample of 60 pages. When we detect a table, we strip it from the page before further processing by setting its pixels to white.

4. We train two layout models to detect individual text columns/blocks: one for standard three-column pages and one for pages containing tables. Each model is trained on a sample of 50 pages with the corresponding structure.

5. We use Google Cloud Vision to perform optical character recognition on each detected text column/block.

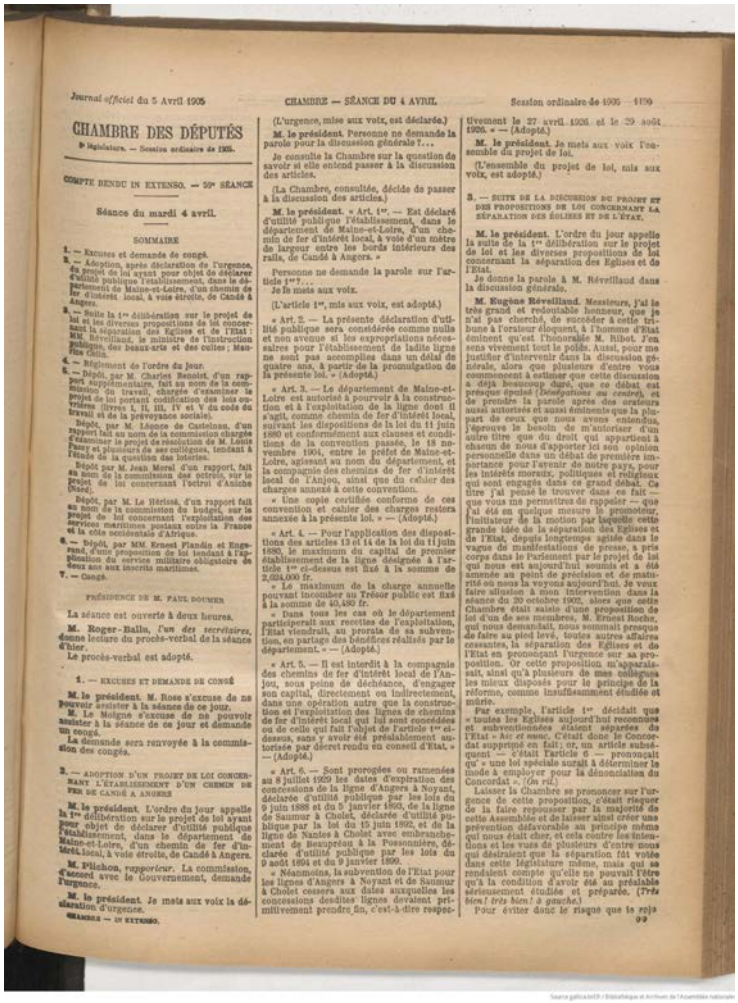
[9] See https://gallica.bnf.fr.

[10] Detectron2 is an image detection library developed by Facebook AI, see https://github.com/facebookresearch/detectron2. For information on Google Cloud Vision, see https://docs.cloud.google.com/vision/docs/ocr.

[11] On Gallica, each issue of the official gazette is associated with its publication date, which typically follows the date of the actual sitting by one or two days. When an issue contains more than one sitting, the transcripts may be split across multiple PDFs or included in a single PDF; typically, these cases represent two sittings held on the same date. By contrast, duplicate PDFs correspond to alternative scans of the same issue; in these cases, we retain only the higher-quality scan.

6. Finally, we reassemble columns/blocks (using their coordinates) to reconstruct each page, and then concatenate pages to reconstruct each document.

Figure A.6 illustrates our application of image-detection algorithms in steps 3 and 4 above both to standard pages and pages with tables.



Journal officiel du 5 Avril 1905 — CHAMBRE — SÉANCE DU 4 AVRIL — Session ordinaire de 1905 — 1190

CHAMBRE DES DÉPUTÉS

9e législature. — Session ordinaire de 1905.

COMPTE RENDU IN EXTENSO. — 29e SÉANCE

Séance du mardi 4 avril.

SOMMAIRE

1. — Excuses et demande de congé.
2. — Adoption, après déclaration de l'urgence, du projet de loi ayant pour objet de déclarer d'utilité publique l'établissement, dans le département de Maine-et-Loire, d'un chemin de fer d'intérêt local, à voie étroite, de Candé à Angers.
3. — Suite de la 1re délibération sur le projet de loi et les diverses propositions de loi concernant la séparation des Églises et de l'État : MM. Réveilland, le ministre de l'instruction publique, des beaux-arts et des cultes ; Maurice Ordinaire.
4. — Règlement de l'ordre du jour.
5. — Dépôt, par M. Charlet Benoist, d'un rapport supplémentaire, fait au nom de la commission du travail, chargée d'examiner le projet de loi portant codification des lois ouvrières (livres I, II, III, IV et V du code du travail et de la prévoyance sociale).

Dépôt, par M. Léonce de Castelnau, d'un rapport fait au nom de la commission chargée d'examiner le projet de résolution de M. Louis Passy et plusieurs de ses collègues, tendant à l'étude de la question des loteries.

Dépôt par M. Jean Morel d'un rapport, fait au nom de la commission des octroi, sur le projet de loi concernant l'octroi d'Aniche (Nord).

Dépôt, par M. Le Hérissé, d'un rapport fait au nom de la commission du budget, sur le projet de loi concernant l'exploitation des services maritimes postaux entre la France et la côte occidentale d'Afrique.

6. — Dépôt, par MM. Ernest Flandin et Rogerand, d'une proposition de loi tendant à l'application du service militaire obligatoire de deux ans aux inscrits maritimes.
7. — Congés.

PRÉSIDENCE DE M. PAUL DOUMER

La séance est ouverte à deux heures.

M. Roger-Ballu, *l'un des secrétaires*, donne lecture du procès-verbal de la séance d'hier.

Le procès-verbal est adopté.

1. — EXCUSES ET DEMANDE DE CONGÉ

M. le président. M. Rose s'excuse de ne pouvoir assister à la séance de ce jour.

M. Le Moigne s'excuse de ne pouvoir assister à la séance de ce jour et demande un congé.

La demande sera renvoyée à la commission des congés.

2. — ADOPTION D'UN PROJET DE LOI CONCERNANT L'ÉTABLISSEMENT D'UN CHEMIN DE FER DE CANDÉ A ANGERS

M. le président. L'ordre du jour appelle la 1re délibération sur le projet de loi ayant pour objet de déclarer d'utilité publique l'établissement, dans le département de Maine-et-Loire, d'un chemin de fer d'intérêt local, à voie étroite, de Candé à Angers.

M. Plichon, *rapporteur*. La commission, d'accord avec le Gouvernement, demande l'urgence.

M. le président. Je mets aux voix la déclaration d'urgence.

CHAMBRE — IN EXTENSO.

(L'urgence, mise aux voix, est déclarée.)

M. le président. Personne ne demande la parole pour la discussion générale ?...

Je consulte la Chambre sur la question de savoir si elle entend passer à la discussion des articles.

(La Chambre, consultée, décide de passer à la discussion des articles.)

M. le président. « Art. 1er. — Est déclaré d'utilité publique l'établissement, dans le département de Maine-et-Loire, d'un chemin de fer d'intérêt local, à voie d'un mètre de largeur entre les bords intérieurs des rails, de Candé à Angers. »

Personne ne demande la parole sur l'article 1er ?...

Je le mets aux voix.

(L'article 1er, mis aux voix, est adopté.)

« Art. 2. — La présente déclaration d'utilité publique sera considérée comme nulle et non avenue si les expropriations nécessaires pour l'établissement de ladite ligne ne sont pas accomplies dans un délai de quatre ans, à partir de la promulgation de la présente loi. » (Adopté.)

« Art. 3. — Le département de Maine-et-Loire est autorisé à pourvoir à la construction et à l'exploitation de la ligne dont il s'agit, comme chemin de fer d'intérêt local, suivant les dispositions de la loi du 11 juin 1880 et conformément aux clauses et conditions de la convention passée, le 18 novembre 1904, entre le préfet de Maine-et-Loire, agissant au nom du département, et la compagnie des chemins de fer d'intérêt local de l'Anjou, ainsi que du cahier des charges annexé à cette convention.

« Une copie certifiée conforme de ces convention et cahier des charges restera annexée à la présente loi. » — (Adopté.)

« Art. 4. — Pour l'application des dispositions des articles 13 et 14 de la loi du 11 juin 1880, le maximum du capital de premier établissement de la ligne désignée à l'article 1er ci-dessus est fixé à la somme de 2,624,000 fr.

« Le maximum de la charge annuelle pouvant incomber au Trésor public est fixé à la somme de 40,480 fr.

« Dans tous les cas où le département participerait aux recettes de l'exploitation, l'État viendrait, au prorata de sa subvention, en partage des bénéfices réalisés par le département. » — (Adopté.)

« Art. 5. — Il est interdit à la compagnie des chemins de fer d'intérêt local de l'Anjou, sous peine de déchéance, d'engager son capital, directement ou indirectement, dans une opération autre que la construction et l'exploitation des lignes de chemins de fer d'intérêt local qui lui sont concédées ou de celle qui fait l'objet de l'article 1er ci-dessus, sans y avoir été préalablement autorisée par décret rendu en conseil d'État. » — (Adopté.)

« Art. 6. — Sont prorogées ou ramenées au 6 juillet 1929 les dates d'expiration des concessions de la ligne d'Angers à Noyant, déclarée d'utilité publique par les lois du 9 juin 1888 et du 5 janvier 1893, de la ligne de Saumur à Cholet, déclarée d'utilité publique par la loi du 15 juin 1892, et de la ligne de Nantes à Cholet avec embranchement de Beaupréau à la Possonnière, déclarée d'utilité publique par les lois du 9 août 1894 et du 9 janvier 1899.

« Néanmoins, la subvention de l'État pour les lignes d'Angers à Noyant et de Saumur à Cholet cessera aux dates auxquelles les concessions desdites lignes devaient primitivement prendre fin, c'est-à-dire respectivement le 27 avril 1926 et le 29 août 1926. » — (Adopté.)

M. le président. Je mets aux voix l'ensemble du projet de loi.

(L'ensemble du projet de loi, mis aux voix, est adopté.)

3. — SUITE DE LA DISCUSSION DU PROJET ET DES PROPOSITIONS DE LOI CONCERNANT LA SÉPARATION DES ÉGLISES ET DE L'ÉTAT.

M. le président. L'ordre du jour appelle la suite de la 1re délibération sur le projet de loi et les diverses propositions de loi concernant la séparation des Églises et de l'État.

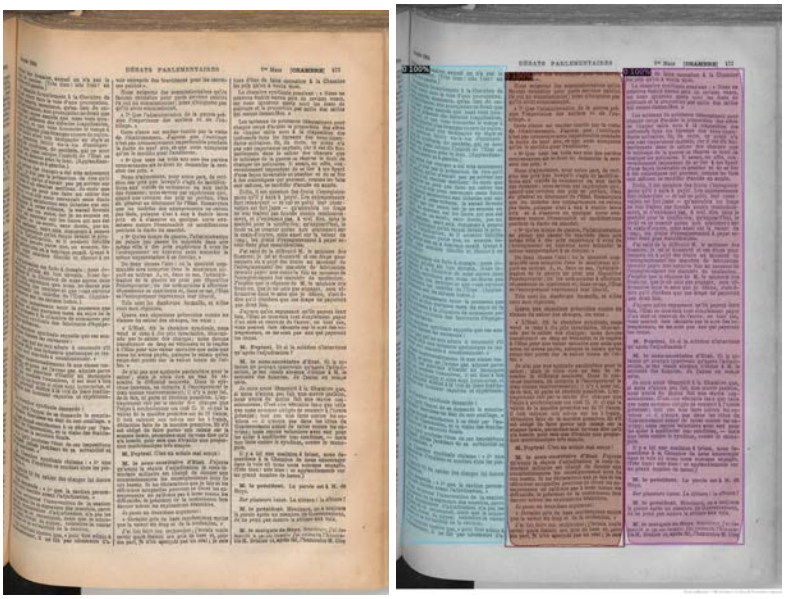
Je donne la parole à M. Réveilland dans la discussion générale.

