Do Elections Moderate or Polarize Political Rhetoric?*

Tito Boeri† Nina Nikiforova‡ Guido Tabellini§

July 27, 2026

Abstract

This paper studies how elections shape political communication. We formulate a theory of electoral competition with costly communication. If opposite candidates reach different audiences with their messages, they diverge and seek to polarize the electorate. We then analyze Twitter/X communication by 367 political leaders in 21 countries during 2013-2022, focusing on competition between populists and non-populists. Following Gentzkow, Shapiro and Taddy (2019), we estimate a structural model of speech and show that political rhetoric becomes more polarized before the election. This reflects two main theoretical mechanisms: politicians increasingly emphasize issues distinctive of their type, and within each issue, they use more extreme language.

Keywords: Electoral competition, Populism, Partisanship, Polarization.
JEL classification: P, H

*We thank Fondazione Rodolfo Debenedetti for access to Twitter data, Leonardo Bianchi, Ginevra Casini, Marco Cerundolo and Gabriele Nespoli for outstanding research assistance, and Elliott Ash, Jim Fearon, Torsten Persson, Andrea Prat, Jesse Shapiro and seminar participants at the Barcelona School of Economics, Bocconi University, Cambridge, CEMFI, King's College, Oxford, the IOG meeting at Berkeley and the CEPR media conference at Bocconi for helpful comments. Tabellini gratefully acknowledges financial support from the Italian Ministry of University and Research, D.D. n. 1802 del 21/11/2024 BANDO FIS 3 - progetto FIS-2024-05869 - CUP J53C25002480001

†Department of Economics and IGIER, Bocconi University; IZA; CEP-LSE
‡Department of Economics, Bocconi University
§Department of Economics and IGIER, Bocconi University; CEPR; CES-Ifo

1 Introduction

Do elections induce moderation or polarization between competing politicians? If competing candidates seek to persuade the same voters, as in standard theories of Downsian or probabilistic voting, then most theories of electoral competition predict policy convergence (Persson and Tabellini, 2002). In this approach, divergence is due to intrinsic partisan ideologies that make candidates accept vote losses to be closer to their preferred positions (Alesina, 1988 and Besley and Coate, 1997). An alternative view is that elections have a polarizing effect. As discussed below, this can happen in a variety of situations, but the most compelling one is that candidates' communication reaches different constituencies. If so, elections create incentives to diverge, as each candidate seeks to please or mobilize their exclusive audiences. (Glaeser, Ponzetto and Shapiro, 2005 and Gennaioli and Tabellini, 2025 and Prummer, 2020) Despite a large empirical literature discussed below, the evidence on which view is closer to the truth is still inconclusive.

To shed light on this question, we study how political communication on Twitter/X is affected by the imminence of elections. During an electoral campaign, communication focuses on the short run goal of persuading and mobilizing voters. In off-election periods, instead, politicians can afford to let their tweets be shaped by their ideology and personal traits, or to react to external events. To guide the empirical analysis, we develop a simple model of electoral competition over two issues and show that, when opposite politicians reach different voters with their messages, elections create an incentive to diverge and polarize the electorate. If the exclusive audiences are sufficiently large, the model generates two main testable predictions: as the election approaches and the goal of winning the election becomes more important, polarization increases both *between* and *within* issues. Between issues, politicians prioritize the topic more distinctive of their political type and more salient for their core supporters; within the same issue, they use more extreme language, both as a signal of their positions and with the goal of polarizing the electorate.

We then take these predictions to the data. We study about 3.4 million tweets sent by 367 political leaders in 21 Western democracies during 2013-2022, and exploit staggered election dates across countries. Messages on Twitter/X, unlike electoral programs or speeches, provide high-frequency data, which allows us to study time variation as the election approaches. Tweets offer more freedom of expression than traditional media, and provide a free platform for politicians challenging the establishment, as with most populist leaders. For these reasons, platforms like Twitter/X have been key arenas of political competition, and affect the salience of topics such as crime, illegal migration or mass shootings, as well as how these topics are discussed (Battiston, 2022 and Zhang et al., 2025). Finally, compared to party manifestos

or political speeches and to other media, tweets can be more easily targeted to specific audiences, and politicians can monitor the feedback received and by which type of voter. As discussed below, this can contribute to explain why our findings differ from those in other similar studies that have focused on candidate websites and manifestos.

Our units of observation are the bigrams (i.e., two-word sequences) used by each politician during a quarter. The main challenge is that the sample of observed bigrams is small, relative to the very large number of bigrams that could be used. As explained by Gentzkow, Shapiro and Taddy (2019) (henceforth GST), simple measures of word counts over-estimate deliberate rhetorical polarization, because many bigrams are randomly chosen. We follow their approach, and estimate a structural model of speech that corrects for finite-sample bias, accounts for differences in verbosity, and allows bigram usage to depend on political type and other observables.

We start by measuring rhetorical polarization by *partisanship*, namely the ease with which one can correctly predict a politician's type based on their words, as in GST. For each politician, partisanship provides a quarterly measure of how polarized their tweets are, relative to politicians with the same observed characteristics but of the opposite political type. Second, using the model's recovered structural objects, we develop an approach to directly test the model's predictions.

Our main focus is on populist vs non-populist politicians, but we also consider left vs right, as well as a four-way partition along both dimensions (left- and right-wing populists, and similarly for non-populists). These classifications are defined ex-ante, based on Funke, Schularick and Trebesch (2023) and by human coders assisted by ChatGPT and party affiliations. There are several reasons for focusing on populist vs non-populist competition. This rivalry is arguably the main dimension of political competition in most Western democracies in this period. Populist leaders are or have been in government in the US, Brazil and Mexico, they are part of governing coalitions in the Netherlands, Austria, Italy and Sweden, influence politics from the outside in France, and command large majorities in Eastern German states. The rise of populist challengers is the other side of the coin of the decline of mainstream parties. The centrality of populist vs non-populist competition is confirmed by our data. Throughout the entire sample, populists vs non-populist rhetorical polarization is about 30% higher than between left vs right. Moreover, populists are more connected across countries (measured by Twitter links), compared to other political groups, and international connections among same-type politicians is greater along the populist vs non-populist divide than among the left vs right divide. Unlike most of the literature on populism, we focus on leaders rather than parties. This seems appropriate when studying the supply of populism, which is often based on the emergence of a charismatic leader.

To assess how proximity to elections shapes political communication, we estimate an event study, comparing rhetorical polarization in a window of 7 quarters around national (legislative and presidential) elections with quarters outside this window. Our main result is that the rhetoric of populist and non-populist politicians becomes significantly more polarized as elections approach. Polarization rises two quarters before the election and persists through the election quarter, relative to its average in more distant quarters, before falling back in post-election quarters. The effect is robust and large: during the election quarter, the predictability of a politician's type exceeds its level in distant quarters by about 15% of a standard deviation of predictability over the whole sample. This is comparable to the peak effect of a major aggregate shock, such as the 2015 European refugee crisis. By contrast, for the left vs right divide we find no significant pre-electoral cycle.

We then unpack the increase in polarization between populists and non-populists and test the more specific predictions. To remain as close as possible to the theory, we measure: (i) two dimensions of competition between populist vs non-populist leaders, each dimension distinctive of one political type; (ii) ideological extremism of bigrams within each dimension.

To construct the two dimensions, we classify tweets among 19 topics chosen a priori, and then assign bigrams to each topic based on the frequency of use in tweets on that topic. Next, we estimate the posterior that the politician is populist, conditional on the bigram's topic. Topics for which the average posterior over the entire sample is above (below) 1/2 are distinctive of populist (non-populist) politicians. As expected, non-populist politicians are more likely to tweet about policy issues, with the exception of immigration and, after COVID, health policy, which instead are distinctively populist. On the other hand, populists are more likely to tweet against the establishment, engage in self-promotion, or to mention specific parties, politicians and elections. To measure bigrams' ideological extremism, we estimate a time invariant measure of each bigram's populist distinctiveness over the entire sample – namely by how much, and in which direction, a neutral observer would change their posterior belief that a politician who used this bigram is populist.[1]

With these two measures, we test the theoretical predictions of between- and within-issue divergence. We cannot identify which political type is responsible for the divergence, but the differences between their language allow us to test the theory. Our results support the theory's implications: closer to the elections, politicians (i) speak more frequently about the issue more distinctive of their type, and (ii) use ideologically more extreme and more distinctive language within each dimension.

Overall, these findings suggest that elections have a polarizing effect on the rhetoric of

[1]This is similar to the measure of bigram distinctiveness used by GST, except that our indicator is time invariant.

populist vs non-populist leaders on Twitter/X. Could this be due to changes in the audience? The number of politicians' followers on Twitter/X does not fluctuate around election dates (Van Kessel et al., 2020). We don't observe the audience composition, but it is not clear why more extreme voters would be more attentive before the election than in off-election periods, compared to moderate voters. A more plausible interpretation is that electoral incentives exert a stronger influence on political communication in pre-election periods. This is consistent with the observed rise in politicians' verbosity before the election – in Section 8 we discuss why verbosity is unlikely to be driving our results. Although there are several reasons why stronger electoral incentives may induce polarization, a prominent one is that populist and non-populist leaders reach different groups of voters, and they seek to mobilize their constituency before the elections. Indeed, the more precise targeting of core supporters afforded by Twitter can explain why our results differ from those on other forms of political communication - cf. Barberá (2015).

We contribute to several lines of research. First, we apply and extend GST's methodology to study different data and a novel question - GST did not study the effect of elections on political rhetoric. Polarization on Twitter between populist and non-populist politicians is larger and more volatile than in congressional speeches of US Republicans and Democrats.[2]

Second, we contribute to the literature on whether elections induce political convergence or divergence by both providing a theory and developing an empirical approach to test it. Besides the research discussed above, on swing voters vs endogenous turnout and on voters' informational asymmetries, other theoretical papers have shown that elections create incentives to diverge, if voters have convex policy preferences (Kamada and Kojima, 2014), or candidates are risk averse and care about their vote share (Bonomi, 2024), or districts are heterogeneous and there is within party disagreement (Polborn and Snyder Jr, 2017).

Empirical research on this is fraught with difficulties, and has not reached conclusive results. In line with our divergence findings, Ash, Morelli and Van Weelden (2017) show that floor speeches of US senators devote more time to divisive issues when they are up for election. On the other hand, using data of candidate websites in US House elections and manifestos for French elections, Di Tella et al. (2025) find evidence of convergence in ideology and rhetorical complexity between primaries / first round elections and final elections. The difference with our results could be due to Twitter enabling better targeting of voters, or to primaries/first round elections inducing even more divergence than final elections. This second interpretation is in line with the findings by Brady, Han and Pope (2007) and Fowler and Fu (Forthcoming).

[2]Fiva, Nedregård and Øien (forthcoming) apply GST's methods to study different dimensions of heterogeneity in legislative speeches of Norwegian MPs.

Our results also speak to a related literature in political science, which has evaluated empirically the tradeoff in vote shares between converging to the center to capture swing voters, vs diverging to mobilize more extreme voters. Here too, results are inconclusive. An earlier literature suggests that pre-electoral divergence in Congressional roll call votes is punished at the election (Canes-Wrone, Brady and Cogan, 2002). Similarly, Hall (2015) and Hall and Snyder Jr (2015) find that, when a more extremist candidate wins the primaries, he is penalized in the general election. On the other hand, Bonica, Rhee and Studen (2025) argue that these findings are confounded by endogenous turnout and ballot composition effects. When these are taken into account, they find that the small electoral penalty for divergence is swamped by the large electoral gains of mobilizing core voters. Hill (2017) reaches similar conclusions with a different methodology.

Third, our paper contributes to the literature on populism and social media. Most of it seeks to explain voters' behavior - cf. Guriev and Papaioannou (2022) and Guriev, Melnikov and Zhuravskaya (2021) and Manacorda, Tabellini and Tesei (2026). But the supply side is equally important, and much less is known about it, particularly on the communication strategies of populists in proximity of elections, and on how non-populist politicians react to these strategies.[3]

Last, but not least, we contribute to research on political polarization. A consensus is emerging that politicians' messages and propaganda influence voters' beliefs and fuel polarization among citizens (see Zhang et al. (2025) and Klein (2020)). Nevertheless, it is less clear why politicians behave in this way, whether for opportunistic reasons or due to their ideological convictions. Our results are consistent with the predictions of Gennaioli and Tabellini (2025) and Bonomi (2024), who show that opportunistic politicians may find it optimal to polarize the electorate.

The paper is structured as follows. Section 2 develops a model of political competition with asymmetric audiences that motivates the empirical analysis. Section 3 describes the methodology. Section 4 describes the data. Section 5 presents aggregate patterns in the data. Section 6 presents the results on partisanship over the electoral cycle, and Section 7 unpacks those results by testing more specific predictions. Section 8 discusses the robustness of our results. Section 9 concludes. The Online Appendix contains additional details.

[3]Research in political science has studied populist communication strategies, but without a focus on the electoral cycle - cf. Cassell (2023) and Aslanidis (2018).

2 Theory

This section develops a model of political competition with costly communication to guide the empirical analysis. The central mechanism is that, when opposing candidates reach different groups of voters with their messages, electoral competition creates incentives to diverge and polarize the electorate.

2.1 The model

There are two political candidates, $P = A, B$, and two voter types, $v = b, a$, respectively with given primitive positions and primitive bliss points in a symmetric two-dimensional space indexed by $k = 1, 2 : \hat{g}_k^A = -\hat{g}_k^B = \varpi > 0$ and $\hat{g}_k^a = -\hat{g}_k^b = \varpi + \beta > \varpi$. Thus, voters are more extreme than politicians. The two dimensions can be interpreted as policies, or attitudes towards elites or institutions, or candidate features that differ between the two candidates. Each voter type has a continuum of voters of size $1/2$, so total population is 1.

Voters' preferences and perceived candidates' positions are affected by political communication, described below. After communication, voters' preferences are (vectors in boldface):

$$V(\mathbf{g}^v, \mathbf{g}^P) = -\frac{1}{2}[\sigma_1^v(g_1^v - g_1^P)^2 + \sigma_2^v(g_2^v - g_2^P)^2]$$

where g_k^v is the voter bliss point, g_k^P is candidate $P's$ position and σ_k^v is dimension k's relative salience, with $\sigma_1^v + \sigma_2^v = 1$. $(g_k^P, g_k^v, \sigma_k^v)$ are affected by candidates' communication.

Political communication Before the election, politicians engage in costly communication of three types: (i) $x_k^P \gtreqless 0$, aimed at changing own perceived positions; (ii) $y_k^P \gtreqless 0$, aimed at changing voters' bliss points; (iii) $s^P \gtreqless 0$, aimed at changing issue salience. We assume:

$$\begin{aligned} g_k^P &= \hat{g}_k^P + x_k^P && (1)\\ g_k^v &= \hat{g}_k^v + \lambda^v(y_k^A + y_k^B), \qquad \lambda^a = -\lambda^b = 1 && (2)\\ \sigma_1^v &= \bar{\sigma}^v + s^A + s^B \qquad \bar{\sigma}^a = 1 - \bar{\sigma}^b = \frac{1}{2} + \epsilon, \quad 0 \leq \epsilon < \frac{1}{2} && (3) \end{aligned}$$

Thus P's perceived positions are only affected by his/her own communication, x_k^P, whereas voters' preferences are affected by communication of both politicians. y_k^P aimed at voters' bliss points, however, shifts them in opposite directions depending on their type. This is due to a backlash on opposite voter types, as in Gennaioli and Tabellini (2025) where propaganda enhances partisan stereotypes. Finally, if $\epsilon > 0$, the model is not perfectly symmetric: with

no effort on salience-communication, issue 1 is more salient for a voters and issue 2 is more salient for b voters. These assumptions are depicted in Figure 1, for a single dimension.

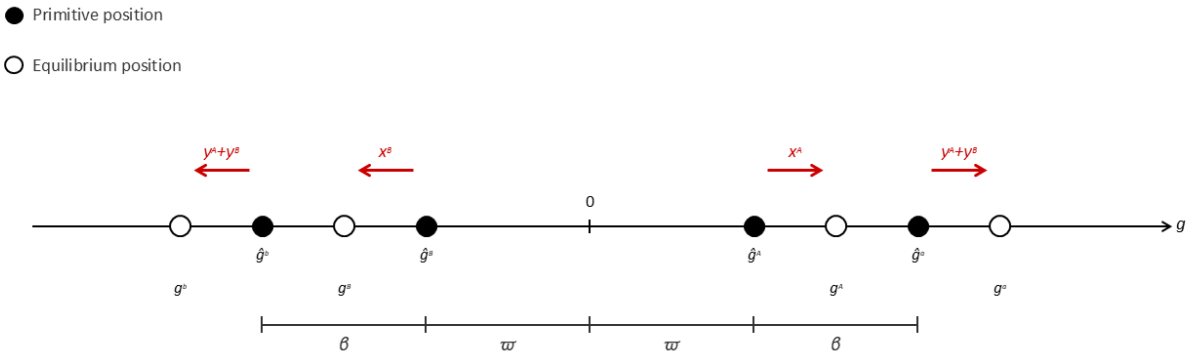


Figure 1: The one-dimensional model. Filled circles: primitive positions of parties ($\hat{g}^A = -\hat{g}^B = \varpi$) and voter bliss points ($\hat{g}^a = -\hat{g}^b = \varpi + \beta$). Open circles: equilibrium perceived party positions $g^P = \hat{g}^P + x^P$ and voter bliss points after communication in the divergent and polarizing regime. Red arrows show the direction of communication efforts.

The cost of communication is $C = \sum_k [C(x_k^P) + C(y_k^P)] + C(s^P)$ where $C(.)$ is strictly convex, twice continuously differentiable and symmetric with a minimum at $C(0) = 0$, and sufficiently convex that the equilibrium is always an interior optimum and salience weights are always between 0 and 1, so that the equilibrium is unique. We distinguish between different types of communications to be transparent on the mechanisms, although in principle a single communication effort could affect voters' beliefs on all objects at once.

Critically, we assume that all voters of type a (resp. b) are reached by the communication of candidate A (resp. B), while only a fraction $0 \leq \alpha \leq 1$ of voters of type b (resp. a) are reached by the communication of candidate A (resp. B). Thus, a fraction $(1 - \alpha)$ of a voters is an exclusive audience of candidate A, and a corresponding fraction of b voters is an exclusive audience of candidate B. The remaining fraction α of voters is reached by both parties. All types of communication obey this assumption. Voters who are not reached by the communication of candidate P behave as if that communication was 0.

Probabilistic voting After communications, voter i of type v votes for candidate A iff:

$$\Delta^v \equiv V(\mathbf{g}^v, \mathbf{g}^A) - V(\mathbf{g}^v, \mathbf{g}^B) > \varepsilon^{iv} \tag{4}$$

where ε^{iv} is independently and uniformly distributed over $[-1/2\phi, 1/2\phi]$, with large enough support that we only consider interior equilibria. Hence, $A's$ vote share is:

$$\pi^A = \frac{1}{2} + \frac{\phi}{2} \sum_v \Delta^v \tag{5}$$

Candidates simultaneously choose their communication strategies $(\mathbf{x}^P, \mathbf{y}^P, s^P)$ to maximize their vote share net of communication costs,

$$U^P(\mathbf{x}^P, \mathbf{y}^P, s^P) = \gamma \pi^P - C \tag{6}$$

taking the opponent's communication as given. Parameter $\gamma > 0$ denotes how much politicians care about their vote share.

Results would be the same if we add independent voters, or if exclusive audiences only choose between voting for their favored candidate or abstaining (rather than voting for either one or the other candidate, as here). The key assumption is informational asymmetry, not the voters' choice margin.

2.2 Equilibrium

Appendix A proves that there is a threshold $0 < \hat{\alpha} < 1$ such that, in equilibrium:

Proposition 1. *(i) If $\alpha = 1$ and $\epsilon = 0$, then candidates moderate their positions ($x_k^A = -x_k^B < 0$) and do not change voters' preferences ($y_k^P = s^P = 0$).*
(ii) If $\alpha < \hat{\alpha}$ and $\epsilon = 0$, then candidates polarize their positions ($x_k^A = -x_k^B > 0$) as well as voters' preferences ($y_k^P > 0$), and do not seek to change issue salience ($s^P = 0$).
(iii) If $\alpha < \hat{\alpha}$ and $\epsilon > 0$ sufficiently small, then candidates: (1) polarize their positions as well as voters' preferences, more so in the dimension that is more salient for their exclusive audiences ($y_1^A = y_2^B > y_2^A = y_1^B > 0$ and $x_1^A = -x_2^B > x_2^A = -x_1^B > 0$); (2) increase the salience of the issue that is already more salient at baseline for their exclusive audiences ($s^A = -s^B > 0$).
(iv) If ϵ is sufficiently small, all non 0 communication effort about candidates' positions and voters' bliss points increases with γ ($\frac{\partial |z_k^P|}{\partial \gamma} > 0 \;\; iff \; |z_k^P| \neq 0$, $for \; z = y, x$). If α is sufficiently small and the cost function C(.) is quadratic, salience effort also increases in γ ($\frac{\partial |s^P|}{\partial \gamma} > 0$ iff $|s^P| \neq 0$).

Consider first uniform information and perfect symmetry ($\alpha = 1$ and $\epsilon = 0$). Because the same voters are contested by both parties, each candidate has an incentive to moderate its perceived position (standard Downsian logic), more so if electoral stakes are higher (larger γ).

There is no incentive to change voters' preferences: with a shared audience and symmetric voter types, any communication effort exactly cancels across groups (since $\lambda^a + \lambda^b = 0$).

Suppose instead that a fraction $(1-\alpha) > 0$ of voters is reached exclusively by their closest candidate. Polarizing this exclusive audience is a one-sided electoral gain: it pushes them further away from both candidates, but with concave voters' preferences this benefits their closest candidate, without any offsetting loss elsewhere. Both candidates therefore invest in voter polarization ($y_k^A = y_k^B > 0$), and more so when electoral incentives are stronger (higher γ). This corresponds to the identity politics mechanism in Gennaioli and Tabellini (2025), where communication exacerbates partisan stereotypes. Whether candidates also diverge in perceived positions depends on a further trade-off. A more extreme position attracts the closest voters (who are more extreme than the candidate), but repels the more distant ones. If the exclusive audience is sufficiently large ($\alpha < \hat{\alpha}$), the first effect dominates and candidates diverge. With perfect symmetry ($\epsilon = 0$), there is no gain in changing issue salience.

If voters also initially disagree on which dimension is more relevant ($\epsilon > 0$), asymmetric communication has two additional effects (if the exclusive audience is sufficiently large).[4] First, communication has stronger electoral effects in the more salient dimension. Hence, both candidates exert greater polarizing effort — on perceived positions and bliss points — along the dimension favored by their exclusive audiences. As a result the equilibrium is no longer symmetric: voters and candidates are more extreme in the dimension which is more salient for captive voters. Second, each candidate benefits from making issue salience even more asymmetric across opposite voter types. Hence, the two candidates champion different issues, generating between-issue divergence in communication.[5]

Since these effects reflect how communication affects vote shares, they are all amplified by parameter γ capturing the strength of electoral motives.

Predictions To take these predictions to our Twitter/X data, we make two further assumptions. First, that the share $(1-\alpha)$ of exclusive audiences is sufficiently large. This is consistent with evidence showing that political communication on Twitter/X is exchanged primarily between individuals with similar political preferences (Barberá, 2015), and that politicians know a lot about their followers.[6] Second, that parameter γ rises in the immi-

[4] As shown in the proof, when $\epsilon > 0$ there are multiple thresholds, one for each aspect of communication, and $\hat{\alpha}$ corresponds to the lowest threshold.

[5] Although not covered by Proposition 1, it can be shown that, if $\alpha = 1$ and $\epsilon > 0$, most statements in part (iii) of the Proposition are reversed: in equilibrium candidate positions converge, they seek to moderate salience asymmetries, and on average voters' bliss points polarization is unaffected by communication. Thus, the critical parameter that induces communication divergence is the size of the exclusive audience, $1-\alpha$, not the asymmetric salience parameter ϵ.

[6] Until 2023 Twitter analytics allowed to recover breakdowns of followers by gender, location, interests (with categories such as "conservative politics"), and device usage. Thus, in addition to relying on self-

nence of elections. This is consistent with the observed pre-electoral rise in verbosity (cf. Appendix Figure H10), and with evidence that during an electoral campaign communication is more focused on the goal of gaining votes, relative to other motives such as expressing personal convictions or reacting to external events.[7]

Under these assumptions, Proposition 1 then predicts that political rhetoric polarizes in pre-electoral periods, compared to post-election periods or dates distant from the election. Specifically, the rest of this paper tests two predictions of Proposition 1. Before the election:

1) *Between-issue* divergence rises. This corresponds to s^A and s^B moving in opposite directions: each candidate speaks more about issues that are distinctively salient to its constituency and over which it enjoys an electoral advantage with its supporters;

2) *Within-issue* divergence rises. This corresponds to $\mathbf{x}^A$ and $\mathbf{x}^B$ moving in opposite directions and $\mathbf{y}^P$ rising. As candidates put more effort to polarize voters and their perceived positions, the way they speak about each issue becomes more distinctive of their political ideology, and more extreme in each dimension.

Of course, the model is very simple: it assumes only two competing candidates, two dimensions and perfect symmetry within each dimension, and some restrictions on parameter values to prove some results. Thus, a rejection can be interpreted that either the share $(1-\alpha)$ of the exclusive audience is small, or parameter γ remains constant over time, or a more general failure of the model.

3 Empirical methodology

This section describes how we take the theory to the data. We rely on the methodology proposed by GST. Relative to other approaches, this method of measuring polarization in political communication has several advantages.[8] First, it is based on a structural model of

selection of followers, leaders could assess whether their communication campaigns were effective in targeting specific audiences.

[7] Severin-Nielsen et al. (2025) interviewed Danish MPs about how their communication strategies differ between election and routine periods. They highlight that, while in routine periods communication is often developed on an ad hoc basis, during an electoral campaign it is much more thoroughly and coherently thought out, also because more resources are devoted to it. This is how a Danish MP described his communication during a routine (as opposed to election) period: "It is a bit more random than you would think. (. . .) It's very focused on issues of the day (. . .). And there are a lot of posts that are based on an impulse".

[8] Alternative methods of text analysis are discussed by Ash and Hansen (2023). Gauthier, Widmer and Ash (2025) propose a more general method for measuring speech polarization based on deep latent variable models, and show that their results are in line with Gentzkow, Shapiro and Taddy (2019). We adopt the latter, whose structural interpretability is central to the tests of our model's predictions in Section 7. Alternative methods based on AI are less transparent, and could attribute observed differences in speech to

speech, that also takes into account other determinants of word usage besides political type, such as nationality, gender, age, or institutional position. This is particularly important in our multi-country setting. Moreover, it allows us to dig deeper into the causes of rhetorical polarization, and directly test the implications of the model, as we discuss below. Second, as explained by GST, this approach substantially reduces finite sample bias and is robust to changes in verbosity across politicians over time. This allows us to exploit all observed communication, without restricting the analysis to a much smaller subset of speech to ensure comparability.

From theory to data We observe the tweets sent over time by several politicians indexed by i. As in the theory, we assume that there are only two political types, say populist ($P_i = 1$) and non-populist ($P_i = 0$), although individual politicians are allowed to differ in other observable features which may also influence communication (below we also allow for four political types). By equation (6), politicians' utility is a function of their communication. Here, following GST and much of the literature, we model communication as a choice of bigrams.[9] Thus, politician i's utility from using bigram j in quarter t is:

$$U_{ijt}^{P_i} = \alpha_{jt} + X_{it}\gamma_j + \phi_{jt}P_i \tag{7}$$

where α_{jt} is a fixed effect for bigram j's popularity in quarter t, γ_j and ϕ_{jt} are parameters to be estimated, X_{it} are dummy variables for other features (not all time varying) of i, namely gender, higher education, being in government in quarter t (directly or by supporting the governing coalition), being a candidate in the closest election, age, and fixed effects for country-by-quarter and for the election window (defined as all quarters closest to the same election) - the latter capture election-specific bigrams.

The resulting probability that i uses bigram j in quarter t is:

$$q_{ijt}^{P_i} = \frac{exp(U_{ijt}^{P_i})}{\sum_k exp(U_{ikt}^{P_i})}, \qquad P_i = 0, 1 \tag{8}$$

We observe the number of times that each bigram j is used by politician i in calendar quarter t, $\mathbf{c_{it}} = \{c_{ijt}\}$, and assume that this count variable has a multinomial distribution:

$$\mathbf{c_{it}} \sim MN(m_{it}, \mathbf{q_{it}^{P_i}}),$$

confounding variables correlated with political type. See Vafa, Athey and Blei (2025) and Gordon et al. (2025) among others for a discussion.

[9] See Gentzkow, Kelly and Taddy (2019) for a discussion of the economics literature using n-grams.

where $m_{it} = \sum_j c_{ijt}$ is the verbosity of i in quarter t and $\mathbf{q_{it}^{P_i}} = \{q_{ijt}^{P_i}\}$.

Thus, a communication strategy $(\mathbf{x}^P, \mathbf{y}^P, s^P)$ is the choice of the entire distribution of bigram-usage probabilities $\mathbf{q_{it}^{P_i}}$ during period t. This formulation entails two simplifying assumptions. First, the utility of using bigram j is allowed to depend on i's observed features, including political type, but not on which other bigrams he/she uses. Second, i's utility from bigram choice depends on the communication of other speakers only through the bigram popularity fixed effects, α_{jt}, and the country by quarter fixed effects. These fixed effects thus capture equilibrium interactions between different politicians.

Estimation For each bigram j we estimate parameters $\{\alpha_{jt}, \gamma_j, \phi_{jt}\}_{t=1...T}$ by minimizing the following penalized objective function, as in GST:

$$\sum_j \Bigg[\sum_t \sum_i m_{it} \exp(\alpha_{jt} + X_{it}\gamma_j + \phi_{jt}P_i) - c_{ijt}(\alpha_{jt} + X_{it}\gamma_j + \phi_{jt}P_i) + \psi(|\alpha_{jt}| + ||\gamma_j||) + \lambda_j|\phi_{jt}| \Bigg] \tag{9}$$

Two aspects of this procedure are worth mentioning. First, in order to make the model computationally feasible, we approximate the likelihood of the multinomial logit with the likelihood of a Poisson model $c_{ijt} \sim Pois(exp(\mu_{it} + u_{ijt}))$ and the plug-in estimator $\hat{\mu_{it}} = log(m_{it})$ for μ_{it} is used. Second, we make use of Lasso shrinkage for populist loadings, $\lambda_j|\phi_{jt}|$. The value of λ_j is chosen separately for each bigram so as to minimize the corrected Akaike information criterion. Overall, this forces the parameter of interest ϕ_{jt} towards 0, reducing finite sample error.[10] For the other parameters, we set the penalty at $\psi = 10^{-5}$. This value controls the extent to which variation in bigram usage across individuals or over time is attributed to covariates rather than being absorbed P_i. This penalty also facilitates numerical convergence. We discuss the choice of the penalty in Appendix B.[11]

Using (7)-(8), we can estimate the probability that a politician of type (P_i, X_{it}) uses bigram j in quarter t, $\hat{q}_{ijt}^{P_i}$, for all j in our sample. The set of covariates is sufficiently rich that this estimated probability is specific to each individual politician.

[10] In our baseline estimates, the estimate of ϕ_{jt} is non-zero for about 2.5% of bigrams per quarter.

