Segregation and clustering of preferences erode socially beneficial coordination

Vítor V. Vasconcelos[a,b,c,d,1], Sara M. Constantino[k,d,e,i], Astrid Dannenberg[f,g], Marcel Lumkowsky[f], Elke Weber[d,e,h], and Simon Levin[j,l,d]

[a]Informatics Institute, University of Amsterdam 1098 XH Amsterdam, The Netherlands; [b]Institute for Advanced Study, University of Amsterdam 1012 GC, Amsterdam, The Netherlands; [c]Princeton Institute for International and Regional Studies, Princeton University, Princeton, NJ 08544; [d]Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ 08540; [e]School of Public and International Affairs, Princeton University, Princeton, NJ 08540; [f]Department of Economics, University of Kassel, Kassel 34109, Germany; [g]Department of Economics, University of Gothenburg 405 30 Gothenburg, Sweden; [h]Department of Psychology, Princeton University, Princeton, NJ 08540; [i]Department of Psychology, Northeastern University, Boston, MA 02115; [j]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544; [k]School of Public Policy and Urban Affairs, Northeastern University, Boston, MA 02115; and [l]High Meadows Environmental Institute, Princeton University, Princeton, NJ 08540

Edited by Charles Perrings, Arizona State University, Tempe, AZ, and accepted by the Editorial Board September 22, 2021 (received for review March 8, 2021)

Polarization on various issues has increased in many Western democracies over the last decades, leading to divergent beliefs, preferences, and behaviors within societies. We develop a model to investigate the effects of polarization on the likelihood that a society will coordinate on a welfare-improving action in a context in which collective benefits are acquired only if enough individuals take that action. We examine the impacts of different manifestations of polarization: heterogeneity of preferences, segregation of the social network, and the interaction between the two. In this context, heterogeneity captures differential perceived benefits from coordinating, which can lead to different intentions and sensitivity regarding the intentions of others. Segregation of the social network can create a bottleneck in information flows about others' preferences, as individuals may base their decisions only on their close neighbors. Additionally, heterogeneous preferences can be evenly distributed in the population or clustered in the local network, respectively reflecting or systematically departing from the views of the broader society. The model predicts that heterogeneity of preferences alone is innocuous and it can even be beneficial, while segregation can hamper coordination, mainly when local networks distort the distribution of valuations. We base these results on a multimethod approach including an online group experiment with 750 individuals. We randomize the range of valuations associated with different choice options and the information respondents have about others. The experimental results reinforce the idea that, even in a situation in which all could stand to gain from coordination, polarization can impede social progress.

polarization | cooperation | coordination | heterogeneity | social change

Different aspects of polarization among partisan elites are growing (1) as is the perception of polarization among the public (2), particularly in the United States (3, 4). Some of the modern challenges we face are collective action problems. These range from action on climate change and ecological collapse mitigation, addressing inclusive growth, food security, inequality, and racial injustice to developing agreements of nuclear proliferation and the ethical use of artificial intelligence. These issues require a critical mass of people or nations to take action or rectify agreements that remove free-riding incentives before others follow suit (5). This dynamic is characteristic of coordination problems—the main focus of this paper—and is evident in many social processes, from elections to flips in social norms (6) or adoption of new technologies. Coordination problems are likely to be affected by the different aspects of polarization, which can be linked to divisions in opinions and segregation among groups. Here, we ask whether the manifestations of polarization have a negative impact in solving these coordination problems.

People's willingness to contribute to the good of society depends on the costs and benefits they face but also on the opinions and actions prevalent in their personal networks (7). A telling example is a recent CNN (Cable News Network) television advertisement reminding people that wearing face masks in times of the COVID-19 pandemic is not a political statement. In the United States, mask-wearing to mitigate the spread of coronavirus quickly came to be perceived as a highly polarized issue. Variation in social distancing behavior is aligned with electoral maps, as well as maps of climate denial, suggesting that the United States is facing a collective action problem that results, in part, from polarization (8). The tendency to conform to the behaviors and views of one's social network can result in endogenous social change processes that can accelerate behavior change. Cessation of smoking in public places or decreased automobile traffic in some areas are examples in which shifting social norms, initiated or supported by political interventions, have successfully transformed behavior (9). However, these same social processes can also lead to entrenched or divergent behaviors, especially in polarized communities (10). At the same time, addressing the most urgent societal problems requires joint action, often beyond the borders of neighborhoods, regions, and nation-states. While

Significance

With different types and manifestations of polarization growing, it is crucial to understand their impact on welfare-promoting processes in society. This understanding is important when polarization interrupts both socially and individually optimal outcomes. We propose a model and framework to study and quantify the relationship between polarization and the ability to coordinate. We show, analytically, that some degree of diversity of opinions may be beneficial to coordination. Then, we explore, both computationally and through experiments, the role that properties of networks, namely segregation and biased perceptions, play in eroding coordination and moving societies away from the social optimum. We expect this work to provide the basis for upcoming research on the connection between the polarization of opinions and social outcomes.

Author contributions: V.V.V., S.M.C., A.D., E.W., and S.L. designed research; V.V.V., S.M.C., A.D., and M.L. performed research; V.V.V., S.M.C., and A.D. contributed new reagents/analytic tools; V.V.V., S.M.C., A.D., and M.L. analyzed data; and V.V.V., S.M.C., and A.D. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. C.P. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: v.v.vasconcelos@uva.nl.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2102153118/-/DCSupplemental.

Published December 6, 2021.

diverse groups show growing grassroots efforts to initiate broad social change (11), polarization may complicate or entirely erode the possibility for society to coordinate on new or welfare-enhancing norms by interrupting the social dynamics that allow norms to spread (10). Understanding the effects of polarization, in its various forms, on the ability of societies to coordinate is, thus, a precursor to understanding how and when polarization can affect positive behavioral change. In the current global context, this can, for instance, inform when social forces can lead to widespread changes in behavior and transitions toward more sustainable lifestyles.

