Polarization, Antipathy, and Political Activism

Jiabin Wu[*]   Hanzhe Zhang[†]

January 26, 2022

Abstract

We present an evolutionary game theory model in which polarization, antipathy, and political activism are simultaneous consequences of the evolution of individuals' ideologies and their attitudes toward other ideologies. We show that the evolutionary process is likely to result in a vicious path with individuals becoming increasingly extreme and polarized on the ideological spectrum and the society ending up with two politically engaged groups sharing no common grounds and strong hatred against each other.

Keywords: Polarization, Political extremism, Antipathy, Political activism, Value formation, Cultural transmission, Evolutionary game theory

JEL: C73, D70, Z13

1 Introduction

(a) Growing minority holds consistent ideological views

(b) A rising tide of mutual antipathy

(c) Political activism gap: right and left are more likely to vote and donate to campaigns

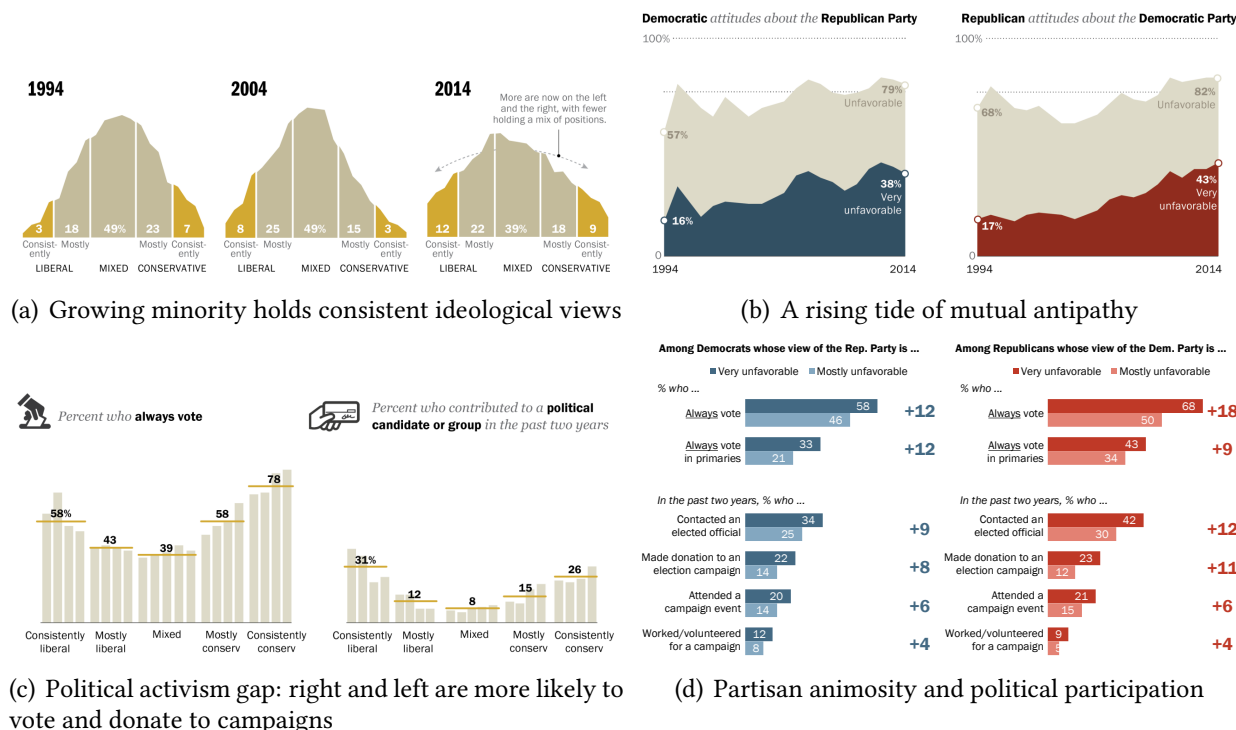(d) Partisan animosity and political participation

Figure 1: Documenting polarization, antipathy, and political activism (Pew Research Center, 2014, 2016, 2017, 2019)

The ongoing American culture war is dividing the country to an unprecedented level. Recent reports issued by Pew Research Center (2014, 2016, 2017, 2019) identify three major trends in the evolution of Americans' ideologies. First, a growing proportion of the population holds extreme ideological views. From 1994 to 2014, the proportion of Americans in the tails of the ideological distribution—consistent liberals or consistent conservatives—has increased from 10% to 23%, and the center has shrunk from 49% to 38% (Figure 1(a)). Second, antipathy has been growing between people holding different ideological views. In 2016, 58% (55%) of Republicans (Democrats) had very unfavorable impressions of the Democratic Party (GOP), while the two percentages were only 21% and 17% in 1994 (Figure 1(b)).[1] The American National Election Studies look at a

---

[1]Similarly, 72% (53%) of consistent conservatives (liberals) had strong negative sentiments toward the Democratic party (GOP) in 2014, while the two percentages were 28% and 23% in 1994. Ideological thinking has become much

longer time span and show that Americans' negative feelings toward opposing parties has been consistently growing for the last 50 years. Third, those with more extreme ideologies and more unfavorable opinions of the opposing party are more politically engaged. 58% (31%) of consistent liberals and 78% (26%) of consistent conservatives always vote (contribute to political candidate or group), compared to 43% (12%), 39% (8%), and 58% (26%) of moderate liberals, centrists, and moderate conservatives, respectively (Figure 1(c)). Compared to those who have mostly unfavorable opinions, Democrats (Republicans) who have very unfavorable opinions of the Republic (Democratic) Party are 12% (18%) more likely to always vote, 12% (9%) more likely always vote in primaries, and more likely to contact an elected official, make donation to a campaign, attend a campaign event, and work/volunteer for a campaign in the past two years (Figure 1(d)).

Aforementioned evidence suggests that Americans have been increasingly polarized on the ideological spectrum, and moreover, their antipathy toward those situated on the opposite end of the spectrum has been growing stronger. At the same time, they are more politically engaged than people close to the center. In this paper, we model the evolution of ideologies and people's attitudes toward other ideologies, and provide an explanation for the increasing polarization, antipathy and political activism in the United States.

We adopt a continuous-time overlapping-generations version of the intergenerational cultural transmission model by Montgomery (2010), which generalizes the seminal model of Bisin and Verdier (2001) from two traits to many traits. Individuals in a large society form and transmit ideologies across generations over time. Each individual becomes a parent during her lifetime and bears one child. A parent naturally has antipathy toward other ideologies.[2] Hence, she has an incentive to transmit her own ideology to her child by exerting an effort. However, she cannot

---

more closely aligned with partisanship: The percentages of Republicans and Democrats holding more extreme values than the median member of the opposing party were 95% and 97% in 2017 and 93% and 94% in 2015, but only 64% and 70% in 1994.

[2]Bisin and Verdier (2001) use the term "cultural intolerance." Antipathy is also related to the notion of oppositional identities discussed by Bisin et al. (2016), Flachea et al. (2017) and Melguizo (2019), by which interacting with individuals with different value systems leads to utility losses.

completely shut her child off from outside influences, but the higher effort she exerts, the smaller the impact her child receives from the society at large. Aggregating the changes in individuals' ideologies, we obtain a dynamic process describing the evolution of the distribution of ideologies in the society.