M. Eugène Réveilland. Messieurs, j'ai le très grand et redoutable honneur, que je n'ai pas cherché, de succéder à cette tribune à l'orateur éloquent, à l'homme d'État éminent qu'est l'honorable M. Ribot. J'en sens vivement tout le poids. Aussi, pour me justifier d'intervenir dans la discussion générale, alors que plusieurs d'entre vous commencent à sollicier que cette discussion a déjà beaucoup duré, que ce débat est presque épuisé (*Dénégations sur plusieurs bancs*), et de prendre la parole après des orateurs aussi autorisés et aussi éminents que la plupart de ceux que nous avons entendus, j'éprouve le besoin de m'autoriser d'un autre titre que du droit qui appartient à chacun de nous d'apporter ici son opinion personnelle dans un débat de première importance pour l'avenir de notre pays, pour les intérêts moraux, politiques et religieux qui sont engagés dans ce grand débat. Ce titre j'ai pensé le trouver dans ce fait — que vous me permettrez de rappeler — que j'ai été en quelque mesure le promoteur, l'initiateur de la motion par laquelle cette grande idée de la séparation des Églises et de l'État, depuis longtemps agitée dans le vague de manifestations de presse, a pris corps dans le Parlement par le projet de loi qui nous est aujourd'hui soumis et a été amenée au point de précision et de maturité où nous la voyons aujourd'hui. Je veux faire allusion à mon intervention dans la séance du 20 octobre 1902, alors que cette Chambre était saisie d'une proposition de loi d'un de ses membres, M. Ernest Roche, qui nous demandait, nous sommait presque de faire au pied levé, toutes autres affaires cessantes, la séparation des Églises et de l'État en prononçant l'urgence sur sa proposition. Or cette proposition m'apparaissait, ainsi qu'à plusieurs de mes collègues les mieux disposés pour le principe de la réforme, comme insuffisamment étudiée et mûrie.

Par exemple, l'article 1er décidait que « toutes les Églises aujourd'hui reconnues et subventionnées étaient séparées de l'État » *ipso et sanc*. C'était donc le Concordat supprimé en fait ; or, un article subséquent — c'était l'article 6 — prononçait qu'« une loi spéciale aurait à déterminer le mode à employer pour la dénonciation du Concordat ». (*On rit.*)

Laisser la Chambre se prononcer sur l'urgence de cette proposition, c'était risquer de la faire repousser par la majorité de cette Assemblée et de laisser ainsi créer une prévention défavorable au principe même qui nous était cher, et cela contre les intentions et les vues de plusieurs d'entre nous qui désiraient que la séparation fût votée dans cette législature même, mais qui se rendaient compte qu'elle ne pouvait l'être qu'à la condition d'avoir été au préalable sérieusement étudiée et préparée. (*Très bien ! très bien ! à gauche.*)

Pour éviter donc le risque que le rejet



(a) Standard Page: Raw (b) Standard Page: Processed

(c) Page with Table: Raw (d) Page with Table: processed

Figure A.6: Parsing the Layout of Parliamentary Transcripts

Note: The figure illustrates our procedure for parsing the layout of the parliamentary transcripts. We first train an image-detection model to identify pages containing tables. For standard pages with a three-column structure (a), we then use an image-detection model to identify column breaks (b). For pages containing a table (c), we strip the table by setting its pixels to white and use an image-detection model to detect the remaining text regions (d)

A.2.2 Parsing Parliamentary Speeches

Our next task is to split the digitized transcripts into individual speeches. First, we separate distinct sittings within each transcript document. Second, we partition each sitting into consecutive sections of the debate, which we refer to as discussions, each centered on a specific issue or bill that Parliament was called upon to examine. Finally, we split each discussion into individual speeches.

As illustrated in Figure A.5, each sitting begins with a short agenda summary under the header "SOMMAIRE." The summary is followed by a header naming the presiding officer (e.g., "PRESIDENCE DE M. PAUL DOUMER"), after which the chair announces the start of proceedings with the formula "the sitting is open" (*la séance est ouverte* in French). Each agenda item is then introduced by an uppercase header (for instance, the second item in Figure A.5 concerns the layout of a railway line between Cand'e and Angers). Finally, each intervention starts on a new line, typically after a period or an uppercase header, and the speaker is introduced by "M." followed by a surname (e.g., M. Plichon) or full name (e.g., M. Eugène Réveillaud), ending with a period. We leverage these conventions to construct regular expressions that identify sitting boundaries, section headers, and to extract individual speeches along with the reported name of the speaker. At the end of each sitting (though not shown in Figure A.5), the chair declares that the sitting is adjourned ("*la séance est levée*"), and the transcript concludes with the signature of the head of the stenographic service ("*le chef du service sténographique*"). We use these markers to detect the end of the debate section and to exclude supplementary material that is sometimes appended to the transcript. Our final sample comprises 1,660,083 speeches, distributed across 82,364 discussions and 8,191 parliamentary sittings; 1,173,687 of these speeches -— 70% of the total —- were delivered in the Chamber of Deputies.

A.2.3 Cleaning Speaker Names

After extracting individual speeches, we standardize the speaker labels recorded in the transcripts to obtain cleaned name fields. We use these cleaned names to link speakers to the biographical dataset of elected MPs, as described in Section A.5.2 below.

As a first step, we flag speeches delivered by (i) rapporteurs, (ii) members of government, (iii) the President of the Chamber, and (iv) other state officials (e.g., secretaries and commissioners), because the speaker label does not always report their personal name, alongside their role. We detect their role using simple keyword patterns (e.g., *rapporteur*, *ministre*, *président*). Rapporteurs are typically identified by name in their first intervention within a debate section (e.g., "M. Allain-Targé, rapporteur") and then referred to with the shorthand "M. le rapporteur" in subsequent interventions. Accordingly, we propagate the rapporteur's name downward within the same discussion. The president and (in roughly 90% of all cases) the members of government) are identified only by their official role and follow a different matching pipeline. By contrast, we do not attempt to recover the identities of other

state officials and match them to the biographical dataset.

Second, we clean speaker names by removing accents, converting text to lower case, and stripping military ranks and aristocratic titles that do not appear in the name fields of the biographical dataset.

Finally, we use the cleaned speaker label to construct two name fields used in the matching step: (i) a cleaned surname, where the surname is defined as the last capitalized name component, including apostrophes and hypens, in the original speaker label; and (ii) a list of all cleaned name components, even when separated by hyphens, and dropping common particles such as *de*, *du*, and *d'*. To illustrate, for M. Clémenceau," both fields take the value `clemenceau`. For "M. Georges Perin," the cleaned surname is `perin` and the list of name components is `georges, perin`. Finally, for "M. Laroche-Joubert", these fields are `laroche-joubert` and `laroche, joubert`.

When multiple components are present, it is difficult to determine whether they correspond to a first name and a surname (as in "M. Georges Perin") or to a multi-part surname (as in "M. Laroche-Joubert"). Moreover, the same individual may appear under variants such as "M. Laroche Joubert" (no hyphen). In most cases, the cleaned surname is sufficient to obtain a unique match with the biographical dataset; the full list of name components helps resolve ambiguous cases and improve robustness to such spelling variants.

A.3 Members of Parliament

Data on the personal and political biographies of all deputies and senators of the French Third Republic are available on the institutional websites of the *Assemblée nationale* (today's lower chamber) and the *Sénat*, respectively.[12] Each MP has a dedicated page reporting his name, birth and death details, along with his political biography. For deputies, the pages list each term served in the Chamber of Deputies, including start and end dates, department of election, and political affiliation. For senators, the pages report election dates and the date on which the senator left office, and we use these election dates to reconstruct the start and end dates of each senatorial term. By contrast, political affiliation is not systematically available for senators, but can only infer from the biographical information for part of them. Finally, each page includes one or two biographical excerpts drawn from the following references: the *Dictionnaire des parlementaires français de 1789 à 1889* by Adolphe Robert and Gaston Cougny, and the *Dictionnaire des parlementaires français de 1889 à 1940* by Jean Jolly. Lanzara et al. (2024) scraped these pages and extensively reviewed and standardized these biographical fields, correcting minor inconsistencies in birth information and in mandate start and end dates. Importantly for our purposes, the name field for each MP is parsed in three separate components: first names, surname, and a list of all name components.

[12] See www2.assemblee-nationale.fr/sycomore/recherche for the *Assemblée nationale* and www.senat.fr/senateurs-3eme-republique/senatl.html for the Senate.

We complement this dataset in two ways. First, we map the parliamentary groups reported for the Chamber of Deputies into five broad political areas: left, center-left, center, center-right, and right. This mapping relies on the classifications in Cagé and Piketty (2025) and is supplemented with information from https://www.france-politique.fr/ for groups that appear in our data but are not covered in Cagé and Piketty. Importantly, the assignment of political groups to broad areas is time-varying. For instance, the Republican Radicals are classified as left in 1881 but as center-left in 1885, reflecting a shift in the political spectrum following the rise of the Socialist party.

Second, we use ChatGPT to infer each MP's occupation from the available biographical excerpts, focusing on the following pre-defined list of broad professional groups: other government officials, other skilled professionals, legal professionals, skilled manual workers, industrial owners, education professionals, senior government officials, military, clergy, agricultural owners, unskilled workers, traders, and NA (when no occupation is reported). Specifically, we define the system role as follows:

> You are a classification engine. Given a text that contains one or more biographical excerpts of a French MP, return the most appropriate occupation category or categories from the following list: {categories}. Return one or more categories when multiple occupations are clearly indicated in the text. Separate multiple categories using a semicolon (;) and no other characters. Do not include any explanations, descriptions, or extra formatting. Output only the category or categories exactly as listed.

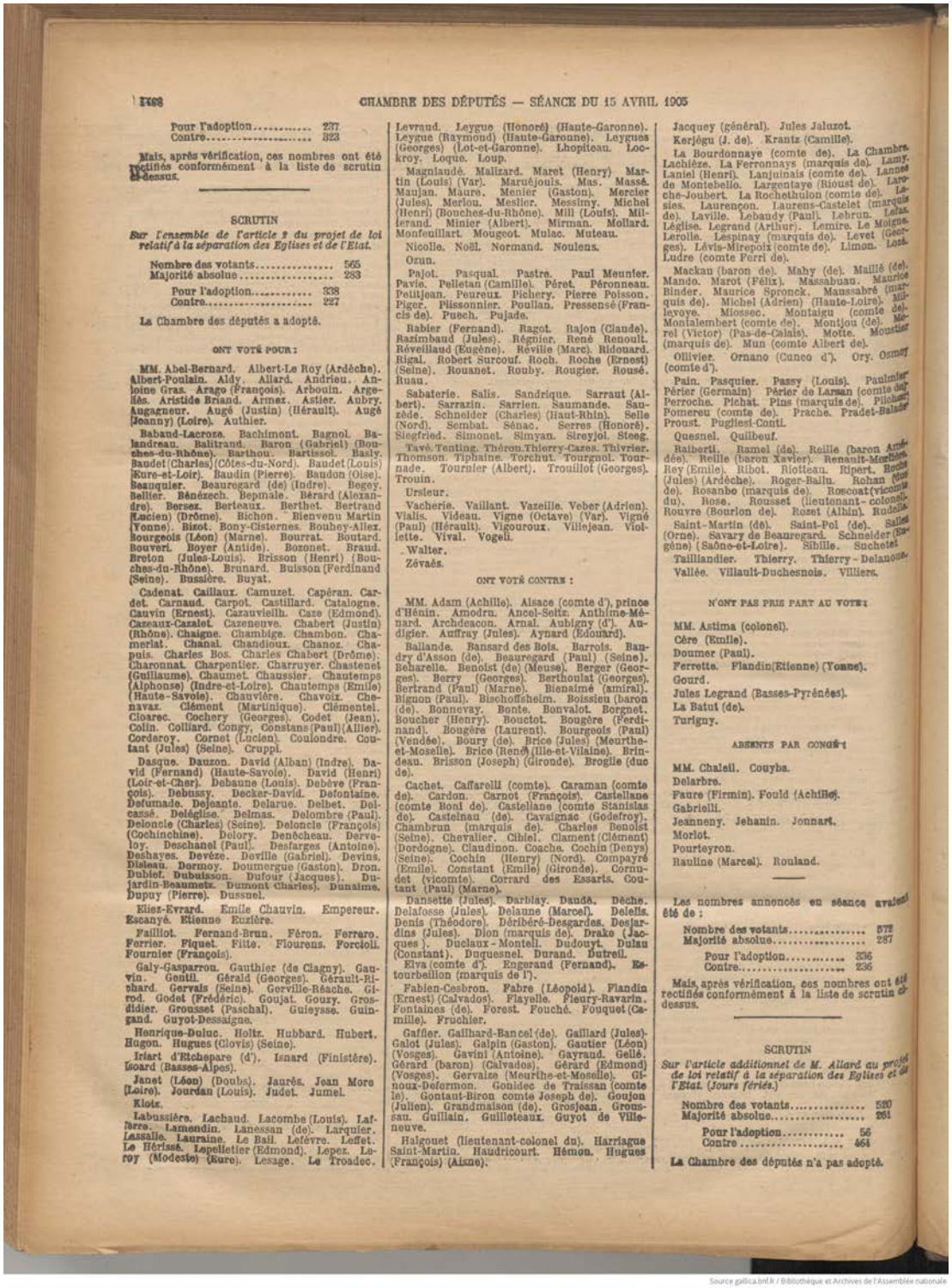
The user prompt then reads:

```
Based on the following biographical excerpt, classify the MP's
primary occupation(s) in one or more of the following
categories: {categories}.
Separate multiple occupations with a semicolon (;). Here is the
excerpt: {excerpt}.
```

When ChatGPT detects multiple occupations, we assign the first one.