In this paper, we develop a theoretical model to investigate the influence of polarization on societies' ability to coordinate. We consider a simple, stylized, *N*-person coordination game in which individuals face two options: a higher-value risky option that only materializes if there is sufficient investment by others and a lower-value safe option that does not depend on others' choices. This coordination dilemma has two Nash equilibria (in pure strategies), one that is efficient and another one that is not. The players can maximize their payoffs by coordinating on the welfare-superior equilibrium. Our game departs from many other related coordination games (12–20) in that the collective benefit is received only by those who have invested in the risky project. Examples are investments in new technologies whose benefits accrue primarily to the investors but are distributed unevenly. It can also mean participating in crowdfunding for a project that is only successful when a critical value is raised and the value of each project is subjective. Other analogies regard the coordination between civil and private sectors, implying a critical mass of adopters/users of the product and that its distribution infrastructure is widely available. Cross-sector coordination includes dietary changes, with convenience increasing nonlinearly with availability of different types of food, or electric vehicles and their charging stations, which benefit only those who own an electric car and for which payouts differ in, for example, distance to home. We allow not only for heterogeneity in valuations of the two outcomes and variation in the amount of information about others but also the possibility of biased information about others' valuations.

Polarization has many aspects, and here, we focus on 1) how it can create or emerge from divergent positions on issues and 2) how it can also reshape social networks and lead to opinion segregation, possibly exacerbated by some reenforcing features of those networks (21, 22). Our model of polarization is inspired by recent empirical work on different approaches to measuring and characterizing polarization. We first assume that individuals are heterogeneous in how they value the return of a particular choice or outcome. We interpret this issue-based heterogeneity as divergence in ideological or political positions ("ideological polarization," or IP) (2), which, according to some, has risen over the last decades (1). Furthermore, an individual's social network may define an in-group. The preferential treatment and assessment of the in-group relative to out-groups has been described as "affective polarization" (AP) and is, in Western democracies, usually measured as animosity among supporters of different political parties (3, 4, 23–27). While a moderate degree of IP seems tolerable or even desirable for a functioning democracy, AP is seen mostly as harmful also because the adverse effects of AP can spill over from the political arena to individual (micro)economic decision-making (28) and interpersonal relationships (29). Like IP, AP has been increasing during the last decades, especially in the United States (3). One consequence of AP is that it can lead to circumscribed in- and out-groups, which can reinforce the adverse effects of AP. This segregation into subgroups can result in individuals making decisions based on incomplete or biased information, in particular, if heterogeneity in ideological positions of a certain issue maps onto these segregated social networks. A large body of literature describes mechanisms for the emergence of modular connection structures in social networks and the clustering of preferences and beliefs in those social networks. For example, research in this special issue shows how polarization can lead to modular social structures and how multimodal distributions of opinions can emerge (30–32). We draw on this literature to inform our analysis of how heterogeneity in preferences or values interacts with segregated social networks to produce sorting of preferences into distinct local clusters, creating social spheres with divergent preferences from society at large. Such networks, in turn, affect perceptions and attitudes and vice versa. Locally circumscribed beliefs and preferences can result from the tendency of individuals to adopt opinions that they perceive to be dominant and to associate themselves with like-minded individuals (33). Extrapolation from limited social spheres of like-minded individuals can lead to biased perceptions about the attitudes and preferences of society as a whole (34). Biased network perceptions may be further exacerbated by the selective disclosure of political opinions with certain groups in order to avoid disagreements (35). In turn, biased expectations about the behavior of others can distort behavioral dynamics, especially when individuals base their own choices on those expectations. In extremis, siloed information sources, providing incomplete or strategically curated information, can distort election outcomes (36).

In our model, we mechanistically derive these biases from the overlap of segregation of network connections and clustering of opinions. Work relating the impact of network structure to the propagation of social processes has shown that multiple overlapping relations, or overlapping groups, dramatically increase the diffusion of norms (37, 38) and increase the perceived marginal gains of cooperation in information-poor environments (39). On the other hand, highly segregated (modular) networks create balkanized communities with distinct norms and little scope for interaction, interrupting the diffusion of new behaviors (38). Attachment to a particular behavior can also lead to the emergence of homogeneous groups (40). Interaction between the propagation of opinions and the network structure creates distinct temporal opinion patterns but also affects their distribution in the network—resulting in either clustered or unclustered distributions of opinions in modular networks (41). Overall, the literature identifies a range of mechanisms that can lead to heterogeneous preferences becoming clustered into distinct social modules. In this paper, we take these patterns of overlapping IP, AP, and sorting into networks as given and ask what these social configurations imply for coordination problems.

Many of the major social challenges we are confronting as a society require coordination, but they are also characterized by uncertainty and risk. For example, large investment, dietary and lifestyle changes, or sustainability transitions in general will require far-reaching and costly behavioral changes—these may only be worth engaging in if a critical mass is doing the same (9). In contexts in which there is risk associated with a particular option, especially strategic risk that depends on the actions of others, individuals' choices depend on their expectations about others (42). Thus, how those expectations are affected by IP, AP, and the intersection of the two is crucial for understanding how coordination on "virtuous" or welfare-enhancing equilibria may be interrupted or facilitated in polarized societies. In our work, we focus on random patterns—in which heterogeneous opinions on an issue are evenly distributed across social networks—and clustered patterns—in which individuals with identical opinions tend to be connected, potentially biasing their view of the broader society. We study the likelihood of coordination in different societies using a model in which behaviors are based on opinions. Individuals have heterogeneous rewards associated with the socially optimal outcome—a

proxy for differing opinions or proclivities in the population toward certain coordinated investments, which we associate here with IP. Furthermore, individuals may only have access to local information about their neighbors or their connections. We look at how different network configurations, both in terms of the amount and representativeness of information, which we associate with AP, affect the ability to reach the social optimum. The model results show, unsurprisingly, that successful coordination is highly dependent on the initial intentions with which the players start the game. For this reason, we use the result of an experiment as input to specify the initial intentions in the different polarization conditions and compare the model outcomes across the different conditions.