We can alternatively interpret the intergenerational cultural transmission model as an intra-generational model in which an individual chooses political efforts to (myopically) maximize her own utility in the current period. An individual has an incentive to maintain her current ideology by exerting an effort because of her antipathy toward other ideologies. One can interpret the effort as the time she spends on watching news aligned with her ideology, interacting with like-minded people, and actively participating in political activities.

We show that the key for ideological extremism and polarization to emerge is that an individual's antipathy toward another individual is increasing and convex in the distance between their ideological positions. That is, people are increasingly more intolerant toward others holding more distanced views from them. Because of this, people situated on both ends of the ideological spectrum have a stronger incentive to be more politically engaged to maintain their ideologies than those close to the center, which drives the center to shrink and the tails to grow. When antipathy is instead concave, diversification emerges as the stable prediction of the dynamic.

Evidence suggests that the American society is becoming more polarized, which matches our model's prediction under convex antipathy. The question is, why is antipathy convex? To answer this question, we further generalize the model to allow individuals with the same ideology to have different levels of antipathy.[3] We show that strong (convex) antipathy eliminates weak (concave) antipathy and at the same time drives ideologies to the extremes. Therefore, convex antipathy can be a result of selection through the dynamic process of ideology formation. This also matches the fact that antipathy between the two extremes is growing stronger in the American society.[4]

---

[3]We distinguish polarization and antipathy, in contrast to Iyengar et al. (2012) who define antipathy as an alternative definition of polarization.

[4]Agents with single-peaked concave utility functions over ideologies in the literature of spatial voting starting

4

Previous works have been done to understand the evolution of political ideologies in two-trait cultural transmission models (Bisin and Verdier, 2000; Melindi-Ghidi, 2012; Ticchi et al., 2013; Besley and Persson, 2017, 2019). The evolution of tolerance and the transmission of extremism have also been separately studied by Correani et al. (2010) and Cerqueti et al. (2013), and Epstein (2007), respectively. Compared with these papers, we make two contributions. First, we utilize the techniques from conventional evolutionary game theory to obtain the conditions under which polarization can arise in a multi-trait cultural transmission model. Second, we allow for different degrees of antipathy and show how antipathy coevolves with ideologies and relates to political activism.

Political scientists have been studying political polarization for decades (Poole and Rosenthal, 1984; Hetherington and Weiler, 2009; Sides and Hopkins, 2015); see Layman et al. (2006) for a thorough survey on the literature. The literature has categorized political polarization into party polarization and popular popularization and debated about the causal relationship between the two; see Hunter (1991), Frank (2004), and Fiorina et al. (2005) for book treatments on the topic of popular popularization. One view suggests that the growing ideological divergence on the electorate level causes polarization among the political elites. Another view instead argues that the mass public has a lower level of attention to and knowledge about politics. Hence, the elites are the more likely culprit. A third view supported by Layman et al. (2006) considers political activists as a main spring of party polarization. They argue that the ideologically more extreme individuals are more politically engaged. As the grassroots-level opinion leaders, these individuals can exert considerable influences on party politics. Although our model mainly focuses on the popular level of political polarization, our results on increasing political activism by those holding more extreme views provide support for the third view.

Alternative mathematical models based on Bayesian updating (Dixit and Weibull, 2007) and network theory (Marvela et al., 2011; Bolletta and Pin, 2020) are also proposed to explain political

from Hotelling (1929), Black (1948) and Downs (1957) considers can be thought to have convex antipathy.

polarization. In addition, Gradstein and Justman (2002, 2005) consider the role of education and Michaeli and Wu (2021) consider policies aiming to increase parents' awareness of their children's peer pressure in fighting cultural polarization in dynamic models.

The paper is organized as follows. Section 2 provides the basic dynamic ideology formation model. Section 3 generalizes the model to include antipathy as an evolving trait. Section 4 concludes.

2   A Dynamic Ideology Formation Model

A unit mass of agents constitutes a population. Each agent has an ideological trait from set $T = \{L, C, R\}$, where L represents the liberal left, R represents the conservative right, and C represents the center.[5] The population state can be described by a vector in $\mathbb{R}^3$, $x = (x_L, x_C, x_R)$, with $x_L + x_C + x_R = 1$, and the collection of the population states is a simplex denoted by X.

We lay out both the intergenerational and the intragenerational interpretations of the population and ideology transmission. Subsequently, we mainly narrate with the intergenerational overlapping-generations interpretation. The notations will be the same for the two interpretations. In the several instances the descriptions differ, we put the alternative intragenerational interpretation in italics in parentheses.

First, consider the intergenerational interpretation. At each time t, agents in the population are selected uniformly at random, and the selected agent becomes a parent and bears a child. Assume that the child does not have a defined trait, and the parent has to decide how much effort to exert to inculcate her own trait into her child. Such socialization within the family is called "direct vertical" socialization. If the parent fails to inculcate her own trait in her child, the child will randomly search for a role model in the society and adopt the trait of the role model. Such socialization by the society is called "oblique" socialization. After the child's trait is formed, the

---

[5]The model can be easily generalized to more than three traits to include moderate ideologies. Nevertheless, the three-trait model is sufficient to deliver our main theoretical insights.

parent is replaced by the child in the current population.

Second, consider the intragenerational interpretation. At each time t, agents in the population are selected uniformly randomly, and the selected agent will receive an impact from the society, which may influence her ideology. The agent can exert an effort to limit herself from the outside influences.[6] The higher effort she exerts, the higher is the probability that she can maintain her ideology. However, there is always a chance that she fails. In this case, she will be affected randomly by another agent in the society.

Now let us consider the effort choice of an agent with ideology $a \in T$. Let $e_a \in \mathbb{R}+$ denote her effort. Define $d : \mathbb{R}_+ \to [0, 1]$ as the probability of success function, where $d(e_a)$ is the probability that she successfully transmits her ideology to her child (*resists the outside influences*) given that she exerts an effort of e_a. Assume that $d(0) = 0$, $d' > 0$ and $d'' \leq 0$. Let $c : \mathbb{R}_+ \to \mathbb{R}_+$ be the cost function of effort. Assume that $c(0) = 0$, $c'(0) = 0$, $c' \geq 0$ and $c'' > 0$. For simplicity, we will adopt the commonly used functional form in the literature for d and c: $d(e) := e$ and $c(e) := e^2/2$.

Let V_{ab} denote the subjective assessment of an agent with ideology $a \in T$ of her utility if she changes her ideology to $b \in T$. Assume that $V_{aa} > V_{ab}$ for any $b \in T \backslash \{a\}$. This reflects that an agent always believes that her child (*she herself*) is better off by inheriting (*maintaining*) her current ideology.[7] We normalize V_{ab} to be in $[0, 1]$ for all $a, b \in T$. Let $\Delta_{ab} := V_{aa} - V_{ab}$ denote the *antipathy* of an agent with ideology $a \in T$ toward ideology $b \in T$. Observe that $\Delta_{aa} = 0$ for any $a \in T$ and we have $\Delta_{ab} \in (0, 1]$ for any $a \neq b$ and $a, b \in T$. Hence, an agent holds no animosity toward those sharing the same ideology with her in the society, but has negative sentiments toward the rest.