A.4 Voting

The PDFs of the *Journal Officiel* available on Gallica contain not only the transcript of parliamentary debates, but also an appendix section reporting additional materials, including the results of all roll-call votes that took place during the corresponding sitting. As illustrated in Figure A.7, for each vote the documents report the title of the legislative act under consideration, the total number of votes in favor and against, the outcome of the vote, and, finally, the list of parliamentarians associated with each voting position: in favor (*pour*), against (*contre*), abstaining (*abstenu*), or absent (*absent*).



1488 CHAMBRE DES DÉPUTÉS — SÉANCE DU 15 AVRIL 1905

Pour l'adoption..... 237
Contre..... 323

Mais, après vérification, ces nombres ont été rectifiés conformément à la liste de scrutin ci-dessus.

SCRUTIN

Sur l'ensemble de l'article 2 du projet de loi relatif à la séparation des Eglises et de l'Etat.

Nombre des votants..... 565
Majorité absolue..... 283

Pour l'adoption..... 338
Contre..... 227

La Chambre des députés a adopté.

ONT VOTÉ POUR :

MM. Abel-Bernard. Albert-Le Roy (Ardèche). Albert-Poulain. Aldy. Allard. Andrieu. Antoine Gras. Arago (François). Arbouin. Argelliès. Aristide Briand. Armez. Astier. Aubry. Augagneur. Augé (Justin) (Hérault). Augé (Joanny) (Loire). Authier.

Babaud-Lacroze. Bachimont. Bagnol. Balandreau. Balitrand. Baron (Gabriel) (Bouches-du-Rhône). Barthou. Bartissol. Basly. Baudet (Charles) (Côtes-du-Nord). Baudet (Louis) (Eure-et-Loir). Baudin (Pierre). Baudon (Oise). Beauquier. Beauregard (de) (Indre). Begey. Bellier. Bénézech. Bepmale. Bérard (Alexandre). Bersez. Berteaux. Berthet. Bertrand (Lucien) (Drôme). Bichon. Bienvenu Martin (Yonne). Bizot. Bony-Cisternes. Bouhey-Allex. Bourgeois (Léon) (Marne). Bourrat. Boutard. Bouvert. Boyer (Antide). Bozonet. Braud. Breton (Jules-Louis). Brisson (Henri) (Bouches-du-Rhône). Brunard. Buisson (Ferdinand (Seine). Bussière. Buyat.

Cadenat. Caillaux. Camuzet. Capéran. Cardet. Carnaud. Carpot. Castillard. Catalogne. Cauvin (Ernest). Cazauvieilh. Caze (Edmond). Cazeaux-Cazalet. Cazeneuve. Chabert (Justin) (Rhône). Chaigne. Chambige. Chambon. Chamerlat. Chanal. Chandioux. Chanoz. Chapuis. Charles Bos. Charles Chabert (Drôme). Charonnat. Charpentier. Charruyer. Chastenet (Guillaume). Chaumet. Chaussier. Chautemps (Alphonse) (Indre-et-Loire). Chautemps (Emile) (Haute-Savoie). Chauvière. Chavoix. Chenavaz. Clément (Martinique). Clémentel. Cloarec. Cochery (Georges). Codet (Jean). Colin. Colliard. Congy. Constans (Paul) (Allier). Corderoy. Cornet (Lucien). Coulondre. Coutant (Jules) (Seine). Cruppi.

Dasque. Dauzon. David (Alban) (Indre). David (Fernand) (Haute-Savoie). David (Henri) (Loir-et-Cher). Debaune (Louis). Debève (François). Debussy. Decker-David. Defontaine. Defumade. Dejeante. Delarue. Delbet. Delcassé. Deléglise. Delmas. Delombre (Paul). Deloncle (Charles) (Seine). Deloncle (François) (Cochinchine). Delory. Denêcheau. Derveloy. Deschanel (Paul). Desfarges (Antoine). Deshayes. Devèze. Deville (Gabriel). Devins. Disleau. Dormoy. Doumergue (Gaston). Dron. Dubief. Dubuisson. Dufour (Jacques). Dujardin-Beaumetz. Dumont (Charles). Dunaime. Dupuy (Pierre). Dussuel.

Elies-Evrard. Emile Chauvin. Empereur. Escanyé. Etienne Euzière.

Failliot. Fernand-Brun. Féron. Ferrero. Ferrier. Fiquet. Flitz. Flourens. Forcioli. Fournier (François).

Galy-Gasparrou. Gauthier (de Clagny). Gauvin. Gentil. Gérald (Georges). Gérault-Richard. Gervais (Seine). Gerville-Réache. Girod. Godet (Frédéric). Goujat. Goury. Grosdidier. Grousset (Paschal). Guieysse. Guingand. Guyot-Dessaigne.

Henrique-Duluc. Holtz. Hubbard. Hubert. Hugon. Hugues (Clovis) (Seine).

Iriart d'Etchepare (d'). Isnard (Finistère). Isoard (Basses-Alpes).

Janet (Léon) (Doubs). Jaurès. Jean Morc (Loire). Jourdan (Louis). Judet. Jumel.

Klotz.

Labussière. Lachaud. Lacombe (Louis). Laffarre. Lamendin. Lanessan (de). Larquier. Lassalle. Lauraine. Le Bail. Lefèvre. Leffet. Le Hérissé. Lepelletier (Edmond). Lepez. Leroy (Modeste) (Eure). Lesage. Le Troadec. Levraud. Leygue (Honoré) (Haute-Garonne). Leygue (Raymond) (Haute-Garonne). Leygues (Georges) (Lot-et-Garonne). Lhopiteau. Lockroy. Loque. Loup.

Magniaudé. Malizard. Maret (Henry) Martin (Louis) (Var). Maruéjouls. Mas. Massé. Maujan. Maure. Menier (Gaston). Mercier (Jules). Merlou. Meslier. Messimy. Michel (Henri) (Bouches-du-Rhône). Mill (Louis). Millerand. Minier (Albert). Mirman. Mollard. Monfeuillart. Mougeot. Mulac. Muteau.

Nicolle. Noël. Normand. Noulens.

Ozun.

Pajot. Pasqual. Pastre. Paul Meunier. Pavie. Pelletan (Camille). Péret. Péronneau. Petitjean. Peureux. Pichery. Pierre Poisson. Piger. Plissonnier. Poullan. Pressensé (Francis de). Puech. Pujade.

Rabier (Fernand). Ragot. Rajon (Claude). Razimbaud (Jules). Régnier. René Renoult. Réveillaud (Eugène). Réville (Marc). Ridouard. Rigal. Robert Surcouf. Roch. Roche (Ernest) (Seine). Rouanet. Rouby. Rougier. Rousé. Ruau.

Sabaterie. Salis. Sandrique. Sarraut (Albert). Sarrazin. Sarrien. Saumande. Sauzède. Schneider (Charles) (Haut-Rhin). Selle (Nord). Sembat. Sénac. Serres (Honoré). Siegfried. Simonet. Simyan. Sireyjol. Steeg.

Tavé. Tenting. Théron. Thierry-Cazes. Thivrier. Thomson. Tiphaine. Torchut. Tourgnol. Tournade. Tournier (Albert). Trouillot (Georges). Trouin.

Ursleur.

Vacherie. Vaillant. Vazeille. Veber (Adrien). Vialis. Videau. Vigne (Octave) (Var). Vigné (Paul) (Hérault). Vigouroux. Villejean. Violette. Vival. Vogeli.

Walter.

Zévaès.

ONT VOTÉ CONTRE :

MM. Adam (Achille). Alsace (comte d'), prince d'Hénin. Amodru. Ancel-Seitz. Anthime-Ménard. Archdeacon. Arnal. Aubigny (d'). Audigier. Auffray (Jules). Aynard (Edouard).

Ballande. Bansard des Bois. Barrois. Bandry d'Asson (de). Beauregard (Paul) (Seine). Béharelle. Benoist (de) (Meuse). Berger (Georges). Berry (Georges). Berthoulat (Georges). Bertrand (Paul) (Marne). Bienaimé (amiral). Bignon (Paul). Bischoffsheim. Boissieu (baron de). Bonnevay. Bonte. Bonvalot. Borgnet. Boucher (Henry). Bouctot. Bougère (Ferdinand). Bougère (Laurent). Bourgeois (Paul) (Vendée). Boury (de). Brice (Jules) (Meurthe-et-Moselle). Brice (René) (Ille-et-Vilaine). Brindeau. Brisson (Joseph) (Gironde). Brogile (duc de).

Cachet. Caffarelli (comte). Caraman (comte de). Cardon. Carnot (François). Castellane (comte Boni de). Castellane (comte Stanislas de). Castelnau (de). Cavaignac (Godefroy). Chambrun (marquis de). Charles Benoist (Seine). Chevalier. Cibiel. Clament (Clément) (Dordogne). Claudinon. Coache. Cochin (Denys) (Seine). Cochin (Henry) (Nord). Compayré (Emile). Constant (Emile) (Gironde). Cornudet (vicomte). Corrard des Essarts. Coutant (Paul) (Marne).

Dansette (Jules). Darblay. Daudé. Déche. Delafosse (Jules). Delaune (Marcel). Delelis. Denis (Théodore). Déribéré-Desgardes. Desjardins (Jules). Dion (marquis de). Drake (Jacques). Duclaux-Monteil. Dudouyt. Dulau (Constant). Duquesnel. Durand. Dutreil.

Elva (comte d'). Engerand (Fernand). Estourbeillon (marquis de l').

Fabien-Cesbron. Fabre (Léopold). Flandin (Ernest) (Calvados). Flayelle. Fleury-Ravarin. Fontaines (de). Forest. Fouché. Fouquet (Camille). Fruchier.

Gaffier. Gailhard-Bancel (de). Gaillard (Jules). Galot (Jules). Galpin (Gaston). Gautier (Léon) (Vosges). Gavini (Antoine). Gayraud. Gellé. Gérard (baron) (Calvados). Gérard (Edmond) (Vosges). Gervaize (Meurthe-et-Moselle). Ginoux-Defermon. Gonidec de Traissan (comte le). Gontant-Biron comte Joseph de). Goujon (Julien). Grandmaison (de). Grosjean. Grousau. Guillain. Guilloteaux. Guyot de Villeneuve.

Halgouet (lieutenant-colonel du). Harriague Saint-Martin. Haudricourt. Hémon. Hugues (François) (Aisne).

Jacquey (général). Jules Jaluzot.

Kerjégu (J. de). Krantz (Camille).

La Bourdonnaye (comte de). La Chambre. Lachièze. La Ferronnays (marquis de). Lamy. Lanel (Henri). Lanjuinais (comte de). Lannes de Montebello. Largentaye (Rioust de). Laroche-Joubert. La Rochethulon (comte de). Lasies. Laurençon. Laurens-Castelet (marquis de). Laville. Lebaudy (Paul). Lebrun. Lelas. Léglise. Legrand (Arthur). Lemire. Le Moigne. Lerolle. Lespinay (marquis de). Levet (Georges). Lévis-Mirepoix (comte de). Limon. Lost. Ladre (comte Ferri de).

