PNAS
www.pnas.org

Supplementary Information for

The nonlinear feedback dynamics of asymmetric political polarization

Naomi Ehrich Leonard[a], Keena Lipsitz[b], Anastasia Bizyaeva[a], Alessio Franci[c], Yphtach Lelkes[d]

[a] Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544

[b] Department of Political Science, Queens College and The Graduate Center, City University of New York, Flushing, NY 11367

[c] Department of Mathematics, National Autonomous University of Mexico, 04510 Mexico City, Mexico

[d] Annenberg School for Communication and Department of Political Science, University of Pennsylvania, Philadelphia, PA 19104

To whom correspondence should be addressed.
E-mail: naomi@princeton.edu and keena.lipsitz@qc.cuny.edu

This PDF file includes:

Supplementary text
Figs. S1 to S12
Tables S1 to S2
SI References

Supporting Information Text

S1. Model analysis

A. Derivation from General Model. In this section we derive the model for evolution of the ideological positions of two political party elite populations by specializing a general model of opinion formation recently proposed in (1); see also (2). Suppose that each political party elite population forms positions on n mutually exclusive ideological dimensions. We define the real-valued variable z_{ij} to be party i's position on ideological dimension j, where $i = 1, 2$. In this notation $z_{ij} > 0$ (< 0) corresponds to party i favoring (disfavoring) policy positions that align with ideological dimension j. Additionally $z_{ij} = 0$ corresponds to a neutral stance along ideological dimension j. Mutual exclusivity of the ideological dimensions places a constraint on each party's positions:

$$\sum_{j=1}^{n} z_{ij} = 0 \quad \text{for } i = 1, 2. \tag{1}$$

Let $\mathbf{Z}_i = (z_{i1}, \ldots, z_{in})$ be the vector of ideological positions of party i and define $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$. We can then define the drift in party i's ideological position along dimension j as

$$F_{ij}(\mathbf{Z}) = -d_{ij} z_{ij} + u_i \left(\tanh\left(\alpha_i z_{ij} - \gamma_i z_{kj}\right) + \sum_{\substack{l \neq j \\ l=1}}^{n} \tanh\left(-\beta_{il} z_{il} + \delta_{il} z_{kl}\right) \right) + b_{ij}, \qquad k \neq i, \qquad i, k = 1, 2 \tag{2}$$

where tanh is the hyperbolic tangent function. The drift Eq. (2) contains a number of real-valued parameters, which we interpret in the following way:

1. $d_{ij} > 0$ is the resistance of party i to forming a non-neutral position along the ideological dimension j;
2. $u_i \geq 0$ is the level of attention party i pays to its within-party and cross-party interactions;
3. b_{ij} is the party's intrinsic bias in favor of or against ideological dimension j;
4. $\alpha_i \geq 0$ is the amount of self-reinforcement in party i's ideological positions;
5. γ_i captures the influence of party k on party i along the same ideological dimension;
6. β_{il} and δ_{il} capture the influence of positions along ideological dimension l on party i's position along dimension j.

The evolution over time of the ideological position of party i along ideological dimension j is then summarized by the ordinary differential equation

$$\tau_z \frac{dz_{ij}}{dt} = F_{ij}(\mathbf{Z}) - \frac{1}{n} \sum_{p=1}^{n} F_{ip}(\mathbf{Z}) \tag{3}$$

where subtracting the average drift in ideological position in Eq. (3) models the mutual exclusivity of the ideological dimensions.

For this paper we further specialize this model to two ideological dimensions, conservative and liberal. With this simplification, each party's ideology is captured by a single variable which we define as

$$x_r := z_{11} = -z_{12} \tag{4}$$

for the Republican party elites and accordingly,

$$x_d := z_{21} = -z_{22} \tag{5}$$

for the Democratic party elites. Additionally, we assume $\beta_{il} = \delta_{il} = 0$ and normalize $d_{ij} = 1$, $u_i = 1$ for all $i, j, l = 1, 2$. Finally, we relabel $\tau_z = \tau_x$, $\alpha_1 = \alpha_r$, $\alpha_2 = \alpha_d$, $\gamma_1 = \gamma_r$, $\gamma_2 = \gamma_d$, and define

$$b_r := \frac{1}{2}(b_{11} - b_{12}), \qquad b_d := \frac{1}{2}(b_{21} - b_{22}). \tag{6}$$

With these assumptions imposed, we arrive at the model equations [1]-[2] from the main paper:

$$\tau_x \frac{dx_r}{dt} = \tanh\left(\alpha_r x_r - \gamma_r x_d\right) - x_r + b_r, \tag{7}$$

$$\tau_x \frac{dx_d}{dt} = \tanh\left(\alpha_d x_d - \gamma_d x_r\right) - x_d + b_d. \tag{8}$$

Although we have arrived at this model by treating the two parties as two distinct entities, the general modeling framework proposed in (1) can also be utilized to model evolution of ideological positions of many interacting party members. For such an agent-based model, each node would represent an individual policymaker rather than the party as a whole. Clustering results, e.g. (1, Theorem III.5), suggest that with proper parametrization of the inter-agent interactions, an agent-based model can behave in a manner that is formally equivalent to the two-node model Eq. (7),Eq. (8). This means that the average opinions of the nodes comprising each of the respective parties would behave the same as x_r, x_d in the two-party model. In the main paper we perform analysis and numerical experiments with the two-party model. At the end of this supplement we use the general agent-based model to represent 50 Republican elites and 50 Democratic elites and illustrate their dynamics in simulation. We show that even with (small) parametric differences among the 50 Republican elites and among 50 Democratic elites, the average behavior of each population agrees with the behavior of the two-population model.

B. Bifurcation analysis of model with constant parameters. In this section we establish using bifurcation analysis the existence of a critical value in one or more of the parameters of the model Eq. (7),Eq. (8). Consider the model with $b_r = b_d = 0$. The Jacobian matrix of this system evaluated at the origin $(x_r, x_d) = (0, 0)$ is

$$J = \frac{1}{\tau_x}\begin{pmatrix} -1+\alpha_r & -\gamma_r \\ -\gamma_d & -1+\alpha_d \end{pmatrix}. \qquad [9]$$

The eigenvalues of Eq. (9) are

$$\lambda_{1,2} = -1 + \frac{1}{2}(\alpha_r + \alpha_d) \pm \frac{1}{2}\sqrt{(\alpha_r - \alpha_d)^2 + 4\gamma_r\gamma_d} \qquad [10]$$

and one of the two eigenvalues is zero whenever

$$(-1+\alpha_r)(-1+\alpha_d) = \gamma_r\gamma_d. \qquad [11]$$

The origin of the nonlinear system Eq. (7), Eq. (8) is stable whenever $\mathrm{Re}(\lambda_{1,2}) < 0$, which is true whenever one of the following conditions is met:

1. $\alpha_r + \alpha_d < 2$ and $\gamma_r\gamma_d < -\frac{1}{4}(\alpha_r - \alpha_d)^2$;
2. $\alpha_r + \alpha_d < 2(1 + \alpha_r\alpha_d - \gamma_r\gamma_d)$ and $\gamma_r\gamma_d \geq -\frac{1}{4}(\alpha_r - \alpha_d)^2$.

Whenever Eq. (11) is satisfied, the Jacobian Eq. (9) is singular. As one or more of the parameters $\alpha_r, \alpha_d, \gamma_r, \gamma_d$ is varied near this singularity, the origin can lose stability and new branches of steady-state solutions can emerge in a nonlinear phenomenon called a steady-state bifurcation. These new solutions will appear along the kernel of J at the singularity. Next we illustrate how this bifurcation analysis specializes to two scenarios examined in the main paper and predicts the emergence of polarized outcomes. Although we only formally handle these two cases, their results (namely, the appearance of a pitchfork bifurcation) apply more generally in Eq. (7),Eq. (8) with heterogeneous parameters.

B.1. Case I: $\gamma_r = \gamma_d = 0$**.** Without party interactions, the evolution in time of each party's ideological position is decoupled from the other and summarized by a one-dimensional equation of the form

$$\tau_x \frac{dx_i}{dt} = -x_i + \tanh(\alpha_i x_i) + b_i. \qquad [12]$$

By (1, Proposition IV.1) when $b_i = 0$, Eq. (12) exhibits a supercritical pitchfork bifurcation at $\alpha_i = 1$. For $\alpha_i < 1$ the neutral state ($x_i = 0$) is stable, and for $\alpha_i > 1$ it is unstable. Two non-neutral stable branches of steady-state solutions appear for $\alpha_i > 1$, one corresponding to a right-leaning position, and the other to a left-leaning position - see Figure S1(A). These ideological positions rapidly become polarized as α_i increases in value.