Results

Consider a population in which individuals decide between two actions: a safe action, S, which provides a fixed benefit, b, and a risky action, R, with a higher potential payoff, o_i, but only if enough, M, others also choose this option. If the number of individuals choosing action R falls below M, those who have taken the risk receive no benefit. Here, we focus on the case in which a large fraction of the population is required for the investments to be successful, representing a need for society-level coordination. This potentially large benefit of individual i, o_i, may differ between players. It captures the subjective valuation of coordinating toward action R, which may vary with risk preferences but also ideological positions or moral values, among other factors. Individuals can, thus, be heterogeneous or homogeneous in their preferences, which is reflected in the distribution of subjective valuations (Fig. 1*A*). We focus on the cases in which heterogeneity, as a result of IP, is reflected as bimodality in the valuations of R. Coordinating toward states in which either no one or everyone chooses R are both Nash equilibria. However, the equilibrium in which everyone chooses R is both an individual and social optimum. In this sense, choosing R is a cooperative act, as it allows others to achieve the best outcome. Doing so, however, depends on the expectations one has that others will also choose R. In our model, we allow individuals to signal their intended strategy and to update it based on the signals they receive from others. We establish a simple relationship between the confidence level that others will play R and the information individuals get about others. We set the expectation about others' behaviors as the subjective probability that each of the others will choose R given by the fraction of R signals an individual observes. Individuals use this probability to estimate the expected payoff of R versus S. They signal the strategy with the highest expected payoff, though there is some noise in this signal. Because each player has a different valuation, and possibly a different neighborhood, the signals they send and receive will differ from those of others. Apart from errors, all signaling is honest. Individuals' valuations, o_i, are derived from a distribution that is characterized in Fig. 1*A*, according to its bimodality and variances. See *Materials and Methods* for additional details.

Social Network Structure. We define two types of social networks in our society that vary in their degree of cohesion or segregation (Fig. 1*B*). In cohesive societies, individuals are well connected and thus receive signals about the preferences of a significant portion of the group, which they use to update their beliefs and expected payoffs of playing R or S. In segregated societies, individuals are connected to a circumscribed group and thus update their expectations about playing R or S based on the limited (and potentially biased) information from that group. Here, we focus on polarization that leads to the segregation of the population into two subgroups of equal size ($z = 1/2$).

Mapping of Valuations onto the Social Network. Finally, we consider how the distribution of subjective valuations interacts with the network structure. In particular, in segregated networks, valuations can be either randomly distributed in the population or be biased by the network (neighbors tend to have similar

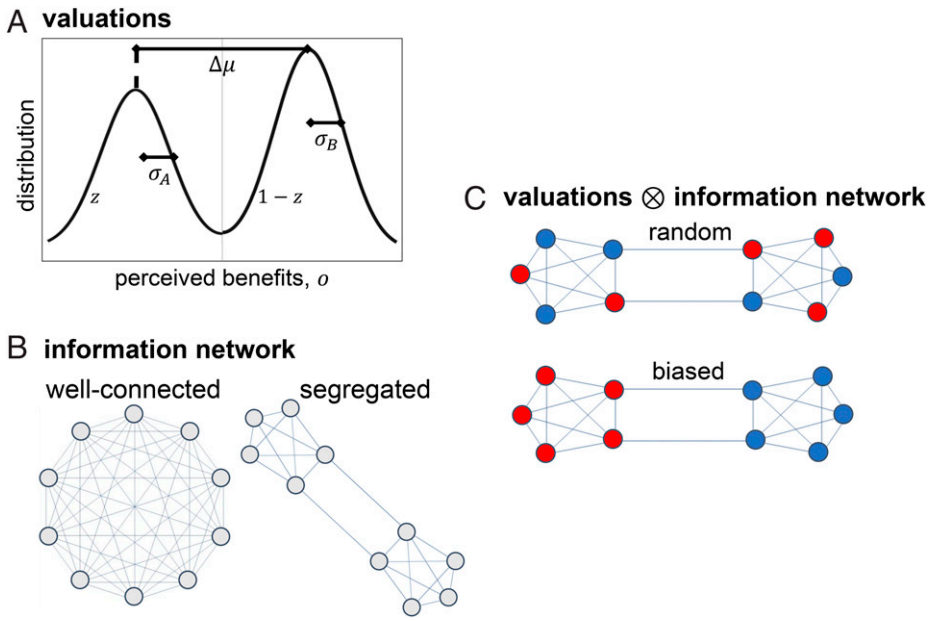

Fig. 1. Configurations of polarization. (*A*) The distribution of valuations is a consequence of the perceived benefits. This distribution is characterized by a mean, μ, a distance between peaks, $\Delta\mu$, and the weight, z or $1-z$, and variance, σ_A and σ_B, of each peak. Thus, the distribution can be homogeneous ($\Delta\mu = \sigma_A = \sigma_B = 0$) or heterogeneous ($\Delta\mu, \sigma_A = \sigma_B = \sigma > 0$). This is a proxy for IP. (*B*) Information about what others do reaches individuals through their social networks. One aspect of AP is that it leads to bottlenecks in information flow. (*C*) When valuations are heterogeneous, they can be random (randomly distributed in the population) or biased (clustered in neighborhoods with similar perceived benefits). This captures the interaction of IP and AP, which, together with sorting into networks, results in groups of like-minded people.