Mackau (baron de). Mahy (de). Maillé (de). Mando. Marot (Félix). Massabuau. Maurice Binder. Maurice Spronck. Maussabré (marquis de). Michel (Adrien) (Haute-Loire). Mille voye. Miossec. Montaigu (comte de). Montalembert (comte de). Montjou (de). Morel (Victor) (Pas-de-Calais). Motte. Moustier (marquis de). Mun (comte Albert de).

Ollivier. Ornano (Cunco d'). Ory. Osmoy (comte d').

Pain. Pasquier. Passy (Louis). Paulmier. Périer (Germain). Périer de Larsan (comte du). Perrochs. Pichat. Pins (marquis de). Pilchery. Pomereu (comte de). Prache. Pradet-Balade. Proust. Pugliesi-Conti.

Quesnel. Quilbeuf.

Raiberti. Ramel (de). Reille (baron Armédée). Reille (baron Xavier). Renault-Morlière. Rey (Emile). Ribot. Riotteau. Ripert. Roche (Jules) (Ardèche). Roger-Ballu. Rohan (duc de). Rosanbo (marquis de). Roscoat (vicomte du). Rose. Rousset (lieutenant-colonel). Rouvre (Bourlon de). Rozet (Albin). Rudelle.

Saint-Martin (de). Saint-Pol (de). Salles (Orne). Savary de Beauregard. Schneider (Eugène) (Saône-et-Loire). Sibille. Sochetet.

Tailliandier. Thierry. Thierry-Delanoue. Vallée. Villault-Duchesnois. Villiers.

N'ONT PAS PRIS PART AU VOTE :

MM. Astima (colonel).
Côte (Emile).
Doumer (Paul).
Ferrette. Flandin (Etienne) (Yonne).
Gourd.
Jules Legrand (Basses-Pyrénées).
La Batut (de).
Turigny.

ABSENTS PAR CONGÉ :

MM. Chalell. Couyba.
Delarbre.
Faure (Firmin). Fould (Achille).
Gabrielli.
Jeanneny. Jehanin. Jonnart.
Morlot.
Pourteyron.
Rauline (Marcel). Rouland.

Les nombres annoncés en séance avaient été de :

Nombre des votants..... 572
Majorité absolue..... 287

Pour l'adoption..... 336
Contre..... 236

Mais, après vérification, ces nombres ont été rectifiés conformément à la liste de scrutin ci-dessus.

SCRUTIN

Sur l'article additionnel de M. Allard au projet de loi relatif à la séparation des Eglises et de l'Etat. (Jours fériés.)

Nombre des votants..... 520
Majorité absolue..... 261

Pour l'adoption..... 56
Contre..... 464

La Chambre des députés n'a pas adopté.

Source gallica.bnf.fr / Bibliothèque et Archives de l'Assemblée nationale

Figure A.7: Transcript of Roll-Call Vote – Chamber of Deputies, April 15th, 1905

Notes: The figure shows the outcome of the roll-call vote on Article 2 of the Law of Separation of Church and State, held during the parliamentary sitting of April 15, 1905, in the Chamber of Deputies.

Unfortunately, the structure of these pages makes it difficult to automatically link voting MPs to our dataset of parliamentarians in a sufficiently complete and reliable way. Furthermore, it is challenging to connect roll-call votes that occurred on different dates – sometimes spanning several weeks or months – into a coherent sequence corresponding to the same bill. These challenges motivated us to focus on a single piece of legislation (the Separation Law) and digitize the relevant roll-call votes manually.

First, we identify 27 roll-call votes concerning the approval of individual articles (such as the roll call illustrated in Figure A.7), paragraphs within articles, as well as the full text of the bill. The first roll call in this sequence took place on April 9, 1905, with a vote on the first article of a counter-proposal presented by parliamentarian Urbain Sénac, while the final roll call concerned the full text of the bill took place on July 3, 1905.

Second, we use information on political mandates from our biographical dataset to construct the list of all MPs who were in office on the day of each roll-call vote, and then manually digitize their voting positions using the lists illustrated in Figure A.7.

Third, for each roll-call vote, we determine whether a recorded "for" vote reflects *support* for the overall agenda of the Separation Law. Specifically, we classify a "for" vote as supporting secularization if left-wing parliamentarians were, on average, at least 25% more likely than right-wing parliamentarians to vote in favor. We adopt this procedure because the political objective of a roll call is often difficult to infer from its title alone (e.g., "On the additional article by M. Allard"), and the full text of the proposed measure is not consistently available in the parliamentary records.

A.5 Linking Datasets

A.5.1 Matching Students with Parliamentarians

We match students from Jesuit schools and parliamentarians based on their names and year of birth. For parliamentarians, we know the exact date of birth. For students, we approximate their year of birth using either information on their imputed age, where available (see Section A.1.3), or the number of years during which they are observed in a Jesuit school. Note that grade-based age information is available for approximately 90% of all records and students.

When age information is available for a specific record, we estimate the implied year of birth for record r as $y_r^{\min} - \text{age}_r$. We then aggregate all records linked to the same student (see Section A.1.2 for details on the linkage procedure) and set $ybirth_i^{min}$ and $ybirth_i^{max}$ equal to, respectively, the minimum and the maximum implied years of birth across all records for the same students.

For the remaining 10% of students whose age information is missing from all linked records, we impute a plausible birth year range based on the years during which the student is observed in a Jesuit school. Jesuit students typically attended school between the ages of 8 and 18, with a full trajectory

lasting 11 years. Based on this, we compute a conservative interval for the year of birth:

$$ybirth_i^{\text{min}} = \min(y_i^{\text{min}} - 19 + n,\ y_i^{\text{max}} - 7 - n),$$
$$ybirth_i^{\text{max}} = \max(y_i^{\text{min}} - 19 + n,\ y_i^{\text{max}} - 7 - n),$$

where n is the number of years the student appears in the data, capped at 11. This formula incorporates the idea that the more years a student is observed, the more accurately we can infer their age. To see this, consider two edge cases:

- If a student is observed for the full eleven years ($n = 11$), then we can assume that the earliest observation corresponds to age 8 and the latest to age 18. This implies a year of birth between $y_i^{\text{min}} - 8$ and $y_i^{\text{max}} - 18$.
- If a student is observed for only one year ($n = 1$), this record may correspond either to the beginning or the end of their school career, leading to a wider range of possible birth years.

For each parliamentarian, we then look for a match in the dataset of linked students. Our matching procedure consists of four steps:

- First, as a minimal consistency requirement, we restrict attention to students who appear in a record when the parliamentarian was older than 4 and younger than 22.
- Second, we only retain students whose birth window overlaps with the MP's birth year, within a margin of 4 years on either side.
- Third, we identify students whose last name is similar to the MP's last name using the Jaro string distance metric, with a cutoff of 0.1. Because, after linking the student records, we may observe different variants for a student's last name, we consider each of the variants.
- Fourth, for students whose last names matched the MP's last name in step 3, we apply a similar fuzzy matching procedure to the first names.[13]

Finally, we manually review all fuzzy matches and retain only those where discrepancies clearly stem from minor typos or spelling variations that do not affect French pronunciation (e.g., *Louis Du Bochet* and *Louis Dubochet*). This procedure identifies 133 parliamentarians – 123 exact matches and 10 fuzzy matches – who served in the French Parliament between January 1881 and April 1914. Of these, 104 served only in the Chamber of Deputies, 16 served only in the Senate, and 13 served both in the Chambre of Deputies and in the Senate.

[13] For parliamentarians with multiple given names, we consider only the first given name.

A.5.2 Matching Speakers with Parliamentarians

We link the extracted speeches to our biographical dataset of MPs using the speaker name and the date of the parliamentary sitting. For each sitting, we first identify the set of unique *speakers* who took the floor, then compare these speakers only to *MPs* who were in office in that branch of Parliament on that date, allowing a small buffer around mandate boundaries to accommodate minor inconsistencies in recorded mandate dates. Then, we apply a fuzzy name-matching algorithm to link speakers to MPs, based on Levenshtein distance. More specifically, our matching procedure consists of six steps:

1. First, for each branch–date combination, we extract the list of unique speakers appearing in the transcript and retain a single observation per speaker.

2. Second, for each branch–date, we retrieve from the biographical dataset the MPs who were in office on that date, using the start and end dates of their mandates. To account for small recording errors, we apply a seven-day buffer on each side of the mandate interval.

3. Third, for speakers whose cleaned name consists of a single component, we attempt to match directly to MPs' cleaned surnames. We keep candidate matches when the Levenshtein distance is small (e.g., distance ≤ 1), including ties, and we remove successfully matched speakers from subsequent steps.

4. Fourth, for remaining speakers with one name component, we attempt matches against MPs' multi-part surnames. This step accounts for cases where speaker appears with a single surname in the transcripts, but this surname is not the one that we identified, for parliamentarian with multi-part surnames, from the biographical dataset.

5. Fifth, for speakers with multi-part names, we match the full lists of name components across the two datasets using Levenshtein distance. For each candidate speaker–MP pair, we first compute the matrix of pairwise Levenshtein distances between all speaker name components and all MP name components. We then construct a one-to-one assignment by iteratively selecting the smallest-distance entry in this matrix and removing the corresponding row and column, so that each component can be used at most once. The overall match distance for the pair is defined as the average Levenshtein distance across these matched component pairs.

6. Sixth, we compute tie diagnostics based on (i) the number of equally good distance matches and (ii) the number of matched name components between the speaker and the MP.

This procedure returns one link per speaker, along with the Levenshtein distance of the match and the number of ties. We classify a speaker–MP link as a successful match when it is both sufficiently close in distance and unambiguous, i.e., no ties among best matches.

B Text Analysis

B.1 Text Preprocessing

We preprocess the text to reduce dimensionality (Gentzkow et al., 2019). First, we assign part-of-speech (POS) tags [14], retaining only adjectives, adverbs, nouns, and verbs from each speech. Second, we normalize the tokens by removing accents, converting all text to lowercase, stripping punctuation and digits, and keeping only words composed entirely of alphabetic characters. Third, we apply the French implementation of the Porter2 stemming algorithm provided in the R package `SnowballC` Porter (2009) to reduce each word to its morphological root.

B.2 Removing Stopwords and Procedural Words

We further preprocess the text data to remove uninformative words. First, we remove short words consisting of fewer than three characters. Second, we filter out rare words that occur in five ore fewer documents or ten or fewer times in the entire corpus. Third, we remove stopwords using a custom list of French stopwords.[15] Next, we remove *procedural* language that is specific to parliamentary discourse. Following Gentzkow et al. (2019), we digitized the Rules of Procedure (*Règlements*) of the Chamber of Deputies and the Senate during the Third Republic, as compiled in *Les Règlements des Assemblées Législatives de France depuis 1789* by Bonnard (1926). We use these texts to identify a set of procedural stopwords – terms frequently used in the context of legislative procedure. Specifically, we implement the following steps: (i) We pre-process the text of the Rules of Procedure using the same pipeline described in Section B.1 (except for the POS tagging). (ii) We train an auxiliary Word2Vec model on the full corpus of parliamentary speeches and use it to generate vector embeddings for each word in the Rules of Procedure. (iii) Using these embeddings, we construct a frequency-weighted centroid vector that captures the semantic content of procedural language. (iv) We compute the cosine distance between each (vectorized) word in the corpus of speeches and the procedural centroid. (v) We rank all words by this distance and remove those falling in the bottom 1% (i.e., most similar to the procedural language centroid). Figure A.8 shows a word cloud of the 100 words in the corpus with the smallest cosine distance to the centroid of procedural terms.

After removing short words, rare words, stopwords, and procedural words, our sample comprises 1,468,790 speeches with nonempty text.

[14] We use the French language model (`fr_core_news_lg`) from the Python library `SpaCy`

[15] We use the list of French stopwords available in the R package `stopwords`, and supplement it with the days of the week, months, and names of French departments, plus 12 frequent low-information terms: (e.g., *faire*, *parl*, *temps*, *nombre*, etc.)



Figure A.8: Procedural Lexicon Removed from the Corpus

Note: The figure shows a word cloud of the 100 words in the corpus with the smallest cosine distance to the centroid of procedural terms, computed using an auxiliary word-embedding model. The procedural centroid is defined as a frequency-weighted average of the (vectorized) words contained in the Rules of Procedure of the French Parliament during the Third Republic. After removing these procedural terms from the corpus, we train the main word-embedding model.