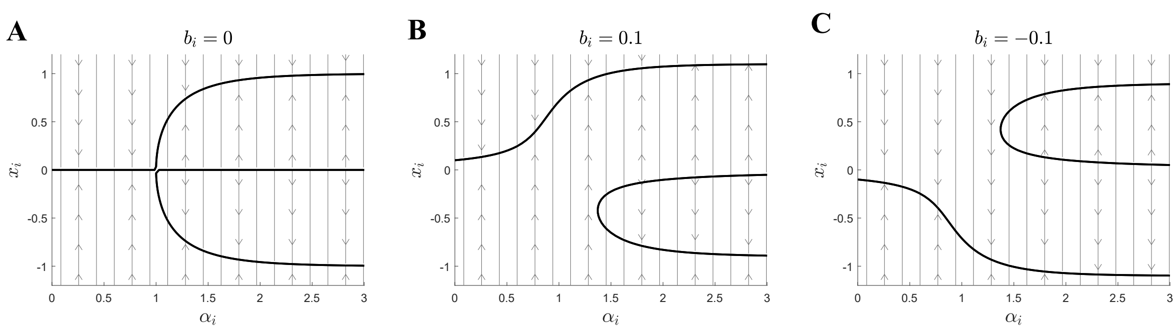


Fig. S1. Bifurcation diagrams for Eq. (12) with (A) no bias, (B) positive bias, and (C) negative bias. Black lines plot steady-state solutions (nullclines) and gray arrows are streamlines showing direction of the flow.

When $b_i \neq 0$, *unfolding theory* (3, Chapter III) predicts that the general shape of the bifurcation diagram resembles the unbiased case pictured in Figure S1(A), but, near the singularity, the equilibrium favored by the additive bias b_i is selected - see Figure S1(B),(C). In the context of political polarization, this means that the ideological position of each party can become strongly polarized in the direction of a small bias, as long as the party's self-reinforcement is sufficiently strong. The degree of polarization increases monotonically with magnitude of α_i, becoming steepest when α_i approaches the critical value of 1. Thus, when α_r and α_d have different values, the party with the greater self-reinforcement is more polarized in its ideological position. This difference in the degree of polarization is particularly strong when one of the α_i coefficients is below its critical value of 1, and the second one is above 1.

B.2. *Case II:* $\alpha_r = \alpha_d = 0$. Let $b_r = b_d = 0$. In this case, the Jacobian in Eq. (9) simplifies to

$$J = \frac{1}{\tau_x}\begin{pmatrix} -1 & -\gamma_r \\ -\gamma_d & -1 \end{pmatrix} \quad [13]$$

and is singular whenever

$$\gamma_r \gamma_d = 1. \quad [14]$$

This corresponds to two potential scenarios:

1. $\gamma_r > 0$ and $\gamma_d > 0$ (reflexive partisanship): At the singularity the kernel of J is

$$\text{span}\{(\sqrt{\gamma_r}, -\sqrt{\gamma_d})\} \quad [15]$$

and therefore new steady-state solution branches appear along a space tangent to

$$x_d = -\sqrt{\frac{\gamma_d}{\gamma_r}} x_r. \quad [16]$$

Qualitatively these solutions correspond to ideological positions of the two parties diverging, with one party taking on a left-leaning stance and the second taking on a right-leaning stance. Additionally when $\gamma_d \neq \gamma_r$, the party with a stronger degree of reflexive partisanship takes on a stronger ideological position. Restricting Eq. (7),Eq. (8) to the kernel of J, the equilibria are fully described by the one-dimensional equation

$$\frac{dx_r}{dt} = -x_r + \tanh(\sqrt{\gamma_r \gamma_d} x_r) \quad [17]$$

coupled with Eq. (16), which is the same equation as Eq. (12) with $\sqrt{\gamma_r \gamma_d}$ acting as a bifurcation parameter. Figure S1(A) illustrates the structure of the equilibria of this equation, if the variable along the horizontal axis is replaced with $\sqrt{\gamma_r \gamma_d}$. The bifurcation point corresponds to $\sqrt{\gamma_r \gamma_d} = 1$, and the two parties' ideological positions become polarized, satisfying Eq. (17) for $\sqrt{\gamma_r \gamma_d} > 1$.

Addition of small nonzero biases b_r, b_d to the model Eq. (7),Eq. (8) will have a two-fold effect: 1) qualitatively changing the structure of the equilibria near the singular point, as pictured in Figure S1(B)),(C), and perturbing the solution vector (x_r, x_d) slightly away from the manifold defined by Eq. (16). When $b_r > 0$ and $b_d < 0$, the branch of equilibria that is selected, for γ_r, γ_d near the singular point, corresponds to $x_r > 0$ and $x_d < 0$. Overall, this analysis means the two parties can develop polarized and asymmetric ideological positions in the direction of their respective small biases. Whether or not the polarization occurs depends on the product of γ_r and γ_d whereas the degree of asymmetry in the ideological positions is determined by their ratio. A much more significant level of difference between γ_r and γ_d is necessary in order to capture a similar level of asymmetry to the $\gamma_r = \gamma_d = 0$ case with α_d slightly below its critical value of 1 and α_r slightly above it.

2. $\gamma_r < 0$ and $\gamma_d < 0$ (bipartisan cooperation): at the singularity the kernel of J is

$$\text{span}\{(\sqrt{\gamma_r}, \sqrt{\gamma_d})\}. \quad [18]$$

Following the same analysis as was carried out in the positive γ_i case, we find that the model undergoes a pitchfork bifurcation at $\sqrt{\gamma_r \gamma_d} = 1$, and whenever $\sqrt{\gamma_r \gamma_d} > 1$, equilibrium solutions appear near the manifold defined by

$$x_d = \sqrt{\frac{\gamma_d}{\gamma_r}} x_r. \quad [19]$$

This analysis predicts that when both parties exhibit bipartisan cooperation, they can overcome the differences in their intrinsic biases b_r, b_d and develop an ideological position that leans in the same direction. In order for this to happen, one (or both) parties must put in sufficient effort to cooperate. This means that it is possible for a party which is putting in a lot of effort to be more cooperative with the other party to entirely switch its ideological leaning.

C. Dynamic parameters. Analysis performed in Section B assumes that parameters of the model Eq. (7),Eq. (8) are static. More generally, equilibria of the static-parameter model inform the behavior of the state trajectories when the parameters become dynamic. For example when there is a slow drift in the bifurcation parameter, geometric singular perturbation theory (4) predicts that trajectories of the system state will remain close to a normally hyperbolic manifold defined by the nullclines of the static-parameter problem. Therefore we can use the analysis in Section B to gain intuition about the dynamic-parameter simulation studies in this paper. In particular we can deduce from this analysis the degree of asymmetry in the trajectories of the ideological positions of the two parties, as well as the parameter values that define the critical point beyond which ideological positions will rapidly polarize.

S2. Numerical experiments

All simulations of the model are run using Python. In each of the simulations in the main text and in this section where policy mood input is used, we first run a short simulation, with all parameters held constant, that starts before the initial time t_0. This allows the dynamics to settle to an equilibrium at the initial time t_0 and prevents initial inadvertent transients. The initial values $x_r(1959) = 0.3$, $x_d(1959) = -0.3$ are chosen to resemble those in the DW-NOMINATE scores.

A. Sensitivity study: asymmetry in γ_r, γ_d growth rate. Figure 4(B) of the main paper shows how increasing reflexive partisanship levels, γ_r and γ_d, yields polarization. However, that polarization exhibits much less asymmetry between the elite ideological positions, when γ_r increases at twice the rate as γ_d, as compared to the asymmetry between elite ideological positions in the case α_r increases at twice the rate as α_d, as shown in Figure 4(A). Here we examine even greater differences between γ_r and γ_d.

As in Figure 4(B), let $x_r(1959) = 0.3$, $x_d(1959) = -0.3$, $\tau_x = 1$ year, $b_r = 0.05$, $b_d = -0.05$, and $\alpha_r = \alpha_d = 0$. Let γ_r and γ_d increase over time (in years) at rate r_r and r_d, respectively:

$$\gamma_r(t) = r_r(t - 1959) + 0.3, \quad \gamma_d(t) = r_d(t - 1959) + 0.3.$$

In Figure 4(B), $r_d = 0.01$ and $r_r = 0.02$. Figure S2(A) shows that there is not much more asymmetry between elite ideological positions, even when $r_d = 0.01$ and $r_r = 0.04$, i.e., when γ_r increase four times as quickly as γ_d. In Figures S2(B) and (C), $r_d = 0$, i.e., γ_d is kept constant at $\gamma_d = 0.3$, and $r_r = 0.04$ and $r_r = 0.06$, respectively. These extreme cases are sufficient to yield asymmetry between elite ideological positions.