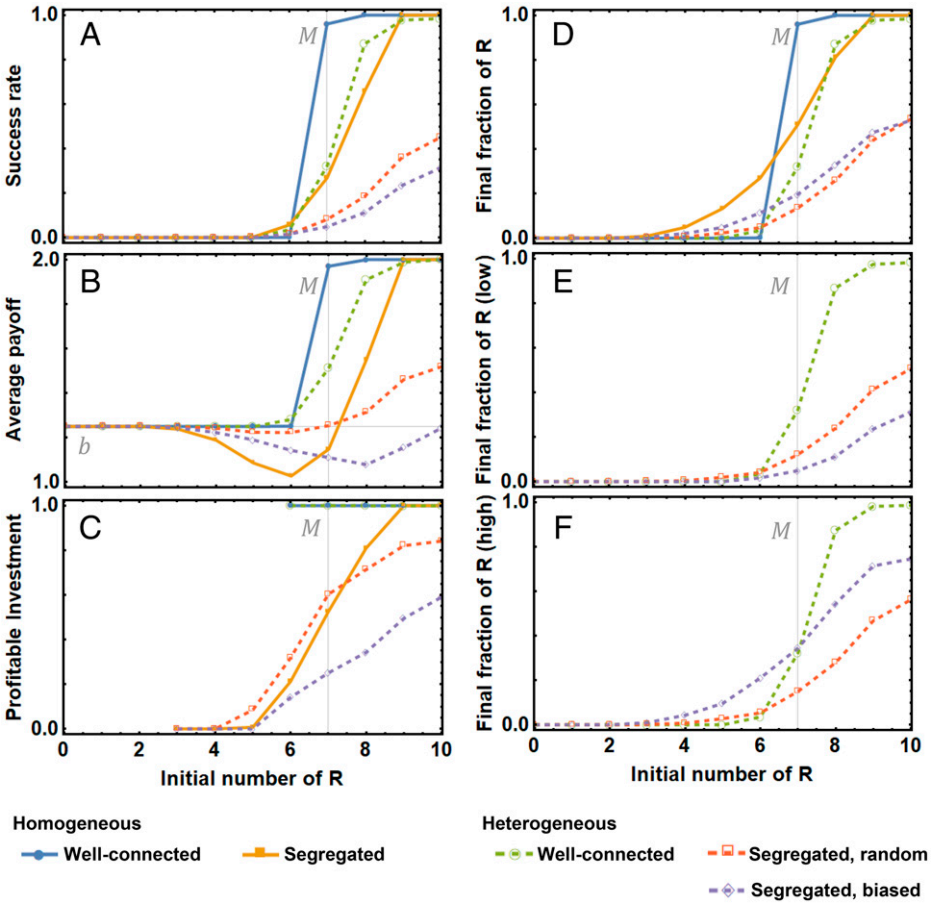

Fig. 2. Impact of polarization on coordination. (*A*) Success rate, computed as the fraction of simulations that terminate with at least M R-players, making R profitable. (*B*) Average payoff, calculated over all players and simulations. (*C*) Profitable investment is the fraction of R-players who achieve their best outcome, o_i. Notice that some points do not exist, as the simulation did not register any investments for those values. (*D*) Fraction of R-players in the population. (*E*) Fraction of R-players in the half of the population with the lowest valuations for R. (*F*) Fraction of R-players in the half of the population with the highest valuations for R. Quantities are the average final outcome of 10^4 simulations with different initial distributions of the location of R-players in the network, of sampled o_i, and networks. Equilibrium was assumed after 10^3 time steps. Simulations run for population size $Z = 10$, baseline benefit, $b = 1.25$, a threshold for achieving the collective outcome of $M = 7$ players, average valuation of collective outcome μ=2, and, for the heterogeneous valuations, $\Delta\mu = 1$, $z = 1/2$, and $\sigma = 0.01$ (Fig. 1). Individual errors in determining the optimal strategy, depending on what the others are signaling, are controlled by $\beta = 20$. In *SI Appendix*, we provide an analytical analysis of the model in the absence of mistakes ($\beta \to \infty$) and large populations (*SI Appendix*, Sections 1–4), and we explore variations of M and β (*SI Appendix*, Section 7).

perceived valuations). If the valuations are heterogeneous and biased, then, although the society as a whole has diverse valuations, each subgroup is more homogeneous in their valuations (Fig. 1*C*).

Overall, we interpret the segregation of the social network as a consequence of AP, with, for example, identity defining tightly connected groups of individuals. High levels of AP and IP in the absence of sorting produce random segregated networks. In turn, clustering of ideological preferences occurs when there is also sorting, which is presumed the case when ideological preferences align with group identity. Thus, segregated biased networks are the reflection of IP, AP, and sorting.

Behavior for Different Manifestations of Polarization. Fig. 2 shows the impact of the different variants of polarization (from Fig. 1) on different success metrics for the population. We chose three metrics: The success rate in coordination (Fig. 2*A*), which is the fraction of groups with successful investments (i.e., with a minimum number M of R-players); the average payoff of individuals (Fig. 2*B*); and the fraction of risky investments in R that are successful (Fig. 2*C*), which corresponds to the fraction of R-players who receive their valuation (as opposed to a return of 0). Fig. 2 *D*–*F* represent the equilibrium behavior of the whole and, for the heterogeneous populations, subparts of the population. We consider these success metrics as a function of the initial number of R-players, which we distribute randomly in the network. Notice that, as players reevaluate their signals, correlations between the signaled strategies and valuations, o_i, are created. The results reflect the coordination nature of the problem, with high dependence on initial conditions.