B.3 Word Embedding

We use this sample of preprocessed speeches to train a Word2vec model using the Python `gensim` package (Mikolov et al., 2013). We train the model with 300 dimensions and using a context window of 8 words. As a result, we obtain vector representations of all words in the corpus.

We then compute a document embedding for each speech as the weighted average of its word vectors. We apply inverse-frequency weighting to down-weigh frequent words in the corpus. Specifically, for a given speech $\mathcal{D}_i$, the document vector $\vec{d_i}$ is computed as:

$$\vec{d_i} = \frac{1}{|\mathcal{D}_i|} \sum_{w \in \mathcal{D}_i} \frac{\alpha}{\alpha + p(w)} \vec{w} \tag{A.1}$$

where $\vec{w}$ is the vector representation of word w, $p(w)$ is the relative frequency of the word in the corpus, and $\alpha = 0.001$ is a smoothing parameter [16].

B.4 Topic Analysis

We apply Latent Dirichlet Allocation (LDA) to uncover the latent topics in our corpus of speeches (Blei et al., 2003). The LDA is an unsupervised topic modeling technique that identifies recurring semantic patterns across documents. Each document is modeled as a mixture of topics and each topic as a distribution over words.[17] Figure A.9 shows the word-topic matrix from our preferred specification with 20 topics. Each row corresponds to a topic, represented as a probability distribution

[16] We set the value of α following Gennaro and Ash (2022)

[17] We use the implementation provided in R package `topicmodels` (Grün and Hornik, 2011) with Gibbs sampling. Following Griffiths and Steyvers (2004), we set the Dirichlet prior $\alpha = 2.5$, and $\delta = 0.1$. Following Hansen, McMahon, and Prat (2018), we run the model for 4,000 iterations with a burn-in of 4,000 and a thinning interval of 50, and we average document-topic and word-topic distributions over iterations.

over the entire vocabulary. The ten columns display the ten words with the highest probability weights in each topic, with darker shades indicating higher weights. Most topics have a clear interpretation and correspond to major policy domains. To mention just three examples, Topic 4 is associated with schooling, Topic 8 with labor, and Topic 11 with military affairs.

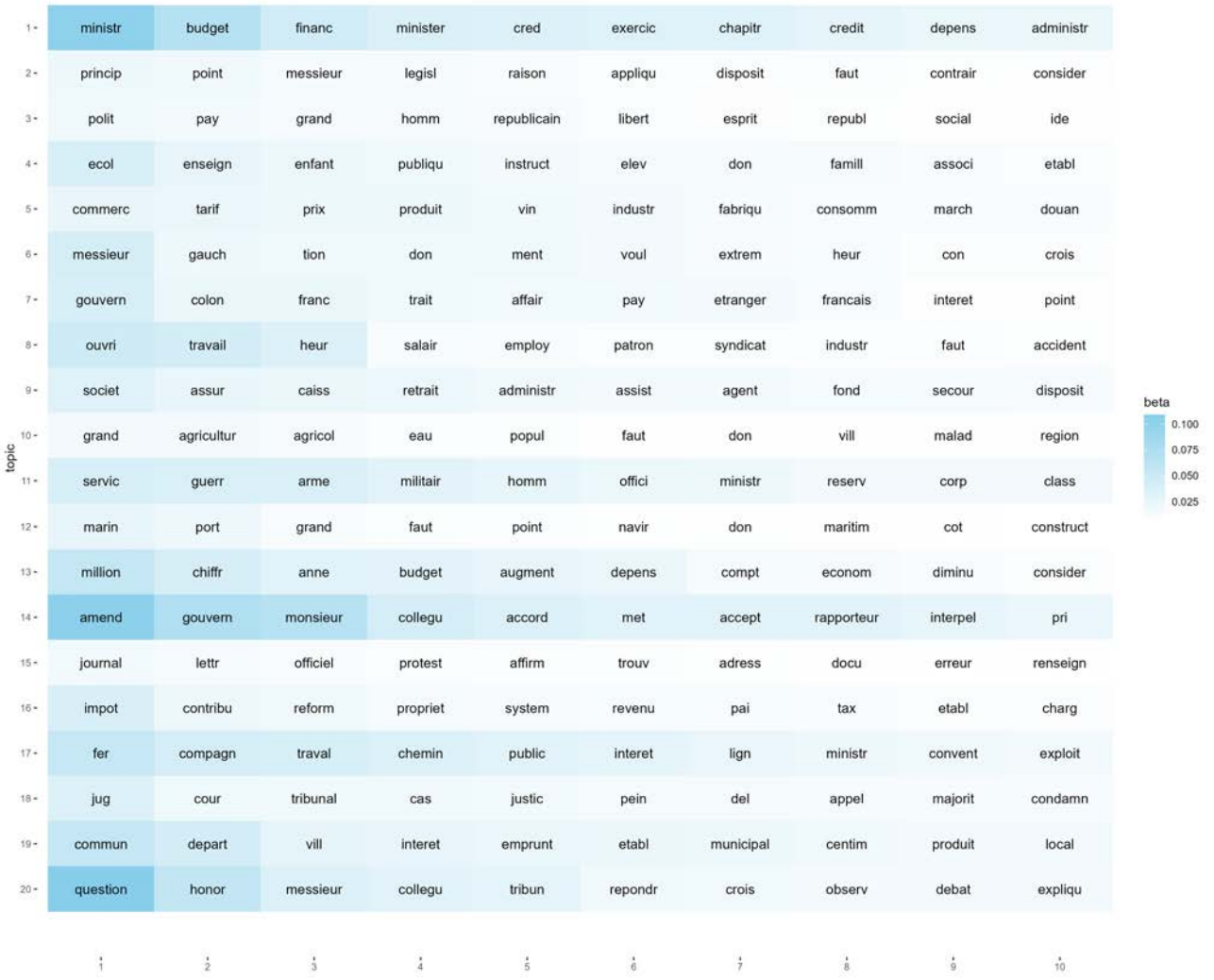


Figure A.9: Word-Topic Distributions

Note: The figure displays the 20 estimated topics from our Latent Dirichlet Allocation analysis. Topics are modeles as probability distributions over words in the vocabulary. Each row corresponds to a topic, and the ten columns display the ten words with the highest probability weights in each topic, with darker shades indicating higher weights.

B.5 Emotionality Score

We follow Gennaro and Ash (2022) to build a measure of emotionality in political speeches. First, we use the French version of the Linguistic Inquiry and Word Count (LIWC) to construct dictionaries of emotional and rational language Tausczik and Pennebaker (2010). LIWC dictionaries were developed to identify linguistic markers of emotions, cognition, and social processes. They are based on expert annotation and psycholinguistic validation of thousands of English words grouped by semantic and psychological categories. We use the French version of these dictionaries available in the 2007 release. Following Gennaro and Ash (2022), we focus on the "Affective Processing" and "Cognitive Processing" categories to capture emotionality and reasoning, respectively. The Affective category includes 1,445 words related to positive and negative emotions and moods, such as anxiety, anger, and sadness. The Cognitive category includes 799 words related to causation, comparison, uncertainty, inclusion, and exclusion. We refine the dictionaries through the following steps: (i) wildcard entries are converted to regular expressions and matched to French lemma names linked to WordNet synsets using NLTK's (Python) multilingual WordNet resources (e.g., wildcard "pens*" is matched to "penser", "pensif", "penseur", etc.) ; (ii) words that appear in both affective and cognitive categories are removed; (iii) for each category, we compute the centroid of the word vectors and exclude the 25% of terms that are most semantically distant from the centroid; (iv) we normalize the remaining words by lower-casing, removing diacritics, and applying stemming (and remove again words shared between the two lists following this step); (v) finally, we eliminate all words found in our list of French stopwords.

Second, we compute the centroid vector of of the emotional and cognitive dictionaries in our word embedding, as the inverse-frequency weighted average of the individual word vectors. Let $\mathbf{E}$ and $\mathbf{C}$ denote the centroids of the emotional and cognitive dictionaries within the word embedding. Using these centroids, we compute the emotionality score of document $\mathcal{D}_i$ as:

$$Escore_i \;=\; \frac{\mathrm{sim}(\vec{d_i}, \mathbf{E}) + 1}{\mathrm{sim}(\vec{d_i}, \mathbf{C}) + 1} \tag{A.2}$$

where $\vec{d_i}$ denotes the vector representation of document $\mathcal{D}_i$ as defined in (A.1) and sim denotes the cosine similarity. As an illustration, Figure A.10 shows the word clouds of the 100 words with the smallest cosine distance to, respectively, the emotional and rational centroids, excluding the words that directly appear in the LIWC dictionaries.



(a) Emotionality



(b) Rationality

Figure A.10: Emotional and Rational Lexicon

Note: The figure shows the word clouds of the 100 words in the corpus with the smallest cosine distance to the centroids of, respectively, the emotional (A.10a) and cognitive (A.10b) LIWC dictionaries, computed with our main word-embedding model. We exclude from the wordclouds the terms that appear in the corresponding dictionary.

B.6 Polarization Score

We adapt Gennaro and Ash (2022)'s methodology to construct a measure of political polarization in language. Broadly speaking, we use the Word2Vec model to compute a "right-wing" alignment score that quantifies the similarity of each speech to right-aligned versus left-aligned political discourse. We then define our *polarization score* as the absolute deviation from the median of this alignment score. This measure captures how much a speech departs from the median rhetorical position, whether toward the left or the right of the ideological spectrum.

Because no categorized dictionaries of politically charged language exist for the French Third Republic, we construct them from our own data, using the speeches of left- and right-aligned parliamentarians. Moreover, since polarized language likely varies over time and across topics, we allow these dictionaries to be year- and topic-specific. We assign speeches to a single topic based on the topic with the highest weight (see Section B.4 for details on the topic analysis).

First, for each year-topic group g, we select speeches by MPs with political affiliation "left" or "right", excluding the president of the chamber and members of government. Second, we compute the centroid vector of the left-wing and right-wing dictionaries for each group g in our word embedding, as the inverse-frequency-weighted average of the individual word vectors. Let $\mathbf{L}_g$ and $\mathbf{R}_g$ denote the inverse-frequency-weighted averages of word vectors appearing in left- and right-wing speeches within group g. Using these centroids, we compute the right-wing alignment score of document $\mathcal{D}_i$ in group g as:

$$rscore_i = \frac{\text{sim}(\vec{d_i}, \mathbf{R}_g) + 1}{\text{sim}(\vec{d_i}, \mathbf{L}_g) + 1}, \tag{A.3}$$

where sim denotes cosine similarity. The polarization score is then defined as the absolute deviation of this alignment from the group median:

$$pscore_i \;=\; \Big|\, rscore_i - \mathrm{median}(rscore) \,\Big|. \tag{A.4}$$



Left-wing Lexicon Right-wing Lexicon

Year 1909, Topic 4: Schooling



Left-wing Lexicon Right-wing Lexicon

Year 1881, Topic 8: Labor



Left-wing Lexicon Right-wing Lexicon

Year 1894, Topic 11: Military Affairs

Figure A.11: Word Clouds for Left- and Right-Wing Lexicons across Selected Years and Topics

Note: Note: The figure shows word clouds of the 100 words in the corpus with the smallest cosine distance to the centroids of left- and right-wing language in some selected years and topics, computed using our main word embedding model. The top panel shows the word clouds for Topic "Schooling" in year 1909; the middle panel shows the word clouds for Topic "Labor" in year 1881; finally, the bottom panel shows the word clouds for Topic "Military Affairs" in 1894.

B.7 Support for Secularization Policies

We apply the same embedding-based methodology used to construct the emotionality and polarization scores to build a measure of support for secularization policies. We focus on the parliamentary debates

on the Law of Separation of churches and State. These debates comprise 12,010 speeches delivered between February and July 1905 in 60 distinct parliamentary discussions. One key advantage of focusing on a single piece of legislation is that roll-call votes allow us to observe each MP's stance on the bill. We leverage this information, in combination with our Word2Vec model, to compute a score that quantifies the similarity of each speech to the average speech in support of, versus in opposition to, the law.

First, we identify MPs who voted consistently in favor of the Secularization Law and MPs who voted consistently against it in both the first and final roll-call votes, thereby excluding MPs who changed position (and may have shifted their language) during the debate. Roll-call votes on the provisions that would define the Law of Separation began on April 16, 1905, when the Chamber approved Article 2, and concluded with the final vote on the full text of the bill on July 3, 1905.[18]

Second, we build two dictionaries from the speeches of these pro-secularization and anti-secularization MPs and represent each dictionary by its centroid in the word-embedding space, denoted $\mathbf{P}$ (pro-law) and $\mathbf{A}$ (anti-law).