[11] Penalizing non-zero coefficients γ_j implies that around 30% of all the country by calendar quarter fixed effects are included on average (the rest are estimated to be 0). Given the importance of distinguishing between different countries and quarters, we thus also include in X_{it} separate fixed effects for each country and each calendar quarter (without the penalty, this would be redundant, since these fixed effects would all be absorbed by the country -by-calendar-quarter fixed effect). The estimates of the event studies described below are not significantly affected by this inclusion (results available upon request).

Partisanship Using $\hat{q}_{ijt}^{P_i}$, the posterior that an observer with neutral priors assigns to politician i having type $P_i = 1$, conditional on i having used bigram j in quarter t, is:

$$\hat{\rho}_{ijt} = \frac{\hat{q}_{ijt}^{1}}{\hat{q}_{ijt}^{1} + \hat{q}_{ijt}^{0}} \tag{10}$$

Following GST, our core measure of rhetorical polarization by politician i in quarter t is *partisanship*, defined as:

$$\hat{\pi}_{it} = \frac{1}{2} \sum_{j} [\hat{q}_{ijt}^{1} \hat{\rho}_{ijt} + \hat{q}_{ijt}^{0} (1 - \hat{\rho}_{ijt})] \tag{11}$$

This variable measures the average predictability of i's type, given a single bigram. It is the average posterior probability with which an observer with neutral priors expects to correctly classify politician i's type, conditional on a single bigram. The average is computed over all possible bigrams, weighted by the estimated probabilities of their use by i in quarter t.

Any specific politician is either $P_i = 1$ or $P_i = 0$. The variable $\hat{\pi}_{it}$ averages the true type P_i and his "clone" $1 - P_i$ with neutral priors. In other words, partisanship measures the average predictability of two specific politicians with features (X_{it}) who are identical in everything but their political type. If opposite types always use the same bigrams, then $\hat{\pi}_{it} = 1/2$. If instead they always use different bigrams, then $\hat{\pi}_{it} = 1$.

We also consider a more granular partition of four political types, defined on two dimensions: populist / non-populist and right / left. Appendix C provides more details on a construction of this measure.

Despite its useful properties, partisanship has two limitations. First, it measures rhetorical divergence, but not who is responsible for it. The reason is that the model of bigram choice is identified only up to *differences* in utilities between the two political types, $U_{ijt}^{1} - U_{ijt}^{0}$, within a given quarter.[12] As a result, the difference in bigram frequencies between opposite political types in a given quarter is a well-defined object, $\mathbf{q}_{it}^{1} - \mathbf{q}_{it}^{0}$, whereas the evolution of $\mathbf{q}_{it}^{P_i}$ over time conflates the effects of political types and of the other covariates entering equation (9). In other words, the specification in (7) does not allow us to tell apart whether one type uses a given bigram more or the other uses it less.

Second, partisanship is silent about the form taken by rhetorical divergence. Yet, the theory emphasizes that divergence is due to both political types: i) speaking more frequently about the issues which are more salient for their core supporters; ii) using ideologically more extreme and more distinctive language within each issue. To overcome this gap, in Section 7 we show that both points can be formulated as testable predictions about specific components of $\mathbf{q}_{it}^{1} - \mathbf{q}_{it}^{0}$, which allows us to test the more precise predictions of the theory.

[12] This is a well-known feature of discrete choice models; see, e.g., Gandhi and Nevo (2021) for a discussion.

4 Data

4.1 Tweets and Politicians

Data on tweets of politicians were collected through Twitter Academic API, version 2. The sample covers the period 2013-22 in 21 democracies selected also based on the prominence of populist political leaders (Australia, Austria, Belgium, Brazil, Czech Republic, Denmark, Finland, France, Germany, Hungary, Italy, Mexico, Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden, United Kingdom, United States).[13]

Politicians were selected based on their leadership positions, according to the following criteria: (i) All candidates for Prime Minister/President in general elections between 2001 and 2022. (ii) Leaders of main political parties with vote share above 5% in at least one general election between 2010 and 2022. (iii) Influential political leaders at the local level or leaders of niche parties, even if the national vote share of their party is below 5 % (e.g., Nicola Sturgeon in the UK, Basque and Catalan political leaders in Spain, Giorgia Meloni in Italy). These criteria identified 496 leaders, of whom 367 had an active Twitter account, for a total of about 3.4 million tweets written between January 2013 and June 2022. In addition to the text of each tweet, we have the exact date and time in which they were issued.

A distinguishing feature of our paper is that we focus on political leaders rather than parties. We classify politicians along two dimensions: populist (P)/non-populist (NP) and left (L)/right (R).

4.1.1 Populist vs. Non-Populist

To identify populist leaders, we draw on the classification of Funke, Schularick and Trebesch (2023) for those few leaders in their dataset who were active during the period covered by our analysis. For the remaining politicians, we conducted a manual classification informed by GPT-4o mini responses to the following question: *Is leader X from country Y commonly considered a populist leader?* A human coder then interpreted the lengthy and detailed answer produced by ChatGPT and used it to classify the leader as populist, non-populist or ambiguous. Key identifiers of populist leaders included the cleavage between the people and the elite, an anti-establishment stance, and a strong emphasis on national sovereignty. The ambiguous category includes genuinely unclear cases, situations where ChatGPT was uncertain, or cases where leaders appeared to shift over time between

[13]We thank the Rodolfo De Benedetti Foundation for giving us access to these data that they had collected.

populist and non-populist positions.[14]

After this step, 29 politicians were classified as *ambiguous*. We classify 16 of these as populist because their party is considered populist by at least one of Norris (2019), Rooduijn (2019) and Guriev and Papaioannou (2022). For the remaining 13 politicians we check that our results are robust to classifying them either way, and in our main analysis we include them among populist types. Examples of leaders belonging to this residual group are Lula da Silva in Brazil, Nick Xenophon in Australia, and Marjan Sarec in Slovenia. Importantly, there are at most four politicians that, according to ChatGPT, may have switched from NP to P or vice versa in the period covered by our data. For these four politicians, we retain throughout the most recent classification. Appendix F.1 describes how we established that. Appendix F.6 lists all politicians and their political type. In total, we classify 84 politicians as populist and 283 as non-populist.

We validate our classification by comparing it to two benchmarks: (a) 150 random binary divisions of politicians and (b) a fully automated classification of politicians into populists and non-populists produced by the Claude Sonnet 4.6 model, with 84 out of 367 populist politicians in all cases. For each partition we computed a fixation index based on a χ^2 measure of speech heterogeneity. The fixation index measures the share of total variation in the language patterns among all politicians due to between group variation (cf. Desmet, Ortuño-Ortín and Wacziarg (2025)). Our classification outperforms 99% of all random partitions, while the classification by Claude outperforms 87% of them. This suggests that our baseline classification captures meaningful latent structures in politicians' language, and does so better than a fully automated AI-based classification. Appendix F.2 provides more details.[15]

Our leader-based classification of populism differs from a party-based classification, and adds relevant information to it. About 86 % of populist leaders according to our definition operate in parties classified as populist in the Global Party Survey (Norris, 2020). However, there are also several *NP* politicians belonging to populist parties, such as Mitt Romney, Mike Pence and George Bush in the US, Gordon Brown, David Cameron, and Theresa May in the UK. Instead, Alexandria Ocasio-Cortez in the US and Mark Latham in Australia are examples of populist leaders in parties that are classified as non-populist.

Figure 2 illustrates the number of active politicians in each quarter disaggregated by their type, and the number of tweets they sent. There are politicians of both types always active

[14]We also validated our classification against the fully automated classification produced by ChatGPT. There are only nine cases of complete disagreement between the two approaches – that is, politicians whom we classify as populists but whom the automated ChatGPT classification identifies as non-populists. Further details on this comparison are provided in Appendix F.1.

[15]Among the politicians unambiguously classified by both Claude and us, only 6 politicians are classified differently by us and Claude.

in all quarters (the minimum number of active populist politicians is always above 40).

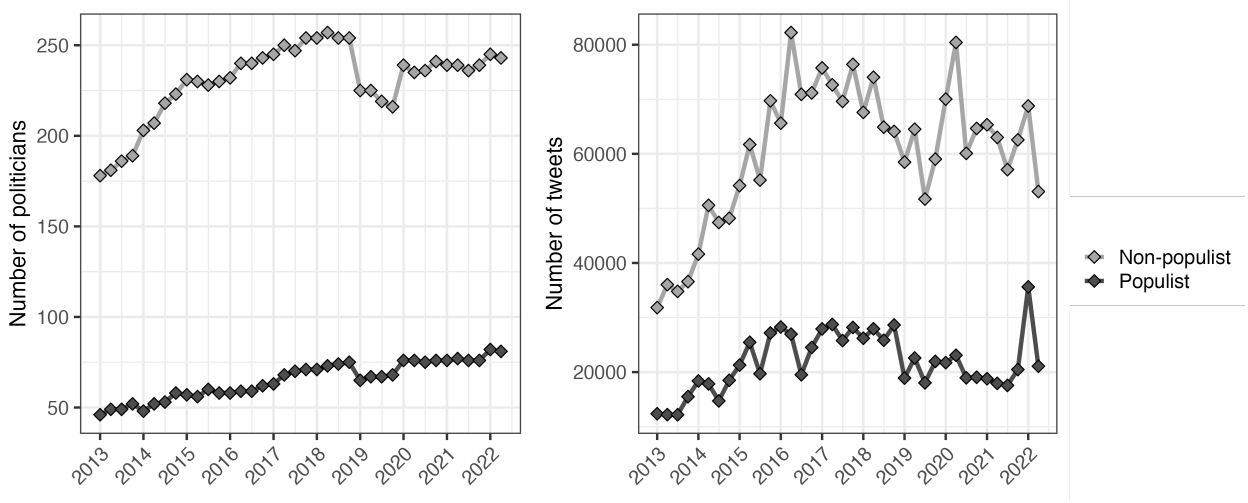


Figure 2: Number of politicians active on Twitter per calendar quarter, and number of tweets

The first three columns of Table 1 provide information on the characteristics of the two sets of politicians. P and NP leaders have approximately the same age, 3 out of 4 are males, and, on average, they began tweeting in 2012. What distinguishes P vs NP is their experience (P entered the national legislature at a later stage) and exposure to government positions (higher for NP leaders).[16] Some 80 % of NP leaders belong to mainstream parties, defined as parties that were in government (or lent support to a government) between 1990 and 2010.[17] NP leaders also display, on average, higher educational attainments than P leaders, but the difference is not statistically significant at conventional levels. Almost 50 % of populist leaders belong to niche parties, based on a GPT-4o model classification that defines as niche parties those emphasizing a limited set of issues that do not coincide with the predominant economic Left/Right division (Meguid, 2005) (see Appendix F.3 for the ChatGPT prompt). Overall, compared to NP leaders, several P leaders are *challengers*, because either they have not been in Government or they belong to parties representing issues not typically dealt with by the establishment. This is why, to avoid confounding political types with this feature, in estimating $\mathbf{q_{it}^{P_i}}$ we always control for being in government or supporting the governing coalition.

The analysis that follows implicitly assumes that the distinctive rhetorical features of politicians of the same type are similar across countries. To gauge the plausibility of this assumption, we study how connected political leaders are with same-type leaders in other coun-

[16]The variable years in Govt measures the number of years in which the leader was either in a party in Government, or supporting the Government, or was herself a Minister

[17]A leader is defined as mainstream even without government experience if she belonged to a party for a short period of time and that party was in government in a different period.

tries. Appendix F.4 shows that cross-border ties are homophilous along the populist/non-populist divide (that is, ties form disproportionately between same-type leaders), and that this homophily is stronger for populist leaders. Although populism often endorses nationalism and hostility to multiculturalism, its leaders behave as part of an international movement. Ties are also homophilous along the left–right divide described below, although more weakly than along the populist/non-populist divide.

Table 1: Differences between Populists and Non-Populists, and Right and Left (means; difference with p-value in parentheses).

	Populist vs Non-Populist			Right vs Left		
Variable	P	NP	Δ(p)	R	L	Δ(p)
Age (in 2010)	44.43	48.32	-3.89	46.31	48.15	-1.85
			(0.01)			(0.13)
Male	0.73	0.76	-0.03	0.78	0.74	0.05
			(0.59)			(0.28)
Higher Education	0.86	0.91	-0.05	0.91	0.89	0.02
			(0.26)			(0.50)
Year of First Tweet	2013.0	2012.4	0.58	2012.7	2012.4	0.39
			(0.08)			(0.14)
Years in Govt	6.44	9.79	-3.35	8.56	9.28	-0.72
			(0.00)			(0.48)
Legislative Experience	0.26	0.39	-0.14	0.40	0.34	0.07
			(0.00)			(0.04)
Mainstream	0.58	0.82	-0.24	0.83	0.72	0.10
			(0.00)			(0.02)
Niche Party	0.43	0.17	0.26	0.28	0.20	0.08
			(0.00)			(0.07)

Notes. Legislative Experience: years passed in 2010 since they first entered the national legislature. *Mainstream*: equals 1 if the leader has at any point in their career belonged to a party which was in government or supported the government coalition in the years 1990-2010. *Niche parties* classification is produced by GPT-4o, and based on the definition proposed in Meguid (2005). See Appendix F.3 for more details.

4.1.2 Left vs Right

We relied on the GPT-4o mini model to classify leaders along the L/R divide. We performed 4 iterations with the prompt displayed in Appendix F.5. Politicians belong to the "Right" group if GPT always unambiguously classified the politician in this category,

and the same with "Left". For 82 politicians, one of the 4 GPT iterations said that the classification is ambiguous. To classify these 82 ambiguous cases, we relied on their party affiliations, using the most recent RILE index of the Manifesto project, which measures how much a party covers left-wing or right-wing issues.[18] Based on this, we assigned 64 out of 82 leaders to the left and the remaining 18 to the right. Appendix F.5 discusses the differences between our classification and the one that would result from only using party affiliations. Appendix F.6 lists all politicians along the L-R divide. L vs R have similar features - cf. the three right-most columns in Table 1.

Appendix Figure H1 depicts the number of active L and R politicians in our sample and their activity in each quarter. Left-wing politicians are more numerous than those on the right – 223 politicians are classified as L and 144 classified as R. We also classify leaders along the two dimensions simultaneously, to obtain four political types (LP, RP etc.). Appendix Table H1 reports the cross tabulation of the four types. Non-populist politicians in our sample tend to be left-wing, while populists are about equally split between left and right.

4.2 Bigrams

All tweets were translated into English using the open-access Python library googletrans, and then decomposed into bigrams.[19] To do so, all special symbols were removed (punctuation marks, hyphens and apostrophes, emojis, user mentions and other hyperlinks, while hashtags were kept) and words were reduced to their stems.

The final dataset comprises around 40 mln bigrams, of which there are 15 mln unique bigrams. The vast majority of bigrams were used just once. Estimations drawing on such a large number of bigrams would be computationally infeasible and in the remainder we restrict the analysis to the most frequent bigrams only. For our benchmark specification, only bigrams that were spoken in at least 10 calendar quarters and that have more than 10 occurrences in at least one quarter were used. We refer to this dataset restriction as 10-10 (Appendix Figure F2 presents the histograms for these variables). This left us with about 25,000 unique bigrams and about 65% of the tweets. In section 8 we assess the robustness to different thresholds (5-10, 5-20, 10-25).

[18] There are no iterations when GPT-4o mini model classification switched from classifying a politician from right to left or vice-versa, but just observations where the classification switched between right and ambiguous or between left and ambiguous or vice versa. The source for the manifesto data is: https://manifesto-project.wzb.eu

[19] To assess the quality of English translations, a sample of 50 tweets per country was translated back into their original languages. The similarity between each back-translated tweet and its original version was then evaluated using multiple text similarity metrics. Cosine similarity is above .9 for all languages, denoting a high degree of semantic similarity between the original and the translated texts. See Appendix F.9 for the results of this exercise.

4.3 Topic classification

We classified bigrams using GPT and BERT into 18 topics chosen a priori (business & industry, climate & environment, energy policy, EU, foreign policy, public health, immigration, inequality, inflation, labor market, rights, security, taxation, trade policy, anti-establishment, connection & personal communication[20], elections, self-promotion). These topics were chosen to achieve a good balance between granularity and scale. Too broad topics would be uninformative, while too granular classifications would not provide a sufficiently large number of tweets per topic for estimation.

We then proceeded in four steps: 1) we used GPT-5 mini to label 100,000 tweets into these 18 topics (see Appendix F.8 for the prompt used in instructing ChatGPT); 2) we then trained BERT on these labeled tweets; 3) next we used the trained BERT to classify all 3.4 million tweets into topics with a threshold probability of .4;[21] 4) finally, we classified bigrams based on their relative recurrence in tweets belonging to the different categories, calculating a simple measure of bigram specificity for each topic: bigrams exceeding the 20 % threshold in the ratio of the number of tweets containing bigram j on topic k to the total number of tweets containing bigram j were assigned to the respective topics.[22] In total, 22,627 bigrams (out of 25,819) were assigned to at least one out of these 18 topics.

We additionally created a 19th manually coded topic, *Specific Parties & Politicians*, which consists of bigrams that include the names of individual politicians or parties. As robustness check, we carried out all our estimations ignoring tweets belonging uniquely to this topic. Results were very similar, and throughout we only report estimates for all bigrams, including those unique of this 19th topic.

As a check of the validity of the classification provided by BERT, we asked two master students to manually (and blindly) classify a sample of 510 tweets, which includes tweets on all the topics. Appendix Table H4 displays the accuracy and F1 score of the different topics compared to the manual classification.

4.3.1 Populist and Non-populist Topics

In the model of Section 2 there are only two issues, each distinctive of one political type. To bring the data as close as possible to this structure, we group these 19 topics into two sets, distinctive of non-populist and populist speech respectively.

[20] Tweets only about personal communication such as greetings, weather or holidays

[21] For the anti-establishment topic, we retained only tweets with negative sentiment.

[22] Relatively low thresholds allow us to capture the fact that some bigrams belong to several topics. For example, "sustainable economy" belongs to both climate and business; "world economy" belongs to business, trade, and foreign policy. Higher thresholds miss this multidimensionality across topics.

To do so, let $\hat{q}_{ikt}^{P_i} = \sum_{j \in k} \hat{q}_{ijt}^{P_i}$ denote the probability that i speaks about issue k in quarter t. Let ρ_{ikt} be the posterior that speaker i is populist, conditional on having spoken about topic k, computed with topics (rather than bigrams) as units of speech. Thus, ρ_{ikt} is defined as in equation (10), except that on the right hand side we have $\hat{q}_{ikt}^{P_i}$ rather than $\hat{q}_{ijt}^{P_i}$. A topic is considered populist if the average ρ_k of ρ_{ikt} over the entire sample exceeds 0.5, and non-populist otherwise. We allow for one exception, public health, which exhibits a clear break during the COVID epidemic. We thus split health into two separate topics, before and after COVID (i.e. after Q1 2020).

Table 2 reports, for each topic, the average posterior ρ_k together with the number of bigrams and the standard deviation of ρ_{ikt} across politicians and quarters. Apart from immigration and post-COVID health, the topics most typical of populist politicians are non-policy oriented: mentions of specific parties and politicians, self-promotion, anti-establishment (the idea that the current political system is broken, corrupt, rigged and undemocratic), and elections (topics related to elections, voting or political campaigns). By contrast, the topics most distinctive of non-populist politicians are predominantly policy-oriented, with climate and environment, trade and foreign policy, and inflation among the most pronounced. These patterns are consistent with the idea that populism is a *thin-centered ideology*, with few clearly defined policy positions other than opposition to immigration and globalization (Mudde and Kaltwasser, 2012), and they map onto the model: k_1 topics are plausibly those more salient to populist voters, and k_0 to non-populist voters. The heterogeneity of ρ_{ikt} within a topic is non-trivial, so the precise ranking of individual topics should not be over-interpreted; what matters for the analysis below is only the binary split into a populist set k_1 ($\rho_k > 0.5$) and a non-populist set k_0 ($\rho_k < 0.5$).[23]

This aggregation is also substantively natural. Boundaries between narrowly defined topics are often blurry, and topics within each set can be viewed as close substitutes: politicians may choose bigrams from one topic or another with similar considerations in mind. Within the non-populist set, for instance, topics such as climate and energy policy, or taxation and inequality, often share messages and bigrams. The same applies to the populist set: self-promotion may substitute for election-related content, while immigration often overlaps with anti-establishment rhetoric, as it frequently invokes similar attitudes.

[23]Looking at median ρ_{ikt}, rather than average, gives very similar results. Also, omitting seven quarters around the election (those that are the focus of the event studies below) when computing posterior means yields a similar classification of topics, with one exception: the security topic, currently just below the 0.5 threshold, lies just above it. Results reported in Section 7 are robust to this change.

Table 2: Average partisanship by topic: number of bigrams, mean, and standard deviation

Populist topics

Topic	N bigrams	Mean ρ_J	SD
Specific parties & politicians	764	0.544	0.126
Immigration	503	0.522	0.083
Anti-establishment	127	0.520	0.099
Health (post-COVID)	1718	0.513	0.060
Elections	5488	0.513	0.056
Self-promotion	7282	0.512	0.046

Non-populist topics

Topic	N bigrams	Mean ρ_J	SD
Security	2206	0.498	0.062
Rights	2934	0.494	0.057
Inequality	804	0.492	0.066
Labour market	1217	0.489	0.064
European Union	1642	0.488	0.065
Health (pre-COVID)	1718	0.487	0.067
Taxation	2499	0.487	0.061
Business & industry	2438	0.485	0.062
Energy policy	937	0.485	0.063
Connection	2284	0.483	0.061
Inflation	358	0.480	0.060
Foreign policy	2741	0.480	0.063
Trade policy	427	0.477	0.061
Climate & environment	1357	0.476	0.060

Note: Each row corresponds to a topic. *N bigrams* is the number of distinct bigrams assigned to the topic. *Mean* ρ_k is the average ρ_{ikt} for all politician-quarter observations. *SD* is the standard deviation of ρ_{ikt} across politician-quarter observations within each topic. The left panel reports the six *populist* topics ($\rho_k > 0.5$); the right panel reports the remaining *non-populist* topics ($\rho_k < 0.5$).

5 Partisanship over time

This section illustrates the main properties of our measures of polarization, and describes how they evolved over time.

5.1 Most distinctive bigrams

We begin by examining which words are most indicative of populist versus non-populist rhetoric in each quarter t. Following GST, we assess the distinctiveness of each bigram based on its contribution to our partisanship measure. Specifically, define the populist partisanship of bigram j in quarter t for politician i as the extent to which an observer with neutral priors would change her expected posterior that politician i is populist if that bigram was removed

from the vocabulary, unbeknownst to the observer.[24] The populist partisanship of bigram j in quarter t is the median of this measure across all politicians active in quarter t. Positive values of this measure denote bigrams typically used by populist politicians, negative values by non-populists.

Table 3 below lists the 15 most distinctively populist and non-populist bigrams (i.e. with the largest positive and negative median values) in 2015 Q3 and 2020 Q1. These dates roughly correspond to the inflow of refugees from Syria to Europe and the COVID shock. During the refugee crisis (2015 Q3), non-populist politicians were more likely to refer to immigrants as *refugees*, and mention topics such as *"climate change"* and *"sustainable development"*. Populist politicians instead frequently emphasized phrases such as *"asylum seekers"*, explicitly mentioned that immigration is *illegal*, or mentioned *"border control"*. During the COVID-19 epidemic (2020 Q1), the bigram *"public health"* was distinctive of non-populist discourse, while populist politicians emphasized connection with people (*"take care"*) and continued to focus on immigration. More generally, populists use direct language and emphasize communication ("*good morning*", "*via youtube*"), whereas non-populists refer to the establishment ("*Prime Minister*", "*The Anniversary*").

5.2 Average partisanship

Next, consider average partisanship in calendar quarter t, defined as:

$$\bar{\pi}_t = \frac{1}{I_t} \sum_i \hat{\pi}_{it} \tag{12}$$

where I_t is the number of politicians active in quarter t.

The solid pink and green lines in Figure 3 display average partisanship for P vs NP ($\bar{\pi}_{t,PNP}$) and L vs R ($\bar{\pi}_{t,LR}$) respectively.[25] P/NP partisanship peaks during the European immigration crisis (2014-2016), and it is much higher than for L vs R. This could suggest that competition between P and NP is more central than the L vs R divide. It could also be due to significant variation in how L – R positions are defined and understood across national contexts, providing further evidence of the dominant importance of the populist

[24] Following GST, the populist partisanship of bigram j used by politician i in quarter t is:

$$\xi_{jit} = \frac{1}{2} - \frac{1}{2} \sum_{h \neq j} \left(\frac{q^1_{iht}}{1 - q^1_{ijt}} + \frac{q^0_{iht}}{1 - q^0_{ijt}} \right) \rho_{iht}$$

[25] Appendix Figure H2 reports the results with estimated confidence intervals for average partisanship. Since standard nonparametric bootstrap is invalid for lasso regression, we use subsampling instead, though it is considerably less efficient when the number of units (politicians) is relatively small.

Table 3: 15 most partisan populist and non-populist bigrams in 2015 Q3 and 2020 Q1

2015 Q3		*2020 Q1*	
Non-Populists	**Populists**	**Non-Populists**	**Populists**
Young People	Asylum Seeker	Young People	Take Care
Prime Minister	Greek People	Prime Minister	Good Morning
Climate Change	Good Morning	Will Continue	European Parliament
Sustainable Development	Via Youtube	Public Health	Fake News
The Anniversary	Press Release	Climate Change	Will Never
Economic Growth	Will Never	The Anniversary	Million People
Good Luck	Take Care	Continue Work	Health Care
Will Continue	Border Control	Rule of Law	Asylum Seeker
Work Together	Illegal Immigration	Work Together	People Will
Will Work	Today PM	President Republic	Bernie Sanders
Developed Countries	Jeremy Corbyn	Will Work	People Want
Well Done	I'm Going	Member State	Help You
Civil Society	Don't Want	Protect Health	Stay Home
Syrian Refugee	Common Sense	Economic Growth	Via Youtube
Renewable Energy	Get Rid	Last Year	Work People

Notes. The table reports the bigrams with the largest and smallest median values (taken over i) of the bigram partisanship measure ξ_{jit} for $t \in \{2015\ \text{Q3}, 2020\ \text{Q1}\}$ (footnote 24 defines ξ_{jit}). Larger median values correspond to populist bigrams, smaller median values – to non-populist bigrams.

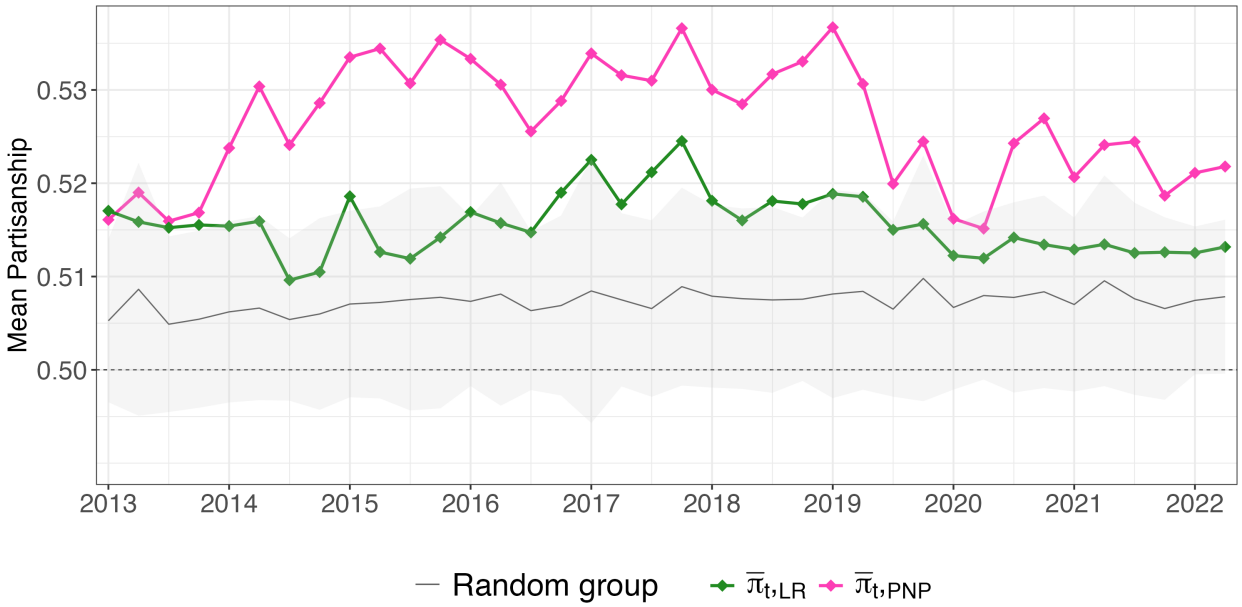
divide in our international sample of politicians.

Our estimate of populist partisanship is notably higher and more volatile than in GST's analysis of U.S. Congressional speeches over the past 35 years. As discussed in Appendix Section E, our estimates of $\hat{\pi}_t$ imply that, upon observing 5 bigrams (the typical content of a single tweet), the posterior of a politician being populist reaches about 0.6 during the refugee crisis. By contrast, the posterior of a speaker being Republican rather than Democrat in GST's sample, conditional on 5 bigrams, peaks at 0.55.[26] The larger and more volatile partisanship in our sample is likely to reflect multiple factors. Tweets represent a high-frequency, low-cost, and highly flexible medium of communication, allowing politicians to react immediately to unfolding events and tailor messages to specific audiences. In contrast, Congressional speeches are more static, tend to follow more standardized formats, with topics and rhetoric constrained by institutional norms. Additionally, our estimates are computed at the quarterly level, rather than every two years as in GST, and politicians differ in their levels of activity across time. These factors jointly contribute to the elevated and more volatile patterns of populist partisanship observed in our setting.

Finally, populist partisanship is substantially higher than that estimated when we randomly divide politicians into two groups reflecting the observed frequencies of true P /NP

[26] In GST's data, partisanship based on their preferred penalized estimator, which is comparable to the one we use, peaks at approximately 0.512 in 2010.

Figure 3: Average partisanship and χ^2 by calendar quarter



Notes. $\bar{\pi}_{t,LR}$ refers to average partisanship defined in the model with two political types: left and right; $\bar{\pi}_{t,PNP}$ refers to average partisanship defined in the model with two political types: populist and non-populist. Random group refers to average partisanship between two randomly defined political types, keeping the relative group sizes similar to those of the populist and non-populist groups; the shaded area is the 95% confidence interval. Partisanship is defined as in Equation 11. Average values are defined as in Equation 12.

types in our sample, repeating this procedure 100 times (the black line is the average, the shaded area is the 95% confidence interval). This random assignment serves as a benchmark to quantify the extent of remaining finite-sample bias in our measure of average partisanship. By construction, partisanship within random groups should be around 0.5. In our case, average random partisanship values are generally just below 0.51, although never significantly above 0.5. This is probably due to the relatively small number of politicians in our sample, and to the possible presence of unobserved political types. We discuss this point more extensively in the next section and in the online Appendix G.

Figures H3 and H4 in the online annex also document a peak in populist partisanship within the topic of immigration for the sample of European countries beginning in the fourth quarter of 2014, coinciding with the onset of the mass inflow of Syrian refugees; and a steep increase in populist partisanship within the topic of public health at the onset of the COVID-19 pandemic. The magnitude of partisanship within the immigration topic significantly exceeds that observed for aggregate partisanship, confirming that immigration is a highly salient and mobilizing topic, especially for populist politicians.