Let us assume that the initial behavior does not depend on our variants of polarization (valuations distribution, information network structure, and sorting into networks) and that we can thus compare the different scenarios for identical initial conditions. As illustrated in Fig. 2*A*, we find that heterogeneous valuations decrease the chance of successful coordination. Not only is coordination achieved less often, but total welfare drops as compared to the homogeneous scenarios, as in Fig. 2 *B* and *C*. The exception is when groups have insufficient initial R-players, which hampers coordination in all scenarios. When valuations are heterogeneous, the clustering of valuations in the social networks becomes a crucial determinant of whether the welfare-superior equilibrium can be achieved. Societies that

have heterogeneous and segregated preferences end up with lower welfare, even below the baseline value *b,* the dashed line in Fig. 2*B*. This can be understood in light of the fraction of successful investments, represented in Fig. 2*C*, which drops even for high initial values of investment signals. This is exacerbated when diverse valuations are also agglomerated in segregated networks so that, within circumscribed groups, there are systematically similar valuations for *R* compared with those of the society as a whole. In Fig. 2 *D–F*, we show the fraction or R-players in Fig. 2*D* for the whole population. The fractions for heterogeneous populations are split between the subset of players with low valuations of *R,* shown in Fig. 2*E*, and, in Fig. 2*F*, the subset with high valuations. We see how these last two highlight miscoordination as the main driver of lower welfare, particularly in the case of segregated individuals in biased networks, with the purple line being much higher for high valuation individuals than for low valuation individuals.

In *SI Appendix*, we verify the robustness of these results across different parameter values, analyzing the outcomes of the model analytically and computationally. First, we quantify analytically the extent to which the level of diversity of valuations can affect the dynamics in fully connected populations for any distribution of valuations. We show that the distribution of valuations translates (monotonically but nonlinearly) into a distribution of behavioral change thresholds (*SI Appendix*, Section 1). We show that some degree of diversity of preferences can be beneficial. Diversity can lead to a distribution of thresholds that triggers a cascade of social change—in particular, a sequence of thresholds in which at least one individual requires just one more individual to change behaviors in order to consider changing their own behavior. This means that small initial shifts in behavior can "tip" society from the all-S to the all-R equilibrium. In this perspective, the optimal distribution of thresholds is not, however, bimodal, but closer to uniform. Diversity can also reduce the costs of interventions aimed at shifting society from one equilibrium to another (*SI Appendix*, Sections 3 and 4). We consider a situation in which an "innovator" or policymaker can intervene in a system to make a nondominant but desirable option more appealing—for example, through the introduction of a new option, subsidies, or strategic messaging. In a context with heterogeneity in valuations, there will be some individuals or groups who prefer the nondominant option but who, due to social influence and pressures to conform, persist on the prevailing social norm. At the most extreme end, an intervention targeting a single individual can trigger a cascade that sets in motion a sequence of behavioral changes without any additional costs or need to intervene on any other individuals. To borrow terminology from the diffusion of innovations literature (43, 44), it may take only an initial and modest intervention to drive these "early adopters" to shift their behaviors. Through this endogenous social change process, these early adopters may "tip" an "early majority," followed by the "late majority" and, finally, the "laggards." Consistently, in the space of bimodal distributions characterized in *Results* in the main text, we show the existence of an optimal distribution of valuations with large variance (*SI Appendix*, Section 5). Then, we allow that distribution to shape the connections between individuals. The benefit of heterogeneity decreases once we allow networks to dynamically become segregated, either through differences in opinions/valuations or actions. These results suggest that the positive effects of IP disappear if IP coexists or even leads to the network effects of AP or vice versa (*SI Appendix*, Section 6). For low thresholds (i.e., thresholds that do not require coordination at the societal level but only at a subgroup level [$M/Z < z$]) sorting of valuations in the segregated networks can be beneficial, as it facilitates the coordination of the subgroup.

Assessing Initial Intentions. The analysis so far assumes that initial conditions do not vary with the different manifestations of polarization. However, in reality, individuals might anticipate different levels of willingness to invest in risky options depending on their local context. To incorporate more realistic initial intentions in the model and make the different scenarios directly comparable, we performed an online experiment with a sample of 750 individuals drawn from the Prolific platform. Participants play a one-shot coordination game in groups of 10 players. In each game, players choose between playing a black card and a red card. Playing the black card corresponds to *S* (the safe option) and returns £1.25 no matter what the other players do. Playing the red card returns o_i if at least 7 out of the 10 players play the red card and zero otherwise, corresponding to *R* (the risky option). A player's o_i depends on the experimental treatment. In the homogeneous treatments, all players in the group have the same o_i, which is either low (£1.50), medium (£2.00), or high (£2.50). In the heterogeneous treatments, there are five high-value players in the group who have an o_i of £2.50 and five low-value players who have an o_i of £1.50. Notice that these values correspond to the parameters of the model in Fig. 2.

We further distinguish between different levels of information available to players when making their decision. In the full information condition, players know all the o_i values in the group, including their own, which represents the scenario in which individuals are "well connected." In the partial information condition, players are informed about the o_i values of 4 out of the 10 players, including their own. The information is representative of the actual distribution of the o_i values in the group. This implies that participants in the heterogeneous treatment are informed that two players in the group have a high o_i value and two other players have a low o_i value, while the valuations of the remaining six players remain unknown to them. This condition represents the "segregated random" scenario because individuals have only partial information, but it can be used to predict the actual distribution of o_i values in the group. In the clustered partial information condition, players are again informed about four o_i values, including their own, but the information is strongly biased in the direction of their own o_i. Specifically, high-value players are informed that there are four high-value players in the group, including themselves, while low-value players are informed that there are four low-value players in the group, including themselves. The valuations of the remaining six players remain unknown. This condition represents the "segregated biased" scenario and is only applicable in the heterogeneous treatments. Members of the same group get different information about others' valuations, which is not possible in the homogeneous treatments in which all players have the same valuation. All participants play three games, keeping the same o_i value but with varying levels of information and different groups. We randomly varied the order of the three games across participants to control for spillover effects. One game was chosen at random to determine their earnings (see *Materials and Methods* and *SI Appendix* for further details and related experimental results from the previous literature).