Third, and finally, For each speech $\mathcal{D}_i$ with embedding $\mathbf{d}_i$, we compute our *pro-secularization score* as

$$\text{Proscore}_i \;=\; \frac{\text{sim}(\vec{d_i}, \mathbf{P}) + 1}{\text{sim}(\vec{d_i}, \mathbf{A}) + 1}, \tag{A.5}$$

where sim denotes cosine similarity. In analogy with the emotionality and polarization scores, this measure captures how closely a speech aligns with the rhetoric of MPs supporting the law relative to the rhetoric of MPs opposing it.

In the context of the Separation Law case study, the smaller sample size also allows us to complement this measure with a large language model (LLM) assessment of each speech's pro-secularization. Specifically, we asked the GPT4o-mini model (OpenAI) to score the intensity of support for the Separation Law on a five-point scale, with values -2 (strongly against), -1 (moderately against), 0 (neutral), 1 (moderately in favor), and 2 (strongly in favor), allowing for missing values in case of purely procedural or irrelevant interventions. To help the model interpret nuances and implicit references, we include in the prompt, alongside the target speech, up to four immediately preceding speeches from the same discussion, as well as a small set of illustrative examples. Figure A.12 in the appendix shows that the LLM coding is also highly correlated with our Word2Vec-based Proscore defined in equation (5). The full text of the prompt is reproduced below.

[18] Article 2 of the Law of Separation stated: "The Republic does not recognize, pay, or subsidize any religious sect. Accordingly, from 1 January following the enactment of this law, there will be removed from state budgets, departments and municipalities, all expenses related to the exercise of religion".

LLM prompt

```
<instructions>
You are an expert research assistant specializing in French Third Republic parliamentary speech
annotation.

Return both scores *only when* the speech contains an explicit stance on the law or the principle.
If a stance is absent for a score, output "." and supply an uncertainty reason.

1. law_score - Position on the specific law under debate: Law Concerning the Separation of Churches
and State (Secularization Law 1905 - establishing the principle of secularism)
2. principle_score - position on the general principle behind the law (secularism, etc.)

Valid values
  2  = strongly supports
  1  = somewhat supports
  0  = explicitly neutral
 -1  = somewhat opposes
 -2  = strongly opposes
 "." = no explicit stance -> choose an uncertainty reason

When you output ".", choose one reason
  * "lack of content"  - no mention of the law/principle
  * "lack of context"  - reference too vague or external

General guidelines:
- Only assign scores if the speech clearly refers to the **law** or the **principle**.
- If one score is assigned but the other cannot be inferred, assign "." with a reason.
- When in doubt, prefer "." with the appropriate uncertainty reason.
- Return both scores only when the speech contains a clear evaluative judgment about the law or
principle.
- Do not assign a score for mentions, procedural discussions, or rhetorical questions without
explicit support or opposition.
- Neutrality ("0") should only be assigned if explicitly stated.
- Do not infer neutrality from tone or balance.
- Procedural or administrative remarks (e.g., "au moins...") do not justify a score unless
evaluative and clearly tied to law or principle.
- If the speaker proposes an amendment or seeks to change the article, assign a negative law_score.
- If the speech discusses an amendment:
   - Support for the amendment = negative law_score
   - Opposition to the amendment = positive law_score
- Criticizing wording but supporting the idea -> law_score = negative, principle_score = positive.
- Accepting a law but rejecting the principle -> law_score = positive, principle_score = negative.
- A ``support'' or ``oppose'' score requires **explicit evaluative language** (e.g., ``je soutiens'',
``je m'oppose'', ``inadmissible'', ``n\'ecessaire'').
- Mere **mention or description** of the law or its consequences is **not sufficient** for scoring.

</instructions>

<format>
Reply using the following format:
{
  "law_score": [value from -2 to 2 or "."],
```

```
  "law_score_uncertainty_reason": ["" or "lack of content" or "lack of context"],
  "principle_score": [value from -2 to 2 or "."],
  "principle_score_uncertainty_reason": ["" or "lack of content" or "lack of context"]
}
</format>

<examples>
<example>
<speech>
Messieurs, le principal argument de M. le rapporteur, celui qu'il vous presentait en premier lieu
est celui-ci : en prenant les biens des fabriques vous allez empecher l'exercice du culte au
lendemain de la separation. Je n'ai qu'un mot a repondre: c'est que nous ne sommes pas charges
d'assurer, au lendemain de la separation, cet exercice du culte. Cela appartient aux fideles
seuls...
</speech>
<answer>
{
  "law_score": "-2",
  "law_score_uncertainty_reason": "",
  "principle_score": "2",
  "principle_score_uncertainty_reason": ""
}
</answer>
</example>

<example>
<speech>
Ce n'est nullement la ma pensee. J'accepte le texte de la commission avec le revenu global.
</speech>
<answer>
{
  "law_score": "2",
  "law_score_uncertainty_reason": "",
  "principle_score": ".",
  "principle_score_uncertainty_reason": "lack of context"
}
</answer>
</example>

<example>
<speech>
Dans quelques jours on va celebrer de grandes fetes pour l'inauguration de la statue de Gambetta.
[...] Vous le voyez, la foi religieuse n'est pas un obstacle au courage militaire, et bien qu'on se
soit agenouille avant la bataille, on peut se battre bravement ensuite ! Eh bien! je vous demande
d'accorder les secours de la religion a ceux qui, ayant la foi, vont se battre pour la patrie.
</speech>
<answer>
{
  "law_score": ".",
  "law_score_uncertainty_reason": "lack of context",
  "principle_score": "-2",
  "principle_score_uncertainty_reason": ""
}
```

</answer>
</example>

<example>
<speech>
La difference est dans ce fait que nous vivons a une heure ou le but unique de toutes les organisations est la federation; dans tous les sens, politique, social, mutualiste, les efforts tendent a substituer a l'initiative des associations individuelles un organisme central qui reunit toutes les forces isolees et en forme un faisceau. Cette force que vous estimez excellente quand il s'agit de groupements socialistes ou politiques, ne trouvez-vous pas qu'elle peut presenter des inconvenients quand il s'agit de l'Eglise? Je dis que vous aller creer un danger. Le resultat de votre interruption a ete de demontrer qu'il y a un danger a donner la personnalite civile aux petites associations; je le reconnais avec vous, mais le danger est diminue la par ce fait qu'il y en a quatre-vingts et qu'elles n'ont pas un organe central pour se reunir et s'entendre. Vous decuplez le danger si vous creez vous-meme cet organisme central, ce comite qui aura le droit de parler au nom de tous et qui disposera du capital commun. En donnant la personnalite civile a des associations depassant le departement, vous donnez a l'Eglise un instrument qu'elle n'avait jamais eu, vous creez un organisme qu'elle ne reclamait pas, qu'elle repudiait il y a quelques jours, et vous augmentez pour les adversaires de l'Eglise la possibilite des poursuites. J'estime qu'il est necessaire de prendre les precautions que je demande dans mon paragraphe additionnel et je demande a la Chambre de le voter.
</speech>
<answer>
{
  "law_score": "-1",
  "law_score_uncertainty_reason": "",
  "principle_score": "1",
  "principle_score_uncertainty_reason": ""
}
</answer>
</example>

<example>
<speech>
Quand je dis: "...un fonctionnaire... ", je crois employer un mot beaucoup plus precis que celui que vous employez vous-meme, monsieur le rapporteur.
</speech>
<answer>
{
  "law_score": ".",
  "law_score_uncertainty_reason": "lack of content",
  "principle_score": ".",
  "principle_score_uncertainty_reason": "lack of content"
}
</answer>
</example>

<example>
<speech>
M. le ministre a rappele les termes de l'article 45 de la loi de germinal an X; si les termes de cet article semblent indiquer que les manifestations sont interdites sur la voie publique, toutes les fois qu'il y a plusieurs cultes dans une meme commune, en fait, dans beaucoup de villes, a Orleans, par exemple, les manifestations exterieures sont tolerees sur la voie publique.

</speech>
<answer>
{
  "law_score": ".",
  "law_score_uncertainty_reason": "lack of context",
  "principle_score": ".",
  "principle_score_uncertainty_reason": "lack of context"
}
</answer>
</example>

<example>
<speech>
Je regrette que l'accord qui s'etait etabli entre mon honorable collegue M. Lacombe et moi ait si peu dure. M. Lacombe avait une hate extreme de souligner les mots que j'avais prononces et de constater que de ce cote de la Chambre on voulait collaborer a la redaction d'un texte commun. Il semblait que nous etions tous d'accord; je regrette qu'on ait fait de cette question tres simple une question de parti. La commission a propose d'affecter les revenus de tous les biens que l'Etat reprendra, au payement des dettes des fabriques, en admettant que les biens propres de ces fabriques, tombes en desherence, n'etant recueillis par aucune association cultuelle, ne suffiraient pas a cette liquidation. J'ai fait une premiere objection: j'ai dit que peut-etre le revenu pourrait etre insuffisant. M. le ministre des cultes a bien voulu sur ce point nous donner des assurances; je vais y revenir. J'ai fait une seconde objection: c'est qu'il est contraire aux principes de la gestion financiere de l'Etat de constituer ainsi une masse de biens inalienables au ministere des finances, et dont on affecterait le revenu au payement de certaines dettes. Je crois qu'il est plus simple que l'Etat aliene la portion de ces biens necessaire pour assurer ce passif et qu'il reprenne la libre disposition du surplus. Personne de nous n'a interet a creer ce compte special dont la liquidation peut demander un assez grand nombre d'annees. Si M. le ministre des finances etait a son banc, je discuterais la question avec lui; j'ai ete, moi aussi, ministre des finances et je crois que nous tomberions tout de suite d'accord. Mais si le Gouvernement n'a pas d'hesitation, je ne veux pas etre plus ministeriel que les ministres en ce qui concerne la suffisance de ce revenu global qui, d'apres le texte de la commission, doit etre affecte au payement des dettes des fabriques. M. le ministre des cultes engage sa parole et dit qu'il a etudie les chiffres.
</speech>
<answer>
{
  "law_score": "1",
  "law_score_uncertainty_reason": "",
  "principle_score": "1",
  "principle_score_uncertainty_reason": ""
}
</answer>
</example>