6 Electoral Cycles in Partisanship

In this section we exploit the fact that election dates differ across countries, as shown in Appendix Table H3, to describe how partisanship evolves over the electoral cycle. The election date refers to national parliamentary elections for parliamentary regimes. For presidential and mixed systems, the election date refers to all parliamentary elections and presidential elections in which more than two candidates are present in our sample of politicians (France, Mexico, the US, Slovenia, Portugal, Finland, Czech Republic).[27]

6.1 Average partisanship and distance from the election

The left hand panel of Figure 4 plots P/NP partisanship averaged across politicians by quarterly distance d from the election, d=0 being the election quarter. On average, populist vs non-populist partisanship increases in the run-up to the election, reaching a peak in the election quarter, and drops fast right after. In other words, as the election date approaches, the language of populist and non-populist politicians becomes more distinctive. Appendix Figure H5 shows that, when conditioning on calendar-quarter fixed effects, the pre-electoral increase in partisanship is even more pronounced. The shaded light pink area represents 90% point-wise confidence intervals of populist vs non-populist partisanship based on subsampling. Appendix D discusses how the confidence intervals are constructed.

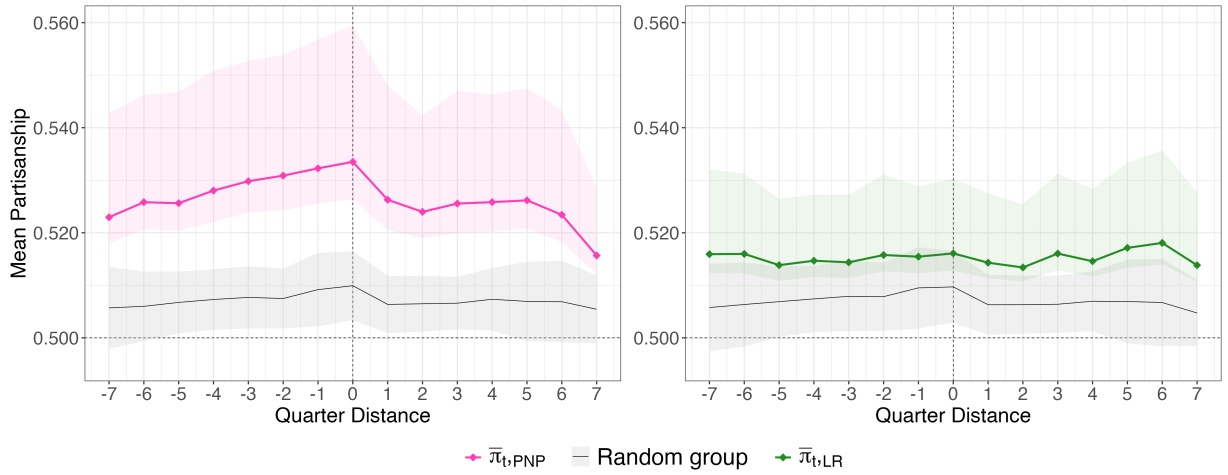
The right panel of Figure 4 plots average Left-Right partisanship over the electoral cycle, and the random group benchmark with L/R proportions. There is no evidence of electoral dynamics. This confirms the dominance of the populist vs non-populist dimension of electoral competition in our sample.

Note that all these patterns refer to aggregate measures of polarization, so they are also affected by changes in the composition of active politicians near the elections. We deal with this issue in Subsection 6.2 below, where we estimate an event study at the level of individual politicians.

The solid line at the bottom of the figure depicts partisanship for randomly grouped politicians, and the shaded gray area is the 95% confidence interval. P/NP partisanship is always well above the random group benchmark. Importantly, the difference between populist partisanship and the random group benchmark increases before the election and peaks at the election quarter, suggesting that the populist dimension of rhetorical polarization has a significant electoral component.

[27] We exclude elections at levels different than national. In our politician-election panel, the mean of the indicator for participating in an election (either directly as a candidate, or indirectly as the member of a party participating in the election) is 0.6.

Figure 4: Average estimated π_{it} per quarter difference from elections.



Notes. $\bar{\pi}_{t,PNP}$ refers to average partisanship defined in the model with two political types: populist and non-populist. $\bar{\pi}_{t,LR}$ – in the model with Left and Right political types. Shaded colored areas behind $\bar{\pi}_{t,PNP}$, $\bar{\pi}_{t,LR}$ are 90% pointwise CI based on subsampling. Random group refers to average partisanship between two randomly defined political types, with group sizes matching those of the corresponding political groups; the shaded grey area behind the random group is the 95% confidence interval. Partisanship is defined as in Equation 11. Average values are defined as in Equation 12.

When looking at the average random-group benchmark over the electoral cycle, it lies significantly above 0.5 and exhibits some pre-electoral dynamics. To explain these patterns, Appendix G runs Monte Carlo simulations using synthetic samples of politicians and bigram usage. We explore two issues: we vary the number of politicians in the sample, and we allow for the presence of unobserved political types. Both aspects are relevant. As the number of politicians increases, the random group benchmark converges towards 0.5. But the presence of electorally relevant unobserved groups also matters and can explain the pre-electoral patterns. The reason is that, in the presence of prominent unobserved types, their bigram usage may be spuriously attributed to random groups by the penalized estimator, resulting in partisanship estimates that systematically deviate from the 0.5 benchmark (although we find this effect to be small in practice). Appendix G provides further details on the simulation design and results.[28]

[28]The penalized estimate of partisanship is bounded below by 0.5, but could exceed 0.5 in the random sample of politicians due to its non-linearity and convexity. The reason is that, despite the Lasso penalization, in a relatively small sample some bigrams will, by pure chance, correlate with the random political label. We verify that this is indeed the case by implementing the leave-out estimate of partisanship proposed by GST, which is an out-of-sample estimator, and as such it is much less subject to the positive bias due to Jensen's inequality. The leave-out estimate recovers the 0.5 benchmark for random-group partisanship even when unobserved types are present, in line with the findings of GST.

Magnitudes Recall that partisanship measures the average predictability of a politician's type, conditional upon observing a single bigram. According to Figure 4, on average for populist vs non-populist politicians this predictability increases from 0.522 seven quarters before the election to 0.535 in the election quarter, and then declines by about the same amount two quarters after the election. How can this magnitude be interpreted?

To address this question, Appendix E computes speech informativeness conditional on observing n bigrams, for different values of n, following GST. These computations imply that, upon reading a single tweet two years before the election, an observer with neutral priors would raise its posterior probability of correctly classifying the politician as P or NP by about 7 percentage points (from 0.5 to about 0.57). During the election quarter, instead, the increase in predictability from reading one tweet jumps to almost 12 percentage points (from 0.5 to 0.62), almost doubling the informativeness of a single tweet. After reading two tweets (10 bigrams) during the election quarter, predictability increases by 17 percentage points (from 0.5 to 0.67), in contrast to an increase of 12 percentage points (from 0.5 to 0.62) if two tweets are read two years before the election, an increase of about 50%. Note that the largest marginal increases in informativeness are obtained after just a few bigrams.

We can compare these estimates to the informativeness of text in 2015 Q3, during the first acute phase of an immigration crisis and when partisanship reached a maximum in our sample. The texts from 2015 Q3 are slightly less informative about populist partisanship than the texts written during the election quarter.

6.2 Event study

The estimates displayed in Figure 4 could reflect the coincidence of elections with other relevant events, or changes in the composition of active politicians. To control for these potential confounders, we exploit staggered election dates across countries and estimate an event study on i's partisanship, π_{it}, with fixed effects for politician (γ_i), calendar date (δ_t) and country by election-window (defined as country-quarters closest to each election, El_{it}):

$$\pi_{it} = \sum_{d(i,t)} \beta_{d(i,t)} D_{d(i,t)} + \gamma_i + \delta_t + El_{it} + \epsilon_{it}, \tag{13}$$

where $D_{d(i,t)}$ are indicators for quarterly distance from the election - which in our baseline specification include 3 quarters before and after the election as well as the election quarter. Country by election-window fixed effects, El_{it}, are included to capture country specific shocks occurring during the election-window, but results are robust to dropping El_{it}. Thus, the estimated coefficients $\beta_{d(i,t)}$ measure the average effect on rhetorical polarization of being at

distance $d(i,t)$ from the election, relative to the average quarters at a greater distance from the election within each election-window, for the same politician. Errors are clustered at the (closest) election level - the treatment variable of interest.[29]

The event study for P/NP partisanship, depicted in Figure 5, confirms that the language used by politicians becomes more polarized two quarters before the election and remains so until the election date. Once the election is over, however, rhetorical heterogeneity declines to its average level outside of the election window. In terms of magnitudes, the estimated rise in populist partisanship during the election quarter relative to non-election quarters is about 15% of one standard deviation of the populist partisanship measure across the full sample (0.008/0.054).

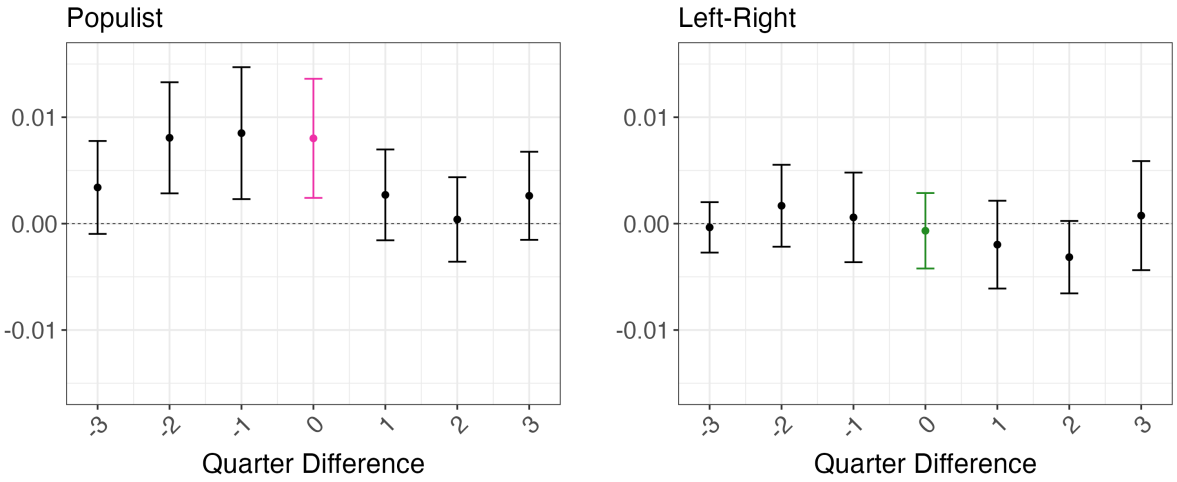
In Appendix Table H2, we report a Wald test for the null that the estimated coefficient β for the election quarter ($d(i,t) = 0$) is equal to those after the election. The null is rejected with p-values of 0.05, 0.03 and 0.3 for 1, 2 and 3 quarters after the elections respectively.

The right panel of Figure 5 depicts the event study for L vs R partisanship. Rhetorical polarization between these two groups does not rise before the elections. Appendix Figure H6 depicts the event study on partisanship when the probabilities q are estimated allowing for four political types, as in Appendix C. The left-most panel refers to populist vs non-populist partisanship, the center panel to left-right partisanship, and the right-most panel refers to partisanship between four types (note that the scale here is smaller, since with four types partisanship ranges between 0.25 and 1). The pre-election pattern is present only for populists vs non-populists and for the four types, but not for the left-right division.

Overall, these findings suggest that the pre-electoral increase in rhetorical polarization is only present for the populist dimension of political competition. From here on we focus in more detail on this dimension only.

[29] That is, clustering is among observations sharing the same fixed effect El_{it}. Thus, we allow arbitrary correlations in residuals among observations closest to the same country-election date. To define El_{it}, for each country and calendar quarter, we find which national election is closest to the middle day of that quarter. This creates an electoral-cycle fixed effect by country, with no ties because closest is defined using the exact middle day of the quarter.

Figure 5: Partisanship event study



Notes. Left panel plots the event study results for the populist partisanship estimates; Right panel – the Left-Right estimates. Each figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from specification 13. Outcome variable is populist/Left-Right partisanship π_{it}. Errors are clustered at the elections level, 95% CI are reported. Average populist partisanship is 0.523, average Left-Right partisanship is 0.516. Partisanship is defined as in Equation 11.

7 Between-issue and Within-issue Divergence

So far we have established that, as predicted by the theory, rhetorical polarization between P and NP politicians rises in the quarters before an election and reverts afterwards (Section 8 shows that this pattern is very robust). This result establishes *that* competing politicians diverge, but not *how*. In this section we test the two sharper predictions about the anatomy of this divergence. To do so, we focus on the electoral cycle in specific components of $\mathbf{q}^1_{it} - \mathbf{q}^0_{it}$, measuring the difference in the probability of bigram usage between opposite political types.

7.1 Between-issue divergence

Prediction 1 says that, as the election approaches, both political types speak more often about the issues more salient to their constituency, and hence distinctive of their type—the *extensive margin* of bigram choice across topics. To test it, we exploit the two sets of topics listed in Table 2, namely we pool all populist topics (with $\rho_k > 0.5$ in Table 2) into a single broader populist set indexed by k_1, and all non-populist topics (with $\rho_k < 0.5$) into a broader non-populist set indexed by k_0.[30]

[30] In principle the analysis could be run separately for a larger number of topics, but in our setting many individual topics are too small to support this. See however the robustness analysis in section 8.

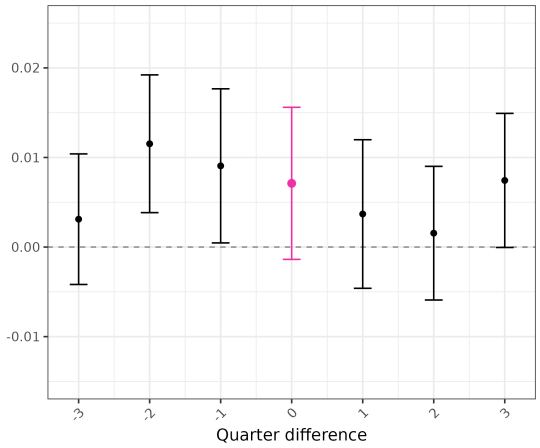
Let $q^1_{i,k_1,t} = \sum_{j \in k_1} q^1_{ijt}$ and $q^1_{i,k_0,t} = \sum_{j \in k_0} q^1_{ijt}$ respectively denote the probability that politician i uses bigrams drawn from populist (resp. non-populist) topics, if i is of type P. Variables $q^0_{i,k_1,t}$ and $q^0_{i,k_0,t}$ are similarly defined, for type NP. Because the type-specific shares are not separately identified, we work with their difference,

$$\Delta q_{i,k_1,t} = q^1_{i,k_1,t} - q^0_{i,k_1,t},$$

namely the gap in the frequency with which P and NP politicians (with the same observable characteristics) discuss the populist set k_1. We normalize bigram frequencies as if speech only consisted of classified bigrams, so that by construction $\Delta q_{i,k_1,t} = -\Delta q_{i,k_0,t}$.[31]

The prediction is that $\Delta q_{i,k_1,t}$ increases before the election and then reverts to the off-election average. We test it with the event-study specification of Equation (13), using $\Delta q_{i,k_1,t}$ as the outcome. Figure 6 presents the results. In the quarters preceding the election, P types talk more frequently about populist topics, compared to NP types with the same features—and less frequently about non-populist topics. When electoral stakes rise, politicians tilt the agenda toward the issues that are more distinctive of their type over the entire sample, as predicted by the theory. The effect is sizable: the pre-electoral coefficient estimates are around 15% of the standard deviation of the dependent variable.

Figure 6: Predicted difference in frequencies of populist-topic speech between populist and non-populist politicians



Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from specification 13 with $\Delta q_{i,k_1,t} = q^1_{i,k_1,t} - q^0_{i,k_1,t}$. $\Delta q_{i,k_1,t}$ is a predicted difference in frequencies of speech for the populist topic between populist and non-populist politicians. Errors are clustered at the elections level, 95% CI are reported.

[31] About 12% of bigrams remain unclassified, meaning that the two dimensions do not cover the entire vocabulary.

7.2 Within-issue divergence

Prediction 2 concerns the *intensive margin*: as the election approaches, politicians use bigrams more extreme and more distinctive of their type *within* a given issue. To test it, we use a time-invariant measure of bigram populist distinctiveness, $\xi_j = \text{median}_{i,t}\, \xi_{ijt}$, where ξ_{ijt} is the bigram-level contribution to partisanship defined in footnote 24 and obtained from a specification of (7) in which we constrain ϕ_j to be constant over time.[32] The index ξ_j has both sign and magnitude: positive values identify populist bigrams, negative values non-populist ones, and the absolute value measures how strongly the bigram is associated with one type. Because ξ_j is fixed over time, it provides a measure of bigram *extremism* over the entire sample period in the ideological dimensions distinguishing populist and non-populist politicians.

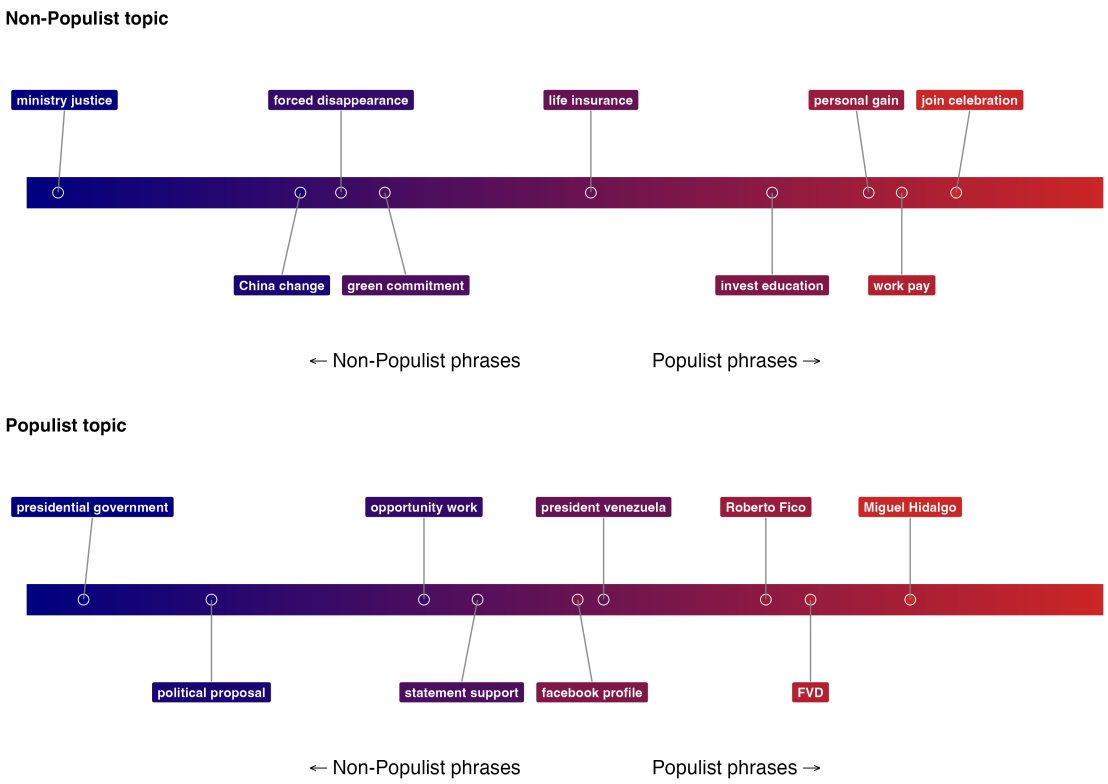
Figure 7 plots examples of bigrams with different values of ξ_j, within the set of populist (k_1) and non-populist topics (k_0) listed in Table 2. While these dimensions do not map onto the traditional Left-Right cleavage, they meaningfully organize the competition between populist and non-populist politicians. Non-populist topics are more policy-oriented, with distinctively non-populist bigrams referring to foreign policy or the environment, among others, while populist bigrams focus on issues and rhetoric closer to that of populist politicians. By contrast, within populist topics, distinctively populist bigrams refer to populist politicians or their parties, such as Roberto Fico or FvD, and invoke national historical figures, such as Miguel Hidalgo, whereas non-populist bigrams concern democratic institutions and bodies of the establishment. Overall, the figure highlights substantial differences in language across political types within each set of topics, and provides guidance for interpreting the within-topic divergence results that follow.

We normalize $\mathbf{q}^0_{iKt}$ and $\mathbf{q}^1_{iKt}$ to sum to one within each set of topics $K = k_1, k_0$, quarter t and politician i, so that the analysis isolates the composition of language within a set of topics. As before, only the difference between the two political types is identified. Define $Y_{iKjt} = q^1_{i,K,j,t} - q^0_{i,K,j,t}$. This variable measures the gap between P and NP types in the frequency of use of bigram j assigned to topics K. Prediction 2 says that, before the election, Y_{iKjt} *increases* for bigrams more distinctively populist (i.e. with higher values of ξ_j) within each set of topics $K = k_0, k_1$, and *decreases* for less distinctively populist bigrams (i.e. with lower ξ_j) within each K.

To test this without imposing a functional form on how the gap varies with distinctiveness, we partition the bigrams of each set of topics K into overlapping 20-percentile windows of ξ_j,

[32] This enables us to measure how *distinctive* is bigram j of each political type over the entire sample period and not just in quarter t, allowing a comparison of bigram usage over time. The results are robust to taking median over ξ_{ijt} using only quarters outside the election window, or a random subsample of quarters.

Figure 7: Examples of bigrams by ideological extremism across both sets of topics



Notes. The figure shows the original, non-stemmed versions of randomly selected bigrams from topics distinctive of non-populist and populist politicians, based on their distinctiveness ξ_j, defined as in footnote 24. Bigrams to the left (dark blue) are more distinctive of non-populist politicians, while bigrams to the right (bright red) are more distinctive of populist politicians.

stepped by 10 percentiles, yielding 9 percentile windows: $[0, 20\%)$, $[10, 30\%)$, ..., $[80, 100\%]$. Letting $\mathcal{Q}_{K,b}$ be the set of bigrams in window b within topics K, we define, for each politician i, quarter t, set of topics K and window b,

$$Y_{iKbt} = \sum_{j \in \mathcal{Q}_{K,b}} Y_{ijt},$$

and estimate the event-study specification of Equation (13) separately for each window b, with Y_{iKbt} as the outcome. We obtain a set of estimated coefficients, $\hat{\beta}_{Kdb}$, for each set of topics K, each quarterly distance d from the election and for each window of populist distinctiveness b. These estimated coefficients capture how frequently bigrams in window b and topics K are used by politicians of type P (relative to type NP) at elections distance d, compared to their relative frequency of use in quarters more distant from the election.

Figure 8 plots these estimated coefficients $\hat{\beta}_{Kdb}$ and their 95% confidence intervals, for each quarterly distance to the election d and for each quantile window b ordered from the

most non-populist bigrams (dark blue, left) to the most populist bigrams (bright red, right), following the same color scale as in Figure 7. Darker colors denote statistically significant estimates $\hat{\beta}_{Kdb}$. Panel (a) refers to non-populist topics, panel (b) to populist topics. The pattern matches the prediction. Within both sets of topics, starting two or three quarters before the election and until the election date, P politicians use more frequently the more distinctively populist bigrams—and less frequently the less distinctive ones— relative to NP politicians, compared to quarters distant from the election. The estimated values of $\hat{\beta}_{Kdb}$ in pre-electoral quarters are about 10-30% of the standard deviation of the dependent variable, depending on windows and distance from election.

Results are especially pronounced for non-populist bigrams within populist topics, which, as noted above, mostly concern democratic institutions and bodies of the establishment. While we cannot tell whether this is due to P types using these bigrams less or NP types using them more, the imminence of elections clearly widens the pre-existing gap in their usage (the mean of Y_{iKbt} for the bigrams in the lowest five windows b is below 0). Paradoxically, a byproduct of this particular form of rhetorical divergence during an electoral campaign could be to erode trust in democratic institutions and undermine their perceived legitimacy.

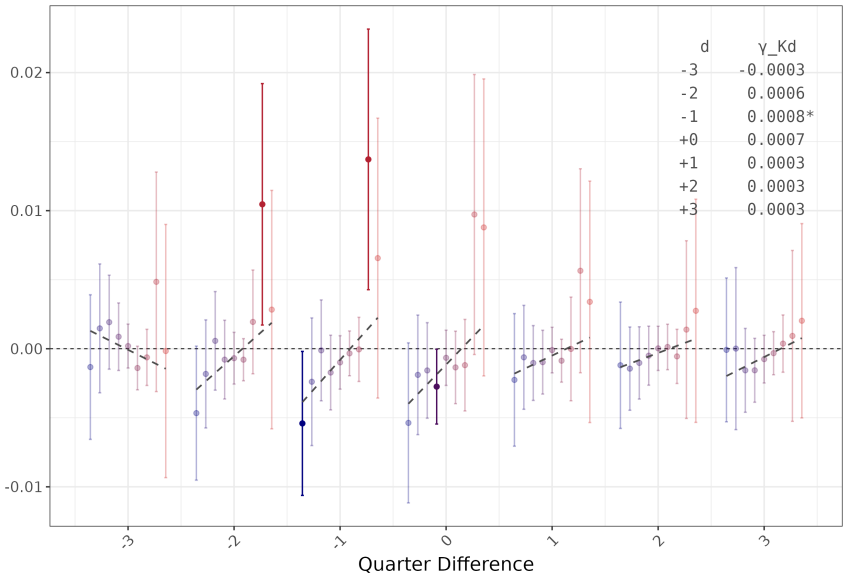
The gradient of how the event-study estimates $\hat{\beta}_{Kdb}$ vary with populist distinctiveness (i.e. with window b) is broadly monotonic: $\hat{\beta}_{Kdb}$ is lowest for the bottom windows (more non-populist bigrams) and highest for the top windows (more populist bigrams). After the election, the pattern in the $\hat{\beta}_{Kdb}$ coefficient is much less pronounced and the estimated $\hat{\beta}_{Kdb}$ are never statistically significant, for all windows b.

To summarize this gradient while accounting for estimation uncertainty, we fit a linear trend across windows by weighted least squares. Namely, for each quarterly distance $d \in \{-3, \ldots, 3\}$ we estimate the cross-window slope

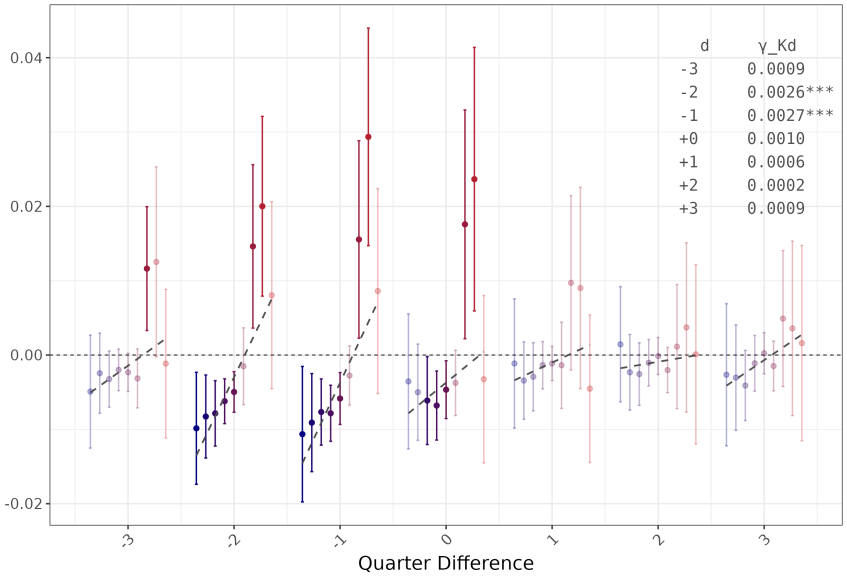
$$\hat{\beta}_{Kdb} \;=\; \alpha_d + \gamma_{Kd}\, b + u_{db}, \qquad \text{weights} = \frac{1}{\widehat{SE}^{\,2}_{Kdb}}.$$

where $\widehat{SE}_{Kdb}$ is the estimated standard error of $\hat{\beta}_{Kdb}$. The standard error of $\hat{\gamma}_{Kd}$ accounts for both the estimation uncertainty in the $\hat{\beta}_{Kdb}$ and their correlation across windows. For each electoral distance d, the slope coefficient γ_{Kd} captures how strongly the relative tendency of P to use a bigram rises with its populist distinctiveness. The gray dashed line in each panel plots the fitted trend, and the slope γ_{Kd} and its significance are reported in the upper-right corner. For both sets of topics the slope is largest in the two quarters immediately preceding the election, and the effect is more pronounced for populist topics. The latent measure ξ_j, recovered from overall speech, thus shapes communication over the electoral cycle in the way the theory predicts, and especially so on the issues that define populist identity.

Figure 8: Event-study estimates for within-topics divergence separately for populist and non-populist dimensions.



(a) Non-Populist Topics, $K = k_0$



(b) Populist topics, $K = k_1$

Notes. Colors in the output plots correspond to windows, ranging from dark blue (most non-populist bigrams) to bright red (most populist bigrams). Non-significant coefficients ($p > 0.05$) are shown in faded colors. The figure plots the event study estimates from the specification 13 with an outcome being differences in bigram usage between populist and non-populist politicians, $Y_{ijt} = q^1_{i,j,t} - q^0_{i,j,t}$, across bigrams with varying distinctiveness ξ_j. $\mathbf{q}^0_{iKt}$ and $\mathbf{q}^1_{iKt}$ are normalized to sum to one within each set of topics $K = k_1, k_0$, quarter t and politician i. Bigrams are grouped into overlapping 20-percentile windows of distinctiveness values ξ_j, with a 10-percentile overlap, separately for non-populist and populist topics (J_0 and J_1).

Taken together, the two tests are in line with both predictions of the model. As the election approaches, populists and non-populists diverge along the *extensive* margin—each type tilts its agenda toward the issues that favor its own constituency (Prediction 1)—and along the *intensive* margin—each type frames those issues in increasingly distinctive language (Prediction 2). This is the anatomy behind the aggregate rise in partisanship of Section 6: the pre-electoral divergence is not confined to topic selection or to slogans and party names, but extends to how policy issues themselves are discussed.

8 Robustness

We now discuss a number of possible concerns with our estimates of partisanship, and show their robustness.