Given a population state x, an agent with ideology $a \in T$ tries to maximize her child's (*her*

---

[6]Socialization effort is arguably a more important factor than inheritance in determining one's partisan and ideological identity: "Just 31% of Democrats cite long-standing ties with the party—'ever since I can remember I've been a Democrat'—as a major reason for identifying with the party. Even fewer Republicans (23%) cite this as a major reason they belong to the GOP." (Pew Research Center, 2016, Page 20)

[7]In the context of intergenerational cultural transmission, this assumption is called the "imperfect empathy" in Bisin and Verdier (2001).

own) utility by taking into account the possibility that she may successfully transmit her own ideology to her child *maintain her own ideology*) and the possibility that she fails to do so and her child (*she herself*) gets influenced by others in the society. The maximization problem for any agent with an ideology $a \in T$ is given as follows:

$$\max_{e_a} \; e_a V_{aa} + (1 - e_a)(x_a V_{aa} + x_b V_{ab} + x_c V_{ac}) - \frac{1}{2}e_a^2, \tag{1}$$

for $b, c \in T\backslash\{a\}$ and $b \neq c$. If she exerts effort e_a, with probability e_a, her child (*she herself*) successfully resists outside influences and inherits (*maintains*) her ideology; and with probability $1 - e_a$, she fails to do so and randomly her child (*she herself*) switches to a new ideology according to population distribution of ideologies. The optimal solution is given in the following lemma.

Lemma 1. *The optimal effort exerted by an agent with ideology $a \in T$ is $e_a^*(x) = x_b\Delta_{ab} + x_c\Delta_{ac}$.*

Proof. See Appendix A. □

One can observe that the optimal effort increases as the aggregate antipathy she has toward the other two ideologies $x_b\Delta_{ab} + x_c\Delta_{ac}$ increases.[8] Such a property is called "cultural substitutability" defined in Wu and Cheung (2018), which generalizes the two-trait version of the same term in Bisin and Verdier (2001). This property is independent of any assumption imposed on the values of Δ_{ab} and Δ_{ac} such as convexity and concavity, which we will impose later on in the subsequent analysis.

The evolutionary dynamic of the distribution of ideologies in the society is characterized by the following system of differential equations:

$$\dot{x}_a = x_b(1 - e_b)x_a + x_c(1 - e_c)x_a - x_a(1 - e_a)(1 - x_a), \tag{2}$$

---

[8]Importantly, although Δ_{ab} and Δ_{ac} are assumed to be constant in our model, her aggregate antipathy does change with the endogenous sizes of the different ideological positions.

for any $a \in T$, $b, c \in T \backslash \{a\}$, and $b \neq c$. The rate $\dot{x}_a$ is the rate of change in the mass of agents with ideology a. The first two terms on the RHS of the dynamic represent the "inflow" of agents with non-a ideologies who switch to ideology a, and the last term is the "outflow" of agents with ideology a who switch to non-a ideologies.

To further understand the evolutionary dynamic, we utilize the insight of Montgomery (2010) that it is essentially equivalent to a replicator dynamic (RD) by Taylor and Jonker (1978) on a population game. More specifically, consider a population of agents who are randomly matched in pairs to play a two-player 3-by-3 symmetric game with strategy set T (coinciding with the set of ideologies) and the antipathy Δ_{ab} is the single match payoff of an agent playing strategy a against an opponent playing strategy b. The game is given as follows in Table 1.

Table 1: Two-players 3-by-3 symmetric game

		Player 2		
		L	C	R
	L	$0, 0$	Δ_{LC}, Δ_{CL}	Δ_{LR}, Δ_{RL}
Player 1	C	Δ_{CL}, Δ_{LC}	$0, 0$	Δ_{CR}, Δ_{RC}
	R	Δ_{RL}, Δ_{LR}	Δ_{RC}, Δ_{CR}	$0, 0$

Define $F_a(x) := \sum_{b \in T} x_b \Delta_{ab}$ as the expected payoff of an agent playing strategy a for any $a \in T$. The evolutionary dynamic can be rewritten as:

$$\dot{x}_a = x_a \left[F_a(x) - \sum_{b \in T} x_b F_b(x) \right] \text{ for any } a \in T, \tag{RD}$$

which is exactly the replicator dynamic.

We are interested in how antipathy affects the trajectory of the evolution of ideology. To obtain meaningful analytic results, we impose the specific payoff structure for Table 1 (we only show Player 1's payoffs given symmetry of the game) in Table 2.

Assume that $h > 0$, $\alpha > 1$ and $\beta > 1$. Parameter h denotes the symmetric antipathy between the extremists and the centrists. Parameter α (respectively, β) measures how much more antipa-

Table 2: Antipathy as an increasing function of ideological difference.

	L	C	R
L	0	h	αh
C	h	0	h
R	βh	h	0

thy the left (the right) has toward the right (the left) than the centrists. The payoff structure captures that antipathy increases in ideological difference and allows asymmetry in antipathy between the two extreme ideologies L and R. We have the following result:

Proposition 1. *Every trajectory in the interior of X converges to the state*

1. $(x_L^*, x_C^*, x_R^*) = (\frac{\alpha}{\alpha+\beta}, 0, \frac{\beta}{\alpha+\beta})$ *if $(\alpha - 1)(\beta - 1) \geq 1$, and*

2. $(x_L^*, x_C^*, x_R^*) = (\frac{\alpha}{2\alpha+2\beta-\alpha\beta}, \frac{\alpha+\beta-\alpha\beta}{2\alpha+2\beta-\alpha\beta}, \frac{\beta}{2\alpha+2\beta-\alpha\beta})$ *otherwise.*

Proof. See Appendix A. □

Proposition 1 characterizes all the globally stable states of the evolutionary dynamic.[9], [10] There are several immediate implications.

First, when $\alpha \geq 2$ and $\beta \geq 2$, that is, agents with ideology L and R have convex antipathy ($\Delta_{LR} \geq 2\Delta_{LC}$, $\Delta_{RL} \geq 2\Delta_{RC}$), $x_C^* = 0$. Hence, convex antipathy is a sufficient condition for the rise of political extremism. The rationale is as follows. Because antipathy is increasing and convex, agents with ideology L or R have a higher incentive to maintain their ideology by exerting higher efforts (being more politically engaged) than those with ideology C. Hence, x_C always decreases and $x_L + x_R$ always increases along the evolutionary trajectory.

---

[9]Here global stability of a population state means that every trajectory in the interior of X converges to the state. If the initial state is on the boundaries of X, i.e., one of the ideologies is initially "zeroed out", the dynamic cannot leave that edge because of the special property of the replicator dynamic.

[10]The more general results for the replicator dynamic on 3-by-3 games have been established in earlier works by Zeeman (1980) and Bomze (1983). See Montgomery (2010) for a more detailed discussion.