<example>
<speech>
Je vous donne egalement satisfaction puisque j'abandonne aux associations l'usufruit d'une maniere constante.
</speech>
<answer>
{
  "law_score": ".",
  "law_score_uncertainty_reason": "lack of context",

```
  "principle_score": "1",
  "principle_score_uncertainty_reason": ""
}
</answer>
</example>

<example>
<speech>
Mais, messieurs, laissez-moi continuer la question en vaut la peine. Il est bien entendu, n'est-ce pas, monsieur le ministre, que vous exigez, et c'est votre devoir, de tous les fonctionnaires, la fidelite absolue aux institutions republicaines; il ne peut pas en etre autrement, mais vous ne les accuserez pas j'espere, d'etre de mauvais republicains si par hasard ils vont a la messe; qu'ils auront le droit d'y aller? M. le ministre de l'instruction publique et des cultes. J'ai declare et je repete que le Gouvernement entend respecter de la maniere la plus absolue la liberte de conscience des fonctionnaires.
</speech>
<answer>
{
  "law_score": "1",
  "law_score_uncertainty_reason": "",
  "principle_score": "2",
  "principle_score_uncertainty_reason": ""
}
</answer>
</example>

<example>
<speech>
Je viens demander aujourd'hui a la gauche de sortir de cette impasse. Il est impossible de trouver une solution a la question telle qu'elle est posee aujourd'hui. Il n'y a pas de devolution de biens possible avec quelque systeme que ce soit. Celui de MM. Caillaux et Leygues ne nous presente pas plus d'avantages que celui de la commission. Il offre meme de graves inconvenients, mais il avait cependant le merite de reserver mieux que le systeme de la commission la propriete collective des biens des fabriques et des menses, que le systeme de M. Briand va individualiser.
</speech>
<answer>
{
  "law_score": "-2",
  "law_score_uncertainty_reason": "",
  "principle_score": "-2",
  "principle_score_uncertainty_reason": ""
}
</answer>
</example>

<example>
<speech>
Je ne veux pas combattre l'amendement qui vous a ete presente par M. Rouanet, je tiens seulement a faire observer a la Chambre que l'article qu'elle vient de voter edicte des penalites en cas d'alienation des objets classes et que, dans ces conditions, il serait bon de viser l'application de l'article 463 du code penal.
</speech>
<answer>
```

```
{
  "law_score": "-1",
  "law_score_uncertainty_reason": "",
  "principle_score": ".",
  "principle_score_uncertainty_reason": "lack of content"
}
</answer>
</example>

<example>
<speech>
Il est excessif de plaider une pareille these !
</speech>
<answer>
{
  "law_score": ".",
  "law_score_uncertainty_reason": "lack of context",
  "principle_score": ".",
  "principle_score_uncertainty_reason": "lack of context"
}
</answer>
</example>

<example>
<speech>
Je crois que le mot "fonctionnaire" a un sens plus restreint que tout autre. Et precisement vous et
vos amis, monsieur le rapporteur, vous avez la preoccupation de distinguer les fonctionnaires
veritables de ceux qui ne le sont pas. Ce souci, je le partage, et comme vous j'ai depose une
proposition de loi pour etendre le droit de se syndiquer a tous employes de l'Etat qui ne sont pas
detenteurs d'une portion de l'autorite publique.
</speech>
<answer>
{
  "law_score": "-2",
  "law_score_uncertainty_reason": "",
  "principle_score": ".",
  "principle_score_uncertainty_reason": "lack of content"
}
</answer>
</example>

<example>
<speech>
Il n'est pas possible d'entendre sans protester un ministre produire une pareille affirmation.
</speech>
<answer>
{
  "law_score": "-2",
  "law_score_uncertainty_reason": "",
  "principle_score": "-2",
  "principle_score_uncertainty_reason": ""
}
</answer>
```

```
</example>

<example>
<speech>
Evidemment la commission a voulu, dans son article 28, remplacer l'article 260 du code penal, lequel vise les entraves que l'on peut apporter au libre exercice d'un culte. L'article 260 ne pouvant subsister, il fallait une redaction conforme au principe nouveau que la Chambre consacre dans la loi qu'elle vote. [...] Si M. le rapporteur veut bien donner quelques explications sur ces trois points, je suis tout pret a retirer mon amendement.
</speech>
<answer>
{
  "law_score": "2",
  "law_score_uncertainty_reason": "",
  "principle_score": "2",
  "principle_score_uncertainty_reason": ""
}
</answer>
</example>

<example>
<speech>
Que peut faire l'Etat? Il la vendra puisque vous lui permettez de la vendre, et une association cultuelle se creera qui achetera l'eglise au prix de l'enchere, c'est-a-dire a un prix derisoire. C'est ainsi que par votre nouvelle redaction vous favorisez certainement la mauvaise foi possible des fabriques. Je demande a la Chambre de maintenir le texte primitif de la commission.
</speech>
<answer>
{
  "law_score": "1",
  "law_score_uncertainty_reason": "",
  "principle_score": ".",
  "principle_score_uncertainty_reason": ""
}
</answer>
</example>
</examples>
### === INPUT SPEECH ===

<speech>
{{text}}
</speech>
```

C Additional Figures

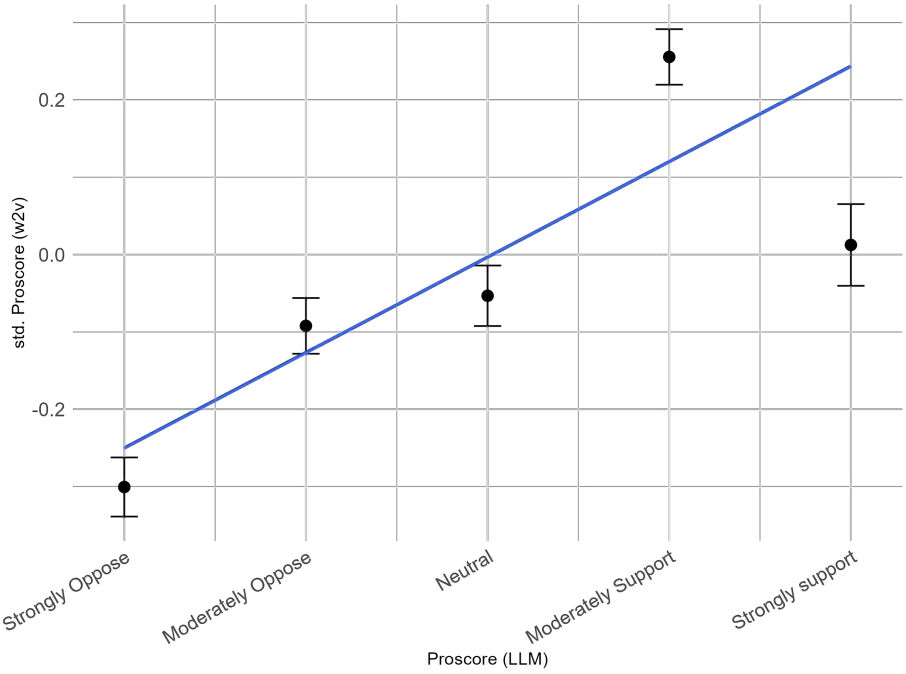
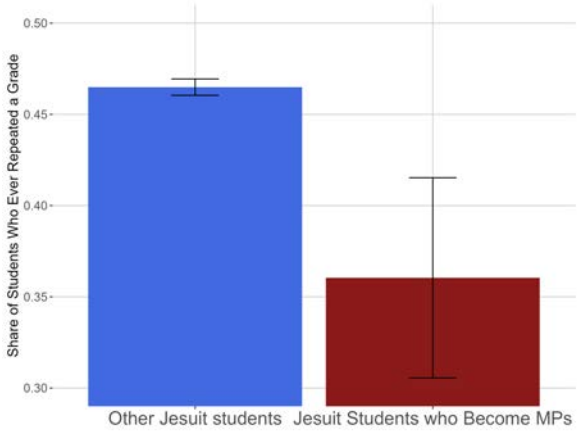
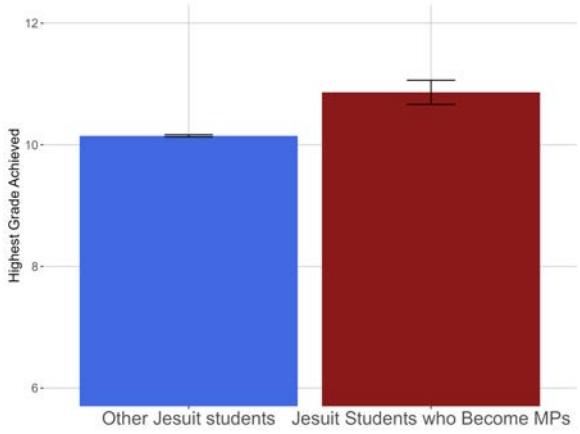


Figure A.12: Correlation between Word2Vec and LLM-based Measures of Support for Secularization

Notes: The figure shows a strong positive relationship between our two main measures of support for secularization policies. The y-axis plots the continuous *Proscore* constructed from our Word2Vec embedding model, while the x-axis shows a categorical score assigned by ChatGPT-4o-mini (OpenAI). The black dots represent the average Word2vec-based Proscore across speeches with the LLM-based Proscore shown on the x-axis. The error bars represent 90% confidence intervals, and the solid blue line is a linear fit. See Section B.7 for details on the construction of these measures.



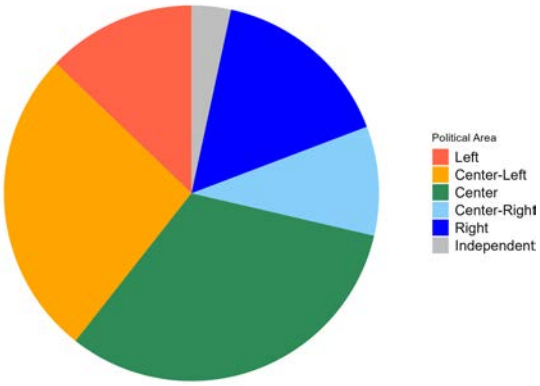
(a) Share of Students Who Ever Repeated a Grade



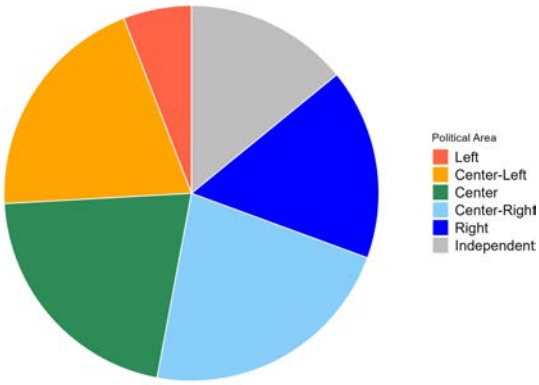
(b) Top Grade Achieved

Figure A.13: School Performance of Jesuit Students who later became MPs

Notes: The figure shows that Jesuit students who later became MPs performed better at school than other Jesuit students. The left panel reports the share of students who ever repeated a grade, whereas the right panel shows the highest grade achieved (with grade 12 being the highest, numbered here in reverse order relative to the French convention of the time). The error bars represent 90% confidence intervals. The sample is restricted to students observed for at least two consecutive years within grades 1 to 12.



(a) Other MPs



(b) Jesuit-educated MPs

Figure A.14: Distribution of MPs across Political Areas

Notes: The figure shows that Jesuit-educated MPs are over-represented among center-right and independent MPs, and under-represented among center-left and left MPs. The sample includes all MPs who served in the Chamber of Deputies between 1881 and 1914 and who are classified in the same political areas across all their mandates.

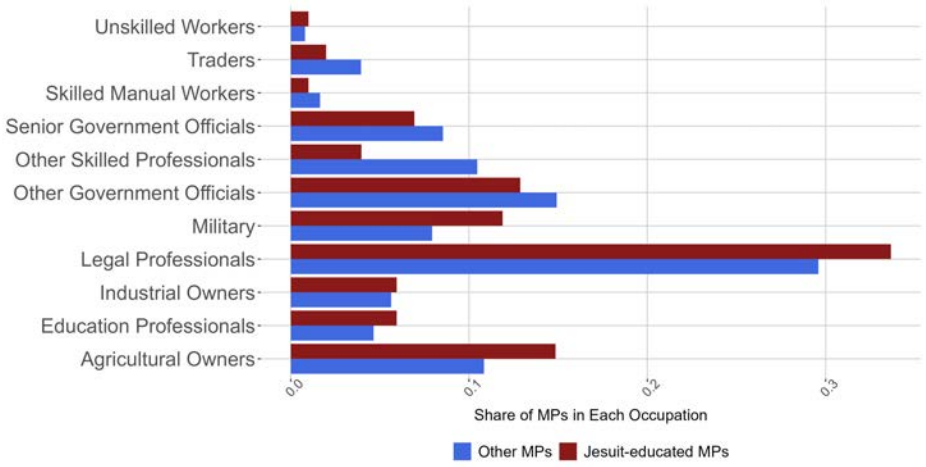
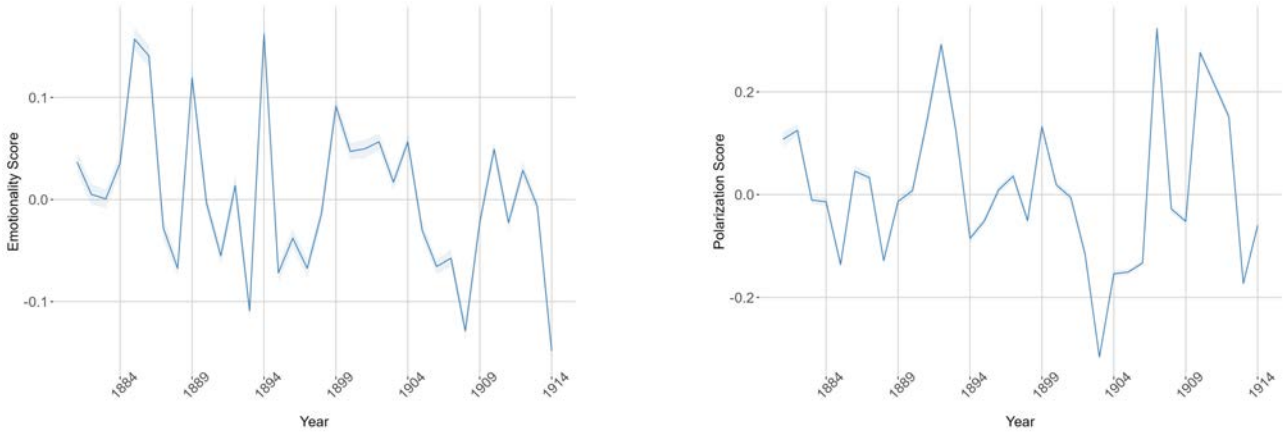


Figure A.15: Distribution of MPs across Occupations

Notes: The figure shows that Jesuit-educated MPs and other MPs had a similar distributions across occupational categories. The sample includes all MPs who served in the Chamber of Deputies between August 1881 and April 1914. The occupational categories were extracted using ChatGPT from biographical summaries available for each MP.