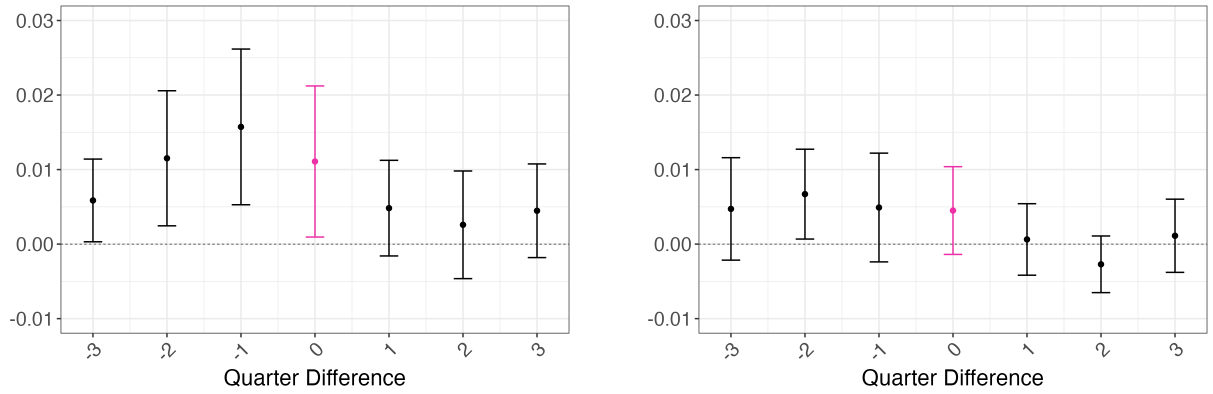
Different electoral systems In standard models of electoral competition where candidates announce their policy positions ahead of the elections, like Downs' or probabilistic voting, candidates' incentives to converge depend on the number of competing candidates. Two vote maximizing candidates who compete for the same voters have strong incentives to converge. With more than two candidates or with the threat of entry by a third candidate, however, convergence is not assured (e.g. Palfrey (1984)). Although we have emphasized the distinction between only two political types, several countries in our sample have more than two parties. Could this account for the observed divergence?

To address this question, we split the sample in two groups of elections: (i) all national elections held in countries with plurality rule and in presidential regimes plus all presidential elections, where competition is typically between two serious parties or candidates; (ii) parliamentary elections in proportional systems, that typically have several competing parties and candidates in the same districts.[33]

Figure 9 reproduces the event study separately for these two groups of elections. While the pre-electoral polarizing effect is present in both groups of elections, it is much larger for plurality rule and presidential elections. Hence, it is unlikely to be driven by competition between more than two parties.

[33]Group (i) includes all elections held in Australia, the UK, the US, France, Brazil, Mexico and Hungary, plus presidential elections with more than one active candidate in our sample of political leaders in Finland, Poland and Slovenia. In Brazil presidential and parliamentary elections are always held in the same dates. In Mexico three congressional elections in our sample were held without a contemporaneous presidential election; we nevertheless include all Mexican elections in group (i) because they tend to be dominated by electoral competition for president. Hungary has a mixed system that strongly favors the largest party.

Figure 9: Partisanship event study estimates for elections under different electoral systems.



(a) Plurality rule and presidential elections

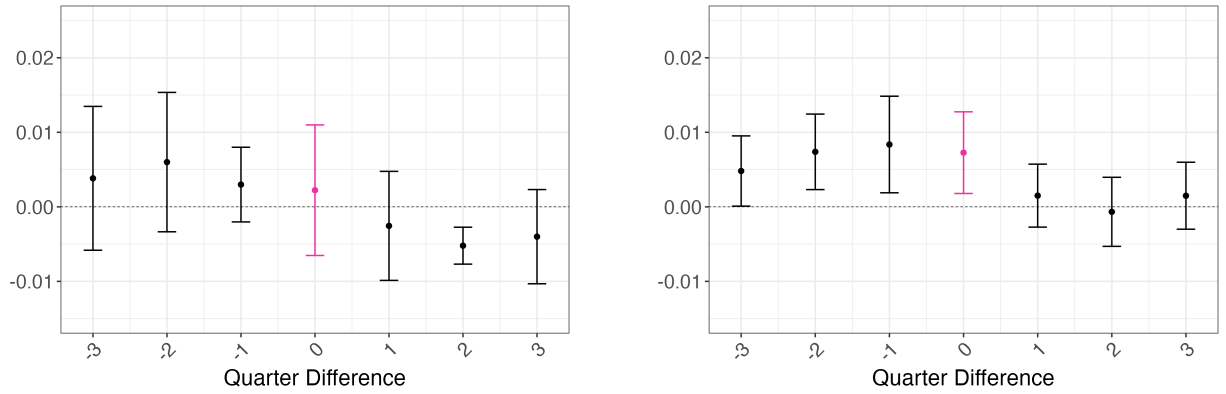
(b) Proportional rule

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from specification 13 separately for the two subsamples of elections held under different institutional rules. Left Panel plots the results for all national elections held in countries with plurality rule and in presidential regimes, as well as for all presidential elections, where competition is typically between two serious parties or candidates; Right Panel – for parliamentary elections in proportional systems, that typically have several competing parties and candidates in the same districts. Outcome variable is partisanship, π_{it}, as defined in Equation 11. Errors are clustered at the elections level, 95% CI are reported.

Snap elections About 20% of elections in our data are snap elections - i.e. called prematurely before the end of the legislature. Although snap elections do not happen without premonitions, they may have a shorter pre-electoral cycle in communications. Moreover, unobserved confounding events could cause both a rise in rhetorical polarization and the occurrence of snap elections. To allow for this, we estimate the specification separately for elections held on time and the elections held ahead of time. Figure 10 shows that indeed snap elections induce a shorter and less precisely estimated cycle of polarization (also because of the smaller sample), which is more focused on the electoral quarter itself. Nevertheless, results remain unaffected for the regular election dates.

Being a candidate Although all politicians in our sample are prominent political leaders with strong stakes in the election outcomes, not all of them stand as candidates in all elections. On the one hand, one would expect that the leaders who are also candidates have sharper electoral incentives, and so their rhetoric should exhibit a stronger pre-electoral cycle. On the other hand, rhetorical polarization could be the result of politicians diverging in their immediate goals – some of them are seeking reelection, while others are continuing with their usual agenda, or supporting running candidates. To explore these conjectures, we analyze how the results differ among politicians who are candidates in given elections, and those who are not. Figure 11 shows that the pre-election cycle is starker among the running

Figure 10: Event study estimates for π_{it} for snap elections (left panel) and elections held in standard time (right panel)

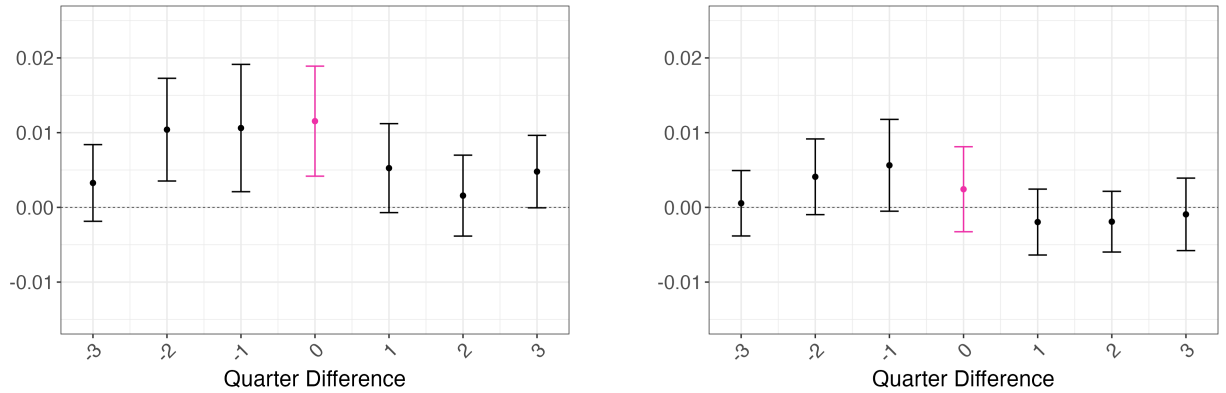


(a) Snap elections (b) Not snap elections

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13 on the subsample of snap elections (left panel) and elections held at a regular time (right panel). Errors are clustered at the elections level, 95% confidence intervals are reported. Partisanship is defined as in Equation 11. Snap elections have been manually classified, with the use of both ChatGPT and Wikipedia. We consider elections to be "snap" if they are held at least 6 months before the natural end of the legislature. Out of 97 unique elections in our sample, 20 elections were classified as "snap".

candidates, in line with the idea that it reflects their stronger electoral incentives, rather than different goals of politicians in more heterogeneous situations.

Figure 11: Event study estimates for π_{it} separately for politicians running for reelection in given elections (candidate, left panel), and those who are not running (not candidate, right panel)



(a) Candidate (b) Not candidate

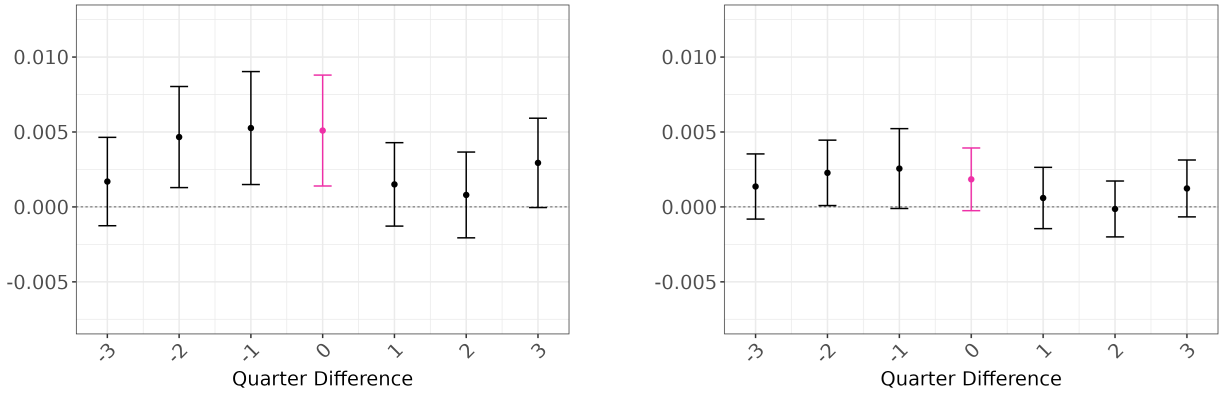
Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13 on the subsample of politicians running for office in a given electoral cycle (left) and those not running (right). Errors are clustered at the elections level, 95% confidence intervals are reported. Partisanship is defined as in Equation 11.

Partisanship between and within topics Our classification of topics into two categories, each one distinctive of a political type, may be subject to measurement error. An alternative approach is to study between- and within-topic divergence on the primitive 19 topics, without attempting to classify them as distinctive of P vs NP types. Although we lose the ability to test the more specific theoretical predictions, we can still explore whether increased rhetorical divergence before the election is due to politicians talking about different issues, or framing each issue differently, or both.

Following GST, define partisanship *within* topic k in quarter t, $\hat{\pi}^k_{it}$ as in (11), except that $\hat{q}^{P_i}_{ijt}$ is averaged only over bigrams that are used to speak about topic k. Within-topic partisanship of i for all 19 topics is the average of $\hat{\pi}^k_{it}$ over topics, weighting each topic k by its frequency of use by i in quarter t. Partisanship of i *between* topics, instead, is constructed using $\hat{q}^{P_i}_{ikt} = \sum_{j \in k} \hat{q}^{P_i}_{ijt}$ computed with topics (rather than bigrams) as units of speech - i.e., using the probability that i speaks about issue k in quarter t. In all cases, $\hat{q}^{P_i}_{ijt}$ is always estimated over the entire sample of bigrams. The between-topic component measures how predictable a politician's type is from the topics he chooses to discuss; the within-topic component measures how predictable his type is from the specific bigrams he uses, given the topic.

Figure 12 reports the event studies. Both between- and within-topic partisanship between P and NP increase before the election, with stronger results within topics. Thus, both the extensive and intensive margin contribute to explain the electoral pattern in partisanship, as predicted by the theory.

Figure 12: Event studies for the between- and within-topics partisanship



(a) Within-topics (b) Between-topics

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from specification 13, using within-topic partisanship (left panel) and between-topic partisanship (right panel) as outcome variables. Errors are clustered at the elections level, 95% CI are reported. Section 8 defines within- and between-topic partisanship.

English speaking countries Given that our analysis is in a multinational context and we translate all the tweets to English, one might suspect that our results are driven more by English-speaking countries. This could also potentially explain why the electoral cycles are more pronounced in the group of plurality rule and presidential elections, where the fraction of English speaking countries is higher. But Appendix Figure H7 shows that this is not the case: partisanship increases in both groups of countries, with the effect actually being more pronounced in non-English-speaking countries. As discussed in Section 4, our back translation metrics also point to a high quality of our translations (in terms of similarity in semantic meaning between the original and the back translated tweets).

Time window We also perform robustness checks, with regard to the length of the time window of the event study. In Appendix Figure H8 we consider longer windows of 4, 5, 6 quarters (rather than 3) around the elections. In principle this could matter because it changes the non-treated observations. But the main results are confirmed, and the election cycle generally begins two quarters before the election, as in the baseline estimates.

Heterogeneous treatment effects As in any two-way fixed-effects design, our event-study estimates could be contaminated by treatment effects that are heterogeneous across elections. Appendix Figure H9 shows that our results are robust to this concern. We re-estimate the specification with leader $\times$ election-window fixed effects, restrict the reference group to a narrow set of pre-election quarters, and implement the interaction-weighted estimator of Sun and Abraham (2021), which interacts the quarter-to-election indicators with election-cycle indicators to recover a separate effect for each cycle at each relative quarter and then aggregates these with non-negative weights.

Verbosity Politicians are more verbose as the election approaches, and verbosity reverts to its average afterwards (Appendix Figure H10). This too is an implication of the theory. Could this electoral pattern in verbosity, rather than the language used, explain our divergence results? We don't think so. First, partisanship measures predictability conditional on a single bigram, not on the entire quarterly sequence of bigrams. Second, verbosity is a parameter in the Poisson model used to estimate bigram frequency. So, unless there is a direct link between verbosity and speech polarization, this should not be a concern (Gentzkow, Shapiro and Taddy, 2019). Third, the model is estimated at the level of calendar quarters, and the political-type coefficient ϕ_{jt} is not individual specific by design. Hence, verbosity of individual politicians cannot directly translate into higher partisanship estimates.

Self-selection into being active A potentially more relevant issue is the extensive (rather than intensive) margin of politicians' activity. Quarters closer to the election could have a larger number of active politicians. By construction, our measure of segregation is based on estimated non-zero coefficients ϕ_{jt} from the Lasso-like optimization problem. Hence, all else equal, the larger is the number of speakers active during the quarter, the more there are non-zero estimated bigram-specific loadings on the populist dummy variable, and the larger is the estimated value of partisanship.

This issue too is not a concern in our sample, however, for several reasons. First, we estimate partisanship over calendar quarters. Hence, the number of speakers could be causing a problem only if those calendar quarters with a larger number of speakers are also more likely to be election quarters. This is hardly the case (see figure H11 in the Appendix).

Second, while our results indicate the profound effect on partisanship in the electoral quarter and in the two quarters immediately preceding it, the number of active speakers does not differ much within a six months window around election time (Appendix Figure H12). The reason is that our sample consists almost exclusively of high-ranking politicians, most of whom use Twitter on a regular basis to communicate with the electorate. This is illustrated in Appendix figure H13, which provides the distribution of the number of politicians corresponding to each maximum absolute quarterly distance from the elections. There are almost no politicians who tweet exclusively around electoral quarters.

Third, in the event study we explicitly control for all quarters in the election window, thus capturing mean partisanship levels during these quarters. Hence, indirectly, this captures the popularity of the quarter too.

Fourth, to make sure that our results are not driven by different numbers of speakers per calendar quarters, we perform a robustness check, keeping 220 speakers[34] in each calendar quarter at random. Note that most quarters in our sample have more than 220 active speakers, as shown in Figure H11, so this simulation could give less precise and smaller estimates of partisanship. The event study results of such robustness analysis are in Appendix Figure H14 and confirm the main results.

Sample of bigrams We also applied different thresholds for bigram selection. Appendix Figure H15 shows that the event studies are unaffected.

Populist classification Our findings remain robust under an alternative classification of populist politicians, in which ambiguous cases are coded as non-populists rather than as

[34] 220 is the minimum observed number of politicians (2013 Q1) that allows us to perform selection without replacement.

populists, as in the main specification – Appendix Figure H16.

Mentions of specific politicians and parties One concern is that our results may be driven by mentions of specific parties and politicians closer to elections. Appendix Figure H17 shows that the results are robust to excluding bigrams of the specific parties & politicians topic. Moreover, if this was driving the result, we should also observe pre-electoral divergence for L vs R politicians, but we don't.

9 Concluding Remarks

This paper studies how electoral competition shapes political communication. First, we formulate a theory of electoral competition with exclusive audiences. Second, using high-frequency data from Twitter/X for 367 leaders in 21 democracies over 2013 - 2022, we document a rise in rhetorical polarization among populist vs non-populist politicians before the elections, and unpack it. The populist dimension plays a central role in this process. Compared to the conventional left-right divide, populist versus non-populist polarization is both more pronounced and more responsive to the electoral cycle.

This electoral cycle suggests that rhetorical divergence is not driven solely by fixed ideological differences. Rather, it reflects an opportunistic response to electoral incentives, as politicians strategically differentiate their communication when electoral stakes are highest. Guided by the theory, we show that this pre-electoral differentiation involves politicians: (i) speaking more about their distinctive issues, and (ii) choosing more extreme and more distinctive vocabulary within each issue. These patterns are consistent with a setting in which populist vs non-populist politicians do not compete for the same swing voters, but instead reach and seek to mobilize different groups of voters.

These findings point to two important directions for future research on polarization. First, is pre-electoral differentiation a general feature of political communication, or is it only present on Twitter/X? The convergence results by Di Tella et al. (2025) on candidates' official positions suggest that social media may play a special role. But this question deserves further attention, both with regard to other social media, and because the convergence result in Di Tella et al. (2025) reflects a comparison of final elections with US primaries or French first round elections, whereas we compare pre-election quarters with post-election or more distant quarters.

Second, greater attention should be paid to how platform design and targeting technologies affect political competition. Much of the literature on social media has emphasized that echo chambers or the diffusion of emotional content may fuel political extremism, although

the evidence points in both directions (Melnikov, 2024 and Allcott et al., 2025). Our results highlight another important mechanism, operating on the supply side. Some social media enable politicians to select their audience. For instance Allcott et al. (2025) show that, in the 2020 Presidential election, political ads on Facebook and Instagram predominantly reached the parties' own supporters. This targeting in turn may profoundly alter candidates' incentives, encouraging differentiation and polarizing propaganda.

References

Alesina, Alberto. 1988. "Credibility and policy convergence in a two-party system with rational voters." *The American Economic Review* 78(4):796–805.

Allcott, Hunt, Matthew Gentzkow, Ro'ee Levy et al. 2025. The Effects of Political Advertising on Facebook and Instagram before the 2020 US Election. Technical report National Bureau of Economic Research.

Ash, Elliott, Massimo Morelli and Richard Van Weelden. 2017. "Elections and Divisiveness: Theory and Evidence." *The Journal of Politics* 79(4):1268–1285.

Ash, Elliott and Stephen Hansen. 2023. "Text algorithms in economics." *Annual Review of Economics* 15:659–688.

Aslanidis, Paris. 2018. "Measuring populist discourse with semantic text analysis: an application on grassroots populist mobilization." *Quality & Quantity* 52(3):1241–1263.

Barberá, Pablo. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *Political analysis* 23(1):76–91.

Battiston, Giacomo. 2022. "Rescue on Stage: Border Enforcement and Public Attention in the Mediterranean Sea." *Available at https://www.dropbox.com/sh/ct8qe4x71l7hjry/AABxUJxsYmQy1Tea2* .

Besley, Timothy and Stephen Coate. 1997. "An economic model of representative democracy." *The Quarterly Journal of Economics* 112(1):85–114.

Bonica, Adam, Kasey Rhee and Nicolas Studen. 2025. "The Electoral Consequences of Ideological Persuasion: Evidence from a Within-Precinct Analysis of US Elections." *Available at SSRN* .

Bonomi, Giampaolo. 2024. "Divide and diverge: Polarization incentives." *Available at https://ssrn.com/abstract=4990250* .

Brady, David W., Hahrie Han and Jeremy C. Pope. 2007. "Primary Elections and Candi-

date Ideology: Out of Step with the Primary Electorate?" *Legislative Studies Quarterly* 32(1):79–105.

Canes-Wrone, Brandice, David W Brady and John F Cogan. 2002. "Out of step, out of office: Electoral accountability and House members' voting." *American Political Science Review* 96(1):127–140.

Cassell, Kaitlen J. 2023. "The Populist Communication Strategy in Comparative Perspective." *The International Journal of Press/Politics* 28(4):725–746.

Desmet, Klaus, Ignacio Ortuño-Ortín and Romain Wacziarg. 2025. Latent Polarization. NBER Working Paper 34229 National Bureau of Economic Research.

Di Tella, Rafael, Randy Kotti, Caroline Le Pennec and Vincent Pons. 2025. "Keep your enemies closer: strategic platform adjustments during US and French elections." *American Economic Review* 115(8):2488–2528.

Fiva, Jon, Oda Nedregård and Henning Øien. forthcoming. "Group Identity and Parliamentary Debates." *Journal of Politics* .

Fowler, Anthony and Shu Fu. Forthcoming. "Do primary elections exacerbate congressional polarization?" *Journal of Politics* .

Funke, Manuel, Moritz Schularick and Christoph Trebesch. 2023. "Populist leaders and the economy." *American Economic Review* 113(12):3249–3288.

Gandhi, Amit and Aviv Nevo. 2021. Empirical models of demand and supply in differentiated products industries. In *Handbook of Industrial Organization*. Vol. 4 Elsevier pp. 63–139.

Gauthier, Germain, Philine Widmer and Elliott Ash. 2025. Deep Latent Variable Models for Unstructured Data. Technical report mimeo.

Gennaioli, Nicola and Guido Tabellini. 2025. "Identity Politics." *Econometrica* 93(6):1937–67.

Gentzkow, Matthew, Bryan Kelly and Matt Taddy. 2019. "Text as data." *Journal of Economic Literature* 57(3):535–574.

Gentzkow, Matthew, Jesse M Shapiro and Matt Taddy. 2019. "Measuring group differences in high-dimensional choices: method and application to congressional speech." *Econometrica* 87(4):1307–1340.

Glaeser, Edward L., Giacomo Ponzetto and Jesse Shapiro. 2005. "Strategic Extremism: Why Republicans and Democrats Divide on Religious Values." *The Quarterly Journal of Economics* 120(4):1283–1330.

Gordon, Matthew, Megan Ayers, Eliana Stone and Luke Sanford. 2025. "Remote Control:

Debiasing Machine Learning Predictions for Causal Inference." *Working paper* .

Guriev, Sergei and Elias Papaioannou. 2022. "The Political Economy of Populism." *Journal of Economic Literature* 60(3):753–832.

Guriev, Sergei, Nikita Melnikov and Ekaterina Zhuravskaya. 2021. "3g internet and confidence in government." *The Quarterly Journal of Economics* 136(4):2533–2613.

Hall, Andrew B. 2015. "What happens when extremists win primaries?" *American Political Science Review* 109(1):18–42.

Hall, Andrew B and James M Snyder Jr. 2015. "Candidate ideology and electoral success." *Manuscript, Stanford University* .

Hill, Seth J. 2017. "Changing votes or changing voters? How candidates and election context swing voters and mobilize the base." *Electoral Studies* 48:131–148.

Kamada, Yuichiro and Fuhito Kojima. 2014. "Voter Preferences, Polarization, and Electoral Policies." *American Economic Journal: Microeconomics* 6(4):203–36.

Klein, Ezra. 2020. *Why We're Polarized*. Simon & Schuster, Inc.

Manacorda, Marco, Guido Tabellini and Andrea Tesei. 2026. "Mobile internet and the rise of political tribalism in Europe." *Review of Economic Studies, forthcoming* .

Meguid, Bonnie M. 2005. "Competition Between Unequals: The Role of Mainstream Party Strategy in Niche Party Success." *American Political Science Review* 99(3):347–359.

Melnikov, Nikita. 2024. "Mobile internet and political polarization." *Available at SSRN 3937760* .

Mudde, Cas and Cristóbal Rovira Kaltwasser. 2012. *Populism in Europe and the Americas: Threat or corrective for democracy?* Cambridge University Press.

Norris, Pippa. 2019. "Varieties of populist parties." *Philosophy & Social Criticism* 45(9-10):981–1012.

Norris, Pippa. 2020. "Measuring populism worldwide." *Party Politics* 26(6):697–717.

Palfrey, Thomas R. 1984. "Spatial Equilibrium with Entry." *The Review of Economic Studies* 51(1):139–156.

Persson, Torsten and Guido Tabellini. 2002. *Political economics: explaining economic policy*. MIT press.

Polborn, Mattias K and James M Snyder Jr. 2017. "Party polarization in legislatures with office-motivated candidates." *The Quarterly Journal of Economics* 132(3):1509–1550.

Prummer, Anja. 2020. "Micro-Targeting and Polarization." *Journal of Public Economics*

188:1–21.

Rooduijn, Matthijs. 2019. "State of the field: How to study populism and adjacent topics? A plea for both more and less focus." *European Journal of Political Research* 58(1):362–372.

Severin-Nielsen, Majbritt Kappelgaard, Sanne Kruikemeier, Jakob Ohme and Kristian Gade Kjelmann. 2025. "From permanent campaign to permanent communication: parliamentarians' cross-media practices in routine and election times." *Journal of Information Technology Politics* 14:1–17.

Sun, Liyang and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of econometrics* 225(2):175–199.

Vafa, Keyon, Susan Athey and David M Blei. 2025. "Estimating wage disparities using foundation models." *Proceedings of the National Academy of Sciences* 122(22):e2427298122.

Van Kessel, Patricia, Regina Widjaya, Sono Shah, Aaron Smith and Adam Hughes. 2020. "The congressional social media landscape." *Available at https://www.pewresearch.org/internet/2020/07/16/1-the-congressional-social-media-landscape/* .

Zhang, Lehan, Gabriele Gratton, Pauline Grosjean and Hasin Yousaf. 2025. "Adding Fuel to the (Gun) Fire: How Politicians Polarize the Public Debate.".

Online Appendix for

Do Elections Moderate or Polarize Political Rhetoric?

Tito Boeri* Nina Nikiforova† Guido Tabellini‡

July 27, 2026

*Department of Economics and IGIER, Bocconi University; IZA; CEP-LSE
†Department of Economics, Bocconi University
‡Department of Economics and IGIER, Bocconi University; CEPR; CES-Ifo

A Proof of Proposition 1

Proof. Candidate A chooses $(\mathbf{x}^A, \mathbf{y}^A, s^A)$ to maximize equation (6), taking B's communication as given. A fraction α of voters of each type is reached by the communication of both candidates, a fraction $1-\alpha$ of a-voters is reached only by A, and a fraction $1-\alpha$ of b-voters only by B; voters not reached by a candidate's communication behave as if it were zero. Writing $\Delta^{v,r}$ for the advantage $\Delta^v = V(\mathbf{g}^v, \mathbf{g}^A) - V(\mathbf{g}^v, \mathbf{g}^B)$ of a type-v voter with reach status $r \in \{rr, rA, rB\}$, the vote share therefore decomposes by audience as

$$\pi^A = \frac{1}{2} + \frac{\phi}{2}\Big\{\alpha \underbrace{\left[\Delta^{a,rr} + \Delta^{b,rr}\right]}_{\substack{\text{voters reached by} \\ \text{both candidates}}} + (1-\alpha) \underbrace{\Delta^{a,rA}}_{\substack{A\text{'s exclusive} \\ \text{audience}}} + (1-\alpha) \underbrace{\Delta^{b,rB}}_{\substack{B\text{'s exclusive} \\ \text{audience}}} \Big\}. \tag{A1}$$

Each audience term is an explicit function of the communication efforts. Since voter utility is quadratic, $V(\mathbf{g}^v, \mathbf{g}^P) = -\frac{1}{2}\sum_k \sigma_k^v (g_k^v - g_k^P)^2$, the difference of squares gives

$$\Delta^{v,r} = \sum_k \sigma_k^{v,r}\, \delta_k^{v,r}, \qquad \delta_k^{v,r} \equiv \frac{1}{2}\left(g_k^A - g_k^B\right)\left(2g_k^v - g_k^A - g_k^B\right), \tag{A2}$$

where all perceptions on the right are evaluated at the values held by audience r : voters reached by both candidates hold the perceptions of the communication technology in the text, while voters in an exclusive audience hold the same perceptions with the unreached candidate's communication set to zero. Explicitly, for $k = 1, 2$ and with $\sigma_2^v = 1 - \sigma_1^v$ throughout:

$$\begin{aligned}
rr\ (v = a, b): &\quad g_k^A = \hat{g}_k^A + x_k^A,\ g_k^B = \hat{g}_k^B + x_k^B,\ g_k^v = \hat{g}_k^v + \lambda^v(y_k^A + y_k^B),\ \sigma_1^v = \bar{\sigma}^v + s^A + s^B; \\
rA\ (v = a): &\quad g_k^A = \hat{g}_k^A + x_k^A,\ \ g_k^B = \hat{g}_k^B,\ \ g_k^a = \hat{g}_k^a + y_k^A,\ \ \sigma_1^a = \bar{\sigma}^a + s^A; \\
rB\ (v = b): &\quad g_k^A = \hat{g}_k^A,\ \ g_k^B = \hat{g}_k^B + x_k^B,\ \ g_k^b = \hat{g}_k^b - y_k^B,\ \ \sigma_1^b = \bar{\sigma}^b + s^B.
\end{aligned}$$

Thus candidate A's instruments enter $\Delta^{a,rr}, \Delta^{b,rr}$ (all three channels) and $\Delta^{a,rA}$ (own position x_k^A, bliss points y_k^A, and salience s^A, with B perceived at its primitive positions), while $\Delta^{b,rB}$ contains none of A's instruments and therefore drops from all of A's first-order conditions.

We first derive the five FOCs at the $\mathcal{T}$-symmetric profile that the equilibrium must satisfy; we then verify each claim of the proposition. The FOCs are obtained by differentiating (A1), using (A2), instrument by instrument.

Setup. The model is invariant under the joint transformation $\mathcal{T} : A \leftrightarrow B,\ a \leftrightarrow b,\ k \leftrightarrow k' = \{1, 2\} \setminus \{k\}$ (the dimension swap is required because $\bar{\sigma}^a + \bar{\sigma}^b = 1$ preserves the salience

pattern when voter types and dimensions are swapped together). Under $\mathcal{T}$, the perceived-position instrument flips sign (since $\hat{g}^A = -\hat{g}^B$), the bliss-point instrument does not (since $y_k^A + y_k^B$ enters symmetrically), and the salience instrument flips sign (since σ_1 maps to $1-\sigma_1$ under the dimension swap). By uniqueness, the transformed equilibrium must coincide with the original one. Hence the equilibrium satisfies

$$x_k^A = -x_{k'}^B, \quad y_k^A = y_{k'}^B, \quad s^A = -s^B. \tag{A3}$$

We write $u \equiv x_1^A$, $w \equiv x_2^A$, $\mu \equiv y_1^A$, $\nu \equiv y_2^A$, and $\tau \equiv \mu + \nu$. The salience weights at the $\mathcal{T}$-symmetric profile are: in the reached-by-both subgroup (rr), $\sigma_1^v = \bar{\sigma}^v$; in the a-only subgroup (rA), $\sigma_1^a = \frac{1}{2} + \epsilon + s^A$; in the b-only subgroup (rB), $\sigma_1^b = \frac{1}{2} - \epsilon - s^A$.