Fig. 3 summarizes the outcomes of the model, taking the experimental results as initial values for the simulation, using the same parameters in the experiment and the simulations. The experimental results are indicated as dashed lines, which are taken as the initial signals and then evolve toward the outcomes of the model, indicated in the bars. First, we notice that the ranking of the experiment, in terms of investment under different polarization conditions, is in line with the ranking predicted by the model from random initial conditions (see Fig. 2 and the direct comparison in *SI Appendix*). In the homogeneous treatments, the investment level is similar in medium and high valuation conditions and lower for the low condition.

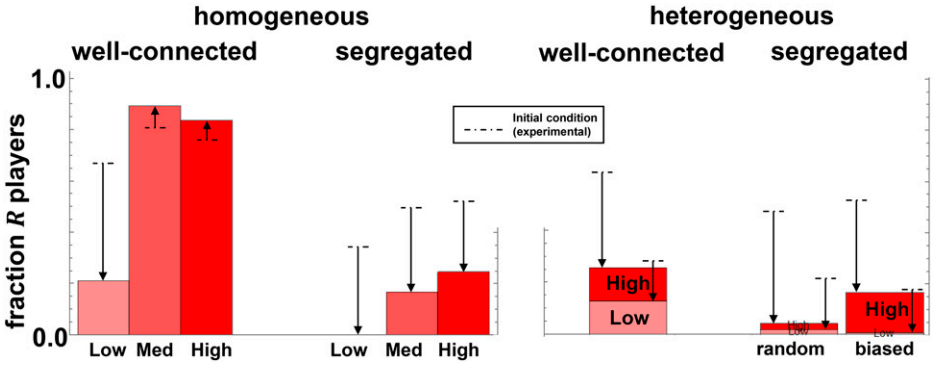

Fig. 3. Comparison across manifestations of polarization using experimental results to set initial signals. Fraction of R-players after many interaction rounds, using the same parameters as in Fig. 2, which match the experimental setup. The initial number of R-players is set as a hypergeometric sampling of the players in the different polarization conditions of the experiments. That corresponds to the average value indicated as a dashed line, which, as the simulation runs, evolves to the corresponding value indicated in the bar.

Participants in the low condition are susceptible to uncertainty about others' valuations. Participants in the medium and high conditions also reduce their investments in the presence of uncertainty, though their behavior is more robust to this change. In the heterogeneous treatments, uncertainty about others' valuations reduces the overall investment level. Also, high-value and low-value players invest similarly when they are randomly segregated, while biased segregation drives a wedge between their investment behavior. Under biased valuation conditions, high-value players tend to invest more and low-value players invest less compared to random segregation.

In order to get a sense of the extent to which the different polarization conditions shaped the expectations about other players, after each game, we included an incentivized elicitation of participants' expectations about others' behaviors and their confidence in this guess. The analysis of these data shows that the available information significantly influences participants' expectations about others and that this moderates the effects of information on their behavior (see *SI Appendix* for details). Individual choices were highly correlated with expectations about others in all experimental conditions. Interestingly, the correlation is somewhat higher for those who are well connected and thus have more information than those who are segregated. We also observe that optimistic players—those who expect many red cards—tend to be more confident in their assessments than pessimistic players, which may help to explain the relatively high investment level in the experiment. These findings generally support the crucial role of expectations and trust, but they also indicate that uncertainty does more than merely reducing expectations—it also disturbs the connection between expectations and action. Another possible reason for the relatively high investment level is due to the sample used. Comparisons between Prolific and other samples, such as MTurk, have shown that Prolific users are more naive about standard experimental games, are more honest in cheating games, use the platform less, and rely less on the earnings (45, 46). In this sense, our results could be considered conservative, as they may overestimate the chance of successful coordination and underestimate the adverse effects of polarization compared to other samples.

In *SI Appendix*, we show a direct comparison between the model with uniform random initial conditions and the model with the initial signals of the experiments. The initial signals we obtain from the experiments and the equilibrium behavior from identical initial conditions show the same directional impacts when changing the manifestations of polarization. This means that taking the experimental results as the initial signal reinforces the differences between different manifestation of polarization found with random initial conditions.

Discussion

Many urgent societal issues are collective action problems, requiring widespread coordination on risky options to achieve welfare-enhancing outcomes. For example, avoiding irreversible impacts of climate change will require a fast global transition to a net-zero carbon emissions economy. This will not only take massive technological changes and financial commitments, but it will also take a shift in prevailing social norms. This paper focuses on the consequences of different forms of polarization for achieving a socially efficient outcome in a context in which potential benefits accrue only if enough individuals commit to a risky high-value option over a safe lower-value option.

Polarization is a multifaceted concept (2). We here model polarization along three central dimensions in the literature and consider only cases in which all members of society gain from social progress. In particular, we consider 1) heterogeneous preferences or policy positions drawn from a bimodal distribution (47), which we take as a proxy for the divergent ideological views suggested by IP, 2) organization of social or information networks into segregated modules, which could emerge from AP, and 3) sorting of policy positions or preferences into segregated modules, which we take as the intersection between IP and AP, in which either the causes of IP on a certain issue and AP are identical or the network structure leads to a clustering of preferences within the network (48–50). Overall, we show theoretically and experimentally that preference heterogeneity does not substantially interrupt coordination, while segregation of social network connections, especially when combined with a clustering of preferences on these networks, does erode coordination. The model and the experiment suggest that the best outcomes arise when societies are well mixed and that this is especially important when some members of society earn limited benefits from coordinating. The erosion of coordination arises from the clustering and segregation of like-minded individuals, which can result in biased expectations about the broader population.