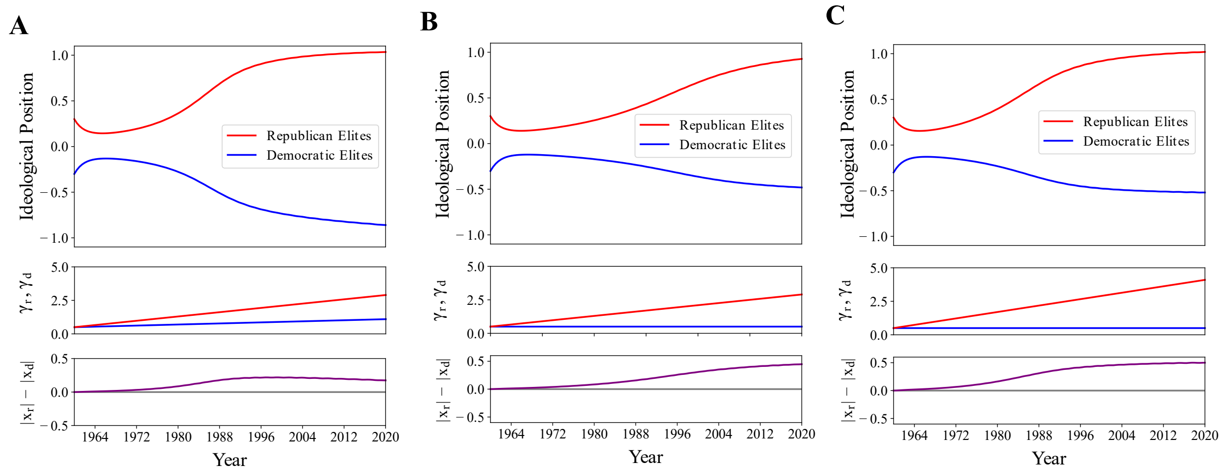


Fig. S2. A. $r_r = 0.04$, $r_d = 0.01$; B. $r_r = 0.04$, $r_d = 0$; C. $r_r = 0.06$, $r_d = 0$.

B. Sensitivity study: time scale τ_x. In all of the time simulations presented in the main text, we let the time scale associated with the evolution of elite ideological position be $\tau_x = 4$ years. In Figure S3, we run the same simulation that produced Figure 6 in the text, but with 1-year, 2-year, 6-year, and 8-year time scales. The runs with 2-year and 6-year time scales perform in a similar fashion to the 4-year scale. We use the 4-year scale, however, because it accords with the length of a presidential term in office and researchers have noted that policy mood inflection points often coincide with party regime change.

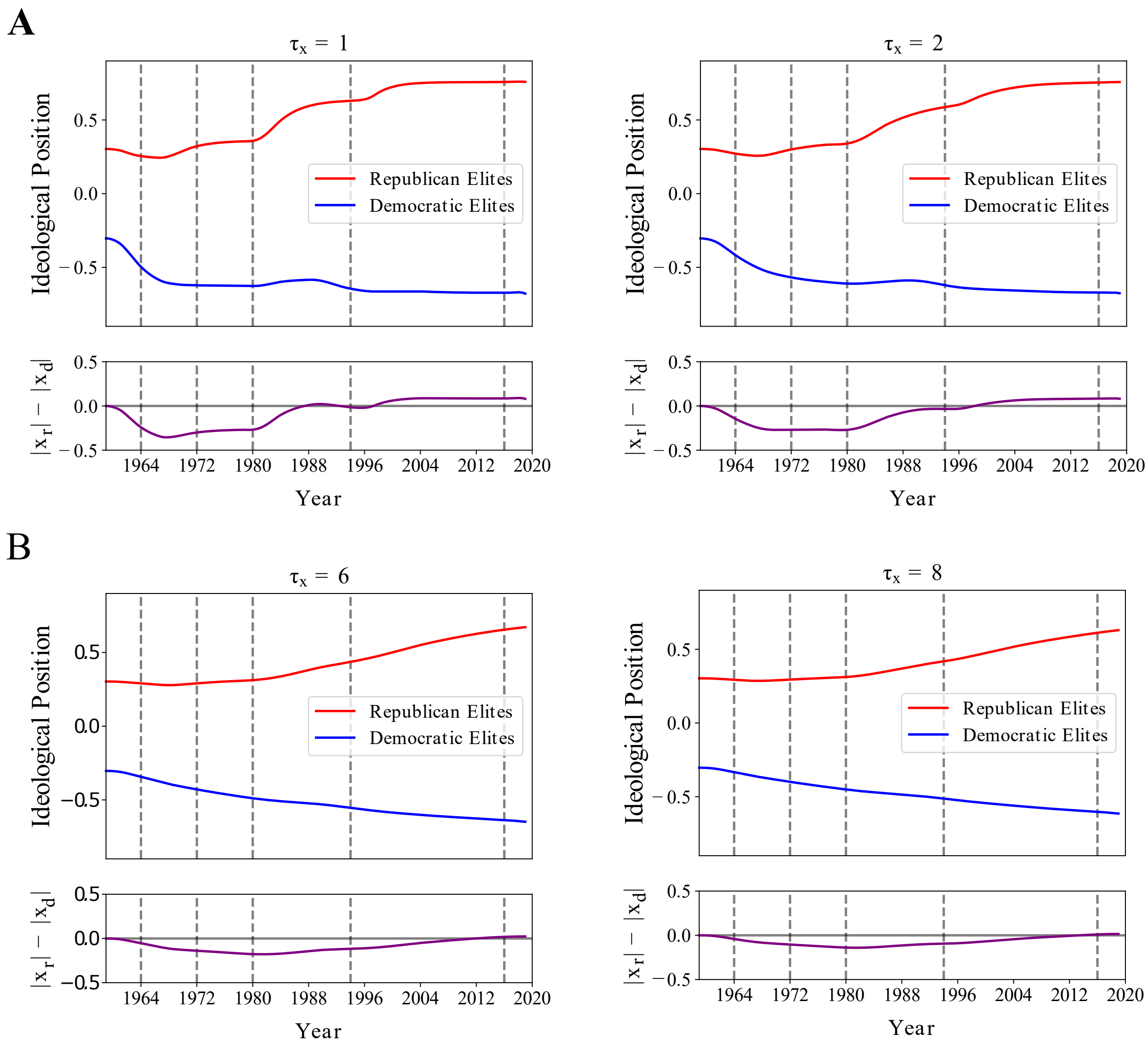


Fig. S3. Same simulation as Figure 6 in main text, with faster (A) and slower (B) time scales.

C. Asymmetric thresholds: Republicans less responsive to public mood swings to the left. In Figure 6 in the main text, we assume Republican and Democratic elites have the same thresholds for responding to swings in PM. Yet, we know elites can have a biased perception of public opinion. While elites of both parties tend to overestimate support for conservative policies, Republicans are particularly prone to making this mistake (5, 6). Research also suggests that the Republican Party is more ideological than the Democratic Party (7) and that Republican members of Congress are more tethered to the national party than their Democratic counterparts (8). This suggests Republicans may be less responsive to leftward swings in policy mood than Democrats to rightward swings. Thus, in Figure S4, we use $U_r = U_d = L_d = 0.45$ as in Figure 6 but set $L_r = 0.55$, increasing the threshold L_r above which the Republican elites will respond to the moderating effect of liberal swings. The simulation results suggest Republicans would have crossed the critical threshold for "run away" polarization much earlier (closer to 1980 than 1990) and, more importantly, that the asymmetric nature of elite polarization overall would have been much worse. Thus, concerns about biased perceptions of public opinion may be over-blown.