Direct computation, using $\sum_v \sigma_k^v \hat{g}_k^v = \pm 2\epsilon(\varpi + \beta)$ (positive on dim 1, negative on dim 2) and $\sum_v \sigma_k^v \lambda^v = \pm 2\epsilon$, yields the five marginal benefits:

$$\frac{\partial \pi^A}{\partial u} = \alpha\frac{\phi}{2}\left[2\epsilon(\varpi + \beta + \tau) - (\varpi + u)\right] + (1-\alpha)\frac{\phi}{2}\sigma_1^a(\beta + \mu - u), \tag{A4}$$

$$\frac{\partial \pi^A}{\partial w} = \alpha\frac{\phi}{2}\left[-2\epsilon(\varpi + \beta + \tau) - (\varpi + w)\right] + (1-\alpha)\frac{\phi}{2}(1-\sigma_1^a)(\beta + \nu - w), \tag{A5}$$

$$\frac{\partial \pi^A}{\partial \mu} = \alpha\phi\epsilon(2\varpi + u + w) + (1-\alpha)\frac{\phi}{2}\sigma_1^a(2\varpi + u), \tag{A6}$$

$$\frac{\partial \pi^A}{\partial \nu} = -\alpha\phi\epsilon(2\varpi + u + w) + (1-\alpha)\frac{\phi}{2}(1-\sigma_1^a)(2\varpi + w), \tag{A7}$$

$$\frac{\partial \pi^A}{\partial s^A} = -\alpha\phi(2\varpi + u + w)(u - w) + (1-\alpha)\frac{\phi}{2}\left[\delta_1^{a,rA} - \delta_2^{a,rA}\right], \tag{A8}$$

where $\delta_k^{a,rA} = \frac{1}{2}(2\varpi + x_k^A)(2\varpi + 2\beta + 2y_k^A - x_k^A)$, using $g_1^A - g_1^B = g_2^A - g_2^B = 2\varpi + u + w$ at the $\mathcal{T}$-symmetric profile, and the rr-term in (A8) uses $\sum_v (\delta_1^v - \delta_2^v)|_{rr} = -2(2\varpi + u + w)(u - w)$, obtained by direct expansion of $\delta_k^v = \frac{1}{2}(g_k^A - g_k^B)(2g_k^v - g_k^A - g_k^B)$ with $g_1^A + g_1^B = u - w$ and $g_2^A + g_2^B = -(u - w)$. The equilibrium satisfies $C'(z_i) = \gamma \cdot \partial\pi^A/\partial z_i$ for each $z_i \in \{u, w, \mu, \nu, s^A\}$. Since C' is odd, strictly increasing, and equals 0 at the origin, z_i has the same sign as the corresponding marginal benefit.

Parts (i) and (ii): $\epsilon = 0$. At $\epsilon = 0$ the model is additionally invariant under the dimension-swap $\mathcal{S} : k \leftrightarrow k'$ alone, which under (3) sends s^A to $-s^A$. At an equilibrium fixed by $\mathcal{S}$, therefore, $u = w$, $\mu = \nu$, and $s^A = 0$; moreover $s^B = -s^A = 0$ by (A3). Substituting $\sigma_1^a = \frac{1}{2}$ and $u = w$, $\mu = \nu$ into (A8) confirms that both the rr- and rA-terms vanish at this point, so the salience FOC is satisfied identically. The remaining FOCs collapse to a

two-variable system for (u, μ) :

$$C'(u) = \gamma\frac{\phi}{2}\left[-\alpha(\varpi + u) + \frac{1-\alpha}{2}(\beta + \mu - u)\right], \qquad C'(\mu) = \gamma\frac{\phi}{4}(1-\alpha)(2\varpi + u). \tag{A9}$$

Part (i): $\alpha = 1$. Only the rr-term survives. The bliss-point FOC gives $C'(\mu) = 0$, hence $\mu = 0$ and $y_k^P = 0$. The position FOC becomes $C'(u) = -\gamma(\phi/2)(\varpi + u)$, whose right-hand side is strictly negative at $u = 0$, zero at $u = -\varpi$, and strictly decreasing in u; together with C' strictly increasing, this gives a unique solution $u \in (-\varpi, 0)$, so $x_k^A = -x_k^B = u < 0$. Implicit differentiation of $C'(u) + \gamma(\phi/2)(\varpi + u) = 0$ yields $\partial u/\partial\gamma = -\frac{\phi}{2}(\varpi + u)/[C''(u) + \gamma\phi/2] < 0$ (using $\varpi + u > 0$), so $\partial|u|/\partial\gamma > 0$, confirming (iv) in this regime.

Part (ii): $\alpha < 1$. The bliss-point FOC $C'(\mu) = \gamma(1-\alpha)(\phi/4)(2\varpi + u)$ determines μ as a strictly increasing function $\mu = h(u)$ of u (since C' is strictly increasing, hence invertible, and the right-hand side is increasing in u), with $h(0) > 0$ because $C'(h(0)) = \gamma(1-\alpha)\phi\varpi/2 > 0 = C'(0)$. Substituting $\mu = h(u)$ into the position FOC yields a single equation in u :

$$F(u) \equiv C'(u) - \gamma\frac{\phi}{2}G_u(u, h(u)) = 0, \qquad G_u(u, \mu) \equiv -\alpha(\varpi + u) + \frac{1-\alpha}{2}(\beta + \mu - u).$$

Implicit differentiation of $C'(h(u)) = \gamma(1-\alpha)(\phi/4)(2\varpi + u)$ gives $C''(h(u))h'(u) = \gamma(1-\alpha)\phi/4$, so

$$F'(u) = C''(u) - \gamma\frac{\phi}{2}\left[\frac{\partial G_u}{\partial u} + \frac{\partial G_u}{\partial \mu}h'(u)\right] = \frac{\det J(u)}{C''(h(u))},$$

where $J(u) \equiv \operatorname{diag}\left(C''(u), C''(h(u))\right) - \gamma\frac{\phi}{2}\nabla G(u, h(u))$ is the same 2×2 Jacobian used below for the comparative statics, here evaluated along the entire curve $\mu = h(u)$ rather than only at the equilibrium. Under the maintained convexity assumption, taken — as is standard for this type of argument — to hold along this curve rather than only at the equilibrium point, $\det J(u) > 0$ everywhere, so F is strictly increasing and has at most one root; by the maintained existence of an interior equilibrium, exactly one.

To locate this root, evaluate F at $u = 0$. Since $C'(0) = 0$ and $h(0) > 0$,

$$F(0) = -\gamma\frac{\phi}{2}G_u(0, h(0)),$$

where

$$G_u(0, h(0)) = -\alpha\varpi + \frac{1-\alpha}{2}\left(\beta + h(0)\right) > -\alpha\varpi + \frac{1-\alpha}{2}\beta = \frac{1}{2}\left[(1-\alpha)\beta - 2\alpha\varpi\right],$$

For $\alpha < \bar{\alpha} \equiv \beta/(2\varpi + \beta)$, we have $G_u(0, h(0)) > 0$, so $F(0) < 0$; since F is strictly increasing,

its unique root satisfies $u^* > 0$. Since h is increasing and $u^* > 0$, $\mu^* = h(u^*) > h(0) > 0$. Hence $x_k^A = -x_k^B = u^* > 0$, $y_k^P = \mu^* > 0$; set $\hat{\alpha} = \bar{\alpha}$ at $\epsilon = 0$.

For the comparative statics of part (iv) discussed below, system (A9) defines (u, μ) implicitly as functions of γ. At $\alpha < \bar{\alpha}$ the right-hand sides G_u and $G_\mu = \frac{1-\alpha}{2}(2\varpi + u)$ are strictly positive at the equilibrium. The Jacobian J introduced above, now evaluated at the equilibrium, is a Z-matrix (both off-diagonal entries equal $-\gamma\phi(1-\alpha)/4 < 0$) with dominant diagonal under the maintained convexity assumption, hence an M-matrix with non-negative inverse; applied to the positive vector $\frac{\phi}{2}(G_u, G_\mu)$ this yields $\partial u/\partial\gamma > 0$ and $\partial\mu/\partial\gamma > 0$, confirming (iv) in this regime.

Part (iii): $\alpha < \hat{\alpha}$ and $\epsilon > 0$ sufficiently small. The $\mathcal{S}$-symmetry breaks because $\bar{\sigma}^a \neq 1/2$. We work with the full five FOCs (A4)–(A8).

Step 1: first-order expansion in ϵ. Under SOC the equilibrium is smooth in ϵ (implicit function theorem). Since the model with $-\epsilon$ is the $\mathcal{S}$-image of the model with ϵ, the differences $\Delta x \equiv u - w$ and $\Delta y \equiv \mu - \nu$ and the salience effort s^A are odd functions of ϵ, hence $O(\epsilon)$, while $u + w$ and $\mu + \nu$ are even, so $u, w = u_0 + O(\epsilon)$ and $\mu, \nu = \mu_0 + O(\epsilon)$ with (u_0, μ_0) the $\epsilon = 0$ equilibrium of part (ii). Linearizing the differences (A4)−(A5) and (A6)−(A7), and equation (A8), around (u_0, μ_0) — using $\sigma_1^a = \frac{1}{2} + \epsilon + s^A$, $\tau = 2\mu_0 + O(\epsilon^2)$, and $\delta_1^{a,rA} - \delta_2^{a,rA} = (\beta + \mu_0 - u_0)\Delta x + (2\varpi + u_0)\Delta y + o(\epsilon)$ — yields, up to $o(\epsilon)$ terms,

$$\begin{aligned}\left[C''(u_0) + \gamma\tfrac{\phi(1+\alpha)}{4}\right]\Delta x - \gamma\tfrac{(1-\alpha)\phi}{4}\Delta y - \gamma(1-\alpha)\phi(\beta + \mu_0 - u_0)\, s^A &= \\ \gamma\epsilon\phi\left[2\alpha(\varpi + \beta + 2\mu_0) + (1-\alpha)(\beta + \mu_0 - u_0)\right],\end{aligned}$$

$$\begin{aligned}-\gamma\tfrac{(1-\alpha)\phi}{4}\Delta x + C''(\mu_0)\Delta y - \gamma(1-\alpha)\phi(2\varpi + u_0)\, s^A &= \\ \gamma\epsilon\phi\left[4\alpha(\varpi + u_0) + (1-\alpha)(2\varpi + u_0)\right],\end{aligned}$$

$$-\gamma\kappa_u\Delta x - \gamma\kappa_\mu\Delta y + C''(0)\, s^A = 0,$$

where $\kappa_u \equiv \frac{(1-\alpha)\phi}{2}(\beta + \mu_0 - u_0) - 2\alpha\phi(\varpi + u_0)$ and $\kappa_\mu \equiv \frac{(1-\alpha)\phi}{2}(2\varpi + u_0) > 0$. Both right-hand sides in the first two rows are strictly positive (recall $\beta + \mu_0 - u_0 > 0$, since $C'(u_0) > 0$ in (A9)). If $\kappa_u > 0$ — which holds for α sufficiently small, since $\kappa_u \to \frac{\phi}{2}(\beta + \mu_0 - u_0) > 0$ as $\alpha \to 0$ — the coefficient matrix is a Z-matrix with dominant diagonal under the maintained convexity assumption, hence an M-matrix with non-negative inverse. Therefore $\Delta x > 0$ and

$\Delta y > 0$, both of order ϵ, and the third row gives

$$s^A = \frac{\gamma}{C''(0)} \left[\kappa_u \Delta x + \kappa_\mu \Delta y\right] + o(\epsilon) > 0. \tag{A10}$$

By (A3) these deliver the orderings $x_1^A = -x_2^B > x_2^A = -x_1^B$, $y_1^A = y_2^B > y_2^A = y_1^B$, and $s^A = -s^B > 0$.

Step 2: positivity of u, w, μ, ν. At $\epsilon = 0$ and $\alpha < \bar{\alpha} = \beta/(2\varpi + \beta)$, part (ii) establishes $u = w > 0$ and $\mu = \nu > 0$. The equilibrium depends continuously on ϵ under SOC, so each of u, w, μ, ν remains strictly positive in a neighborhood of $\epsilon = 0$.

Define

$$\hat{\alpha}(\epsilon) \equiv \sup \left\{ a \in (0, \bar{\alpha}] : u, w, \mu, \nu, s^A > 0 \text{ at every } \alpha \in (0, a) \right\},$$

the largest initial interval of α on which all five inequalities hold, so that the conclusions of part (iii) hold for every $\alpha < \hat{\alpha}(\epsilon)$ by construction. The preceding steps show that this set is nonempty, so $\hat{\alpha}(\epsilon) > 0$: κ_u is continuous in α with $\kappa_u \to \frac{\phi}{2}(\beta + \mu_0 - u_0) > 0$ as $\alpha \to 0$, so $\kappa_u > 0$ on a compact interval $[0, a_0]$ with $0 < a_0 < \bar{\alpha}$; since (u_0, μ_0) and hence the M-matrix bounds of Step 1 and the continuity bounds of Step 2 vary continuously with α, the $o(\epsilon)$ remainders are uniform in α over $[0, a_0]$, and all five inequalities hold at every $\alpha \in (0, a_0)$ once ϵ is sufficiently small. Hence $\hat{\alpha}(\epsilon) \geq a_0$. Note that the requirement $s^A > 0$ binds through the sign of $\kappa_u \Delta x + \kappa_\mu \Delta y$ in (A10), which is independent of ϵ at leading order; hence, unlike the thresholds governing $u, w, \mu, \nu > 0$, $\hat{\alpha}(\epsilon)$ need not converge to $\bar{\alpha}$ as $\epsilon \to 0$.

Part (iv): comparative statics in γ. The equilibrium is defined implicitly by the five-equation system

$$C'(z_i) = \gamma \cdot G_i(z; \alpha, \epsilon), \quad i \in \{u, w, \mu, \nu, s^A\}, \tag{A11}$$

where G_i is the corresponding right-hand side in (A4)–(A8). Under SOC the Jacobian $J \equiv \text{diag}(C'') - \gamma \nabla_z G$ is invertible at the equilibrium, and the implicit function theorem yields

$$\frac{\partial z}{\partial \gamma} = J^{-1} G(z).$$

We verify the claim regime by regime.

Regimes (i) and (ii) ($\epsilon = 0$). Since the model is $\mathcal{S}$-invariant at every γ, the equilibrium path $\gamma \mapsto z(\gamma)$ stays on the symmetric manifold $u = w$, $\mu = \nu$, $s^A = 0$, and the comparative statics reduce to those of the lower-dimensional systems analyzed above: part (i) gives $\partial |u| / \partial \gamma > 0$ by direct implicit differentiation, and part (ii) gives $\partial u / \partial \gamma, \partial \mu / \partial \gamma > 0$ from the two-variable M-matrix argument. Instruments equal to zero remain zero. This confirms (iv)

at $\epsilon = 0$.

Regime (iii) ($\epsilon > 0$ small). For the position and bliss-point instruments, $\partial z_i/\partial\gamma$ is continuous in (γ, ϵ) under SOC and strictly positive at $\epsilon = 0$ by the previous paragraph; hence, on any compact set of γ, for ϵ sufficiently small $\partial u/\partial\gamma, \partial w/\partial\gamma, \partial\mu/\partial\gamma, \partial\nu/\partial\gamma > 0$, and since these four instruments are strictly positive by part (iii), $\partial|z_i|/\partial\gamma > 0$ for each of them. For the salience instrument continuity is uninformative, because $\partial s^A/\partial\gamma \to 0$ as $\epsilon \to 0$. By (A10), $s^A = \epsilon\,\gamma\Psi(\gamma)/C''(0) + o(\epsilon)$, where $\epsilon\Psi(\gamma) \equiv \kappa_u\Delta x + \kappa_\mu\Delta y > 0$ is evaluated at the $\epsilon = 0$ equilibrium associated with γ. Hence $\partial|s^A|/\partial\gamma > 0$ for ϵ small if and only if $\gamma\Psi(\gamma)$ is strictly increasing in γ. This is a joint restriction on the cost function and the remaining parameters of the model; Lemma 1 below establishes it for quadratic costs and α sufficiently small. □

Lemma 1. *Let $C(z) = \frac{c}{2}z^2$ and $\bar\alpha \equiv \beta/(2\varpi + \beta)$. There exists $\breve\alpha \in (0, \bar\alpha]$ such that for every $\alpha \in [0, \breve\alpha)$, at every interior equilibrium satisfying the maintained conditions (positivity of the leading principal minors of the symmetric and antisymmetric Jacobians, $\kappa_u > 0$, $\beta > 0$, and $\sigma_1^a, \sigma_1^b \in (0,1)$), $\gamma\Psi(\gamma)$ is strictly increasing in γ; hence $\partial|s^A|/\partial\gamma > 0$ for ϵ sufficiently small.*

Proof. Step 1: decoupling at $\alpha = 0$. At $\alpha = 0$ only the rA and rB groups matter, so the two dimensions interact only through the salience weight $\sigma \equiv \sigma_1^a$. Given σ, dimension 1's equilibrium (u, μ) solves the linear system $u = \frac{\eta}{2}(\beta + \mu - u)$, $\mu = \frac{\eta}{2}(2 + u)$ with $\eta = \gamma\sigma$, and dimension 2 solves the same system with $\eta = \gamma(1 - \sigma)$. Explicitly,

$$u^*(\eta) = \frac{(\eta/2)(\beta + \eta)}{D(\eta)}, \qquad \mu^*(\eta) = \eta + \frac{\eta}{2}u^*(\eta), \qquad D(\eta) = 1 + \frac{\eta}{2} - \frac{\eta^2}{4},$$

valid on the per-dimension SOC region $D(\eta) > 0$, i.e. $\eta < \bar\eta \equiv 1 + \sqrt{5}$.

Step 2: the salience response as a scalar fixed point. The salience FOC becomes $s = \frac{\gamma}{2}\left[\Phi(\gamma\sigma) - \Phi(\gamma(1-\sigma))\right]$, where

$$\Phi(\eta) \equiv \frac{1}{2}\big(2 + u^*(\eta)\big)\big(2 + 2\beta + 2\mu^*(\eta) - u^*(\eta)\big)$$

is the advantage term of a dimension with weight parameter η, evaluated along the curve (u^*, μ^*). Since $\sigma = \frac{1}{2} + \epsilon + s$, linearizing at $\epsilon = 0$ gives $s = N(\gamma)(\epsilon + s) + o(\epsilon)$ with

$$N(\gamma) \equiv \gamma^2\Phi'(\gamma/2),$$

so that

$$\frac{s}{\epsilon} = \frac{N(\gamma)}{1 - N(\gamma)} + o(1),$$

and the salience second-order condition is exactly $N(\gamma) < 1$. Direct computation gives

$$\Phi'(\eta) = \frac{-4\left[\beta^2(\eta^3 + 6\eta^2 + 8) + 8\beta\eta(\eta^2 + 3\eta + 6) + 4(3\eta^3 + 12\eta^2 + 12\eta + 8)\right]}{(\eta^2 - 2\eta - 4)^3} > 0$$

on the SOC region, since there $\eta^2 - 2\eta - 4 = -4D(\eta) < 0$ and the bracket has all positive coefficients; hence $N > 0$ and $s > 0$, consistent with part (iii). As $\eta \to \bar{\eta}^-$, $\Phi'(\eta) \to +\infty$, so N reaches 1 strictly before the position-block minor D vanishes: the binding constraint defining the upper endpoint $\bar{\gamma}(0)$ of the admissible γ-interval is the salience minor $1 - N$, and the position-block minors have strict slack near $\bar{\gamma}(0)$.

Step 3: monotonicity at $\alpha = 0$. Since $x \mapsto x/(1-x)$ is strictly increasing on $x < 1$, it suffices that N is strictly increasing, i.e. that $\frac{d}{d\eta}\left[\eta^2\Phi'(\eta)\right] = \eta\left[2\Phi'(\eta) + \eta\Phi''(\eta)\right] > 0$. Direct computation gives

$$\begin{aligned} 2\Phi'(\eta) + \eta\Phi''(\eta) \;=\; & \frac{4}{(\eta^2 - 2\eta - 4)^4}\Big[\beta^2 q(\eta) + 8\beta\eta(\eta^4 + 10\eta^3 + 44\eta^2 + 48\eta + 72) \\ & \qquad + 4(3\eta^5 + 36\eta^4 + 120\eta^3 + 224\eta^2 + 128\eta + 64)\Big], \end{aligned}$$

with $q(\eta) = \eta^5 + 16\eta^4 + 32\eta^3 + 128\eta^2 - 16\eta + 64$. The denominator is a fourth power, hence positive. Every coefficient in the bracket is positive except the single term $-16\beta^2\eta$ inside $\beta^2 q(\eta)$, which is dominated because $128\eta^2 - 16\eta + 64$ has negative discriminant ($256 - 4 \cdot 128 \cdot 64 < 0$), so $q(\eta) > 0$ for all $\eta > 0$. Hence N is strictly increasing on all of $\eta > 0$, and s/ϵ is strictly increasing on the entire admissible γ-interval at $\alpha = 0$.

Step 4: extension to α small. With quadratic costs, $\gamma\Psi(\gamma;\alpha)$ is an explicit rational function of (γ, α), jointly continuous (together with its γ-derivative) wherever the binding Jacobian minor is nonzero, and the upper endpoint $\bar{\gamma}(\alpha)$ of the admissible γ-interval (the first zero of that minor) is continuous in α. We cover $(0, \bar{\gamma}(\alpha))$ by three zones and take $\breve{\alpha}$ as the minimum of the three resulting thresholds. (i) *Near $\gamma = 0$:* at $\alpha = 0$, $\gamma\Psi(\gamma)/C''(0) = \frac{1}{2}(\beta^2 + 4)\gamma^2 + O(\gamma^3)$ (using $\Phi'(0) = (\beta^2 + 4)/2$), so its γ-derivative is positive on some $(0, \delta]$; the quadratic coefficient is continuous in α (it equals $\kappa_u^0 f_x^0 + \kappa_\mu^0 f_y^0$ evaluated at the origin, which is positive at $\alpha = 0$ and continuous), and the $O(\gamma^3)$ remainder is uniform for α in a compact neighborhood of 0, so positivity of the derivative on $(0, \delta]$ persists for all α small. (ii) *Compact middle:* on $[\delta, \bar{\gamma}(\alpha) - \delta]$, Step 3 gives $d(\gamma\Psi)/d\gamma \geq m(\delta) > 0$ at $\alpha = 0$, and joint continuity extends strict positivity to all α below some threshold. (iii) *Near the boundary:* write $\gamma\Psi/C''(0) = P/T$ with T the normalized binding minor. At $\alpha = 0$,

$T = 1 - N$ with $T' = -N' < 0$ strictly (Step 3) and $P(\bar{\gamma}(0)) = N(\bar{\gamma}(0)) = 1 > 0$, so $\frac{d}{d\gamma}(P/T) = (P'T - PT')/T^2 > 0$ on a neighborhood $(\bar{\gamma} - \delta, \bar{\gamma})$; the strict inequalities $P > 0$ and $T' < 0$ at $(\bar{\gamma}(0), 0)$ persist, with a uniform neighborhood width, for all α small by continuity. Combining the three zones proves the claim for all $\alpha \in [0, \breve{\alpha})$. □

Remark. (1) The threshold $\breve{\alpha}$ is not explicit, and need not cover the entire range $\alpha < \hat{\alpha}(\epsilon)$ of part (iii); when the two ranges differ, Proposition 1(iv)'s salience clause is established only on their intersection. Some restriction on α is unavoidable: numerically, monotonicity fails in the admissible region for $\alpha \gtrsim 0.62$ (attainable, since $\alpha < \bar{\alpha}$ requires only $\beta \gtrsim 10\varpi$ there); for example, at $\alpha = 0.83$, $\beta = 33.7\varpi$ the full non-linear equilibrium exhibits s^A single-peaked in γ. An explicit uniform bound ($\alpha \leq 1/2$) can be established by a computer-assisted positivity certificate, available on request. (2) The conditions $\sigma_1^a, \sigma_1^b \in (0, 1)$ are not implied by quadratic costs, which impose no barrier as s^A approaches $\pm(1/2 - \epsilon)$. For every α and $\beta > 0$ they hold throughout the admissible γ-range once ϵ is below a strictly positive threshold $\bar{\epsilon}(\alpha, \beta)$; this threshold is largest precisely in the small-α range covered by the lemma (numerically $\bar{\epsilon} \approx 0.04$ at $\alpha = 0.1$, $\beta = 2\varpi$), so the interiority qualifier is least restrictive exactly where the lemma applies.

B Choice of fixed covariate penalty ψ

Following GST, the fixed covariate cost should be chosen so that, when the populist penalty is set to an almost negligible level for all bigrams ($\lambda_j = 10^{-10}$), the resulting partisanship estimates are similar to those obtained under MLE. As discussed in the main text, the role of λ_j in the main specification is to reduce the finite-sample bias of the estimator. Therefore, we want the fixed covariate cost ψ calibrated such that, when we do not attempt to reduce finite-sample bias, the estimates resemble the MLE results. At the same time, we prefer to keep the fixed covariate cost relatively high, so that the covariates do not absorb too much of the variation in bigram usage.

To find an optimal value, we explore different fixed cost values – $\psi \in \{10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$. Both 10^{-5} and 10^{-6} lead to a pattern of partisanship similar to the MLE (results available upon request). We thus choose 10^{-5} as a fixed cost for the rest of the paper.

C Partisanship with four political types

Suppose there are four political types, defined on two dimensions: populist / non-populist and left / right. Hence, rewrite (7) as:

$$U_{ijt}^{P_iR_i} = \alpha_{jt} + X_{it}\gamma_j + \phi_{jt}P_i + \psi_{jt}R_i$$

with $P_i = 0, 1$, $R_i = 0, 1$. We thus have four political types, Populist - Right, Non-Populist - Right, and so on, and estimate a vector of probabilities for each type, $q_{\mathbf{it}}^{\mathbf{P}_i\mathbf{R}_i} = \left\{q_{ijt}^{P_iR_i}\right\}$.

The posterior probability that i is of type, say, Populist - Right, with neutral priors and conditional on bigram j, is:

$$\hat{\rho}_{ijt}^{11} = \frac{\hat{q}_{ijt}^{11}}{\hat{q}_{ijt}^{11} + \hat{q}_{ijt}^{10} + \hat{q}_{ijt}^{01} + \hat{q}_{ijt}^{00}}$$

and similarly for the other types.

Partisanship of i in this more granular model is the weighted average of these posterior probabilities across all bigrams for all possible types, weighted by the probability of their occurrence, namely:

$$\hat{\pi}_{it4} = \frac{1}{4}\sum\nolimits_j \sum_{P=0,1;R=0,1} \hat{q}_{ijt}^{PR}\hat{\rho}_{ijt}^{PR}$$

This measures the predictability of $i's$ (granular) type by averaging his true type P_iR_i with all his possible "clones" with neutral priors. If all types always use the same bigrams, then they are indistinguishable and $\hat{\pi}_{it4} = 1/4$. If instead they always use different bigrams, then $\hat{\pi}_{it4} = 1$.

We can also estimate the predictability of P vs NP, when there are 4 political types as:

$$\hat{\pi}_{it4}^{P/NP} = \frac{1}{4}\sum\nolimits_j [(\hat{q}_{ijt}^{11} + \hat{q}_{ijt}^{10})\hat{\rho}_{ijt}^{1^\circ} + (\hat{q}_{ijt}^{01} + \hat{q}_{ijt}^{00})(1 - \hat{\rho}_{ijt}^{1^\circ}]$$

where $\hat{\rho}_{ijt}^{1^\circ} = \hat{\rho}_{ijt}^{11} + \hat{\rho}_{ijt}^{10}$ is the posterior that i is populist ($P_i = 1$). This expression for partisanship measures the average predictability of $i's$ type being P or NP, when there are four possible political types. Similarly, we can define right-left partisanship when there are four political types as:

$$\hat{\pi}_{it4}^{R/L} = \frac{1}{4}\sum\nolimits_j [(\hat{q}_{ijt}^{11} + \hat{q}_{ijt}^{01})\hat{\rho}_{ijt}^{^\circ 1} + (\hat{q}_{ijt}^{10} + \hat{q}_{ijt}^{00})(1 - \hat{\rho}_{ijt}^{^\circ 1}]$$

where now $\hat{\rho}_{ijt}^{^\circ 1} = \hat{\rho}_{ijt}^{11} + \hat{\rho}_{ijt}^{01}$.

By Bayes rule, the posterior that the speaker has features X_{it}, rather than being some other politician of the same country c, conditional on observing bigram j in quarter t, starting

with neutral prior $(1/n_c)$, is:

$$\hat{\rho}^c_{ijt} = \frac{\hat{q}_{ijt}}{\sum_{i \in c} \hat{q}_{ijt}} = \frac{\hat{q}_{ijt}}{n_c \hat{q}_{cjt}}$$

Define partisanship $\hat{\pi}^c_{it}$ as the weighted average of $\hat{\rho}^c_{ijt}$ across bigrams

$$\begin{aligned} \hat{\pi}^c_{it} &= \sum_j \hat{q}_{ijt} \hat{\rho}^c_{ijt} \\ &= \frac{1}{n_c} \sum_j \frac{\hat{q}^2_{ij}}{\hat{q}_{cjt}} \\ &= \frac{1}{n_c} [1 + \hat{\chi}^2_{ict}] \end{aligned}$$

where for the last step we used:

$$\begin{aligned} \hat{\chi}^2_{ict} &= \sum_j \frac{\hat{q}^2_{ijt} + \hat{q}_{cjt}{}^2 - 2\hat{q}_{ijt}\hat{q}_{cjt}}{\hat{q}_{cjt}} \\ &= \sum_j \frac{\hat{q}^2_{ijt}}{\hat{q}_{cjt}} + \sum_j \hat{q}_{cjt} - 2\sum_j \hat{q}_{ijt} \\ &= \sum_j \frac{\hat{q}^2_{ijt}}{\hat{q}_{cjt}} - 1 \end{aligned}$$

D Construction of confidence intervals

Given that standard nonparametric bootstrap is known to be invalid for lasso regression (Chatterjee and Lahiri, 2011), confidence intervals are estimated via subsampling, similar to the way it is done in GST. To do so, we randomly draw speakers without replacement to create 100 subsamples each containing (up to integer restrictions) one-fifth of all speakers and, for each subsample k, compute the penalized estimate of partisanship π^k_t. The confidence interval (90%) is then constructed according to the following formula around the estimated π^*_t:

$$CI(\pi^*_t) = \frac{1}{2} + [\exp(\log(\pi^*_t - 1/2) - Q^{*k}_{t(95)}/\sqrt{\tau}); \exp(\log(\pi^*_t - 1/2) - Q^{*k}_{t(5)}/\sqrt{\tau})],$$

where τ is the number of speakers in the full sample, $Q^{*k}_{t(b)}$ is the bth order statistic of:

$$Q^{*k}_t = \sqrt{\tau_k}(\log(\pi^k_t - 1/2) - \log([\frac{1}{100}\sum_{l=1}^{100} \pi^l_t] - 1/2))$$

E Speech informativeness

Following the logic of the model in Section 3, and as in GST, for each politician i and calendar quarter t we randomly draw N bigrams from a multinomial distribution with bigram choice probabilities given by the estimated frequencies, $\hat{\mathbf{q}}_{it}^{P_i} = \{\hat{q}_{ijt}^{P_i}\}$ (P_i denotes politician i's true type). Let $\hat{q}_{int}^{P_i}$ denote the frequency of use of the n-th drawn bigram for politician i in calendar quarter t. Then, for any sequence length $n \in \{1, 2, \ldots, N\}$, we compute the following measure of speech informativeness conditional on the observed sequence:

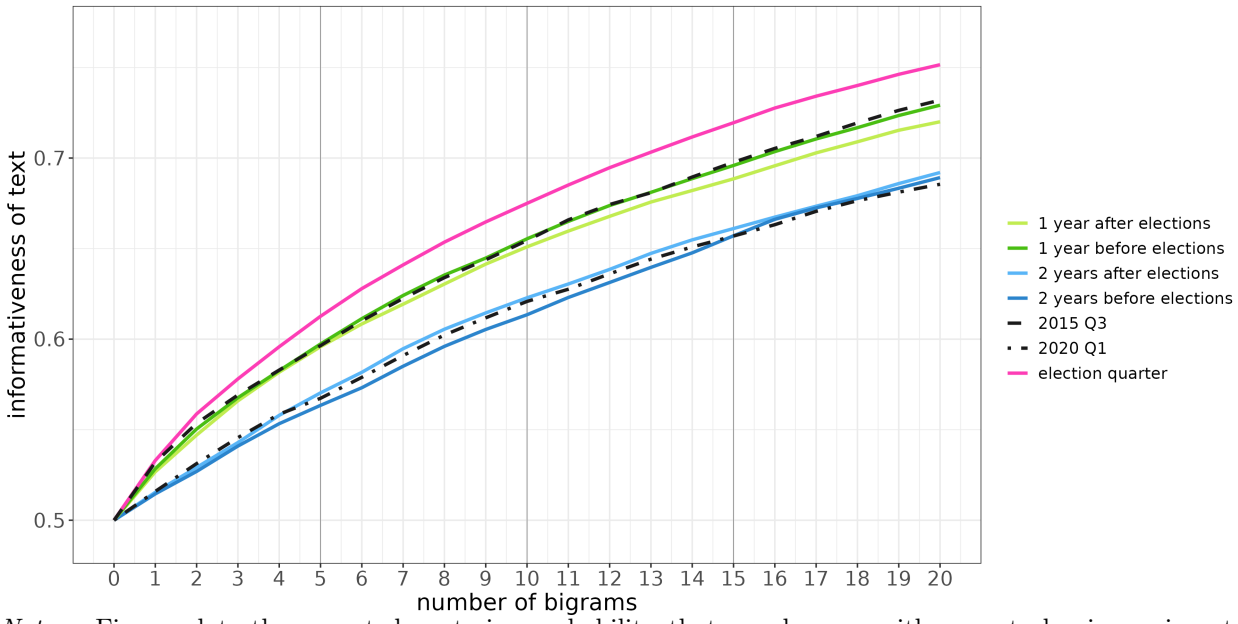
$$\eta_{itn} = \frac{\hat{q}_{int}^{P_i}\eta_{itn-1}}{\hat{q}_{int}^{P_i}\eta_{itn-1} + \hat{q}_{int}^{1-P_i}(1-\eta_{itn-1})}, \tag{E1}$$

where $\eta_{it0} = \frac{1}{2}$ for all i and calendar quarters t (i.e. as before, an observer starts with a neutral prior).[1] Averaging η_{itn} over politicians i in a given quarter t we obtain a measure of speech informativeness η_{tn}. To obtain a similar measure at a given quarter distance d from the election, we take all the politicians-quarter observations at a given quarter distance d, and average $\eta_{id(i,t)n}$ over i to obtain η_{dn}.