Second, when $\alpha, \beta \in (1, 2)$, that is, agents with ideology L or R have strictly concave antipathy, all three ideologies are in the support of the stable state. Hence, concave antipathy leads to diversification.

In Case 1 of Proposition 1, we have $\frac{\partial x_L^*}{\partial \alpha} > 0$, $\frac{\partial x_R^*}{\partial \alpha} < 0$, implying that if the antipathy toward ideology R for the agents with ideology L is stronger, the proportion of agents with ideology L increases and the proportion of agents with ideology R decreases. In Case 2 of Proposition 1, we have $\frac{\partial x_L^*}{\partial \alpha} > 0$, $\frac{\partial x_C^*}{\partial \alpha} < 0$, $\frac{\partial x_R^*}{\partial \alpha} \leq (>) \, 0$ if $\beta \leq (>) \, 2$. Hence, when the antipathy toward ideology R for the agents with ideology L is stronger, the proportion of agents with ideology L increases and the proportion of agents with ideology C decreases. The comparative statics for x_R^* with respect to α depends on the value of β. In particular, when $\beta > 2$, x_R^* increases in α.

Although convex antipathy leads to political extremism, it does not guarantee polarization. According to the classic measurement of polarization proposed by Esteban and Ray (1994) and Duclos et al. (2004), the polarization of a distribution of individual attributes must exhibit three features: 1) there must be a high degree of homogeneity within each group; 2) there must be a high degree of heterogeneity across groups; 3) there must be a small number of significantly sized groups. While convex antipathy guarantees that the first two features are satisfied, the validity of the third feature depends on the values of α and β. When α and β are highly asymmetrical, one of the extreme groups would be significantly larger than the other, leading to a quasi-homogeneous society with one extreme ideology being dominant. Hence, the emergence of polarization requires not only convex antipathy but also more balanced antipathy between the two extreme groups.

Let us demonstrate this point more formally. Esteban and Ray (1994) and Duclos et al. (2004) posit that the measure of polarization should be proportional to what they call the sum of all "effective antagonisms." The higher this sum is, the more likely it results in social unrest. Their idea of antagonism is similar to our definition of antipathy. Hence, we consider the aggregated antipathy of all individuals as a measure of polarization. Let us focus on the case of convex

antipathy ($\alpha, \beta \geq 2$). Consider the following maximization problem of the aggregated antipathy:

$$\max_{x_L, x_R} \quad x_L\big((1 - x_L - x_R)h + x_R\alpha h\big) + (1 - x_L - x_R)\big(x_L h + x_R h\big) + x_R\big(x_L \beta h + (1 - x_L - x_R)h\big) \quad (3)$$

such that $x_L \geq 0$, $x_R \geq 0$, $x_L + x_R \geq 1$.

We have the following result:

Proposition 2. *When $\alpha + \beta < 8$, $x_L = x_R = \frac{1}{2}$ is the unique solution to the maximization problem given in (3).*

Proof. See Appendix B. □

The condition of $\alpha + \beta < 8$ guarantees that the Hessian matrix is negative definite. Note that this condition also makes the game in Table 1 a contractive game defined by Hofbauer and Sandholm (2009) (see the proof of Proposition 1 for the use of contractive game in proving global convergence). Proposition 2 shows that when the society is equally split into two extreme groups, the aggregated antipathy is the highest in the society. Hence, this should be considered as the most polarized state.

The aggregate antipathy corresponding to the most polarized state is $\frac{1}{4}(\alpha + \beta)h$, while the aggregate antipathy corresponding to the equilibrium in Case 1 of Proposition 1 is $\frac{\alpha\beta}{\alpha+\beta}h$. The ratio $\big(\frac{\alpha\beta}{\alpha+\beta}h\big) / \big(\frac{1}{4}(\alpha + \beta)h\big) \leq 1$ and the equality holds when $\alpha = \beta$. Hence, as we have stressed before, convex antipathy and balanced antipathy between the two extreme groups together lead to the most polarized state.

We provide some graphic illustrations of the evolutionary dynamic in difference cases ($h = 0.05$ across all cases) by utilizing the software *Dynamo* (Franchetti and Sandholm, 2013) in Figure 2. Hollow dots represent the unstable rest points and solid dots represent the stable states. The arrows represent the trajectories of the evolutionary dynamic. The different colors of shades in the figures represent different speeds of the dynamic. Warmer colors are associated with faster

speeds. Figures 2(a) and 2(b) show that convex antipathy and balanced antipathy between the two extreme groups leads to polarization and concave antipathy leads to diversification, respectively.

(a) Convexity and Polarization: $\alpha = \beta = 2$.

(b) Concavity and Diversification: $\alpha = \beta = 1.5$.

(c) Weak asymmetry: $\alpha = 2$, $\beta = 1.5$.

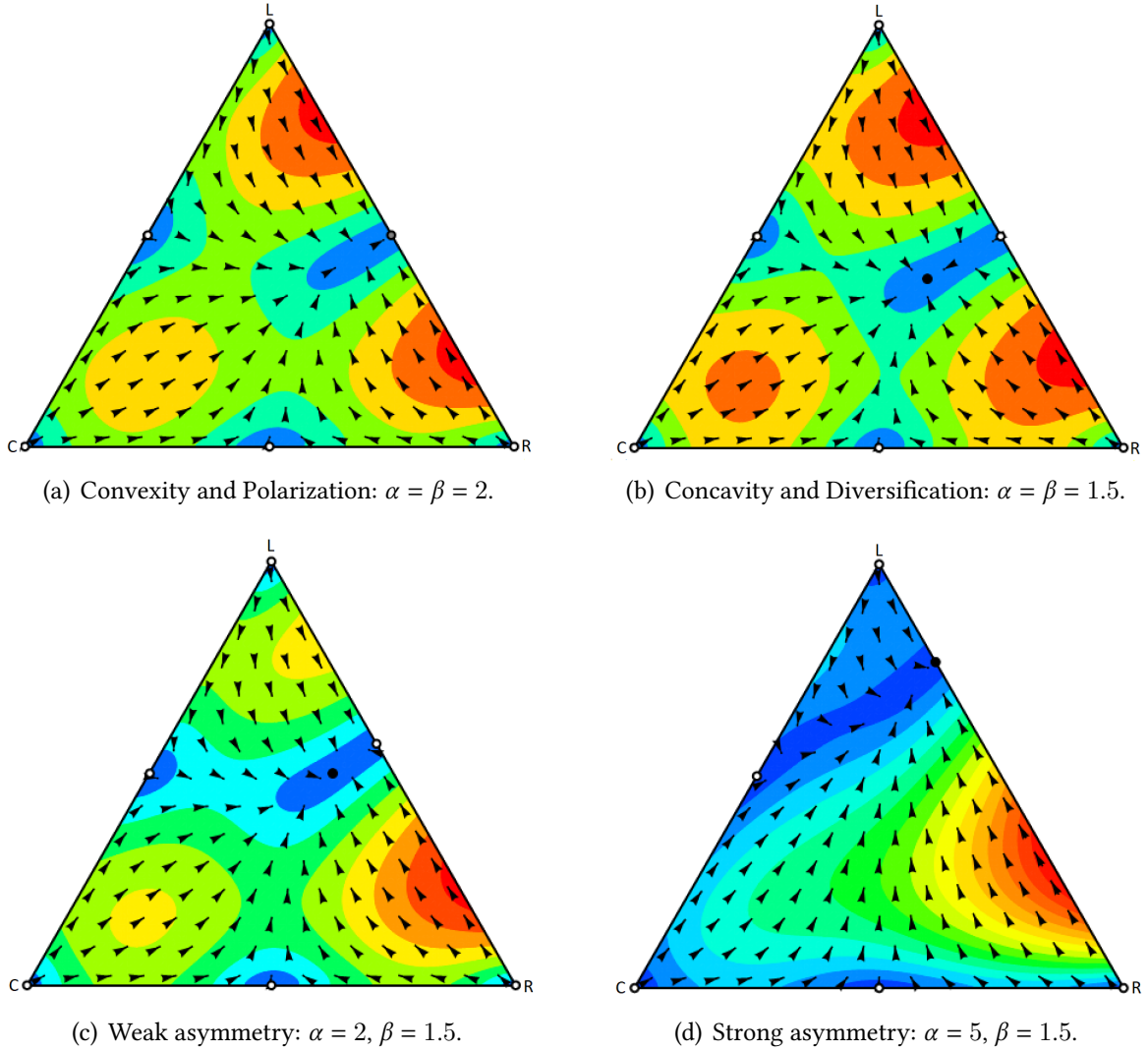(d) Strong asymmetry: $\alpha = 5$, $\beta = 1.5$.