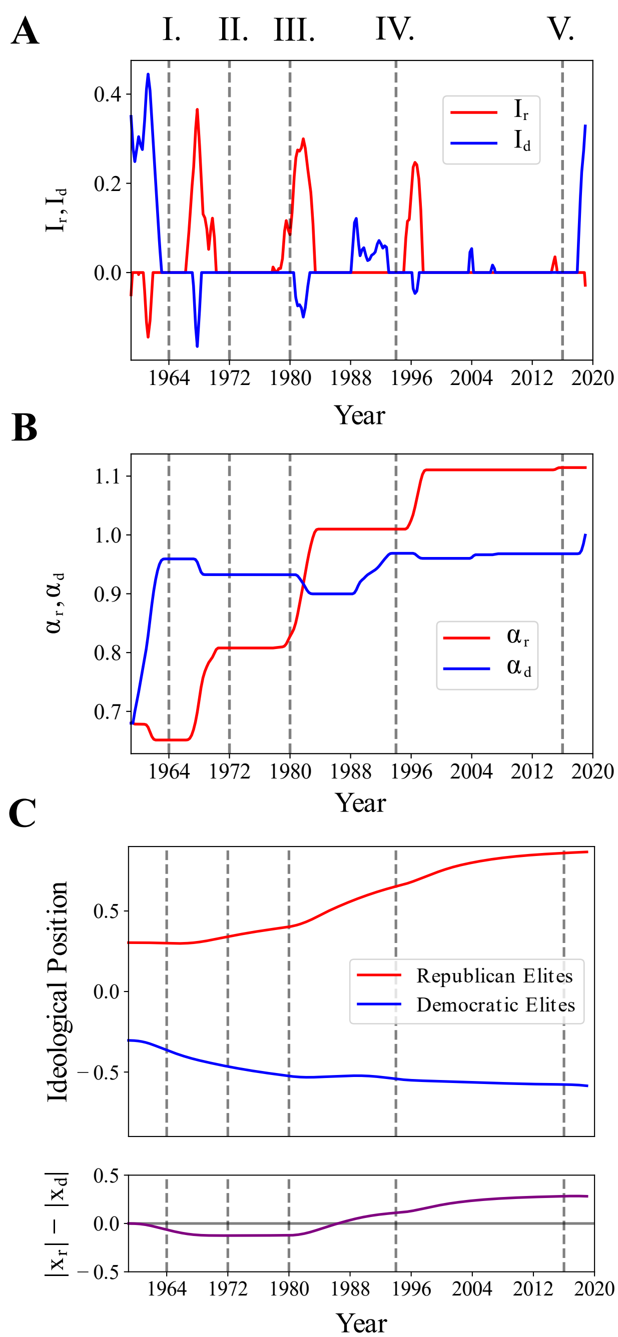


Fig. S4. Asymmetric thresholds in response to policy mood. Model parameters: $U_r = U_d = L_d = 0.45$, $L_r = 0.55$, $I_0 = 0.1$, $\alpha_r(t_0) = \alpha_d(t_0) = 0.68$, $k_\alpha = 0.25$, $x_r(t_0) = 0.3$, $x_d(t_0) = -0.3$, $b_r = 0.1$, $b_d = -0.1$, $\tau_x = 4$, and $t_0 = 1959$.

D. Policy mood drives γ_r, γ_d. Here we test the hypothesis that reflexive polarization levels γ_r and γ_d are driven by policy mood, and not self-reinforcement levels α_r and α_d. All conditions in Figure S5(A) and (B) are the same as in Figures 6 and 7(D), respectively, except with the roles of the γ_r and γ_d swapped with α_r and α_d, respectively. These simulations serve to rule out this hypothesis since there is virtually no asymmetry exhibited in the polarization.

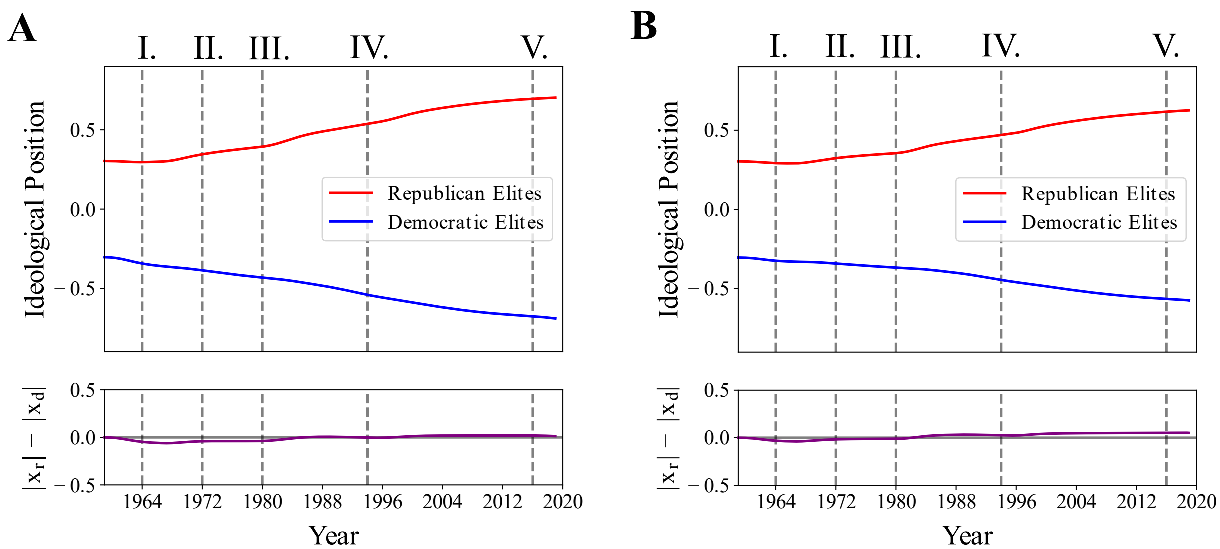


Fig. S5. A. Policy mood drives γ_r, γ_d with symmetric thresholds as in main text Figs. 5 and 6, $\alpha_r = \alpha_d = 0$; B. Policy mood drives γ_r, γ_d with time-varying U_d as in main text Figure 7, $\alpha_r = \alpha_d = 0$.

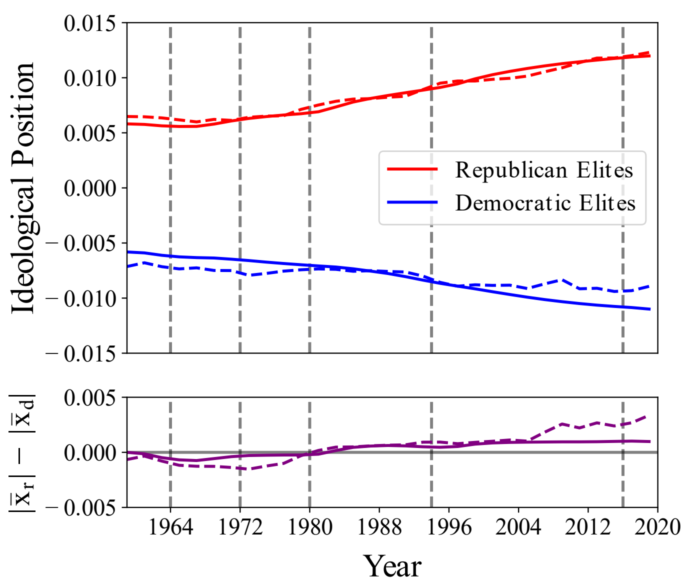


Fig. S6. A. Same as Figure S5(B), with the areas between the curves normalized to 1, and similarly normalized DW-NOMINATE scores (dashed lines) superimposed. The bottom panel shows that the simulated difference in ideological position (solid purple) grossly underpredicts the asymmetry in polarization in the DW-NOMINATE data (dashed purple). See Section S4 for definition of the normalization.

E. Bipartisan cooperation. Here we test whether the break down of bipartisan norms, which began in the 1970s, can account for the rise of asymmetric polarization. We do this by using the same conditions as in Figure 7 but by also applying a small negative γ where $\gamma_r = \gamma_d = -0.1$. As Figure S7 suggests, even if norms of bipartisanship had endured, it would not have prevented the rise of asymmetric polarization.