Figure E1 presents the average results of repeating this procedure 10 times through a Monte Carlo simulation *for each politician* in each quarter. The vertical axis measures speech informativeness about the politician being populist or non-populist, the horizontal axis measures the number of bigrams. One tweet corresponds to about 5 bigrams (the average length of a bigram is about 20 characters and the average length of a tweet is about 100 characters, although tweets also include many rarely used bigrams that we have excluded from the analysis). Different colors correspond to different distances from the election date. The figure yields the estimates described in the text.

[1] This measure corresponds to the expected posterior probability of correctly identifying politician i's type after observing n bigrams, starting with a neutral prior.

Figure E1: Expected posterior belief of an observer with a neutral prior after reading a given number of bigrams



Notes. Figure plots the expected posterior probability that an observer with a neutral prior assigns to politician i's true populist affiliation after hearing n bigrams (η_{itn}), averaged over i. η_{itn} is defined as in Equation E1. Estimated bigram-usage probabilities $\hat{\mathbf{q}}_{it}$ from our benchmark specification are used. Vertical lines indicate approximately one, two, and three tweets (five, ten, and fifteen bigrams, respectively).

F Data

F.1 Classification of Populist Leaders

F.1.1 Switching political types

We directly asked ChatGPT whether the politician changed between being a populist and a non-populist using the prompt below, and repeated the process 4 times. Over the 367 politicians in our sample, the answer is "Unchanged" for 363 of them with full consistency (4/4). We just have 3 cases where ChatGPT answered 2 times over 4 "Changed from NP to P" or "from P to NP". These politicians are Cabo Daciolo from Brazil, Mark Latham from Australia (both from NP to P), and Juliette Boulet from Belgium (from P to NP). For Paweł Piotr Kukiz 3 out of 4 times ChatGPT gave an answer "Changed from NP to P".

The ChatGPT prompt used to assess whether politicians switched from one political type to another was as follows:

*I will give you the name of a present or past leader of a political party and the country they are from. I am interested in whether this person **changed** between being a populist*

and a non-populist, or vice-versa, during the period 2010–2022.

My definition of populism is the following: A leader is defined as populist if he or she divides society into two artificial groups – "the people" vs. "the elites" – and then claims to be the sole representative of the true people. Populists place the alleged struggle of the people ("us") against the elites ("them") at the center of their political campaign and governing style. More precisely, populists typically depict "the people" as a suffering, inherently good, virtuous, authentic, ordinary, and common majority, whose collective will is incarnated in the populist leader. By contrast, "the elite" is an inherently corrupt, self-serving, power-hoarding minority, negatively defined as all those who are not "the people".

*Determine whether the leader **changed** between being a populist and a non-populist, or vice-versa, **between 2010 and 2022** and give me one of these four possible answers:*

*1. **Changed from P to NP**: if the answer is yes and the politician changed from being a Populist to a Non-Populist.*

*2. **Changed from NP to P**: if the answer is yes and the politician changed from being a Non-Populist to a Populist.*

*3. **Unchanged**: if the answer is no, meaning the politician was either always a Populist or always a Non-Populist.*

*4. **Ambiguous**: if the answer is ambiguous because either you have lack of sufficient information or it is truly an ambiguous case.*

Rules: - Output only a single label. - Be concise and accurate.

F.1.2 Comparison of our classification with fully automated ChatGPT classification

We ran an exercise similar to the one above to validate our manual classification of populist politicians. To do so, we asked ChatGPT to classify politicians as either populists, non-populists or ambiguous. The results show a total of 26 cases of disagreement between the fully automated ChatGPT classification and the benchmark classification used in the main text (the classification in which the remaining ambiguous cases are considered populists). In all 26 cases, our benchmark classification identifies the politicians as populists, whereas the fully automated ChatGPT classification labels them as non-populists. Of these 26 politicians, 10 are classified as populists in our data because they belong to a populist party, following our approach for ambiguous cases to reduce the overall number of ambiguities.

Among the remaining 16 politicians, 7 are classified as ambiguous and do not belong to a populist party; therefore, following the approach described in the main text, we treat them as populists in the benchmark specification. The other 9 politicians are considered populists in our classification but are labeled as non-populists by the automated ChatGPT classification. These politicians are Beata Maria Kuśińska, Catarina Martins, Elizabeth Warren, Jeremy Corbyn, Judit Varga, Kristian Thulesen Dahl, Pia Kjærsgaard, Roberto Maroni, and Zoltán Kovács.

F.2 Fixation Index

To validate the classification of P vs NP leaders we measured its performance vis-a-vis alternative classifications. For each politician, Claude 4.6 Sonnet was asked to report a Populist Score by answering the following question: *"I have a sample of politicians and I want you to answer how certain you are that the politician is populist. 0% means definitely non-populist, 100% means definitely populist."* Once this task was completed, partitions of the sample into two groups were created. First, we reconstructed the benchmark classification used in the paper, which has 84 Populists and 283 Non-Populists. A second partition was created by taking the 84 politicians with the highest Populist Score as estimated by Claude as one group. We then computed 150 more random partitions, maintaining the group sizes of 84 and 283.

For each partition m, a χ^2-based measure of the Fixation Index was computed at the quarter level and then averaged across all quarters. Probabilities were estimated with the standard specification, including all covariates. The group dummy was not penalized but included among the covariates and interacted with calendar quarter.

Consider the usual notation, where $q_{i,j,t}$ denotes the usage probability of bigram j by individual i in calendar quarter t. A partition m splits the sample into two groups $g \in \{P, NP\}$, and we let $g_m(i)$ denote the group to which individual i is assigned under partition m. Define the overall mean for each bigram,

$$q_{j,t} = \frac{1}{N_t} \sum_i q_{i,j,t},$$

and the group-specific means for each bigram within each quarter,

$$q_{g,j,t} = \frac{1}{N_{g,t}} \sum_{i\,:\,g_m(i)=g} q_{i,j,t},$$

where N_t is the number of individuals in quarter t and $N_{g,t}$ is the number of individuals

in group g in quarter t. We measure total and between-group heterogeneity in quarter t respectively by:

$$\chi_t^2 = \frac{1}{N_t} \sum_i \sum_j \frac{(q_{i,j,t} - q_{j,t})^2}{q_{j,t}}.$$

$$\chi_{m,t}^2 = \frac{1}{N_t} \sum_{g \in \{P,NP\}} N_{g,t} \sum_j \frac{(q_{g,j,t} - q_{j,t})^2}{q_{j,t}},$$

The Fixation Index was then computed as

$$F_{m,t} = \frac{\chi_{m,t}^2}{\chi_t^2},$$

and the average for partition m is

$$\hat{F}_m = \frac{1}{T} \sum_{t=1}^{T} F_{m,t}.$$

This measure has a clear interpretation: the share of variation in predicted bigram usage due to between-group variation, for the groups defined by partition m.

Figure F1 plots the distribution of $\hat{F}_m$ over the generated partitions of politicians. The left-most vertical line corresponds to the Claude partition, the right-most vertical line to our benchmark classification. Thus, our classification uncovers the latent language patterns in the texts very well, and it outperforms the partition based on the Claude classification.

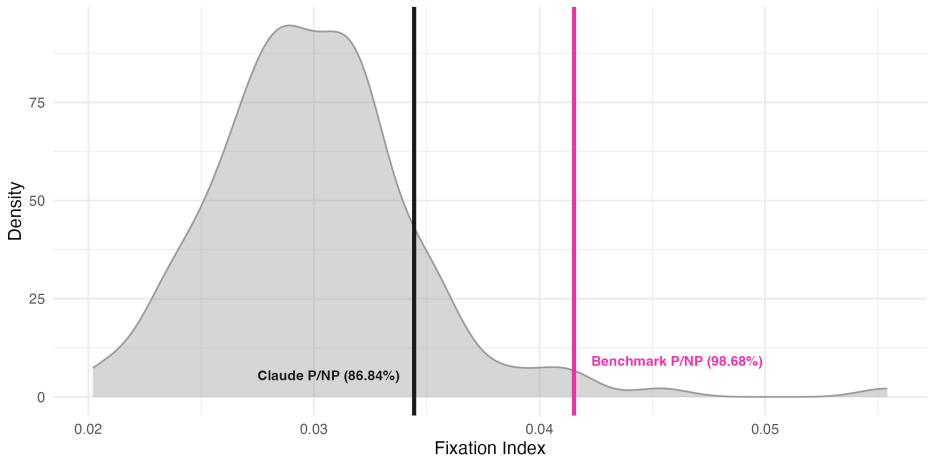


Figure F1: Fixation index distribution over generated partitions of politicians.

Notes. *Benchmark P/NP* refers to the classification used as the benchmark in the paper. *Claude P/NP* refers to a classification that partitions the sample into populist and non-populist groups of the same size as in the benchmark, based on populist scores generated by Claude Sonnet.

F.3 Classification of Niche Parties

To classify niche parties, the following prompt was used:
You are a helpful research assistant. I will give you the name of a past or present political party and what country they are from. I am interested in whether they are a 'niche' political party. In particular, a party is niche if it satisfies all three of the following criteria:

1) Niche parties reject the traditional class-based orientation of politics. So instead of prioritizing demands for more or less redistribution, these parties politicize other issues which were previously not represented in party competition. In many cases, green parties or radical right parties would be examples of this.

2) The issues raised by niche parties are not only novel, but they often do not coincide with existing lines of political division. Niche parties appeal to groups of voters that may cross-cut traditional partisan alignments. As a result, cases of voter defection between "unlikely" party pairs have occurred. The defection of former British Conservative voters to the Green Party in 1989 and former French Communist party voters to the radical right Front National in 1986 are typical examples.

3) Niche parties keep a very narrow policy agenda, focusing only on a few key policies. So for example the Green Party in Germany is not a niche party because it is a mainstream party with a very broad policy agenda.

Respond with 'yes' if the party is niche and 'no' if it is not.

F.4 International connections of Leaders

We explore cross-border connections of political leaders along the P/NP and L/R divides using data on direct followers for 305 (out of 367) leaders across the 21 countries covered by our study. Our goal is to evaluate differential rates of connection between groups, and to do so we adopt a dyadic regression approach similar to Colonnelli, Neto and Teso (2025). This approach allows us to account for group imbalances, as well as to control for confounding factors.

We transform our directed network into a dataset of cross-border dyads: ordered pairs (i, j) in which politician i (the sender) may follow politician j (the receiver) located in a different country. Each dyad is thus a potential tie, realised when i actually follows j; the unit of observation is the dyad. The full dataset of cross-border dyads comprises 88,102 potential ties, of which 2,523 are realized ties. We define homophily as the tendency for cross-border ties to occur disproportionately between politicians of the same type. Using a linear probability model, we compare the probability that a same-type dyad forms a tie with the corresponding probability for a different-type dyad. We then test whether this

same-type selectivity is stronger among populist senders than among non-populist senders, and we repeat the same analysis for the Left/Right divide.

We estimate the following linear probability model over all ordered cross-border dyads:

$$\text{Tie}_{ij} = \alpha + \beta_1 \,\text{SameType}_{ij} + \beta_2 \,\text{Pop}_i + \beta_3 \left(\text{SameType}_{ij} \times \text{Pop}_i\right) + \mathbf{X}'_{ij}\boldsymbol{\gamma} + \mu_{c(i)} + \nu_{c(j)} + \varepsilon_{ij}, \quad \text{(F1)}$$

where i is the potential sender (follower), j the potential receiver (followed). The dependent variable $\text{Tie}_{ij} = \mathbf{1}\{i \text{ follows } j\}$ equals one when the tie is realized and zero otherwise. The regressor $\text{SameType}_{ij} = \mathbf{1}\{\text{Pop}_i = \text{Pop}_j\}$ indicates that sender and receiver share the same populism status, and $\text{Pop}_i = \mathbf{1}\{i \text{ is populist}\}$ indicates a populist sender. The vector $\mathbf{X}'_{ij}$ contains dyad-level indicators for shared gender and shared left–right classification, together with sender and receiver gender, birth year, right-wing status, and indicators for having held a position as head of government or head of state (as proxy for leader popularity). The $\mu_{c(i)}$ and $\nu_{c(j)}$ terms are sender- and receiver-country fixed effects. They absorb differences across sender and receiver countries in the probability of forming a cross-border tie. In the sender-fixed-effects specification, $\mu_{c(i)}$ and all sender-level covariates are absorbed by an individual sender effect ϕ_i.[2] Throughout, standard errors are clustered two-way on sender and receiver identity to account for the dependence among all dyads that share a sender and all dyads that share a receiver. We repeat the same analysis for the Left-Right divide.

The coefficient β_1 measures the increase in tie probability associated to same type $NP \to NP$ match, compared to a $NP \to P$ match (non-populist homophily); β_2 is the difference in tie probability for a $P \to NP$ match compared to a $NP \to P$ match. Finally, β_3 measures the differential homophily of populist compared to non-populist leaders, which estimates the difference in tie probability associated to a $P \to P$ match (wrt a $P \to NP$ match), compared to a $NP \to NP$ match (wrt an $NP \to P$ match).

Table F1 shows the results of our cross-border estimates. In the unconditional specification, the coefficient β_1 is positive and significant, while β_3 is positive, but not statistically significant. β_3 becomes larger and statistically significant controlling for sender-country and receiver-country fixed effects, as well as dummies for Head of Government or Head of State positions. In these specifications populists appear to be more homophilic in their cross-border connections compared to non-populists, and the increase in tie probability associated with a same-type match is around 2.3 percentage points larger among populist senders than among non-populist senders.

We repeat the same analysis along the Left/Right divide in Table F2. We do not find a differential right-wing (vs left-wing) homophily in cross-border ties. The coefficient β_3 is statistically significant only when we exclude from the sample politicians leaning to a central position in the political spectrum (having a RILE index in the (-5,5) interval, Column 4).

[2] We do not include receiver fixed effects, as they would absorb the receiver's populism status, leaving β_1 and β_3 no longer separately identified.

Table F1: Differential populist homophily in cross-border political ties

	(1)	(2)	(3)
$(NP \to NP) - (NP \to P)$ $[\beta_1]$	0.015**	0.006	0.006
	(0.007)	(0.005)	(0.005)
Differential homophily $[\beta_3]$	0.006	0.023**	0.023**
	(0.012)	(0.010)	(0.010)
Sender-country FE	No	Yes	Absorbed
Receiver-country FE	No	Yes	Yes
Sender individual FE	No	No	Yes
Individual-level controls	No	Yes	Yes
HoG/S controls	No	Yes	Yes
Observations	88,102	88,102	88,102
Adjusted R^2	0.003	0.061	0.097
Sender/receiver clusters	305/305	305/305	305/305

Notes: The sample contains all ordered cross-border dyads. The dependent variable equals one if sender i follows receiver j. Differential homophily is $\beta_3 = \big[\Pr(P \to P) - \Pr(P \to NP)\big] - \big[\Pr(NP \to NP) - \Pr(NP \to P)\big]$. Individual-level controls include shared gender, shared right-wing status, and sender and receiver gender, birth year, and right-wing status. HoG/S controls indicate whether the sender and receiver have held a position as head of government or head of state. Sender-level covariates are absorbed by sender fixed effects in column (3). Standard errors, two-way clustered by sender and receiver, are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table F2: Differential right-wing homophily in cross-border political ties

	(1)	(2)	(3)	(4)
$(L \to L) - (L \to R)$ $[\beta_1]$	0.011*	0.017**	0.017**	0.026***
	(0.006)	(0.007)	(0.007)	(0.009)
Differential homophily $[\beta_3]$	−0.002	−0.016	−0.017	−0.031*
	(0.012)	(0.013)	(0.013)	(0.017)
Sender-country FE	No	Yes	Absorbed	Absorbed
Receiver-country FE	No	Yes	Yes	Yes
Sender individual FE	No	No	Yes	Yes
Individual-level controls	No	Yes	Yes	Yes
HoG/S controls	No	Yes	Yes	Yes
Drops $\lvert\text{RILE}\rvert < 5$	No	No	No	Yes
Observations	88,102	88,102	88,102	59,222
Adjusted R^2	0.001	0.061	0.097	0.091
Sender/receiver clusters	305/305	305/305	305/305	250/250

Notes: The sample contains all ordered cross-border dyads. The dependent variable equals one if sender i follows receiver j. A politician is classified as right-wing (R) if the RILE (right–left) score of their party's manifesto is positive, and non-right (L) otherwise. Differential homophily is $\beta_3 = \big[\Pr(R \to R) - \Pr(R \to L)\big] - \big[\Pr(L \to L) - \Pr(L \to R)\big]$. Individual-level controls include shared gender, shared populism status, and sender and receiver gender, birth year, and populism status. HoG/S controls indicate whether the sender and receiver have held a position as head of government or head of state. Sender-level covariates are absorbed by sender fixed effects in columns (3) and (4). Column (4) repeats the specification of column (3) on the subsample that excludes politicians whose RILE score lies in $(-5, 5)$. Standard errors, two-way clustered by sender and receiver, are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Differences in homophily over international connections appear to be stronger along the P/NP divide than along the L/R divide when we compare the two differential effects within the same regression (see Table F3). In particular we run the following regression:

$$\text{Tie}_{ij} = \alpha + \delta_1 \,\text{SameType}_{ij} + \delta_2 \,\text{SameRight}_{ij} + \mathbf{X}'_{ij}\boldsymbol{\gamma} + \mu_{c(i)} + \nu_{c(j)} + \varepsilon_{ij}, \tag{F2}$$

where $\text{SameType}_{ij} = \mathbf{1}\{\text{Pop}_i = \text{Pop}_j\}$ and $\text{SameRight}_{ij} = \mathbf{1}\{\text{Right}_i = \text{Right}_j\}$ indicates that the sender and receiver have the same political ideology (P or NP and L or R respectively). The vector $\mathbf{X}'_{ij}$ now contains main effects for receiver and sender right-wing and populist status, along

with dyad-level indicators for shared gender, and receiver and sender gender and year of birth. All other terms are as in Equation F1.

The results are reported in Table F3. The benchmark specification (4) shows that sharing populist status increases the probability of a cross-border tie by 1.8 percentage points, while sharing left/right-wing alignment increases it by 0.9 percentage points. The difference between these two probabilities is around 1 percentage point and statistically significant at the 5% level. This difference becomes smaller and not statistically significant when politicians close to the ideological center are excluded from the sample, specification (5).

Table F3: P/NP versus L/R homophily in cross-border political ties

	(1)	(2)	(3)	(4)	(5)
P/NP homophily [δ_1]	0.017***	0.017***	0.018***	0.018***	0.016***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
L/R homophily [δ_2]	0.009***	0.009***	0.009***	0.009***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Difference ($\delta_1 - \delta_2$)	0.009**	0.009**	0.009**	0.009**	0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Sender-country FE	No	Yes	Yes	Absorbed	Absorbed
Receiver-country FE	No	Yes	Yes	Yes	Yes
Sender individual FE	No	No	No	Yes	Yes
Demographic controls	No	Yes	Yes	Yes	Yes
HoG/S controls	No	No	Yes	Yes	Yes
Drops |RILE| < 5	No	No	No	No	Yes
Observations	88,102	88,102	88,102	88,102	59,222
Adjusted R^2	0.004	0.040	0.061	0.097	0.091
Sender/receiver clusters	305/305	305/305	305/305	305/305	250/250

Notes: The sample contains all ordered cross-border dyads. The dependent variable equals one if sender i follows receiver j. Row [δ_1] reports the coefficient on an indicator for sender and receiver sharing the same populism status; row [δ_2] the coefficient on an indicator for sharing the same left–right status, where a politician is classified as right-wing if the RILE score of their party's manifesto is positive. Both indicators enter the same regression, each alongside its own sender and receiver main effects (included in all columns). The reported difference tests whether populist homophily exceeds left–right homophily. Demographic controls include shared gender and sender and receiver gender and birth year. HoG/S controls indicate whether the sender and receiver have held a position as head of government or head of state. Sender-level covariates are absorbed by sender fixed effects in columns (4) and (5). Column (5) excludes politicians whose RILE score lies in $(-5, 5)$. Standard errors, two-way clustered by sender and receiver, are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

F.5 Classification of Leaders along Left-Right

F.5.1 ChatGPT Classification

Classification of politicians with ChatGPT was performed using the model GPT-4o-mini and with the following prompt:

I will give you the name of a present or past leader of a political party and what country they are from. I am interested in whether they are a left-wing or right-wing politician in the period 2010-2021. In particular, give me one of these three possible answers:

1. left: commonly considered a left-wing leader 2. right: commonly considered a right-wing leader 3. ambiguous: ambiguous whether they are left- or right-wing because either: (i) they are truly an ambiguous case (ii) they change from being left- to right-wing (or vice-versa) during the period (iii) you do not have enough information about the individual to give an answer

Rules: - Output only a single label. - If the label is ambiguous, specify whether it is because of (i), (ii) or (iii). - Be concise and accurate.

ChatGPT classification exhibits some randomness, therefore results slightly change when the same procedure is repeated. Reassuringly, in the 4 iterations there was no case in which ChatGPT switched a politician's classification from right to left or vice versa; the classification only switched between right and ambiguous, or between left and ambiguous.

To account for this, the classification has been repeated 4 times, and politicians have been classified as "Right" if ChatGPT classified the politician in this category all 4 times, the same with "Left". Otherwise, politicians have been assigned to the "ambiguous" category. This process led to 82 ambiguous cases (later 64 of these politicians were classified as left by the RILE classification, 18 as right). The number of politicians which were positively classified is 285: the confusion matrix for this sample against the RILE classification is reported below – Table F4.

F.5.2 RILE Classification

We also classify all the politicians using the party-level classification. The Manifesto Project Dataset measures left-right positions of political parties through the 'rile' index, which measures how much a party talks about left-wing or right-wing issues. We consider, for each party, the most recent codification of this index and classify politicians as right if they belong to a party whose RILE score is greater than zero, as left otherwise. Politicians that have no match with the dataset have been manually classified (with the use of ChatGPT).

We prefer the ChatGPT classification to the one based on the Manifesto Project Dataset, as we believe that the former could better capture within-party factions which are relevant in classifying individual leaders, and it is more comparable internationally. In Appendix Figure H18 we present our main results for the L-R classification based on the RILE index instead of ChatGPT.

Table F4: Left-Right Classifications Confusion Matrix - Absolute Values

	GPT Left	GPT Right	
RILE Left	147	46	193
RILE Right	12	80	92
	159	126	285

F.6 Summary of leaders classifications

Australia

Andrew Bartlett (NP, L), Bill Shorten (NP, L), Bob Katter (P, R), Campbell Newman (NP, R), Christine Milne (NP, L), Clive Palmer (P, R), Deb Frecklington (NP, R), John Anderson (NP, L), John-Paul Langbroek (NP, R), Julia Gillard (NP, L), Kevin Rudd (NP, L), Kim Beazley (NP, L), Malcolm Turnbull (NP, R), Mark Latham (P, R), Natasha Stott Despoja (NP, L), Nick Xenophon (P, L), Pauline Hanson (P, R), Richard Di Natale (NP, L), Scott Morrison (NP, R), Tim Nicholls (NP, R), Tony Abbott (P, R)

Austria

Alexander Van der Bellen (NP, L), Beate Meinl-Reisinger (NP, L), Christian Kern (NP, L), Heinz-Christian Strache (P, R), Josef Bucher (P, R), Matthias Strolz (NP, L), Norbert Hofer (P, R), Pamela Rendi-Wagner (NP, L), Peter Pilz (NP, L), Sebastian Kurz (P, R), Ulrike Lunacek (NP, L), Werner Kogler (NP, L)

Belgium

Alexander De Croo (NP, R), Bart De Wever (P, R), Benoît Lutgen (NP, R), Caroline Gennez (NP, L), Charles Michel (NP, R), Didier Reynders (NP, L), Elio Di Rupo (NP, L), Guy Verhofstadt (NP, L), Gwendolyn Rutten (NP, R), Herman Van Rompuy (NP, L), Jean-Marc Nollet (NP, L), John Crombez (NP, L), Joëlle Milquet (NP, R), Juliette Boulet (NP, L), Marianne Thyssen (NP, R), Meyrem Almaci (NP, L), Peter Mertens (P, L), Philippe Henry (NP, L), Philippe Lamberts (NP, L), Stefaan De Clerck (NP, R), Tom Van Grieken (P, R), Wouter Beke (NP, R), Wouter Van Besien (NP, L), Yves Leterme (NP, L), Zoé Genot (NP, L)

Brazil

Anthony Garotinho (P, L), Aécio Neves (NP, R), Cabo Daciolo (P, R), Ciro Gomes (P, L), Cristovam Buarque (NP, L), Dilma Rousseff (P, L), Fernando Haddad (NP, L), Geraldo Alckmin (NP, R), Heloísa Helena (P, L), Henrique Meirelles (NP, R), Jair Bolsonaro (P, R), José Serra (NP, R), João Amoêdo (NP, R), Luciana Genro (NP, L), Luiz Inácio Lula da Silva (P, L), Marina Silva (NP, L), Rui Costa Pimenta (NP, L)

Czech Republic

Andrej Babiš (P, L), Ivan Bartoš (NP, L), Jan Farský (NP, L), Jiří Paroubek (NP, L), Jiří Rusnok (NP, R), Karel Schwarzenberg (NP, L), Lubomír Zaorálek (NP, L), Mirek Topolánek (NP, R), Miroslav Kalousek (NP, R), Miroslava Němcová (NP, R), Pavel Bělobrádek (NP, R), Petr Fiala (NP, R), Tomio Okamura (P, R), Vladimír Špidla (NP, L), Vojtěch Filip (NP, L)

Denmark

Anders Fogh Rasmussen (NP, R), Anders Samuelsen (NP, R), Bendt Bendtsen (NP, R), Helle Thorning-Schmidt (NP, L), Holger K. Nielsen (NP, L), Kristian Thulesen Dahl (P, R), Lars Barfoed (NP, R), Lars Løkke Rasmussen (NP, L), Margrethe Vestager (NP, L), Marianne Jelved (NP, L), Mogens Lykketoft (NP, L), Morten Østergaard (NP, L), Pernille Skipper (P, L), Pernille Vermund (P, R), Pia Kjærsgaard (P, R), Pia Olsen Dyhr (NP, L), Søren Pape Poulsen (NP, R), Uffe Elbæk (NP, L)

Finland

Alexander Stubb (NP, R), Anna-Maja Henriksson (NP, L), Anneli Jäätteenmäki (NP, L), Anni Sinnemäki (NP, L), Antti Rinne (NP, L), Carl Haglund (NP, L), Eero Heinäluoma (NP, L), Harry Harkimo (NP, L), Jussi Halla-aho (P, R), Jutta Urpilainen (NP, L), Jyrki Katainen (NP, R), Li Andersson (NP, L), Mari Kiviniemi (NP, R), Matti Vanhanen (NP, R), Osmo Soininvaara (NP, L), Paavo Arhinmäki (NP, L), Pekka Haavisto (NP, L), Petteri Orpo (NP, R), Päivi Räsänen (NP, R), Sari Essayah (NP, R), Suvi-Anne Siimes (NP, L), Tarja Cronberg (NP, L), Ville Niinistö (NP, L)

France

Benoît Hamon (NP, L), Bernard Cazeneuve (NP, L), Dominique Voynet (NP, L), Dominique de Villepin (NP, R), Edouard Philippe (NP, R), Emmanuel Macron (NP, L), Eva Joly (NP, L), François Bayrou (NP, L), François Fillon (NP, R), François Hollande (NP, L), Jean-Luc Mélenchon (P, L), Jean-Marc Ayrault (NP, L), Jean-Marie Le Pen (P, R), Jean-Pierre Raffarin

(NP, R), Manuel Valls (NP, L), Marine Le Pen (P, R), Nicolas Dupont-Aignan (P, R), Nicolas Sarkozy (NP, R), Noël Mamère (NP, L), Ségolène Royal (NP, L)

Germany

Alice Weidel (P, R), Bernd Lucke (P, R), Cem Özdemir (NP, L), Christian Lindner (NP, R), Dietmar Bartsch (NP, L), Gabi Zimmer (NP, L), Gregor Gysi (NP, L), Jürgen Trittin (NP, L), Katrin Göring-Eckardt (NP, L), Martin Schulz (NP, L), Peer Steinbrück (NP, L), Renate Künast (NP, L), Sahra Wagenknecht (P, L)