Figure 2: Simulations of the evolutionary dynamic.

Figures 2(c) and 2(d) characterize cases in which the antipathy of agents with ideology L is convex and the antipathy of agents with ideology R is strictly concave. Figures 2(c) shows that all three ideologies still coexist and L is the majority when L's antipathy toward R is not too strong comparing to R's antipathy toward L. Figure 2(d) implies that as R's antipathy becomes more concave ($\beta \downarrow 1$), we need L's antipathy become much more convex (at the same time, its

13

antipathy toward R becomes much stronger) to maintain polarization.

Note that in Table 2, the antipathies that the centrists have toward the left and the right are both assumed to be h, which equals to the antipathy that the extremists have toward the centrists. Although it is reasonable to assume that the centrists have somewhat mild antipathy, it is still plausible that the centrists have higher contempt for extreme ideologies. Let us consider the following variant of Table 2.

Table 3: Centrists with stronger antipathy.

	L	C	R
L	0	h	αh
C	ρh	0	ρh
R	αh	h	0

For simplicity, we assume that the two extremists have equal antipathy toward each other. We also assume that $1 < \rho < \alpha$. The assumption $1 < \rho$ ensures that the centrists have stronger antipathy toward the extremists than the extremists have toward the centrists. The assumption $\rho < \alpha$ ensures that the centrists' antipathy toward the left (the right) is weaker than the right's (the left's) antipathy toward the left (the right).

One can show that when $\frac{1}{2}\alpha \geq \rho$, all three ideologies coexist. Hence, as long as the centrists' antipathy toward the extremists is more than half as intense as the extremists' antipathy toward each other, polarization would not occur. This suggests that increasing the centrists' antipathy toward the extreme ideologists through education/media can serve as an effective policy tool to fight against polarization.

3   Strong versus Weak Antipathy

In this section, we generalize the previous model to allow agents holding the same ideological position to have different degrees of antipathy.

A unit mass of agents constitutes a population. Each agent has a trait from the set $T = \{L_s, L_w, C, R_w, R_s\}$, where L, C, and R still represent the three different ideologies, the liberal left, the center, and the conservative right, and s and w represent strong and weak antipathy, respectively. Hence, we allow both angry and mild-tempered liberals (conservatives). The population state can be described by a vector in $\mathbb{R}^5$, $x = (x_{L_s}, x_{L_w}, x_C, x_{R_w}, x_{R_s})$, with $x_{L_s} + x_{L_w} + x_C + x_{R_w} + x_{R_s} = 1$. Let X denote the collection of all the population states.

The modeling details are identical to the one described in Section 3. Hence, we skip them and directly focus on the replicator dynamic on the equivalent population game, and we examine the payoff structure in Table 4.

Table 4: Antipathy Matrix

	L_s	L_w	C	R_w	R_s			L_s	L_w	C	R_w	R_s
L_s	0	$\Delta_{L_s L_w}$	$\Delta_{L_s C}$	$\Delta_{L_s R_w}$	$\Delta_{L_s R_s}$		L_s	0	0	h	γh	θh
L_w	$\Delta_{L_w L_s}$	0	$\Delta_{L_w C}$	$\Delta_{L_w R_w}$	$\Delta_{L_w R_s}$	$=$	L_w	0	0	h	τh	γh
C	$\Delta_{C L_s}$	$\Delta_{C L_w}$	0	$\Delta_{C R_w}$	$\Delta_{C R_s}$		C	h	h	0	h	h
R_w	$\Delta_{R_w L_s}$	$\Delta_{R_w L_w}$	$\Delta_{R_w C}$	0	$\Delta_{R_w R_s}$		R_w	γh	τh	h	0	0
R_s	$\Delta_{R_s L_s}$	$\Delta_{R_s L_w}$	$\Delta_{R_s C}$	$\Delta_{R_s R_w}$	0		R_s	θh	γh	h	0	0

Again, parameter h denotes the symmetric antipathy between the extremists and the centrists. Similar to α and β in Table 2, γ, θ, and τ measure the intensity of the extremists' antipathies toward each other compared to their antipathies toward the centrists. We make several assumptions on the payoff structure. First, we assume symmetry, that is $\Delta_{ab} = \Delta_{ba}$ for any $a, b \in T$.[11] Second, we assume that the liberal lefts (conservative rights) do not have antipathy against each other. i.e, $\Delta_{i_s, i_w} = \Delta_{i_w, i_s}$, for $i \in \{L, R\}$, which implies that the agents only have antipathy toward those with different ideologies. This assumption helps to simplify the analysis. Relaxing it would not affect the main result as long as one's antipathy toward another having the same ideology is

---

[11]When a population game has more than three strategies, it is generally impossible to analytically characterize the predictions of the evolutionary dynamic, unless we impose specific properties on the payoff structure. The symmetry assumption imposed makes the game a "double symmetric" game, which is a special case of a potential game (Sandholm, 2001). As one can see from the proof of Proposition 3, a potential game allows us to establish global stability results.

weaker than her antipathy toward another having a different ideology. Third, we assume that agents with ideology C do not distinguish between i_s and i_w, for $i \in \{L, R\}$. Again, this is a simplifying assumption. Relaxing it would not change the main insights, unless we assume that their antipathies toward L_s and R_s are sufficiently stronger than their antipathies toward L_w and R_w, which would help to preserve ideology C as we have discussed at the end of Section 2. Finally, assume that $\tau \leq \gamma \leq \theta$, with at least one inequality being strict. This assumption guarantees that s-type agents have stronger antipathy than w-type agents toward those having a different ideology. In addition, assume $\tau > 1$, which assures that antipathy is increasing in ideological difference. We have the following result:

Proposition 3. *When $\theta > 2$, every trajectory in the interior of X converges to the state* $(x^*_{L_s}, x^*_{L_w}, x^*_C, x^*_{R_w}, x^*_{R_s}) = (\frac{1}{2}, 0, 0, 0, \frac{1}{2})$.