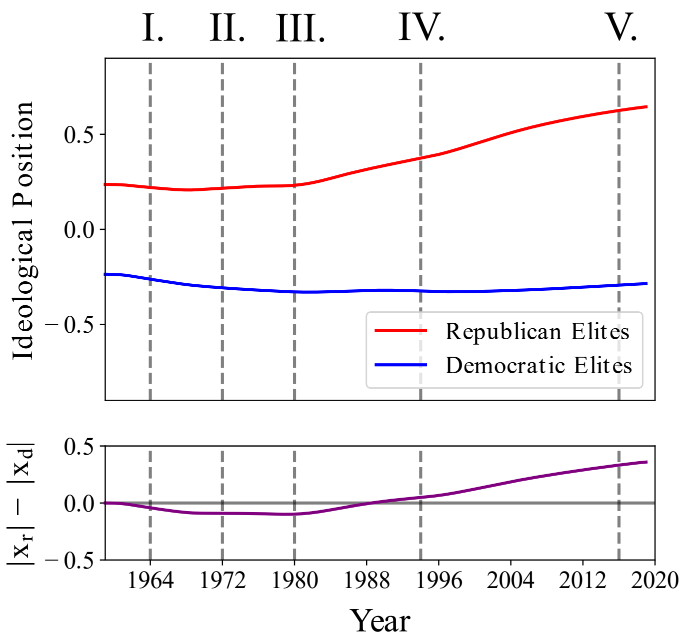


Fig. S7. Same conditions as main text Figure 7 except $\gamma_r = \gamma_d = -0.1$, $x_r(t_0) = 0.24$, $x_d(t_0) = -0.24$.

F. Parameter sweep: finding parameter values that minimize mean square error between model results and data for Hypotheses A, B, and C. To rigorously analyze and compare the simulated trajectories to the DW-NOMINATE score data, we introduce the following normalization:

$$\bar{x}_r = \frac{x_r}{\hat{x}}, \quad \bar{x}_d = \frac{x_d}{\hat{x}}, \tag{20}$$

where the normalization factor $\hat{x}$ is the area between the curves $x_r(t)$ and $x_d(t)$, computed using the trapezoidal rule over the time period of interest, which is from 1979 to 2019 in this section. We then run the model over the range of values defined in Table S1 of four parameters over this time period. We begin the simulations in 1979 because the DW-NOMINATE scores are close in magnitude in 1979, which makes more natural a comparison between the data and the simulated results that start from symmetric initial conditions.

In this section we consider the three separate hypotheses on elite response mechanism:

Hypothesis A. Policy mood drives party self-reinforcement levels α_r and α_d, with $\gamma_r = \gamma_d = 0$ and $b_r = -b_d = 0.1$;

Hypothesis B. Policy mood drives reflexive partisanship levels γ_r and γ_d with $\alpha_r = \alpha_d = 0$ and $b_r = -b_d = 0.1$;

Hypothesis C. Policy mood drives additive inputs b_r and b_d, with $\alpha_r = \alpha_d = \gamma_r = \gamma_d = 0$.

For each of the three hypotheses we simulate the model dynamics over a range of values for each of four parameters: U, L, p_0, and k, where $U_r = U_d = U$, $L_r = L_d = L$, $k_r = k_d = k$, $\alpha_r(1979) = \alpha_d(1979) = p_0$ for Hyp. A, $\gamma_r(1979) = \gamma_d(1979) = p_0$ for Hyp. B , and $b_r(1979) = b_d(1979) = p_0$ for Hyp. C. The range of values simulated are described in Table S1 and the results of the simulation are presented in Table S2.

The first measure of comparison we consider is the mean square error (MSE) between the normalized simulated trajectories ($\bar{x}_r$ and $\bar{x}_d$) and the normalized DW-NOMINATE scores ($\bar{x}_r^{DW}$ and $\bar{x}_d^{DW}$), defined as

$$\text{MSE} = \frac{1}{N}\sum_{n=1}^{N}\left((\bar{x}_r(t_n) - \bar{x}_r^{DW}(t_n))^2 + (\bar{x}_d(t_n) - \bar{x}_d^{DW}(t_n))^2\right), \tag{21}$$

where N is the number of time instances at which the comparison is made and t_n is the set of indexed time instances, where $n = 1, \ldots, N$. The MSE measures how well the normalized trajectories of the simulation agree with the normalized trajectories of the data, such that the smaller the MSE the better the agreement. In Table S2 we present the results for each hypothesis corresponding to the simulation yielding the lowest MSE over all combinations of parameters as listed in Table S1. The results include the corresponding parameter values and two measures of the asymmetry in the simulated polarization. The first is

	Hyp. A (α)	Hyp. B (γ)	Hyp. C (b)
U_{min}	0.1	0.1	0.1
U_{max}	0.7	0.7	0.7
No. values U	10	10	10
L_{min}	0.1	0.1	0.1
L_{max}	0.7	0.7	0.7
No. values L	10	10	10
k_{min}	0.1	0.1	0.1
k_{max}	0.5	0.5	0.5
No. values k	8	8	8
$p_{0,min}$	0.6	0.6	0.1
$p_{0,max}$	0.8	0.8	0.8
No. values p_0	5	5	5

Table S1. Minimum and maximum values and number of values used for each of the four parameters: U, L, k, p_0 in the set of analyses performed for Hypotheses A, B, and C. A total of 4000 simulations were run for each hypothesis, where each simulation used a different combination of parameter values. We present in Table S2 the results of the simulation with the best results, defined as the lowest mean square error (MSE) of modeled ideological positions with respect to the DW-NOMINATE data.

the polarization asymmetry index (PAI), which we define as the ratio of difference between magnitudes of the normalized trajectories to difference between magnitudes of the normalized DW-NOMINATE scores at the end of the simulation:

$$\text{PAI} = \left(\frac{|\bar{x}_r| - |\bar{x}_d|}{|\bar{x}_r^{DW}| - |\bar{x}_d^{DW}|} \right) (2019). \qquad [22]$$

If PAI > 1, the simulation overpredicts the asymmetry polarization in 2019 and if PAI < 1, the simulation underpredicts it. The second measure is the mean square error (MSEdif) between the differences in magnitude of the ideological positions of the parties, defined as

$$\text{MSEdif} = \frac{1}{N} \sum_{n=1}^{N} \left(\big(|\bar{x}_r(t_n)| - |\bar{x}_d(t_n)|\big) - \big(|\bar{x}_r^{DW}(t_n)| - |\bar{x}_d^{DW}(t_n)|\big) \right)^2. \qquad [23]$$

The MSEdif measures how well the simulated asymmetry in polarization resembles the asymmetry in polarization in the data over time, such that the lower the MSEdif the better the resemblance.

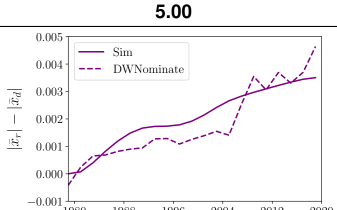
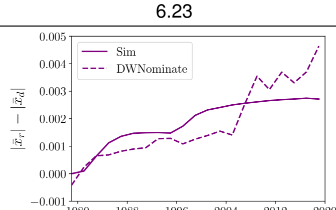
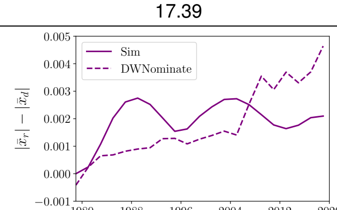
	Hyp. A (α)	Hyp. B (γ)	Hyp. C (b)
MSE ($\times 10^{-7}$)	**4.37**	5.27	9.91
U	0.57	0.57	0.30
L	0.37	0.50	0.17
p_0	0.65	0.60	0.80
k	0.10	0.44	0.10
PAI	**0.80**	0.57	0.45
MSEdif ($\times 10^{-7}$)	**5.00**	6.23	17.39
$\lvert\bar{x}_r\rvert - \lvert\bar{x}_d\rvert$			

Table S2. Parameters and results from the simulation with lowest MSE over the complete set of 4000 simulations, with parameters ranging as described in Table S1, for each of the three hypotheses. $N = 21$ for MSE and MSEdif calculations. Hyp. A performs best over all measures (see bolded values). In particular, we note how well Hyp. A captures the asymmetry in the polarization in the data as illustrated in the plot and in the PAI and MSEdif values. The best simulation for Hyp. B does not do as well with respect to the MSE; however, what is most striking is that even this best run for Hyp. B still underpredicts the asymmetry in polarization in the data. The best simulation for Hyp. C underperforms with respect to MSE and with respect to the asymmetry in polarization. As can be seen in the plot for Hyp. C, $|\bar{x}_r| - |\bar{x}_d|$ tracks the asymmetry in the PM, as predicted by the theory, and does not resemble the asymmetry in the DW-NOMINATE scores. The parameters U, L, and p_0 are quite similar for Hyp. A and Hyp. B. The value $k = 0.1$ for Hyp. A implies relatively slow α_r and α_d dynamics and with it flexibility to match the data. The value $k = 0.44$ for Hyp. B implies relatively fast γ_r and γ_d dynamics and inflexibility to match the data (and notably the asymmetry) since these dynamics saturate early.