Our findings suggest one plausible mechanism for pluralistic ignorance—the misperception that there is broad support for a norm or belief that is actually rejected by most—a widespread phenomenon that has been attributed with inaction on many topics including racial segregation, climate change mitigation, or binge drinking (34, 51–53). The finding that coordination is interrupted by biases that arise from the sorting and clustering of like-minded individuals suggests some pathways for facilitating coordination in polarized societies. For example, highly visible and bipartisan institutions and elites may be avenues for increasing coordination among disparate groups. To the extent that institutions create overlaps or bridges between otherwise segregated social groups, they can act as coordination devices,

correcting the misperceptions that result from local information networks (54–56). However, research suggests that the broad signals provided by these all-encompassing institutions need to be directed to individuals and not groups in order to affect actions (57). To this end, while consumption of partisan media has been shown to increase partisan hostility, sizeable public broadcasting stations tend to be correlated with lower AP (3). Related literature has found that exposure to diverse political views can exacerbate bias and that political cues erode social learning but that structured bipartisan networks that are devoid of political cues can increase social learning across partisan groups (58). In this volume, Santos et al. show using a model that tuning automated link recommendations might correct some of the tendencies for self-sorting and segregation in networks (59).

Our research has shown that initial conditions are also very important for successful coordination, which raises the question of what factors determine initial conditions. In particular, much research has shown that cultural background is an important determinant of cooperation in social dilemmas and economic preferences, including time and risk preferences, reciprocity, altruism, and trust, which are, in turn, correlated with economic outcomes and behaviors (60, 61). This same work also finds substantial within-country individual-level heterogeneity in cooperation and economic preferences, depending on factors like age, gender, and cognitive ability. While some influencing factors, such as cultural history or demographic characteristics, are hardly changeable, others can be influenced by policy. For example, public service broadcasting and education initiatives can improve communication and create shared beliefs and value systems across disparate groups, increasing the perceived benefits and ease of coordination. Furthermore, measures that reduce the cost of investment, for example, through subsidies, or the risk of stranded assets can also facilitate coordination by altering the incentive structure.

While we have attempted to capture some key features and consequences of IP and AP in this first step, there remain several fruitful avenues for future research. The model should be extended to account for additional features of polarization, such as greater intensity of both IP and AP, complex feedback between the two, and social identity or affective components of AP, including out-group animosity (2). Affective responses to in- and out-groups create intergroup conflict, making coordination among multiple groups challenging (62, 63). Out-group hate has been shown to impede cooperation in part by creating grim expectations about others' behaviors (64). Literature in cultural evolution, anthropology, and political psychology suggests that, when actions or beliefs are closely coupled with identities and those actions or beliefs come under threat, there might even be a doubling down on the inefficient equilibrium and related attempts to protect one's worldviews (65–67). Although recent work on the topic finds limited evidence of this sort of backlash in a series of surveys, some suggest that individuals may become distrustful of counter attitudinal arguments in antagonistic contexts (68). One way to elicit stronger affective responses is to consider a societal transition that leaves some groups worse off as compared to the status quo. This may lead some groups to perceive themselves to be under threat and to resist the transition; it may also mean that these groups need to be compensated to ensure an equitable transition (e.g., consider fossil fuel workers in a transition to a net-zero economy) (69, 70). These considerations are important, as perceived inequities can change over time with economic growth and decline (71) and can drive greater AP (4)—just as AP can, in turn, also increase inequities. Indeed, recent work has shown that poverty and inequality undermine cooperation, social cohesion, and trust and that they interact with other group differences such as race or ethnicity (72, 73). Understanding the distribution of the most affected will also be crucial to quantify their overall impact in the coordination process. Other feasible extensions include the modeling of asymmetries—that is, uneven clustering of groups and opinions, resulting in asymmetries in how information is distorted—and greater heterogeneity. Very large inequalities, resulting from greater heterogeneity, could have substantial welfare costs and create moral dilemmas. We have limited these considerations in our settings to initially understand the consequences of even moderate levels of polarization.

A second avenue for future research is extending the social dilemma under consideration. Here, we chose to focus on preference heterogeneity and key structural features of polarization in a stylized coordination problem. As a first pass, we abstract away many of the subtleties of real-world social dilemmas and institutional design. Preference heterogeneity may come not just from variation in opinions but other sources, including variation in self-interest, local conditions, including geographic conditions, and others. Settings in which the socially efficient outcome is not an equilibrium or in which actors who defect to the nonrisky option still benefit from the risks taken by others represent more difficult collective action problems that may be exacerbated by polarization (and increase it). An example, and of particular relevance for sustainability challenges, is recent work showing that uncertainty about when a threshold needs to be reached can result in uneven or "polarized" outcomes (20). In these more complicated settings, strategic signaling might prove particularly relevant. Additionally, while we do explore the impacts of different thresholds in the model (see *SI Appendix* for details), we leave the consideration of nonthreshold systems for future research. Furthermore, the feedback between revealed outcomes, inequality, and polarization, or other social dynamics, is an avenue that could be explored by extending our one-shot coordination game to a repeated game. The model does allow actors to update what they are signaling based on the information in their networks, but it does not currently allow for reciprocity, reflecting the one-shot nature of our analysis. Extending the experiments and the model to allow for this would create the opportunity to study the erosion of coordination—vicious cycles in which those who have taken risks and feel cheated might punish those who opted for the safe option. It would similarly offer the opportunity to understand under what conditions repeat interactions might facilitate the emergence of virtuous norms, for example, through the emergence of leadership or the establishment of trust or institutions (74, 75). Future work might systematically study how sequential behavioral responses can shape preferences, expectations, and even information networks.

In summary, our study reveals important effects of polarization that can serve as a lower bound estimate on the effects of polarization while also providing a framework for further research in this area. Our model and experimental setup provide the scaffolding for future extensions along several dimensions and suggest plausible interventions to help ameliorate the consequences of polarization.