Proof. See Appendix C. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Proposition 3 demonstrates that strong antipathy is able to eliminate weak antipathy and when $\theta > 2$, that is, the type-s agents' antipathy is convex in ideological difference (after eliminating L_w and R_w), ideology C is eliminated as well due to its central position, leaving us with L_s and R_s agents equally splitting the population. Note that the potential function associated with the game is half of the aggregate antipathy, which serves as a measure of polarization as we have discussed in Section 2. In the proof of Proposition 3, we show that the state $(x^*_{L_s}, x^*_{L_w}, x^*_C, x^*_{R_w}, x^*_{R_s}) = (\frac{1}{2}, 0, , 0, 0, \frac{1}{2})$ is the unique global maximizer of the potential function. Hence, it is the most polarized state.

We provide some graphic illustrations of the evolutionary dynamic in a numerical example in Figure 3, utilizing the software *ABED* (Izquierdo et al., 2019). We set $h = 0.05$, $\tau = 2$, $\gamma = 3$, and $\theta = 4$. The initial distribution of traits is $(0.1, 0.2, 0.4, 0.2, 0.1)$. Figures 3(a) and 3(b) provide the time series of the strategy distributions (population states) and the strategies' expected payoffs (the efforts exerted by agents with different traits), respectively.
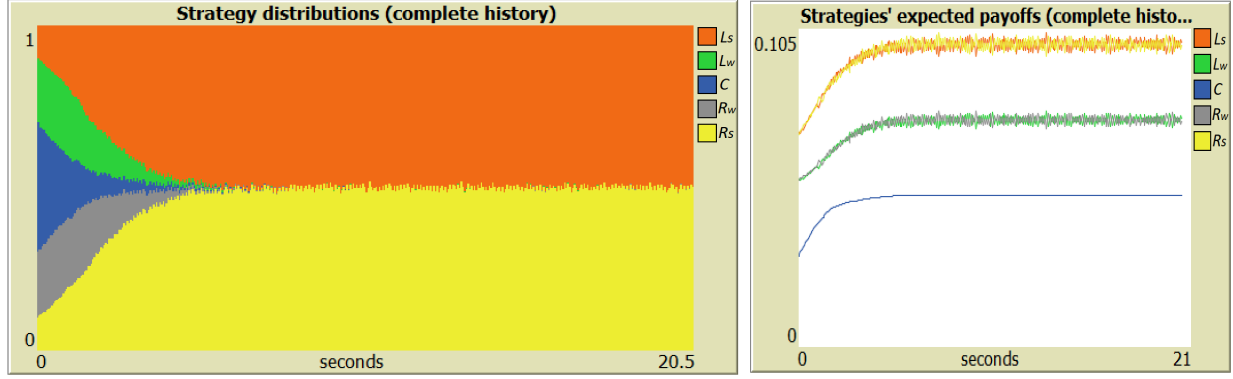
(a) Time series of the population states.　　(b) Time series of the expected payoffs.

Figure 3: Simulations of the evolutionary dynamic.

Figure 3(b) shows that the efforts exerted by all agents are monotonically increasing over time before the dynamic reaches the stable state. In addition, at each time, the efforts of L_s and R_s are larger than those of L_w and R_w, which are larger than that of C, showing that agents with strong antipathy situated on the left and the right are the most politically active.

4 Conclusion

This paper proposes a dynamic model of ideological formation and we show how people's antipathy toward those who are ideologically different from them can lead to polarization even in the absence of social changes such as influx of immigration and refugees, increasing income inequality, and growing segregation along the lines of race and socioeconomic classes. Although the current American society has not reached the extreme polarization predicted by our model, empirical evidence suggests that it is on the path.

Many blame traditional media, social media, politicians, party propaganda, or foreign influences for the current situation in the United States. Our paper suggests that these factors may imperceptibly influence individuals' antipathy and slowly drive them apart.

17

Appendix A: Proof of Lemma 1 and Proposition 1

We first prove Lemma 1. By taking the first order condition of the objective function in the maximization problem in (1), we get

$$
\begin{aligned}
e_a &= (1 - x_a)V_{aa} - x_b V_{ab} - x_c V_{ac} \\
&= (x_b + x_c)V_{aa} - x_b V_{ab} - x_c V_{ac} \\
&= x_b(V_{aa} - V_{ab}) + x_c(V_{aa} - V_{ac}) \\
&= x_b \Delta_{ab} + x_c \Delta_{ac}.
\end{aligned}
$$

Next, we prove Proposition 1. For notation convenience, we denote the payoff matrix by

$$
\Gamma := \begin{bmatrix} 0 & h & \alpha h \\ h & 0 & h \\ \beta h & h & 0 \end{bmatrix}.
$$

Let $z = (z_1, z_2, -z_1 - z_2)$ be a displacement vector. We have

$$
\begin{aligned}
z' \Gamma z &= -h\left[(\alpha + \beta)z_1^2 + (\alpha + \beta)z_1 z_2 + 2z_2^2\right] \\
&= -h\left[(\alpha + \beta)(z_1 + \frac{1}{2}z_2)^2 + (2 - \frac{\alpha + \beta}{4})z_2^2\right].
\end{aligned}
$$

When $\alpha + \beta \le 8$, $z' \Gamma z < 0$ as long as z is non zero, which is the defining property of a strictly contractive game (also called the strictly stable game) by Hofbauer and Sandholm (2009) (see also Sandholm (2010)). According to Hofbauer and Sandholm (2009), a strictly contractive game has a unique Nash equilibrium and every trajectory of the replicator dynamic in the interior of X converges to this unique Nash equilibrium. Hence, the remaining task is to find the Nash equilibrium.

When $(\alpha - 1)(\beta - 1) \geq 1$, for $x^* = (x_L^*, x_C^*, x_R^*) = (\frac{\alpha}{\alpha+\beta}, 0, \frac{\beta}{\alpha+\beta})$, we have $F_L(x^*) = F_R(x^*) >$ $F_C(x^*)$. Hence, it is the unique Nash equilibrium.

When $(\alpha - 1)(\beta - 1) < 1$, for $x^* = (x_L^*, x_C^*, x_R^*) = (\frac{\alpha}{2\alpha+2\beta-\alpha\beta}, \frac{\alpha+\beta-\alpha\beta}{2\alpha+2\beta-\alpha\beta}, \frac{\beta}{2\alpha+2\beta-\alpha\beta})$, we have $F_L(x^*) = F_C(x^*) = F_R(x^*)$. Hence, it is the unique Nash equilibrium.

When $\alpha + \beta > 8$, we still have the same results on Nash equilibrium in both cases. However, we can no longer use the results for the replicator dynamic on contractive games to study the stability of the Nash equilibrium. Nevertheless, we can still establish the global stability results.