Hungary

Bocskor Andrea (NP, R), Deli Andor (NP, L), Deutsch Tamás (P, R), Edina Toth (P, R), Gergely Karácsony (NP, L), Gordon Bajnai (NP, L), István Ujhelyi (NP, L), Judit Varga (P, R), Katalin Novák (P, R), László Andor (NP, L), Márton Gyöngyösi (P, R), Péter Niedermüller (NP, L), Tibor Navracsic (NP, R), Zoltan Kovacs (P, R)

Italy

Alessandro Di Battista (P, L), Beppe Grillo (P, L), Carlo Calenda (NP, L), Daniela Santanché (P, R), Emma Bonino (NP, L), Enrico Letta (NP, L), Francesco Rutelli (NP, L), Giorgia Meloni (P, R), Giuseppe Conte (P, L), Luigi Di Maio (P, L), Mario Monti (NP, R), Matteo Renzi (NP, L), Matteo Salvini (P, R), Paolo Gentiloni (NP, L), Pier Ferdinando Casini (NP, R), Pier Luigi Bersani (NP, L), Pietro Grasso (NP, L), Roberto Maroni (P, R), Silvio Berlusconi (P, R), Walter Veltroni (NP, L)

Mexico

Alberto Anaya Gutiérrez (NP, L), Alejandro Moreno (NP, R), Alfonso Ramírez Cuéllar (P, L), Andrés Manuel López Obrador (P, L), Beatriz Mojica Morga (NP, L), Carlos Puente Salas (NP, L), Carolina Viggiano (NP, R), Clemente Castañeda Hoeflich (NP, L), Dante Delgado (NP, R), Enrique Peña Nieto (NP, R), Felipe Calderón (NP, R), Gabriel Quadri de la Torre (NP, R), Hugo Eric Flores Cervantes (P, L), Héctor Larios (NP, R), Jaime Rodríguez Calderón (P, L), Josefina Vázquez Mota (NP, R), José Antonio Meade (NP, L), Luis Castro Obregón (NP, L), Manuel Granados Covarrubias (NP, L), Marko Cortés (NP, R), Martí Batres (P, L), Patricia Mercado Castro (NP, L), Reginaldo Sandoval Flores (P, L), Ricardo Anaya (NP, R), Roberto Campa Cifrián (NP, L), Yeidckol Polevnsky (P, L)

Netherlands

Alexander Pechtold (NP, L), André Rouvoet (NP, L), Arie Slob (NP, R), Diederik Samsom (NP, L), Emile Roemer (P, L), Geert Wilders (P, R), Gert-Jan Segers (NP, R), Henk Krol (NP, L), Jan Peter Balkenende (NP, R), Jesse Klaver (NP, L), Job Cohen (NP, L), Kees van der Staaij (NP, R), Lodewijk Asscher (NP, L), Marianne Thieme (NP, L), Mark Rutte (NP, R), Paul Rosenmöller (NP, L), Sybrand van Haersma Buma (NP, R), Thierry Baudet (P, R), Thom de Graaf (NP, L), Tunahan Kuzu (P, L)

Norway

Audun Lysbakken (NP, L), Bjørnar Moxnes (NP, L), Dagfinn Høybråten (NP, R), Erna Solberg (NP, R), Hanna E. Marcussen (NP, L), Jens Stoltenberg (NP, L), Jonas Gahr Støre (NP, L), Knut Arild Hareide (NP, L), Kristin Halvorsen (NP, L), Liv Signe Navarsete (P, L), Rasmus Hansson (NP, L), Thorbjørn Jagland (NP, L), Trine Skei Grande (NP, R), Une Aina Bastholm (NP, L), Åslaug Haga (P, L)

Poland

Barbara Nowacka (NP, L), Beata Maria Kusińska (P, R), Donald Tusk (NP, L), Ewa Kopacz (NP, L), Grzegorz Napieralski (NP, L), Janusz Korwin-Mikke (P, R), Janusz Palikot (P, R), Janusz Piechociński (NP, L), Leszek Miller (NP, L), Mateusz Morawiecki (P, R), Małgorzata Maria Kidawa-Błońska (NP, L), Paweł Piotr Kukiz (P, L), Roman Giertych (NP, R), Ryszard Petru (NP, R), Wojciech Olejniczak (NP, L), Władysław Marcin Kosiniak-Kamysz (NP, L), Włodzimierz Czarzasty (NP, L)

Portugal

André Ventura (P, R), António Costa (NP, L), Assunção Cristas (NP, R), Carlos Guimarães Pinto (NP, R), Catarina Martins (P, L), José Manuel Barroso (NP, L), José Sócrates (NP, L), Mário Centeno (NP, L), Pedro Passos Coelho (NP, R), Rui Rio (NP, R)

Slovenia

Alenka Bratušek (NP, L), Borut Pahor (NP, L), Dejan Židan (NP, L), Franc Bogovič (NP, R), Gregor Golobič (NP, L), Gregor Virant (NP, L), Janez Janša (P, R), Janez Podobnik (NP, R), Karl Erjavec (NP, L), Katarina Kresal (NP, L), Ljudmila Novak (NP, R), Luka Mesec (P, L), Marjan Šarec (P, L), Matej Tonin (NP, R), Miro Cerar (NP, L), Zmago Jelinčič Plemeniti (P, R), Zoran Janković (NP, R)

Spain

Albert Rivera (NP, R), Cayo Lara (NP, L), Francesc Homs (NP, L), Gabriel Rufián (P, L), Gaspar Llamazares (NP, L), Joan Ridao (P, L), Josu Erkoreka (NP, L), Mariano Rajoy Brey (NP, R), Oriol Junqueras (P, L), Pablo Casado (NP, R), Pablo Iglesias (P, L), Pedro Sánchez (NP, L), Rosa Díez (NP, L), Santiago Abascal (P, R)

Sweden

Alf Svensson (NP, L), Annie Lööf (NP, R), Ebba Busch Thor (P, R), Gudrun Schyman (NP, L), Göran Hägglund (NP, R), Göran Persson (NP, L), Isabella Lövin (NP, L), Jan Björklund (NP, R), Jimmie Åkesson (P, R), Jonas Sjöstedt (NP, L), Lars Ohly (NP, L), Maria Wetterstrand (NP, L), Stefan Löfven (NP, L), Åsa Romson (NP, L)

United Kingdom

Arlene Foster (NP, R), Boris Johnson (P, R), Charles Kennedy (NP, L), David Cameron (NP, R), Ed Miliband (NP, L), Gordon Brown (NP, L), Jeremy Corbyn (P, L), Jo Swinson (NP, L), John Swinney (NP, L), Nicola Sturgeon (NP, L), Nigel Farage (P, R), Paddy Ashdown (NP, L), Theresa May (NP, R), Tim Farron (NP, L), William Hague (NP, R)

United States[3]

Alexandria Ocasio-Cortez (P, L), Barack Obama (NP, L), Bernie Sanders (P, L), Bill Frist (NP, R), Chuck Schumer (NP, L), Elizabeth Warren (P, L), George Bush (NP, R), Harry Reid (NP, L), Hillary Clinton (NP, L), Joe Biden (NP, L), John Boehner (NP, R), John Kerry (NP, L), John McCain (NP, R), José E. Serrano (NP, L), Kevin McCarthy (NP, R), Mike Pence (NP, R), Mitch McConnell (NP, R), Mitt Romney (NP, R), Nancy Pelosi (NP, L), Paul Ryan (NP, R), Steny Hoyer (NP, L)

F.7 Bigrams selection

Figure F2 restricts the visualization to bigrams that appear in at least two quarters (left panel) and that occur at least ten times in at least one quarter (right panel). The Figure guides us in our choice of bigram-selection thresholds. The benchmark specification uses bigrams used in at least 10 quarters with a minimum of 10 occurrences in at least one quarter. For robustness checks we also focus on bigrams that were used in at least 5/10 quarters and at least 10/20/25 times in at least one of the quarters.

[3]Donald Trump is not in the US list as his Twitter account (together with that of his digital director, Gary Coby) was permanently suspended after his second impeachment.

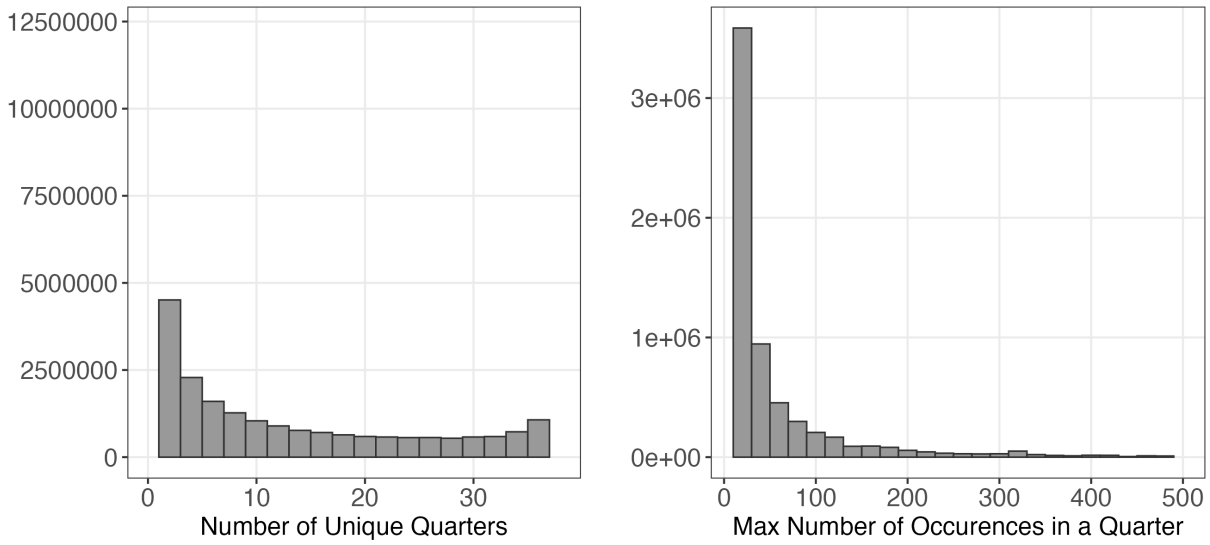


Figure F2: Histograms for the two measures of bigram frequency: No. of unique quarters a bigram was used (left panel), Max no. of occurrences of a bigram in a quarter (right panel). Low values of both measures are omitted.

F.8 Topics Classification Prompt

To create the training sample using ChatGPT the following prompt was used:

"You are a helpful research assistant.
I will give you a tweet from a politician around the world (translated into English). Classify it into one or more specific topics or return "no topic" if none apply. Here are the topics to classify into, along with descriptions. Some topics also include a list of keywords: these are examples of common expressions used for that topic, not strict rules. Use them as hints, but still classify semantically when the topic is implied even without those words. Use only the labels shown in parentheses. Each tweet must match exactly one or more of these labels.

When classifying, return only the labels:
Immigration and integration: topics related to the movement of people across borders, refugee issues, their integration into host countries, acquisition of citizenship rights, treatment of illegal migrants.
Environmental policy and climate change: topics related to the natural environment, pollution, long-term changes in global weather patterns and their environmental impact, and policies addressing climate change and environmental impact.
Foreign policy and defence: topics related to international relations, bilateral or multilateral agreements, military policies, defence spending.
Rights: topics related to gender, civil, human, or political rights. Including but not limited to

gender equality, women's rights, abortion, LGBTQ+ issues, diversity equity and inclusion policies, individual freedoms and rights (including free speech and personal liberties), political freedoms (such as voting, assembly, freedom of the press, and any kinds of political participation), bioethical issues.

Security, crime, and violence: topics related to internal security (e.g. "We need more police patrolling", "There is a security threat"), policing, crime, guns license, acts of violence (e.g. news of a murder).

Public health and healthcare policy: topics related to public health policies, health crises, vaccines, healthcare systems.

European Union affairs: topics specifically related to the European Union, including its institutions and policies.

Industrial policy, business, finance: topics related to regulation, productivity, banks, credit, finance, market competition, antitrust, business policies, small and medium enterprises, industrial subsidies, and topics mentioning any private firm or corporate entity.

International trade and globalization: topics related to international trade, tariffs, globalization, imports, commercial wars, unfair competition from China.

Energy and resources: topics related to energy sources, such as electricity, hydrogen, solar, wind power, nuclear, coal, oil, and natural gas.

Inequality: topics related to disparities in wealth, income, or socio-economic status, poverty, redistribution, policies providing social assistance to poor people.

Fiscal policy and taxation: topics related to taxes, government revenues, budgets, public debt and economic policy related to taxation, government spending, pension systems.

Inflation and monetary policy: topics related to rising price levels, cost of living, and to the related monetary policy measures and instruments (e.g. interest rates, quantitative easing/tightening, central banks policies).

Labour market and workers' rights: topics related to employment, unemployment, work injuries, job markets, labour rights, labour market policies, and labour market institutions (such as minimum wages, unions, employment protection legislation).

Connecting: topics related to non-political but personal communication such as greetings, weather or holidays. Keywords: greetings, holiday, birthday.

Elections and voting: topics related to elections, voting or political campaigns. Keywords: election, vote, campaign, candidate.

Self-promotion: when politicians promote themselves or their work e.g. "I'm on air today", "Check out my interview", "We have accomplished a lot during our term" etc. Keywords: join me, live, watch, interview.

Anti-establishment: Covers tweets which are

(a) anti-system: idea that the current political system is broken, corrupt, rigged and undemocratic.

(b) anti-elites: idea that elites are irresponsible, dishonest, unresponsive to the needs of the people,

guilty of cronyism and only look after their own interests. By elites we mean politicians, bureaucrats and civil servants (including international organizations like EU, IMF, WTO), economic elites (eg. big business, millionaires), central banks, top universities and establishment media.
(c) emphasizing the conflict between "the people" against "the elites/establishment": idea that society is split into two antagonistic groups: "the people" and "the corrupt elite".
Do not tag anti-establishment when:
(i) The target is a single politician or a specific party without generalizing to political class / elites / establishment.
(ii) The target is a foreign leader or regime (use foreign policy and possibly rights/security).
(iii) It is ordinary policy disagreement without elites/system are illegitimate framing.

Rules:
Return only the labels as a comma-separated list (e.g., "labour market, taxes"). If no topic applies, return exactly "no topic". A tweet can belong to multiple topics. If multiple topics apply, list all of them. If the tweet does not fit any of these topics, output only "no topic". Be concise and accurate in matching the topics."

F.9 Back Translation

To assess the quality of English translations, a sample of tweets per country was translated back into their original languages. For countries where many tweets were originally written in English and therefore excluded from the validation sample, larger samples were initially drawn to maintain representativeness. Before being compared, punctuation and language-specific stopwords were removed, and all texts have been lowercased. Tweets have been translated using Google Translate. The similarity between each back-translated tweet and its original version was then evaluated using multiple text similarity metrics.

First, we compute Jaccard Similarity, which measures how many words the two tweets have in common relative to the total number of unique words across both texts:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|},$$

where A is the set of words in the reference text and B is the set of words in the translated text. Second, we compute the BLEU Score: define unigram (bigram) precision as the number of matching unigrams (bigrams) over the number of unigrams (bigrams) in the candidate text. The score is based on modified n-gram precision, where matching n-gram counts are clipped by their maximum frequency in the reference text in order to avoid rewarding repeated words. Let p_1 and

p_2 denote the modified unigram and bigram precision, the BLEU score is computed as:

$$\text{BLEU} = BP \cdot \exp\left(\frac{1}{2}\log p_1 + \frac{1}{2}\log p_2\right) = BP \cdot \sqrt{p_1 p_2},$$

where BP is the brevity penalty, whose goal is to discourage overly short outputs, and is defined as:

$$BP = \begin{cases} 1 & \text{if } c > r, \\ \exp(1 - r/c) & \text{if } c \leq r, \end{cases}$$

with c denoting the candidate length and r the reference length. Third, we compute Cosine Similarity, which measures the semantic similarity between two texts by comparing their vector representations (embeddings) generated using the multilingual SentenceTransformer model `paraphrase-multilingual-MiniLM-L12-v2`, which supports more than 50 languages. Given the two embedding vectors A and B of the reference and translated texts, cosine similarity is computed as

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\|\|B\|}.$$

Table F5: Back Translation Metrics

Jaccard	BLEU (Only Bigrams)	BLEU	Cosine Similarity	Tweets	Language
0.65	0.59	0.65	0.96	106	Spanish
0.62	0.55	0.63	0.97	100	Italian
0.61	0.55	0.62	0.96	98	Swedish
0.56	0.45	0.54	0.96	96	German
0.54	0.46	0.54	0.94	95	Czech
0.65	0.59	0.65	0.96	88	Portuguese
0.56	0.51	0.58	0.95	83	Danish
0.61	0.53	0.61	0.95	83	Slovene
0.60	0.55	0.62	0.95	75	Norwegian
0.54	0.45	0.53	0.95	66	Dutch
0.40	0.31	0.40	0.92	61	Finnish
0.68	0.61	0.68	0.96	61	French
0.59	0.52	0.60	0.96	51	Polish
0.48	0.35	0.44	0.90	48	Hungarian

G Monte Carlo simulation for number of speakers and unobserved groups

Figure 3 raises two concerns: (1) partisanship for the random group exceeds 0.5, (2) partisanship for the random group of politicians exhibits pre-electoral divergence (qualitatively similar to the dynamics for the populist divide). Here we conduct a Monte Carlo simulation to examine whether these patterns could be due to the small number of politicians in the sample (relative to the number of possibly relevant political types).

We find that the observed patterns for the random groups are likely to be driven by both the small number of politicians and the presence of unobserved political types. However, it is important to note that the results obtained in this section reflect the properties of the penalized estimators when being used on different samples, not the properties of the samples. We verify that this is indeed the case by estimating random groups partisanship with the leave-out estimator, which consistently yields partisanship of 0.5 even in the presence of unobserved groups.

We keep the same set of bigrams used in the benchmark specification, but generate random bigram usage along with random samples of politicians for 20 sessions. For the politician samples, we construct a populist dummy, preserving the observed share of populist politicians (approximately 0.225), and vary the number of politicians per sample across $N \in \{50, 250, 500\}$. The median number of active politicians in our sample is 287, so the sample of 250 politicians is the closest to our data configuration. For the bigrams sample, we randomly draw bigram usage using Dirichlet sampling, keeping verbosity close to its observed value. We try to keep the distribution of bigrams such that partisanship matches the one estimated using the real data. Having obtained the simulated sample of bigrams used by each politician: (i) we estimate the parameters of the utility of speech; (ii) compute $\hat{q}_{ijt}^{P_i}$, for each politician i bigram j and session t; (iii) obtain individual - session specific partisanship measure, π_{it}; (iv) compute average partisanship, π_t, over the entire sample of politicians.

The bigrams sample is simulated under two scenarios, which we now describe, to explore the importance of a small number of politicians and of unobserved political types.

Scenario I. Here we explore the consequences of changing the number of politicians in the sample. First, we randomly select 1,000 bigrams and classify them as populist. When generating bigram usage, we boost the frequency of these bigrams for *populist* politicians by a factor of two. Second, among the 20 sessions we simulate, we treat the first five as pre-electoral, i.e., sessions in which the effect of political types on bigram choice (parameter ϕ) is higher. To capture this, we apply an additional boost to the same set of 1,000 bigrams in the first five sessions, of about 20% on average. In other words, populist politicians consistently use this set of 1,000 bigrams more frequently across all 20 sessions, but especially so during the first five.

The results of this simulation are presented in Figure G1. The solid lines show the average partisanship outcome, π_t, for the populist versus non-populist division, while the dotted lines depict

the average π_t for ten random binary partitions that do not correspond to the populist–non-populist split. Vertical segments are 95% confidence intervals.

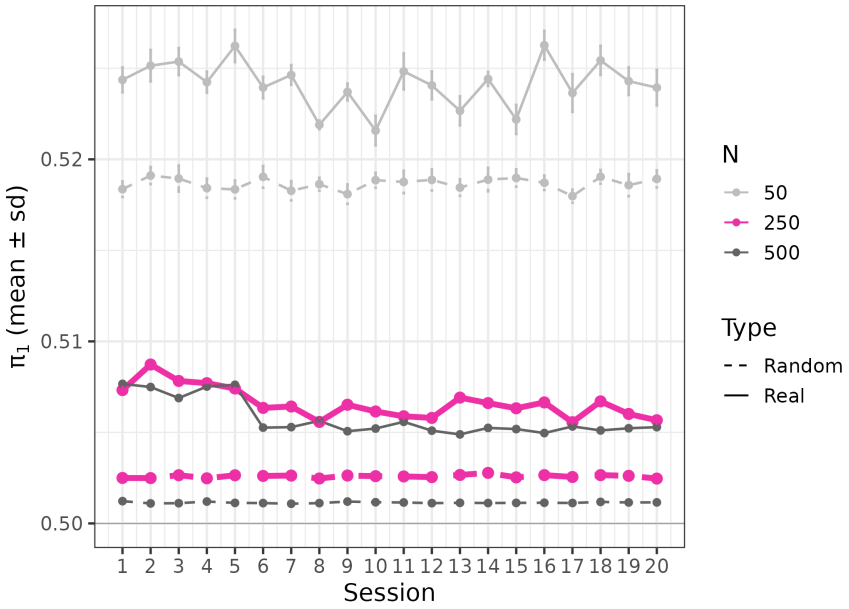


Figure G1: Average partisanship estimates over politicians and sessions for real and random populist groups. No unobserved electorally relevant groups.

As the number of politicians increases, populist partisanship is not affected once it reaches 250 politicians, but partisanship of the random groups (the dotted lines) converges toward 0.5. This addresses concern (1), suggesting that the relatively large observed values of partisanships for the random groups are partly driven by the small number of politicians in our sample.

Nevertheless, in these simulations partisanship of the random groups is not sensitive to varying parameter ϕ, since it has the same pattern in the first five sessions as in the remaining ones (the difference between the first sessions and the rest is statistically significant for populist partisanship, but not for partisanship of the random groups). This suggests that the number of politicians, on its own, cannot explain the observed pre-electoral pattern of the random group partisanship in Figure 4.

Scenario II. To explain this pre-electoral pattern, in Figure G2 we explore the consequences of also having unobserved political types. Here we keep the same data generating process, except that in addition to populist and non-populist, we generate two additional political types within the populist and non-populist subsets, boosting for each type the usage of a specific set of bigrams. Moreover, similar to the additional boost in sessions 1-5 for the populist bigrams, we boost the usage of these specific sets of bigrams in sessions 1-5 too, thus making these four political groups electorally relevant. We then compute average partisanship between two groups of politicians: populist vs non-populist (the solid lines) and an entirely random partition (the dotted lines). Again

we repeat these simulations for three sample sizes of politicians, 50, 250 and 500. The results of these simulations are depicted in Figure G2.

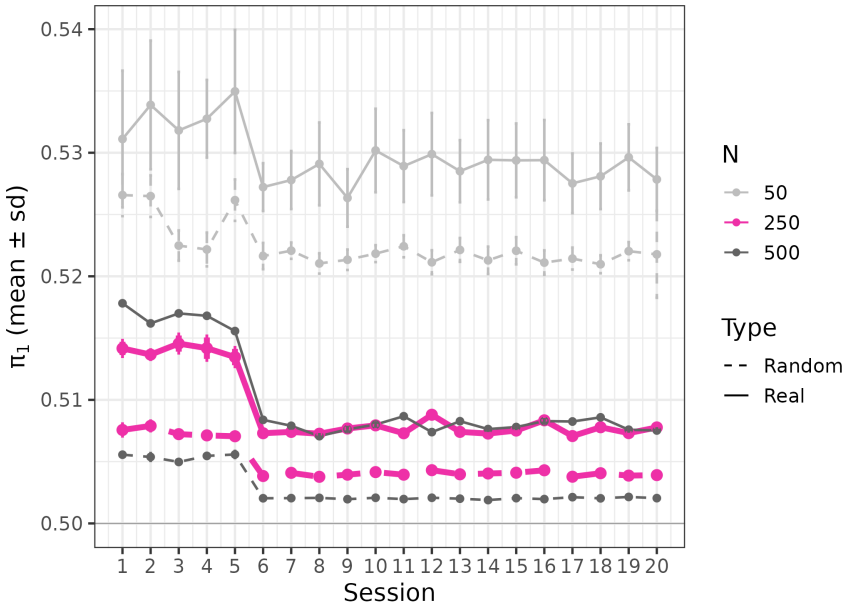


Figure G2: Average partisanship estimates over politicians and sessions for real and random populist groups. Four unobserved electorally relevant groups.

Partisanship in the random groups now is indeed much larger in Sessions 1-5, where unobserved political types have a larger effect on bigram usage, than in the remaining sessions, and this difference is statistically significant. This pattern suggests that some political types are not fully accounted for in our framework. Note that the sample size of politicians continues to be relevant also in these simulations.

This suggests that what matters is not merely the number of politicians, but the number of politicians relative to the number of relevant political types. If the number of politicians is not sufficiently large relative to the number of relevant political types, the two random groups differ systematically in their bigram usage and this is captured by the partisanship statistics.

H Tables and Figures

H.1 Tables

Table H1: Politicians by ideology and populism (counts with row percentages)

	Non-populist	Populist	Total
Left	185 (83.0%)	38 (17.0%)	223 (60.8%)
Right	98 (68.1%)	46 (31.9%)	144 (39.2%)
Total	283 (77.1%)	84 (22.9%)	367 (100%)

Table H2: Wald Tests: Coefficient differences from $\beta_{\text{quarter diff=0}}$

Hypothesis	F-statistic	p-value	Dep. Variable
$\beta_{\text{quarter diff=1}} = \beta_{\text{quarter diff=0}}$	3.775	0.0521	π_{it}
$\beta_{\text{quarter diff=2}} = \beta_{\text{quarter diff=0}}$	4.479	0.0343	π_{it}
$\beta_{\text{quarter diff=3}} = \beta_{\text{quarter diff=0}}$	1.076	0.3000	π_{it}

Table H3: Unique Elections by Quarter

No. of unique countries per election quarter	1	2	3	4	5
No. of occurrences	12	10	3	2	2

Notes. Only calendar quarters used in the main analysis are considered. Elections that fall outside of the study window are not considered.

Table H4: BERT vs Manual Labeling Metrics

	Accuracy	F1 Score
Policy Topics		
Business	0.91	0.64
Climate	0.96	0.83
Energy	0.97	0.87
European Union	0.98	0.89
Foreign Policy	0.92	0.61
Health	0.97	0.86
Immigration	0.98	0.88
Inequality	0.92	0.55
Inflation	0.96	0.55
Labor Market	0.96	0.83
Rights	0.93	0.68
Security	0.94	0.69
Taxes	0.97	0.86
Trade	0.96	0.42
Non-Policy Topics		
Anti Establishment	0.94	0.49
Personal Communication	0.96	0.53
Elections	0.96	0.67
Self Promotion	0.81	0.33

Notes: Accuracy is defined as the number of correct predictions divided by the total number of predictions. The F1 Score is the harmonic mean of precision and recall: it is twice their product divided by their sum. Precision is the number of true positives divided by the number of all predicted positives, while recall is the number of true positives divided by the number of actual positives.

H.2 Figures

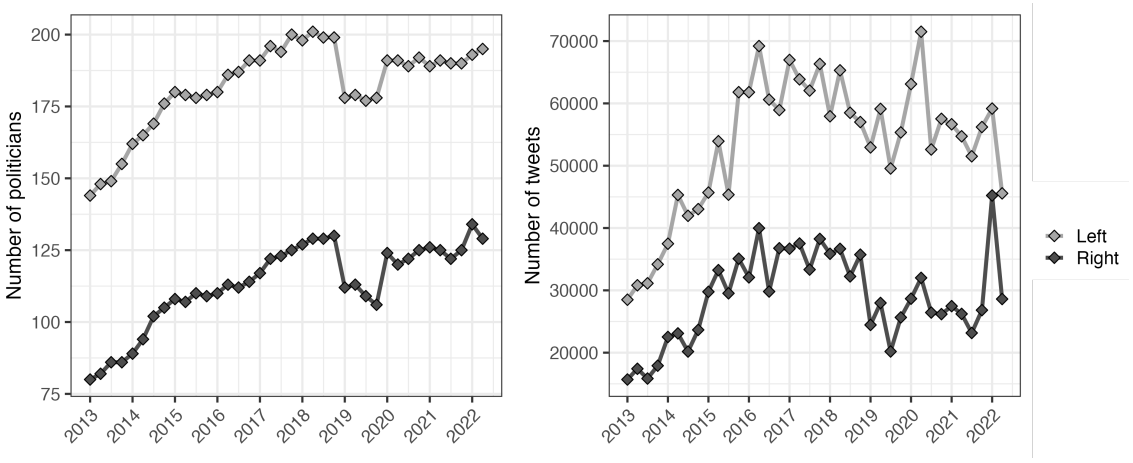


Figure H1: Number of politicians active on Twitter per calendar quarter, and number of tweets, per Left-Right divide

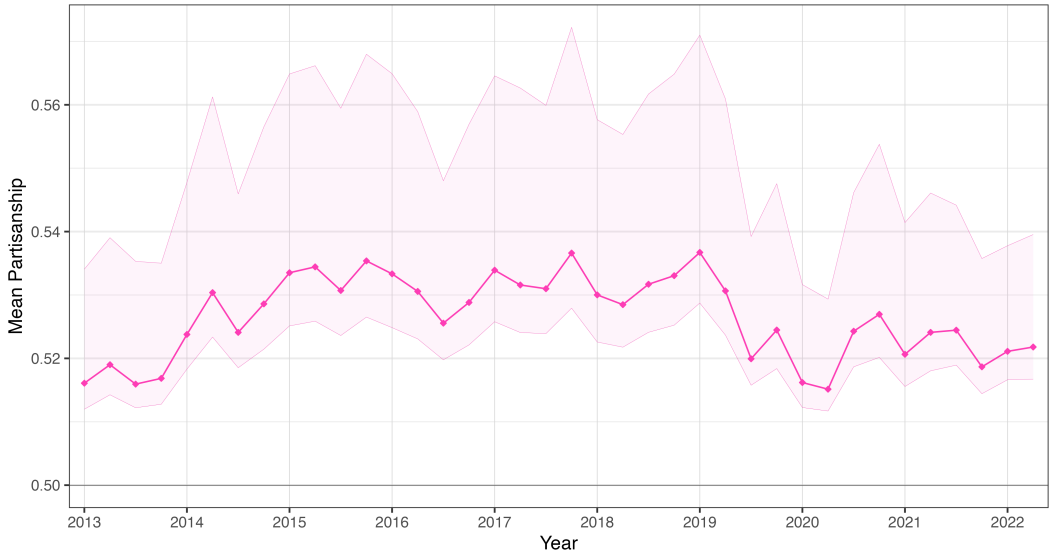


Figure H2: Average partisanship with 90% CI

Notes. Figure plots average values and confidence intervals of populist partisanship over the study period. Partisanship is defined as in Equation 11. Average values are defined as in Equation 12. Confidence Intervals are constructed via subsampling, as described in Section D.