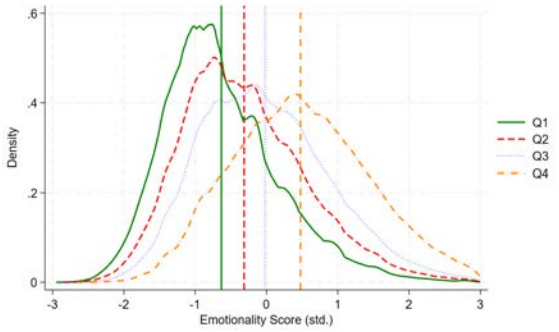


(a) Emotionality

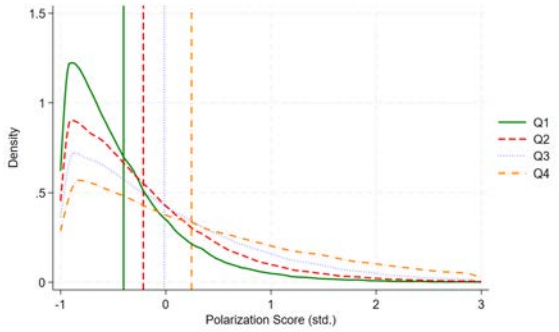
(b) Polarization

Figure A.16: Evolution of Emotionality and Polarization Over Time

Notes: The figures show the evolution of emotionality and polarization in the French Parliament from 1881 to 1914. The solid blue lines represent the annual averages of emotionality (left panel) and polarization (right panel) across all speeches delivered in the Chamber of Deputies. The shaded areas represent 90% confidence intervals.



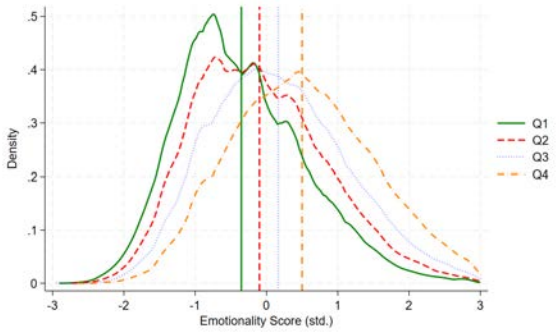
(a) Emotionality



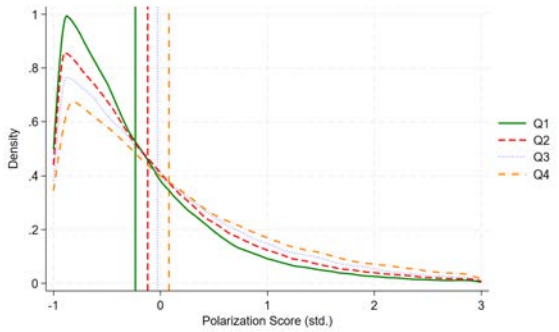
(b) Polarization

Figure A.17: Variation of Emotionality and Polarization within vs. across Discussions

Notes: The figure shows that the lion's share of the variation in the emotionality and polarization of French parliamentary speeches arises within discussions rather than between discussions, where a discussion is defined as a specific topic on the parliamentary agenda in a given sitting. The left panel shows the speech-level densities of the standardized emotionality score across quartiles of the discussion-level emotionality distribution. Specifically, we regress the standardized emotionality score on discussion fixed effects and then assign all speeches from a discussions to the quartile of the discussion's fixed effect. The right panel performs the same exercise for the standardized polarization score.



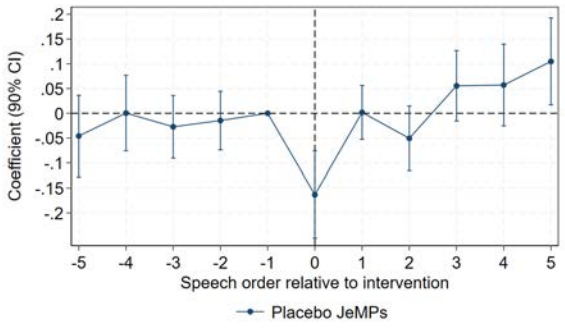
(a) Emotionality



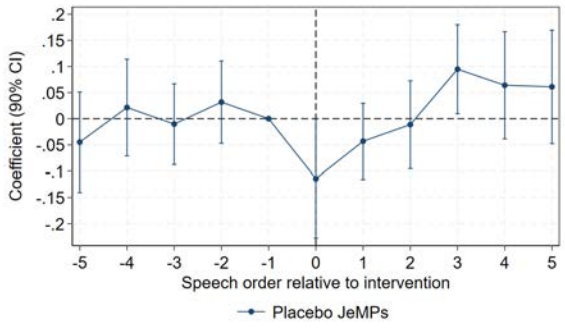
(b) Polarization

Figure A.18: Variation of Emotionality and Polarization within vs. across Speakers

Notes: The figure shows that most of the variation in the emotionality and polarization of French parliamentary speeches arises within speakers rather than between speakers. The left panel shows the speech-level densities of the standardized emotionality score across quartiles of the speaker-level emotionality distribution. Specifically, we regress the standardized emotionality score on speaker fixed effects and then assign all speeches by a speaker to the quartile of this speaker's fixed effect. The right panel performs the same exercise for the standardized polarization score.



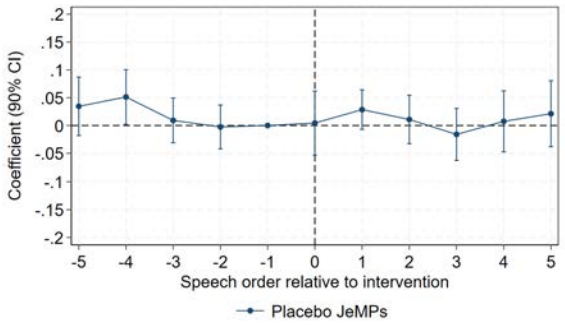
(a) Emotionality Score



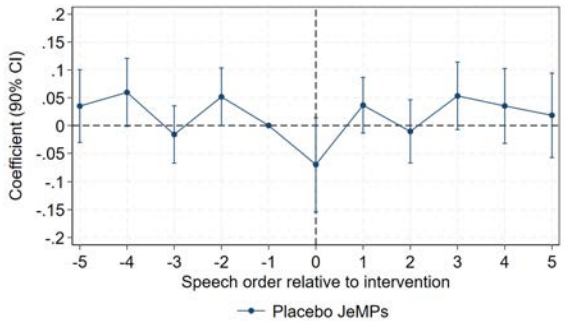
(b) Polarization Score

Figure A.19: Placebo Event Study: Least Emotional or Polarized (non-JeMP) MPs

Notes: The figure repeats the event studies from Figures 7 and 8 in the paper, using low-emotionality and low-polarization "Placebo JeMPs" as treatment. Those are constructed as follows: (i) we exclude actual JeMPs, the President of the Chamber, and members of government; (ii) for each year, we regress the emotionality score and the polarization score on speaker fixed effects, including the controls from col 4 in Table 1; (iii) within each year, we select the 5% MPs with the lowest estimated speaker fixed effects. The left panel in the figure uses "Placebo JeMPs" selected from the emotionality regressions, and the right panel, from the polarization regressions. See the note to Figure 7 for detail on the event study specification.



(a) Emotionality Score



(b) Polarization Score

Figure A.20: Placebo Event Study: (non-JeMP) MPs from Least Emotional or Polarized Departments

Notes: The figure repeats the event studies from Figures 7 and 8 in the paper, using speakers elected in the least emotional or polarized departments as "Placebo JeMPs." Those are constructed as follows: (i) we exclude actual JeMPs, the President of the Chamber, and members of government; (ii) we regress the emotionality score and the polarization score on department fixed effects, including the controls from col 5 in Table 1; (iii) we select the 5% departments with the lowest estimated department fixed effects and use the MPs from those departments to form the "Placebo JeMP" group. See the note to Figure 7 for detail on the event study specification.

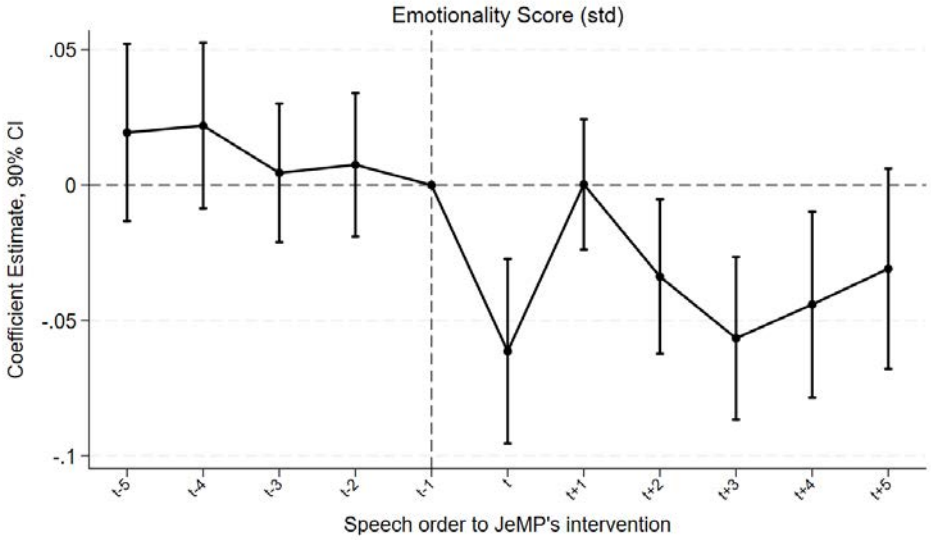


Figure A.21: Event Study: JeMPs Reduce the Emotionality of Subsequent Speakers

Notes: The figure repeats the analysis from Figure 7, but it allows for multiple treatment events in a discussion. As in Figure 7, the first JeMP intervention in a discussion is defined as a treatment; here, we additionally treat as events any JeMP interventions that occur at least 30 speeches after the previous treatment within the same discussion. See the note to Figure 7 for details on the control variables.

References

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

Bonnard, R. (1926). *Les règlements des assemblées législatives de la France depuis 1789: notices historiques et textes*. Paris: Société anonyme du Recueil Sirey.

Cagé, J. and T. Piketty (2025). *A History of Political Conflict*. Cambridge, MA: Harvard University Press.

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

Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as Data. *Journal of Economic Literature 57*(3), 535–574.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica 87*(4), 1307–1340.

Griffiths, T. L. and M. Steyvers (2004). Finding Scientific Topics. *Proceedings of the National Academy of Sciences 101*(suppl_1), 5228–5235.

Grün, B. and K. Hornik (2011). topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software 40*(13), 1–30.

Hansen, S., M. McMahon, and A. Prat (2018). Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach. *The Quarterly Journal of Economics 133*(2), 801–870.

Lanzara, G., S. Lazzaroni, P. Masella, and M. P. Squicciarini (2024). Economic Shocks and Assimilation Policies: Phylloxera and Educational Expansion in French Algeria. BAFFI CAREFIN Centre Research Paper 221.

Mikolov, T., I. Sutskever, K. Chen, G. Corrado, and J. Dean (2013). Distributed Representations of Words and Phrases and Their Compositionality. *arXiv preprint arXiv:1310.4546*.

Porter, M. (2009). The English (Porter2) Stemming Algorithm. http://snowball.tartarus.org/algorithms/english/stemmer.html.

Tausczik, Y. R. and J. W. Pennebaker (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology 29*(1), 24–54.