S3. Filtering of Policy Mood Data

PM data (Figure S8(A)) was filtered through a first-order high-pass filter with transfer function

$$H_{HP}(s) = \frac{s}{\tau_{HP}s + 1}$$

and a first-order low-pass filter with transfer function

$$H_{LP}(s) = \frac{1}{\tau_{LP}s + 1}.$$

Filters with polynomial transfer functions are realizable as simple differential equations and thus they provide simple interpretable models of generic dynamical responses to inputs.

The high-pass filter models the sensitivity of elite response only to variations of PM, i.e., elites are not sensitive to low-frequency components (the zero-frequency average, in particular) of PM variations. The filter time constant τ_{HP} roughly determines the threshold frequency below which PM variations are filtered-out. In our model, $\tau_{HP} = 1.0$ year.

The low-pass filter models memory of elite response to PM variations. Such a first-order filter "forgets" about the input past history exponentially with time-constant τ_{LP} (expressed in years in our model). Input events more recent than τ_{LP} have relatively large influence on the filter response. Input events more remote than τ_{LP} have relatively small influence on the filter response. In our model, $\tau_{LP} = 10.0$ years.

The filtered PM data is shown in Figure S8(B) and also in Figure 5(A) in the main paper.

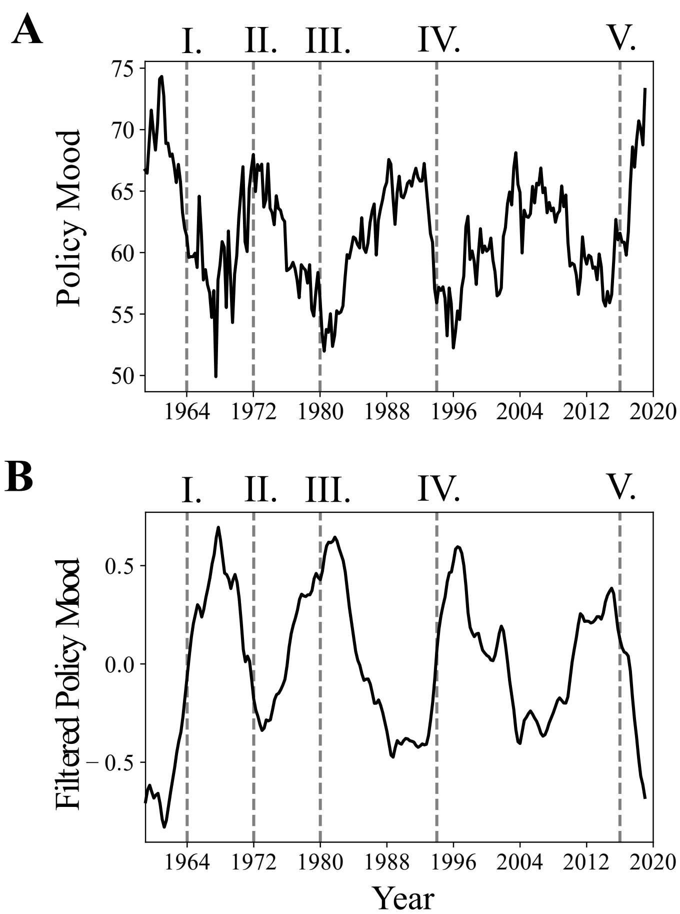


Fig. S8. A. Policy mood data. B. Filtered policy mood data, multiplied by -1 to match the left-right sign convention used throughout this work.

S4. Agent-based model

Consider a group of N agents split into non-overlapping Republican and Democratic elite groups.* We associate to each agent a unique index between 1 and N. Let $\mathcal{R} \subset \{1, \ldots, N\}$ be the set of indices of the agents in the Republican group and $\mathcal{D} \subset \{1, \ldots, N\}$ be the set of indices of the agents in the Democratic group. Then $\mathcal{R} \cap \mathcal{D} = \emptyset$ and $\mathcal{R} \cup \mathcal{D} = \{1, \ldots, N\}$. For simplicity, we assume that these index sets do not change over time.

*Please contact AF (afranci@ciencias.unam.mx) for the Julia code used to run the agent-based simulations.

Let $x_{i_r}(t)$ (resp. $x_{i_d}(t)$) represent the scalar ideological position of agent i_r (resp. i_d) in the Republican (resp. Democratic) elite. The *ideological self-reinforcement level of agent* i is modeled by $\alpha_i(t) \geq 0$. The *level of Republican self-reinforcement of agent* i_r *with respect to the ideological position of agent* j_r is modeled by $\gamma^{rr}_{i_r j_r}(t) \geq 0$, $i_r, j_r \in \mathcal{R}$, $i_r \neq j_r$. The *level of Democratic self-reinforcement of agent* i_d *with respect to the ideological position of agent* j_d is modeled by $\gamma^{dd}_{i_d j_d}(t) \geq 0$, $i_d, j_d \in \mathcal{D}$, $i_d \neq j_d$. The *level of Republican reflexive partisanship of agent* i_r *with respect to the ideological position of agent* j_d is modeled by $\gamma^{rd}_{i_r j_d}(t) \geq 0$, $i_r \in \mathcal{R}$, $j_d \in \mathcal{D}$. The *level of Democratic reflexive partisanship of agent* i_d *with respect to the ideological position of agent* j_r is modeled by $\gamma^{dr}_{i_d j_r}(t) \geq 0$, $i_d \in \mathcal{D}$, $j_r \in \mathcal{R}$. Each agent i possesses an *ideological bias* $b_i(t)$, which is conservative for Republican agents, i.e., $b_{i_r}(t) \geq 0$ for $i_r \in \mathcal{R}$, and liberal for Democratic agents, i.e., $b_{i_d}(t) \leq 0$ for $i_d \in \mathcal{D}$.

The agent-based model equations are

$$\tau_x \frac{dx_{i_r}}{dt} = S\left(\alpha_{i_r} x_{i_r} + \sum_{\substack{j_r \in \mathcal{R} \\ j_r \neq i_r}} \gamma^{rr}_{i_r j_r} x_{j_r} - \sum_{j_d \in \mathcal{D}} \gamma^{rd}_{i_r j_d} x_{j_d}\right) - x_{i_r} + b_{i_r}, \quad i_r \in \mathcal{R}\,, \tag{24a}$$

$$\tau_x \frac{dx_{i_d}}{dt} = S\left(\alpha_{i_d} x_{i_d} + \sum_{\substack{j_d \in \mathcal{D} \\ j_d \neq i_d}} \gamma^{dd}_{i_d j_r} x_{j_d} - \sum_{j_r \in \mathcal{R}} \gamma^{dr}_{i_d j_r} x_{j_r}\right) - x_{i_d} + b_{i_d}. \quad i_d \in \mathcal{D}\,. \tag{24b}$$

Note that using (1, Theorem III.5) this agent-based model can be shown to be formally equivalent to the two-party model Eq. (7),Eq. (8) where each node represents a within-group average opinion. In all simulations $N = 100$, with $\mathcal{R} = \{1, \ldots, 50\}$ and $\mathcal{D} = \{51, \ldots, 100\}$.

In all simulations we let parameter values for individuals vary from the average of the individual's group by a deviation drawn from a Normal distribution. First, we examine the *symmetric parameter case* in which the average parameter magnitudes for the Republicans are the same as for the Democrats. Second, we examine the *asymmetric parameter case* in which average parameter magnitudes for the Republicans are not the same as for the Democrats. In both cases α_i, ideological self-reinforcement for agent i, varies with respect to the same average across Republicans and Democrats.