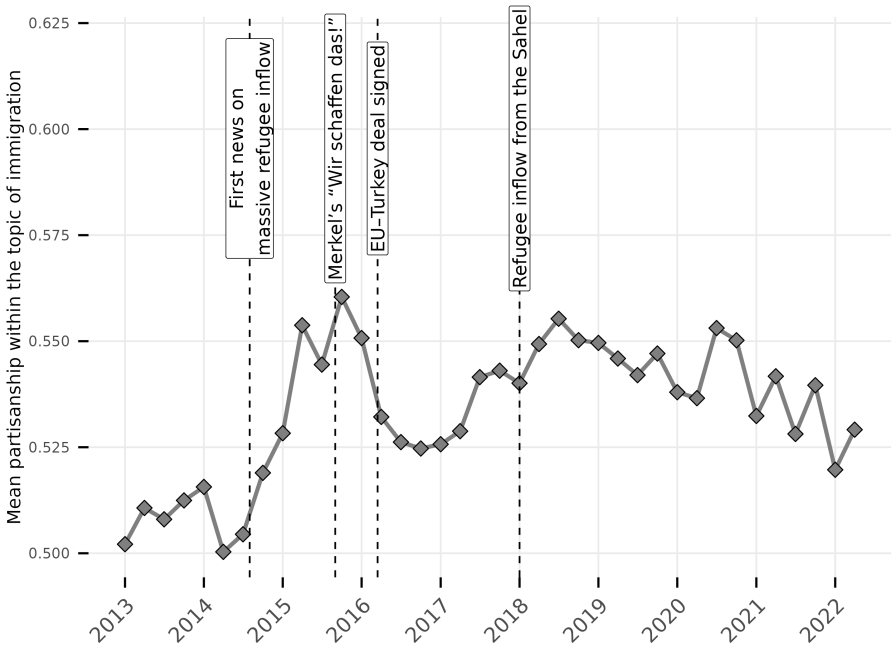


Figure H3: Partisanship within the topic of immigration

Notes. The figure reports the average P–NP partisanship by calendar quarter, computed using the sample of bigrams that belong to the immigration topic and restricting the sample to European countries. Partisanship is defined as in Equation 11. Average values are defined as in Equation 12.

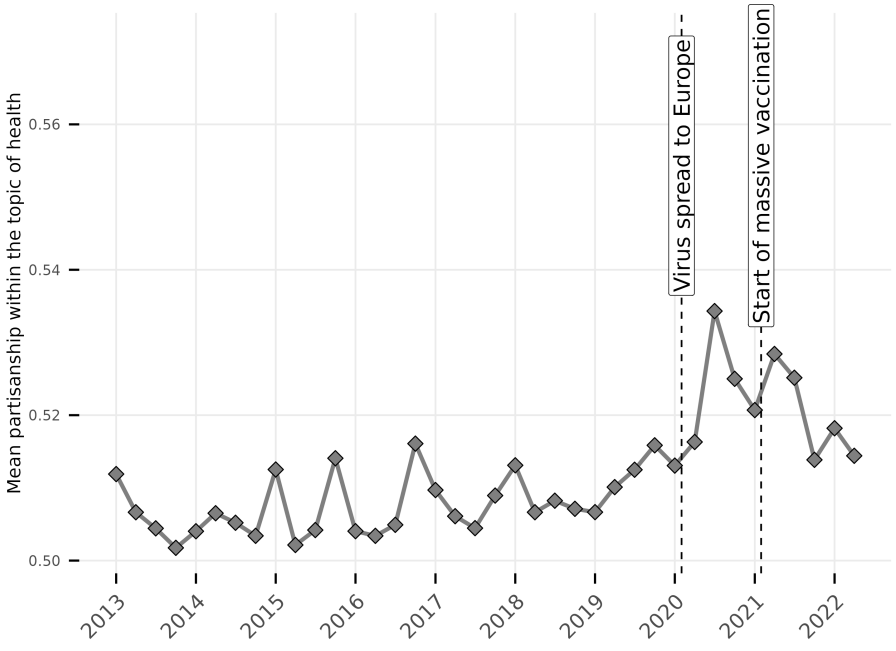


Figure H4: Partisanship within the topic of public health

Notes. The figure reports the average P–NP partisanship by calendar quarter, computed using the sample of bigrams that belong to the public health topic. Partisanship is defined as in Equation 11. Average values are defined as in Equation 12.

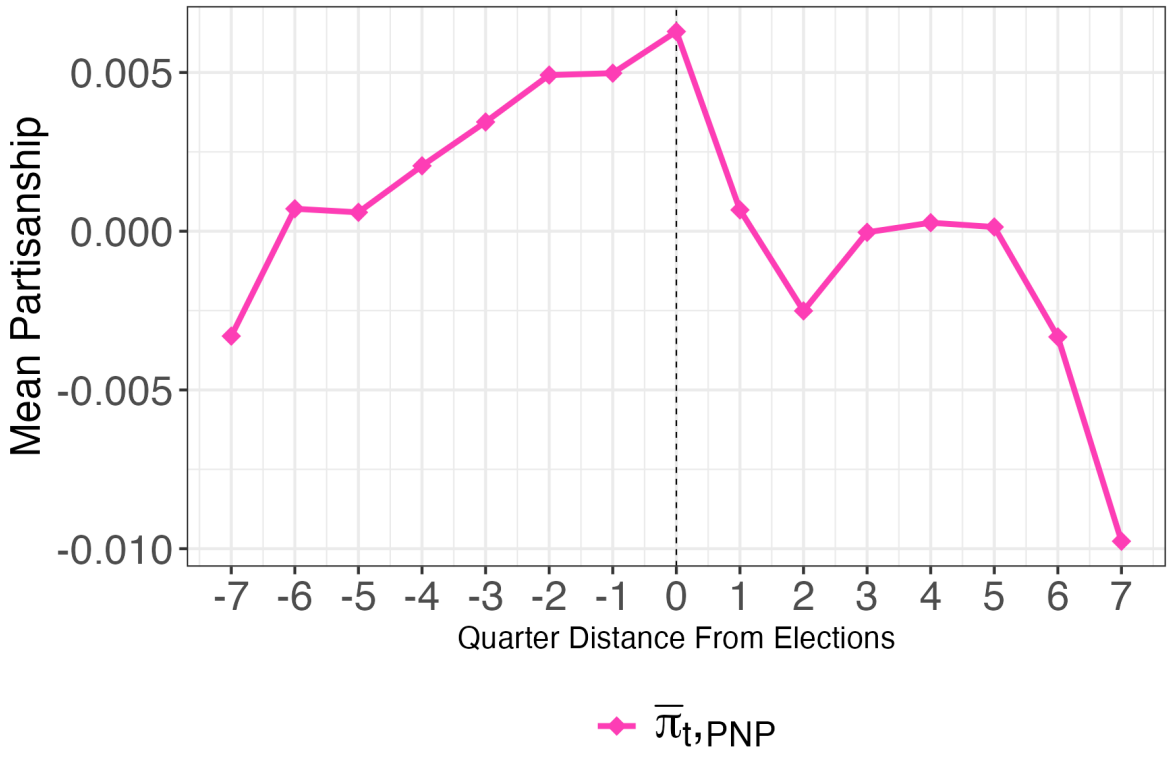
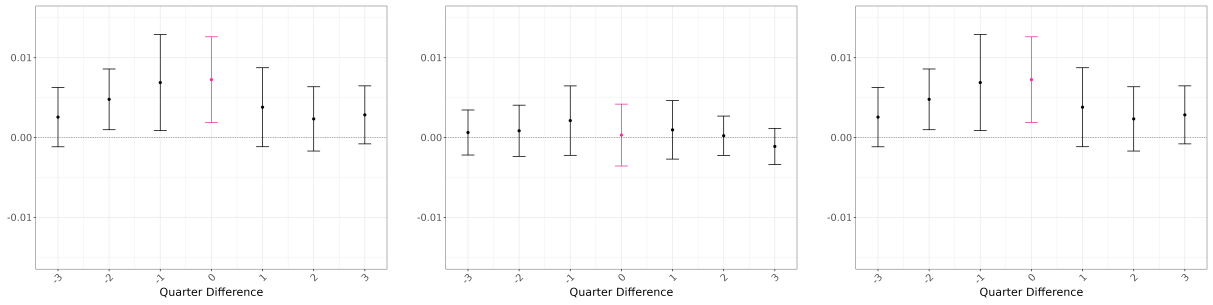


Figure H5: Average estimated π_{it} per quarter difference from elections (net of calendar-quarter fixed effects).

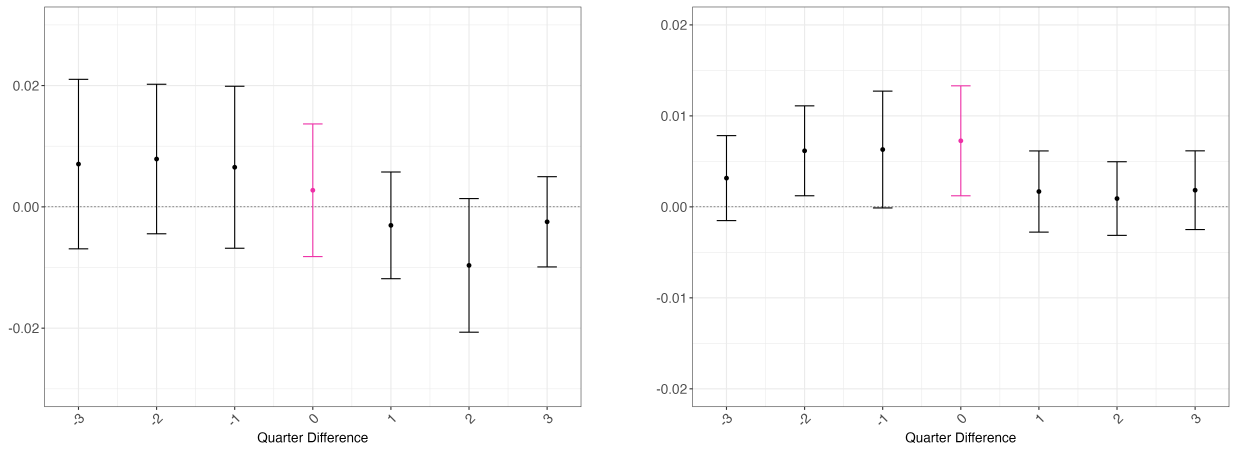
Notes. $\bar{\pi}_{t,PNP}$ refers to average populist partisanship net of calendar-quarter fixed effects. Partisanship is defined as in Equation 11. Average values are defined as in Equation 12.



(a) Populist vs Non-Populist (b) Left vs Right (c) Four political types

Figure H6: Event study estimates from the model with four political types (including both populist-non-populist and left-right dimensions)

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13. Outcome variable is populist partisanship (left panel), left-right partisanship (middle panel) and four types partisanship (right panel). Errors are clustered at the elections level, 95% confidence intervals are reported. Partisanship with several political types is defined as in Section C.



(a) English-speaking countries

(b) Not English-speaking countries

Figure H7: Event study estimates for π_{it} for countries with English and other languages as main official language

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13 on the subsamples of English-speaking countries (left) and not English-speaking countries (right). Errors are clustered at the elections level, 95% confidence intervals are reported. Partisanship is defined as in Equation 11.

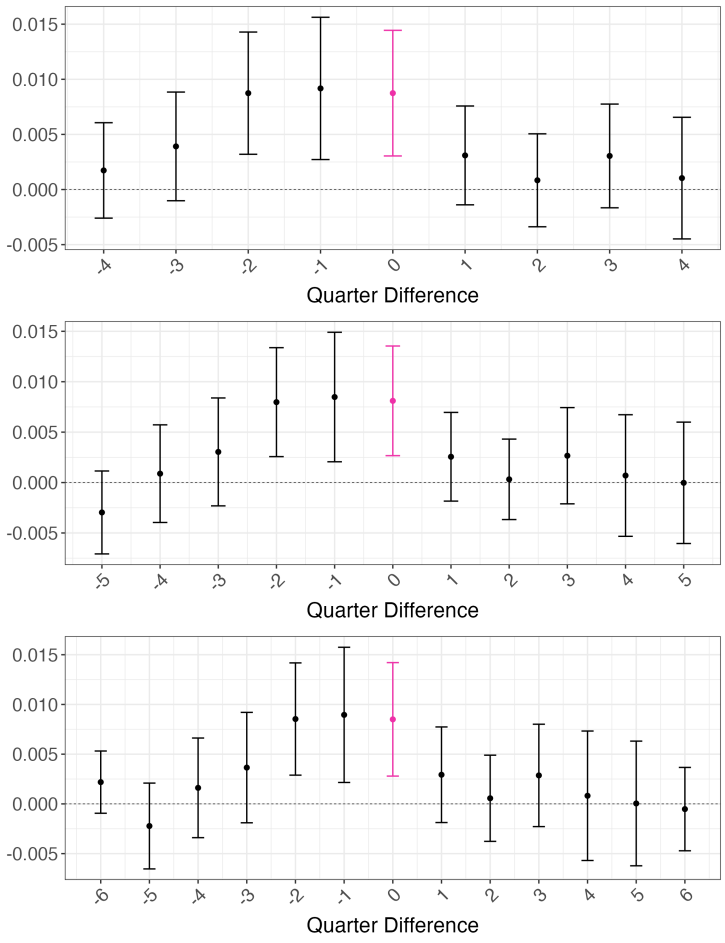


Figure H8: Event study results. Bigram selection threshold is 10-10. Event window is 4 quarters, 5 quarters and 6 quarters.

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13 using different definitions of electoral window – 4 quarters (upper panel), 5 quarters (middle panel), 6 quarters (bottom panel). Errors are clustered at the elections level, 95% confidence intervals are reported. Outcome variable is a populist partisanship, defined as in Equation 11.

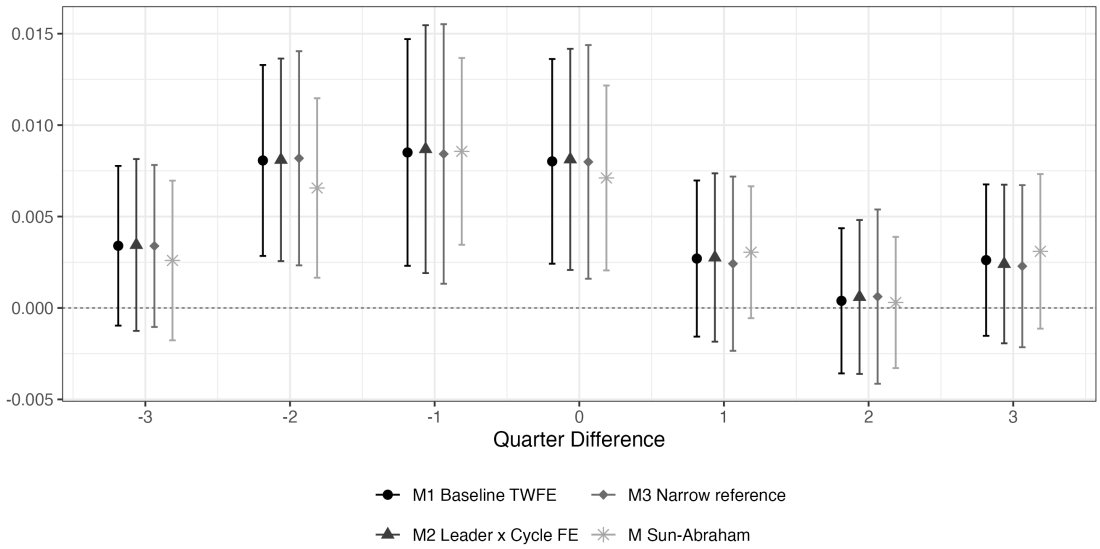


Figure H9: Partisanship event study results, TWFE robustness

Notes. Partisanship is defined as in Equation 11. Each panel plots coefficients from four TWFE event-study specifications. M1 is the baseline (leader, quarter, and country × electoral-cycle fixed effects; errors clustered by electoral cycle). M2 replaces the leader FE with a leader × electoral-cycle FE, restricting identification to within-leader within-cycle variation. M3 uses a narrow reference window (quarters 4–6 from the election only). Bars are 95% confidence intervals. M Sun-Abraham plots an interaction-weighted estimator, which aggregates election-cohort-specific coefficients with non-negative weights to guard against bias from heterogeneous effects across elections (Sun and Abraham, 2021).

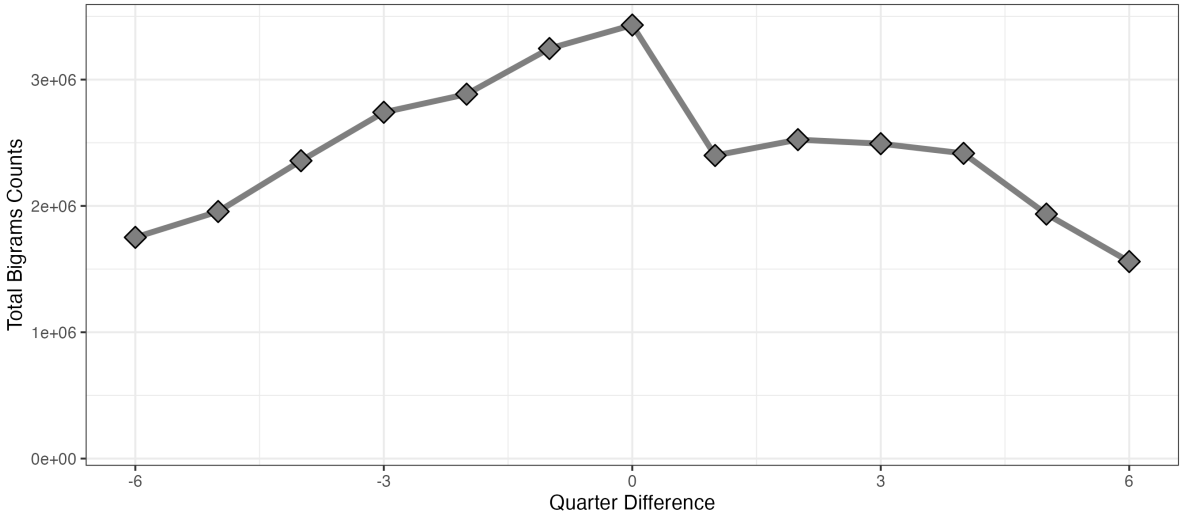


Figure H10: Number of bigrams used per quarter, difference from the elections

Notes. This graph plots the total number of all bigrams used in tweets at a given quarter difference from elections.

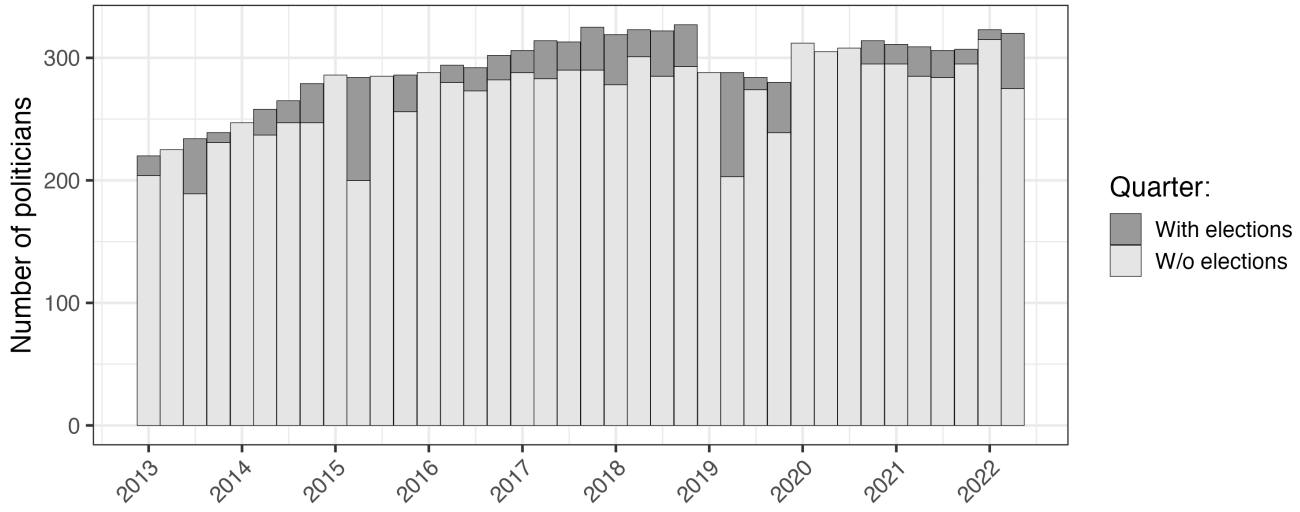


Figure H11: Number of politicians per calendar quarter

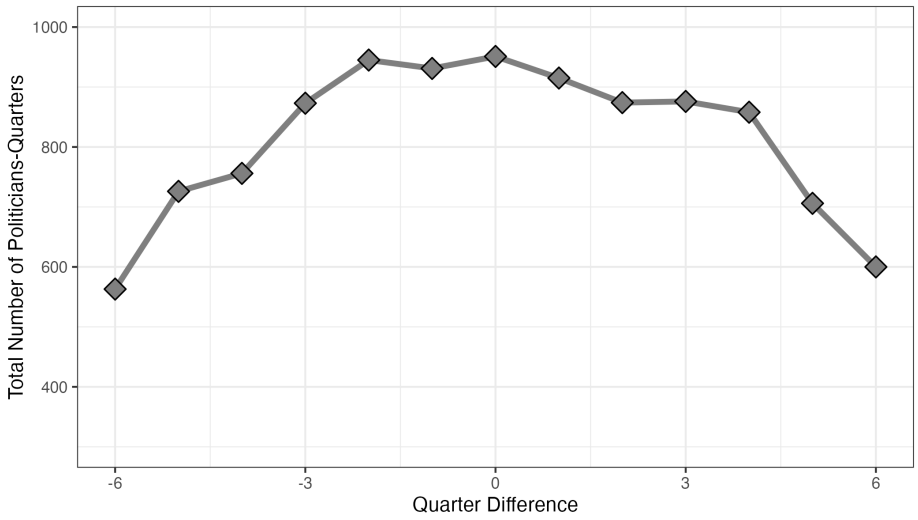


Figure H12: Number of unique quarter-politicians observations across the electoral cycle

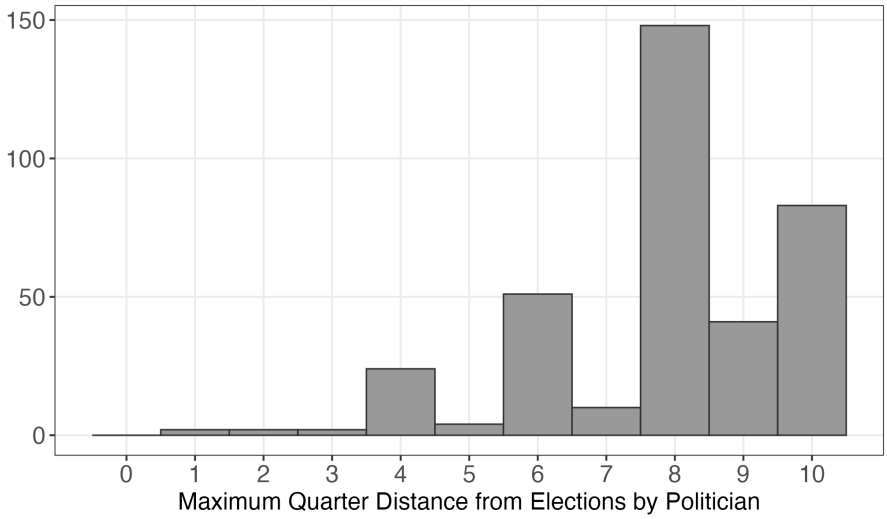


Figure H13: Maximum absolute quarter distance from the elections for politicians

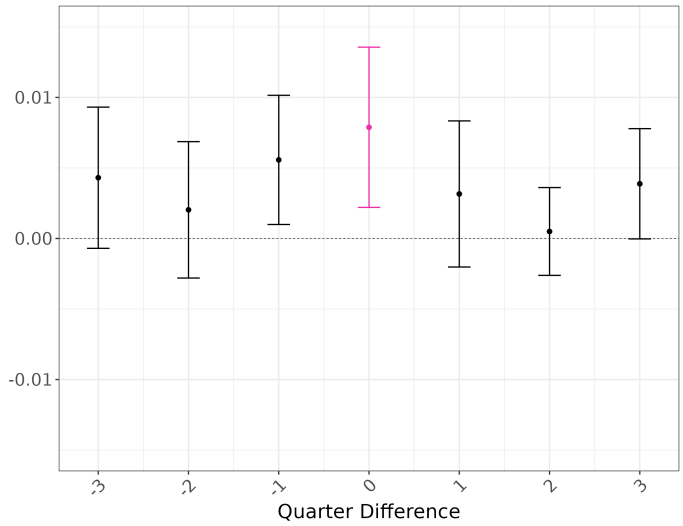
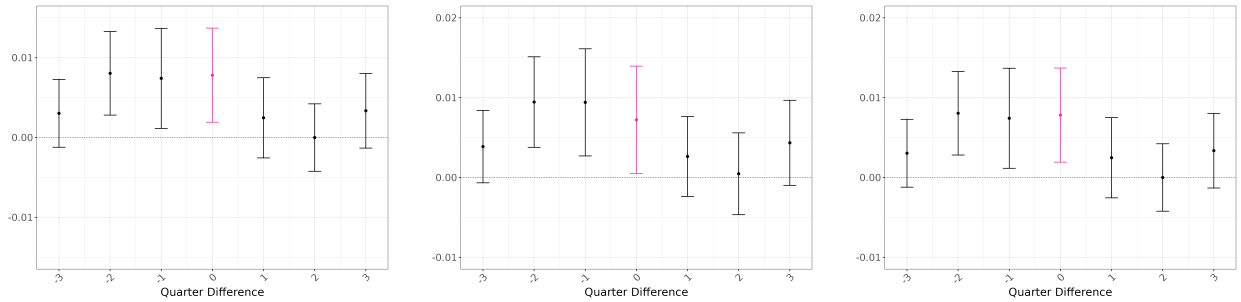


Figure H14: Event study estimates for π_{it} based on a restricted random sample of 220 politicians per quarter.

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13 on the restricted sample of 220 politicians randomly selected in each quarter. Errors are clustered at the elections level, 95% confidence intervals are reported. Outcome variable is populist partisanship, defined as in Equation 11.



(a) 5-10 (b) 5-20 (c) 10-25

Figure H15: Event study estimates for π_{it} using the 5-10 bigram selection threshold (left panel), 5-20 (center panel) and 10-25 (right panel)

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13 on the samples with different bigram selection thresholds. Left panel includes all bigrams used in at least 5 quarters and at least 10 times in at least one quarter. Center panel includes all bigrams used in at least 5 quarters and at least 20 times in at least one quarter. Right panel includes all bigrams used in at least 10 quarters and at least 25 times in at least one quarter. Errors are clustered at the elections level, 95% confidence intervals are reported. Outcome variable is populist partisanship, defined as in Equation 11.

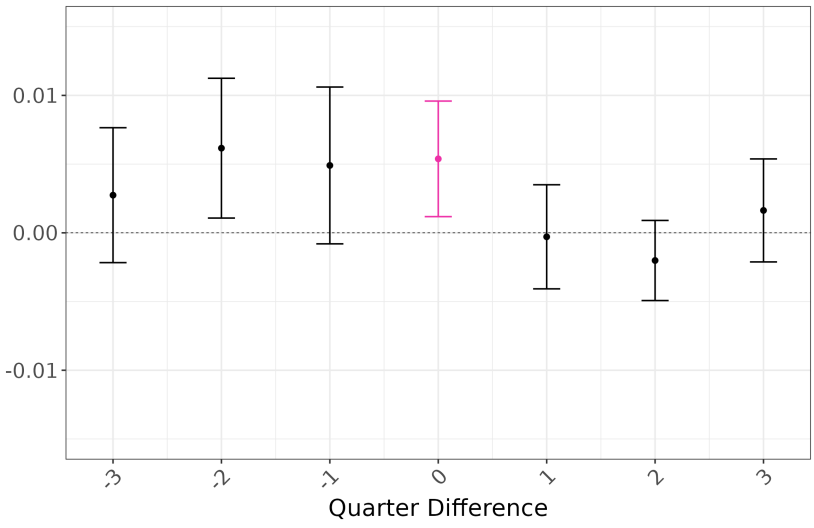


Figure H16: Event study estimates for π_{it} with ambiguous politicians coded as non-populists

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13. Outcome variable is populist partisanship, defined as in Equation 11. Partisanship is estimated by assigning ambiguous politicians to the non-populist group (rather than to the populist group, as in the main text). Errors are clustered at the elections level, 95% confidence intervals are reported.

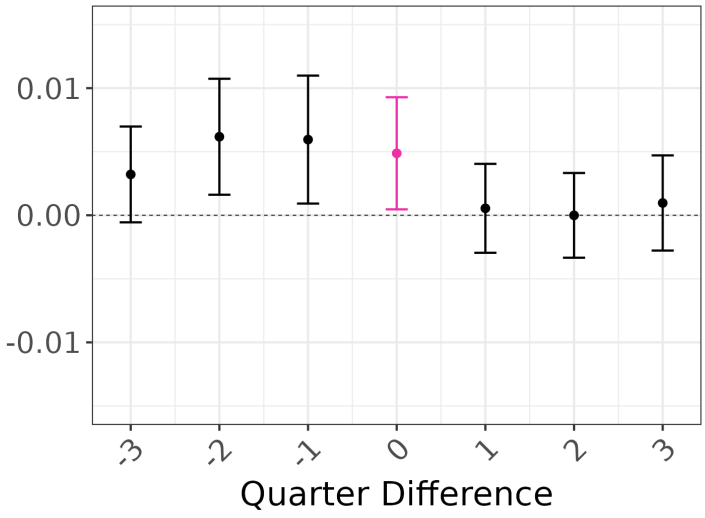


Figure H17: Event study estimates for π_{it} on the sample of bigrams without mentions of specific parties and politicians

Notes. Figure plots point estimates and confidence intervals for $\beta_{d(i,t)}$ from Specification 13. Outcome variable is populist partisanship, defined as in Equation 11. Partisanship is estimated on a subset of bigrams that excludes bigrams mentioning specific parties or politicians. Errors are clustered at the elections level, 95% confidence intervals are reported.

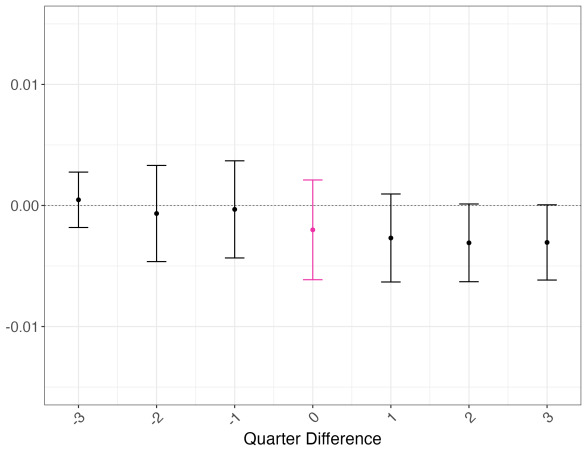


Figure H18: Event studies for π_{it}, Left-Right partisanship measure

Notes. Point estimates and confidence intervals are plotted for $\beta_{d(i,t)}$ from Specification 13. Outcome variable is left-right partisanship, with affiliation defined by the Manifesto RILE variable (instead of the ChatGPT classification). Partisanship is defined as in Equation 11. Errors are clustered at the elections level, 95% confidence intervals are reported.

References

Chatterjee, Arindam and Soumendra Nath Lahiri. 2011. "Bootstrapping lasso estimators." *Journal of the American Statistical Association* 106(494):608–625.

Colonnelli, Emanuele, Valdemar Pinho Neto and Edoardo Teso. 2025. "Politics at work." *American Economic Review* 115(10):3367–3414.

Sun, Liyang and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of econometrics* 225(2):175–199.