First, both the eigenvalues of the Jacobian matrix are negative at the unique Nash equilibrium in both cases. Hence, it is locally stable. Then, for the case $(\alpha - 1)(\beta - 1) \geq 1$, we can show that the other 5 rest points (2 on the edges and the 3 vertices) are unstable by utilizing the Jacobian matrix; and for the case $(\alpha - 1)(\beta - 1) < 1$, we can show that the other 6 rest points (3 on the edges and the 3 vertices) are unstable. Therefore, we conclude that the unique Nash equilibrium is the uniquely locally stable state in both cases.

Since there is no interior stable state when $(\alpha - 1)(\beta - 1) \geq 1$, all trajectories must converge to the boundary of the simplex according to Hutson and Moran (1982). Hence, the unique Nash equilibrium must be globally stable.

Hofbauer and Sigmund (1988) show that the replicator dynamic admits no limit cycles when there is an interior locally stable state in 3 by 3 games. Hence, for the case $(\alpha - 1)(\beta - 1) < 1$, the unique Nash equilibrium is globally stable.

Appendix B: Proof of Proposition 2

The Hessian matrix of the objective function in the maximization problem (3) is given by

$$\begin{bmatrix} -4 & \alpha + \beta - 4 \\ \alpha + \beta - 4 & -4 \end{bmatrix}$$

When $\alpha + \beta < 8$, it is negative definite.

Let $\lambda \geq 0$, $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$ be the Lagrange multipliers for the constraints $x_L + x_R \leq 1$, $x_L \geq 0$ and $x_R \geq 0$, respectively. Then the first order Kuhn-Tucker necessary conditions are given by:

$$-4x_L + 2 + (\alpha + \beta - 4)x_R - \lambda + \gamma_1 = 0,$$

$$-4x_R + 2 + (\alpha + \beta - 4)x_L - \lambda + \gamma_2 = 0,$$

$$\lambda(1 - x_L - x_R) = 0;$$

$$\gamma_1 x_L = 0;$$

$$\gamma_2 x_R = 0.$$

Case 1: Suppose $\gamma_1 = \gamma_2 = 0$, then we must have $x_L > 0$ and $x_R > 0$. This implies that

$$-4x_L + 2 + (\alpha + \beta - 4)x_R = -4x_R + 2 + (\alpha + \beta - 4)x_L = \lambda,$$

which further implies that $x_L = x_R$. Suppose $\lambda = 0$, then we must have $x_L = x_R < \frac{1}{2}$. We also have

$$-4x_L + 2 + (\alpha + \beta - 4)x_R = 2 + (\alpha + \beta - 8)x_R = 0.$$

However, since we assume convex antipathy, that is, $\alpha, \beta \geq 2$, and we know $x_R < \frac{1}{2}$, we must have $2 + (\alpha + \beta - 8)x_R > 0$. Hence, it is impossible to have $\lambda = 0$. Now consider $\lambda > 0$. Then $x_L = x_R = \frac{1}{2}$, which satisfies the Kuhn-Tucker conditions.

Case 2: Suppose $\gamma_1, \gamma_2 > 0$. Then $x_L = x_R = 0$, which also implies that $\lambda = 0$. In this case, we have $2 + \gamma_1 = 0$ and $2 + \gamma_2 = 0$, both of which are impossible.

Case 3: Suppose $\gamma_1 > 0$ and $\gamma_2 = 0$. Then $x_L = 0$ and $x_R > 0$. In this case, we have

$$2 + (\alpha + \beta - 4)x_R - \lambda + \gamma_1 = 0,$$

$$-4x_R + 2 - \lambda = 0.$$

Suppose $\lambda = 0$. Then $x_R = \frac{1}{2}$. However, since $\alpha + \beta \geq 4$, $2 + (\alpha + \beta - 4)\frac{1}{2} + \gamma_1 > 0$. Hence, it is impossible to have $\lambda = 0$. Now consider the case that $\lambda > 0$, then $x_R = 1$. But then we have $-4x_R + 2 - \lambda = -2 - \lambda < 0$. Hence, it is also impossible to have $\lambda > 0$. To conclude, only $x_L = x_R = \frac{1}{2}$ satisfies the Kuhn-Tucker conditions. Since the objective function is strictly concave, $x_L = x_R = \frac{1}{2}$ is the unique global maximizer.

Appendix C: Proof of Proposition 3

We want to show that $(x_{L_s}^*, x_{L_w}^*, x_C^*, x_{R_w}^*, x_{R_s}^*) = (\frac{1}{2}, 0, , 0, 0, \frac{1}{2})$ is the unique Nash equilibrium. First, suppose L_w is in the support of a Nash equilibrium x, we must have $F_{L_w}(x) \geq F_{L_s}(x)$, which requires $x_{R_w} = x_{R_s} = 0$. Then we have $F_{R_w}(x) > F_{L_w}(x)$ and $F_{R_s}(x) > F_{L_w}(x)$, violating the definition of the Nash equilibrium. Therefore, L_w cannot be in the support of a Nash equilibrium. Similarly, R_w cannot be in the support of a Nash equilibrium either.

Second, suppose $x = (x_{L_s}, 0, x_C, 0, x_{R_s})$ with $x_C > 0$ is a Nash equilibrium. Then we have $F_{L_s}(x) = h(x_C + \theta x_{R_s})$, $F_{R_s}(x) = h(x_C + \theta x_{L_s})$ and $F_C(x) = h(x_{L_s} + x_{R_s})$. When $\theta > 2$, $\frac{1}{2}(F_{L_s}(x) + F_{R_s}(x)) = h(x_c + \frac{\theta}{2}(x_{L_s} + x_{R_s})) > F_C(x)$. Hence, at least one of $F_{L_s}(x)$ and $F_{R_s}(x)$ is strictly larger than $F_C(x)$, which violates the definition of the Nash equilibrium. Hence, we must have $x_C = 0$. Then it is straightforward to show that $(x_{L_s}^*, x_{L_w}^*, x_C^*, x_{R_w}^*, x_{R_s}^*) = (\frac{1}{2}, 0, , 0, 0, \frac{1}{2})$ is a Nash equilibrium, and so it is unique.

Since we assume $\Delta_{ab} = \Delta_{ba}$, for any $a, b \in T$, the population game is a potential game for a continuous population of players (Sandholm, 2001). Since there is a unique Nash equilibrium,

then every trajectory of the replicator dynamic in the interior of X converges to this unique Nash equilibrium. The rationale is that when the game has a unique Nash equilibrium, it must be the global maximizer (and also the unique local maximizer) of the potential function of the game, which equals to half of the average payoff $(f(x) = \frac{1}{2} \sum_{a \in T} x_a F_a(x))$ in our model. Hofbauer and Sigmund (1988) show that a rest point is stable if and only if it is a local maximizer of the potential function and (Sandholm, 2001) shows that every trajectories converges to some rest points. Since the unique Nash equilibrium is the only stable rest point, every trajectory in the interior of X converges to it. While the trajectories on the boundaries of the simplex converge to some unstable rest points.