A. Agent-based simulations with symmetric parameters between Republicans and Democrats. In the symmetric setting, we let the average magnitude of Republican self-reinforcement level, Republican reflexive partisanship level, and Republican bias, be equal to the average magnitude of Democratic self-reinforcement level, Democratic reflexive partisanship level, and Democratic bias, respectively. More precisely, in the simulations, we let

- α_i, $i = 1, \ldots, N$, are kept constant and drawn from a Normal distribution with mean $\bar{\alpha}$ and variance $\Delta\alpha$;
- for $i_r, j_r \in \mathcal{R}$, $i_r \neq j_r$, $\gamma^{rr}_{i_r, j_r}(t) = \bar{\gamma}_{self}(t) + \Delta\gamma^{rr}_{i_r, j_r}$, where $\Delta\gamma^{rr}_{i_r, j_r}$ is drawn from a Normal distribution with zero mean and variance $\Delta\gamma_{self}$;
- for $i_d, j_d \in \mathcal{D}$, $i_d \neq j_d$, $\gamma^{dd}_{i_d, j_d}(t) = \bar{\gamma}_{self}(t) + \Delta\gamma^{dd}_{i_d, j_d}$, where $\Delta\gamma^{dd}_{i_d, j_d}$ is drawn from a Normal distribution with zero mean and variance $\Delta\gamma_{self}$;
- for $i_r \in \mathcal{R}$, $j_d \in \mathcal{D}$, $\gamma^{rd}_{i_r, j_d}(t) = \bar{\gamma}_{reflex}(t) + \Delta\gamma^{rd}_{i_r, j_d}$, where $\Delta\gamma^{rd}_{i_r, j_d}$ is drawn from a Normal distribution with zero mean and variance $\Delta\gamma_{reflex}$;
- for $i_d \in \mathcal{D}$, $j_r \in \mathcal{R}$, $\gamma^{dr}_{i_d, j_r}(t) = \bar{\gamma}_{reflex}(t) + \Delta\gamma^{dr}_{i_r, j_r}$, where $\Delta\gamma^{dr}_{i_d, j_r}$ is drawn from a Normal distribution with zero mean and variance $\Delta\gamma_{reflex}$;
- b_{i_r}, $i_r \in \mathcal{R}$, are kept constant and drawn from a Normal distribution with mean $\bar{b}$ and variance $\Delta\bar{b}$;
- b_{i_d}, $i_d \in \mathcal{D}$, are kept constant and drawn from a Normal distribution with mean $-\bar{b}$ and variance $\Delta\bar{b}$.

A.1. Symmetric polarization by symmetric increase in ideological self-reinforcement. Figure S9. Simulation parameters:

- $\bar{\alpha} = 0.05$, $\Delta\alpha = 0.0025$;
- $\bar{\gamma}_{self}(t) = 0.1/50 + (1.1 - 0.1)/50 \cdot t/T$, $\Delta\gamma_{self} = 0.05$;
- $\bar{\gamma}_{reflex}(t) = 0.0$, $\Delta\gamma_{reflex} = 0.0$;
- $\bar{b} = 0.8$, $\Delta\bar{b} = 0.8$;

A.2. Symmetric polarization by symmetric increase in reflexive partisanship. Figure S10. Simulation parameters:

- $\bar{\alpha} = 0.05$, $\Delta\alpha = 0.0025$;
- $\bar{\gamma}_{reflex}(t) = -0.1/50 - (1.1 - 0.1)/50 \cdot t/T$, $\Delta\gamma_{reflex} = 0.05$;
- $\bar{\gamma}_{self}(t) = 0.0$, $\Delta\gamma_{self} = 0.0$;
- $\bar{b} = 0.8$, $\Delta\bar{b} = 0.8$;

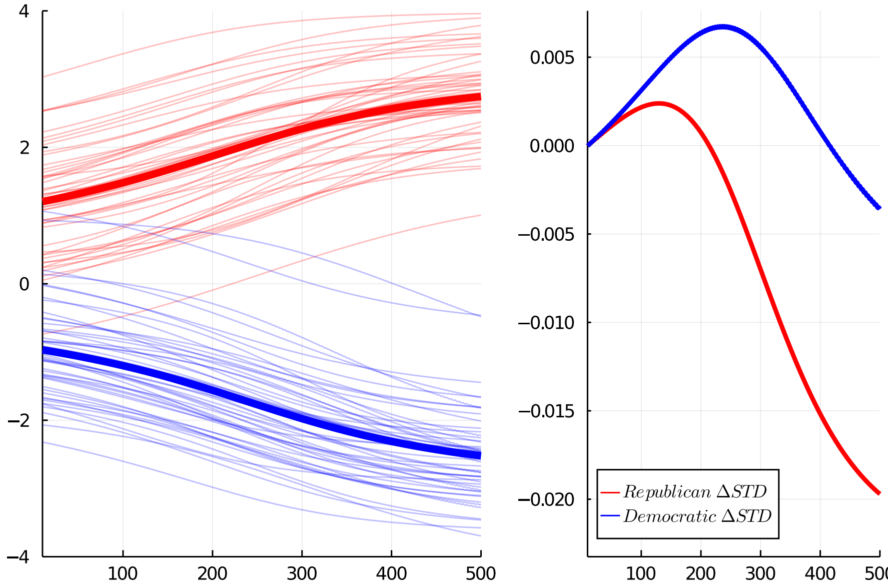


Fig. S9. Left. Thin lines are the evolution of ideological position of each Republican elite (red) and each Democratic elite (blue) as a function of time. Bold lines are the average Republican elite ideological position (red) and average Democratic elite ideological position (blue) as a function of time. Right: Evolution over time of the standard deviation of the ideological positions of Republican elites (red) and Democratic elites (blue) as compared to the standard deviations at the initial time.

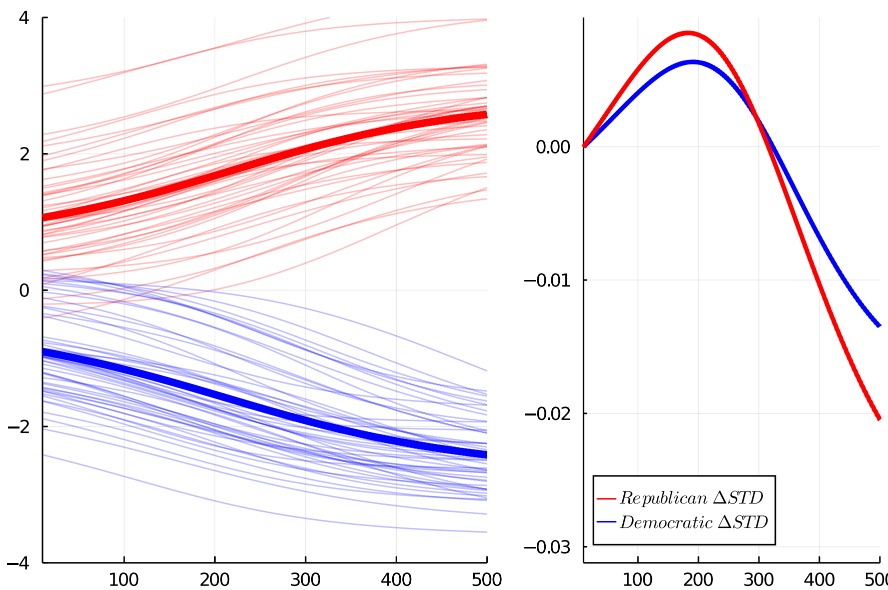


Fig. S10. Left. Thin lines are the evolution of ideological position of each Republican elite (red) and each Democratic elite (blue) as a function of time. Bold lines are the average Republican elite ideological position (red) and average Democratic elite ideological position (blue) as a function of time. Right: Evolution over time of the standard deviation of the ideological positions of Republican elites (red) and Democratic elites (blue) as compared to the standard deviations at the initial time.

B. Agent-based simulations with asymmetric parameters between Republicans and Democrats. In the asymmetric setting, we let the average magnitude of Republican self-reinforcement level and Republican reflexive partisanship level be different from the average magnitude of Democratic self-reinforcement level and Democratic reflexive partisanship level, respectively. We let the average magnitude of the Republican bias and the Democratic bias be the same. More precisely, in the simulations, we let

- α_i, $i = 1, \ldots, N$, are kept constant and drawn from a Normal distribution with mean $\bar{\alpha}$ and variance $\Delta\alpha$;
- for $i_r, j_r \in \mathcal{R}$, $i_r \neq j_r$, $\gamma^{rr}_{i_r,j_r}(t) = \bar{\gamma}^r_{self}(t) + \Delta\gamma^{rr}_{i_r,j_r}$, where $\Delta\gamma^{rr}_{i_r,j_r}$ is drawn from a Normal distribution with zero mean and variance $\Delta\gamma^r_{self}$;
- for $i_d, j_d \in \mathcal{D}$, $i_d \neq j_d$, $\gamma^{dd}_{i_d,j_d}(t) = \bar{\gamma}^d_{self}(t) + \Delta\gamma^{dd}_{i_d,j_d}$, where $\Delta\gamma^{dd}_{i_d,j_d}$ is drawn from a Normal distribution with zero mean and variance $\Delta\gamma^d_{self}$;
- for $i_r \in \mathcal{R}$, $j_d \in \mathcal{D}$, $\gamma^{rd}_{i_r,j_d}(t) = \bar{\gamma}^r_{reflex}(t) + \Delta\gamma^{rd}_{i_r,j_d}$, where $\Delta\gamma^{rd}_{i_r,j_d}$ is drawn from a Normal distribution with zero mean

and variance $\Delta\gamma^r_{reflex}$;

- for $i_d \in \mathcal{D}$, $j_r \in \mathcal{R}$, $\gamma^{dr}_{i_d,j_r}(t) = \bar{\gamma}^d_{reflex}(t) + \Delta\gamma^{dr}_{i_r,j_r}$, where $\Delta\gamma^{dr}_{i_d,j_r}$ is drawn from a Normal distribution with zero mean and variance $\Delta\gamma^d_{reflex}$;
- b_{i_r}, $i_r \in \mathcal{R}$, are kept constant and drawn from a Normal distribution with mean $\bar{b}$ and variance $\Delta\bar{b}$;
- b_{i_d}, $i_d \in \mathcal{D}$, are kept constant and drawn from a Normal distribution with mean $-\bar{b}$ and variance $\Delta\bar{b}$.

B.1. Asymmetric polarization by asymmetric increase in ideological self-reinforcement. Figure S11. Simulation parameters:

- $\bar{\alpha} = 0.05$, $\Delta\alpha = 0.0025$;
- $\bar{\gamma}^r_{self}(t) = 0.1/50 + (1.1 - 0.1)/50 \cdot t/T$, $\Delta\gamma^r_{self} = 0.05$;
- $\bar{\gamma}^d_{self}(t) = 0.1/50 + (0.7 - 0.1)/50 \cdot t/T$, $\Delta\gamma^d_{self} = 0.05$;
- $\bar{\gamma}^r_{reflex}(t) = 0.0$, $\Delta\gamma^r_{reflex} = 0.0$;
- $\bar{\gamma}^d_{reflex}(t) = 0.0$, $\Delta\gamma^d_{reflex} = 0.0$;
- $\bar{b} = 0.8$, $\Delta\bar{b} = 0.8$;

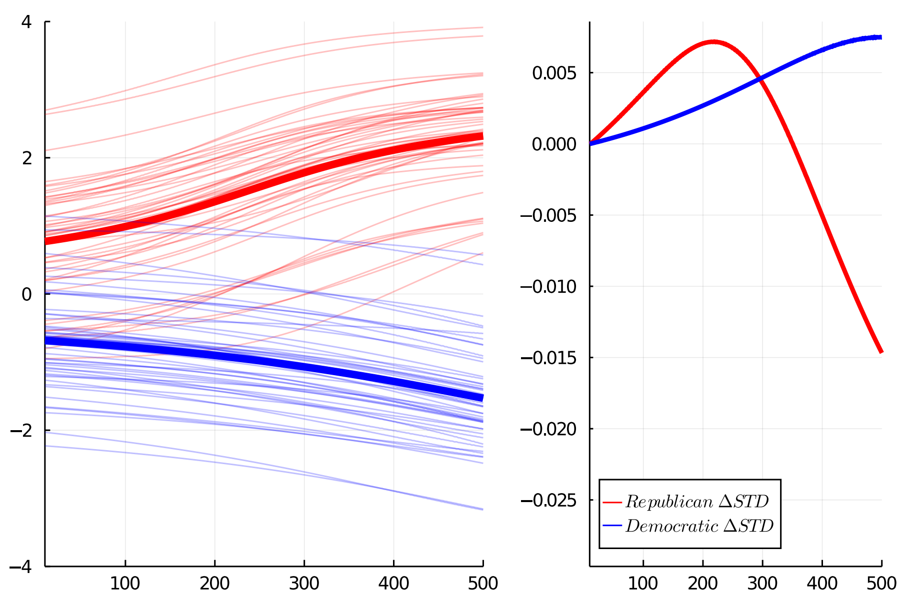


Fig. S11. Left. Thin lines are the evolution of ideological position of each Republican elite (red) and each Democratic elite (blue) as a function of time. Bold lines are the average Republican elite ideological position (red) and average Democratic elite ideological position (blue) as a function of time. Right: Evolution over time of the standard deviation of the ideological positions of Republican elites (red) and Democratic elites (blue) as compared to the standard deviations at the initial time.

B.2. Symmetric polarization by asymmetric increase in reflexive partisanship. Figure S12. Simulation parameters:

- $\bar{\alpha} = 0.05$, $\Delta\alpha = 0.0025$;
- $\bar{\gamma}^r_{reflex}(t) = 0.1/50 + (1.2 - 0.1)/50 \cdot t/T$, $\Delta\gamma^r_{reflex} = 0.05$;
- $\bar{\gamma}^d_{reflex}(t) = 0.1/50 + (0.7 - 0.1)/50 \cdot t/T$, $\Delta\gamma^d_{reflex} = 0.05$;
- $\bar{\gamma}^r_{self}(t) = 0.0$, $\Delta\gamma^r_{self} = 0.0$;
- $\bar{\gamma}^d_{self}(t) = 0.0$, $\Delta\gamma^d_{self} = 0.0$;
- $\bar{b} = 0.8$, $\Delta\bar{b} = 0.8$;

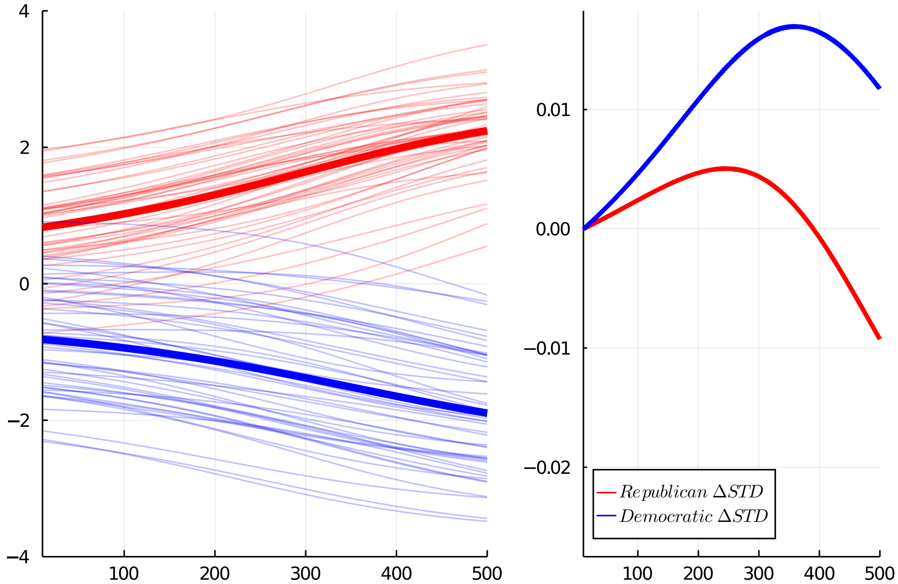


Fig. S12. Left. Thin lines are the evolution of ideological position of each Republican elite (red) and each Democratic elite (blue) as a function of time. Bold lines are the average Republican elite ideological position (red) and average Democratic elite ideological position (blue) as a function of time. Right: Evolution over time of the standard deviation of the ideological positions of Republican elites (red) and Democratic elites (blue) as compared to the standard deviations at the initial time.

References

1. A Bizyaeva, A Franci, NE Leonard, Nonlinear opinion dynamics with tunable sensitivity. *arXiv:2009.04332 [math.OC]* (2020).
2. A Franci, A Bizyaeva, S Park, NE Leonard, Analysis and control of agreement and disagreement opinion cascades. *Swarm Intell.* **15**, 47–82 (2021).
3. M Golubitsky, DG Schaeffer, *Singularities and Groups in Bifurcation Theory*, Applied Mathematical Sciences. (Springer-Verlag, New York, NY) Vol. 51, (1985).
4. N Fenichel, Geometric singular perturbation theory for ordinary differential equations. *J. Differ. Equations* **31**, 53–98 (1979).
5. DE Broockman, C Skovron, Bias in perceptions of public opinion among American political elites. *Am. Polit. Sci. Rev.* **112**, 542–563 (2018).
6. B Highton, Issue accountability in U.S. House elections. *Polit. Behav.* **41**, 349–367 (2019).
7. M Grossmann, DA Hopkins, *Asymmetric Politics: Ideological Republicans and Group Interest Democrats.* (Oxford University Press), (2016).
8. TL Brunell, B Grofman, S Merrill, Components of party polarization in the U.S. House of Representatives. *J. Theor. Polit.* **28**, 598–624 (2016).