Acknowledgements

References

Besley, Timonthy and Torsten Persson, "JEEA-FBBVA LECTURE 2017: The Dynamics of Environmental Politics and Values," *Journal of the European Economic Association*, 2017, *17*, 993–1024.

— **and** — , "Democratic Values and Institutions," *American Economic Review: Insights*, 2019, *1*, 59–76.

Bisin, Alberto and Thierry Verdier, "A model of cultural transmission, voting and political ideology," *European Journal of Political Economy*, 2000, *16*, 5–29.

__ **and** __ , "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 2001, *97* (2), 298–319.

__ , **Eleonora Patacchini, Thierry Verdier, and Yves Zenou**, "Bend it like Beckham: Ethnic Identity and Integration," *European Economic Review*, 2016, *90*, 146–164.

Black, Duncan, "On the Rationale of Group Decision-Making," *Journal of Political Economy*, February 1948, *56* (1), 23–34.

Bolletta, Ugo and Paolo Pin, "Polarization when People Choose Their Peers," 2020. Working Paper.

Bomze, Immanuel M., "Lotka-Volterra Equation and Replicator Dynamics: A Two Dimensional Classification," *Biological Cybernetics*, 1983, *48* (3), 201–211.

Cerqueti, Roy, Luca Correani, and Giuseppe Garofalo, "Economic Interactions and social tolerance: a dynamic perspective," *Economics Letters*, 2013, *120*, 458–463.

Correani, Luca, Fabio Di Dio, and Giuseppe Garofalo, "The Evolutionary Dynamics of Tolerance," *Theoretical and Practical Research in Economic Fields*, 2010, *1* (2), 218–230.

Dixit, Avinash K. and Jörgen W. Weibull, "Political Polarization," *Proceedings of the National Academy of Science of the United States of America*, 2007, *104* (18), 7351–7356.

Downs, Anthony, "An Economic Theory of Political Action in a Democracy," *Journal of Political Economy*, April 1957, *65* (2), 135–150.

Duclos, Jean-Yves, Joan-María Esteban, and Debraj Ray, "Polarization: Concepts, Measurement, Estimation," *Econometrica*, 2004, *72* (6), 1737–1772.

Epstein, Gil S., "Extremism within the Family," *Journal of Population Economics*, 2007, *20*, 707–715.

Esteban, Joan-María and Debraj Ray, "On the Measurement of Polarization," *Econometrica*, 1994, *62* (4), 819–851.

Fiorina, Morris P., Samuel J. Abrams, and Jeremy C. Pope, *Culture War? The Myth of a Polarized America*, Longman, 2005.

Flachea, Andreas, Michael Mäsa, Thomas Feliciania, Edmund Chattoe-Brownb, Guil-laume Deffuantc, Sylvie Huetc, and Jan Lorenz, "Models of Social Influence: Towards the Next Frontiers," *Journal of Artificial Societies and Social Simulation*, 2017, *20* (4).

Franchetti, Francisco and William H. Sandholm, "An Introduction to Dynamo: Diagrams for Evolutionary Game Dynamics," *Biological Theory*, 2013, *8*, 167–178.

Frank, Thomas, *What's the Matter with Kansas? How Conservatives Won the Heart of America*, Picador, 2004.

Gradstein, Mark and Moshe Justman, "Education, Social Cohesion , and Economic Growth," *American Economic Review*, 2002, *92* (4), 1192–1204.

__ **and** __ , "The Melting Pot and School Choice," *Journal of Public Economics*, 2005, *89* (5-6), 871–896.

Hetherington, Marc J. and Jonathan D. Weiler, *Authoritarianism and Polarization in American politics*, Cambridge University Press, 2009.

Hofbauer, Josef and Karl Sigmund, *The Theory of Evolution and Dynamical Systems*, Cambridge University Press, 1988.

__ **and William H. Sandholm**, "Stable Games and Their Dynamics," *Journal of Economic Theory*, 2009, *144*, 1665–1693.

Hotelling, Harold, "Stability in Competition," *The Economic Journal*, March 1929, *39* (153), 41–57.

Hunter, James Davison, *Culture Wars: The Struggle to Define America*, Basic Books, 1991.

Hutson, Vivian and William Moran, "Persistence of Species Obeying Difference Equations," *Journal of Mathematical Biology*, 1982, *15*, 203–213.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes, "Affect, Not Ideology: A Social Identity Perspective on Polarization," *Public Opinion Quarterly*, Fall 2012, *76* (3), 405–431.

Izquierdo, Luis R., Segismundo S. Izquierdo, and William H. Sandholm, "An Introduction to ABED: Agent-Based Simulation of Evolutionary Game Dynamics," *Games and Economic Be-

havior, 2019, *118*, 434–462.

Layman, Geoffrey C., Tomas M. Carsey, and Juliana Menasce Horowitz, "Party Polarization in American Politics: Characteristics, Causes, and Consequences.," *Annual Review of Political Science*, 2006, *9*, 83–110.

Marvela, Seth A., Jon Kleinberg, Robert Kleinberg, and Steven Strogatz, "Continuous-time model of structural balance," *Proceedings of the National Academy of Science of the United States of America*, 2011, *108* (5), 1771–1776.

Melguizo, Isabel, "Homophily and the Persistence of Disagreement," *Economic Journal*, 2019, *129* (619), 1400–1424.

Melindi-Ghidi, Paolo, "A Model of Ideological Transmission with Endogenous Paternalism," *International Journal of Economic Theory*, 2012, *8* (4), 381–403.

Michaeli, Moti and Jiabin Wu, "Fighting Polarization with (Parental) Internalization," 2021.

Montgomery, James D., "Intergenerational Cultural Transmission as an Evolutionary Game," *American Economic Journal: Microeconomics*, November 2010, *2* (4), 115–136.

Pew Research Center, "Political Polarization in the American Public," June 2014.

— , "Partisanship and Political Animosity in 2016," June 2016.

— , "In Polarized Era, Fewer Americans Hold a Mix of Conservative and Liberal Views," October 2017.

— , "Partisan Antipathy: More Intense, More Personal," October 2019.

Poole, Keith T. and Howard Rosenthal, "The Polarization of American Politics," *The Journal of Politics*, November 1984, *46* (4), 1061–1079.

Sandholm, William H., "Potential Games with Continuous Player Sets," *Journal of Economic Theory*, March 2001, *97* (1), 81–108.

— , *Population Games and Evolutionary Dynamics*, The MIT Press, 2010.

Sides, John and Daniel J. Hopkins, *Political Polarization in American Politics*, Bloomsbury Academic, 2015.

Ticchi, D., T. Verdier, and A. Vindigni, "Democracy, dictatorship and the cultural transmission of political values," 2013.

Wu, Jiabin and Man-Wah Cheung, "On the Probabilistic Transmission of Continuous Cultural Traits," *Journal of Economic Theory*, February 2018, *174*, 300–323.

Zeeman, Erik Christopher, "Population Dynamics from Game Theory," *Global Theory of Dynamical System: Lecture Notes in Mathematics*, 1980, *819*, 471–497.