Materials and Methods

Model. Consider a population with Z individuals. Each individual has to decide between actions R or S. The benefit individual i associates with coordinating on R depends on the subjective valuation i has about that action, o_i. In a group of N individuals, if M choose to invest in R, then each earns a large benefit, o_i. Individuals who select the safe option S get a smaller baseline benefit, b. If k is the number of individuals playing R, and $\Theta[x]$ the unit-step function that is 0 for $x < 0$ and 1 otherwise, we can write the payoff of each action as

$$\Pi_i^R[k] = \Theta[k - M]o_i \text{ and} \quad [1.1]$$

$$\Pi_i^S = b. \quad [1.2]$$

The equilibrium in which all players choose R is payoff dominant given $o_i \geq b\ \forall i$, while the equilibrium in which all players choose S is risk dominant if $o_i < 2b\ \forall i$. This sets up a coordination dilemma in which the safe choice is to choose S, but if one expects that a sufficient number of others will choose R, then it is worthwhile to do so too. Thus, playing R is a cooperative action, as it

may offer a larger personal benefit but also makes it more likely that others will achieve their maximum outcome. The best course of action is thus dependent on one's expectations about others taking action R.

At the start, individuals' subjective valuations, o_i, are sampled from a bimodal distribution with peak-to-peak distance $\Delta\mu$ and peak variance σ_A and σ_B for the left and right peaks, respectively (Fig. 1*A*). Peak A contains a weight z, whereas peak B contains a weight $1-z$. To achieve this, we define the distribution of valuations as a mix of two normal distributions, $\rho(o) \sim \text{MixtureDistribution}[z\ \text{Normal}[\mu-\Delta\mu/2, \sigma_A] + (1-z)\text{Normal}[\mu+\Delta\mu/2, \sigma_B]]$. For simplicity, we set $z = 1/2$ and $\sigma_A = \sigma_B = \sigma$.

Let t_i be *i*'s belief that each of the other players will take action R. We assume individuals develop these expectations based only on the information they get from their surroundings and build from it a likelihood that others will choose R. In a population of size Z, the expected payoff of R versus S for this player is

$$\begin{aligned}\Delta_i &= \sum_{k=0}^{Z-1}\binom{Z-1}{k} t_i^k (1-t_i)^{Z-1-k}\left(\Pi_i^R[k+1] - \Pi_i^S\right) \\ &= o_i \sum_{k=M-1}^{Z-1}\binom{Z-1}{k} t_i^k (1-t_i)^{Z-1-k} - b.\end{aligned} \quad [2]$$

Individuals signal their strategy based on what they see around them, such that t_i corresponds to the fraction of neighbors they observe signaling R in their visible network (see Fig. 1*B*). Based on this, they chose to signal R with probability $\left(1+e^{-\beta\Delta_i}\right)^{-1}$ and S with complementary probability. The value of β controls for mistakes in this signaling process, with $\beta \to \infty$ representing a deterministic signal of the best response and $\beta \to 0$ a coin flip. Individuals start with a random signal, according to the set fraction of initial R signals. An individual is selected to update their signal at each time step, and the process is repeated until an equilibrium is reached. Rewards are distributed based on the last signal.

We define a social network in which nodes represent individuals and edges the information flow of signals. Individuals who are connected see each other's signals. Well-connected networks correspond to a complete graph, in which all nodes are connected to all other nodes. The segregated networks are generated by starting with two Barabasi–Albert random graphs of average degree 5 and adding 0.1 Z links between the two. We have two types of segregated networks: in the segregated unbiased network, we distribute o_i randomly; in the segregated biased network, we sort o_i such that $o_i \leq o_{i+1}$ and place the lower half of valuations in the first cluster and the higher half of valuations in the second cluster.

Experimental Design and Procedure. We obtained Institutional Review Board approval from Princeton University, and all participants provided informed consent before participating in the study. We conducted the study online with a sample of Prolific users. Prolific is an on-demand online platform, which allows researchers to run surveys and experiments with a large number of people. Participants played a coordination game in groups of 10. The game was played asynchronously, with anonymous group members and no information about players' identities. Our experiment had two treatment layers: we used a between-subject design for players' o_i values and whether there was homogeneity or heterogeneity in the o_i values (proxy for IP) and a within-subject design for the information levels (proxy for AP). This means that all participants played three different one-shot games in which their own o_i valuation associated with playing R and the homogeneity or heterogeneity in the group remained the same throughout, but the information about the o_i values of the other players in the group varied between games. As biased information is not possible in the homogenous games, participants played one game with no information about the o_i values of the other players (see *SI Appendix* for the results of this condition). We randomly varied the order of the three games across participants to control for spillover effects. Participants were told that the group composition would change between games and that, at the end, one of the three games would be randomly selected for payment. A total of 750 individuals took part in the experiment, ∼150 in each of the five o_i value conditions (high, medium, and low in the homogenous treatments, high and low in the heterogenous treatments). About half of the participants are students, and one-third are women. The average age is 26 y old. The number of participants per value condition was determined using a power analysis based on a pilot experiment with 60 Prolific users in the heterogeneous treatment—these data were not included in the final dataset.

In all treatments, players were aware of the size of the group, the threshold, and their own o_i value. What they knew about the o_i values of the other players differed from game to game. *SI Appendix*, Table S3 shows the available information per player and treatment. After participants had made their decision in a game, they were asked to guess how many of the other players would play the red card and to indicate how confident they felt about their guess. Correct guesses in the selected game were incentivized with an additional bonus of £0.50. Further details on the experimental design and an instruction sample are provided in *SI Appendix*.

Data Availability. Anonymized data have been deposited in the publicly accessible database Harvard Dataverse (https://doi.org/10.7910/DVN/PMN61R) (76). All other study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We would like to acknowledge The College of Liberal Arts and Sciences at Arizona State University for providing the funding for the workshops that led to this publication as well as the constructive feedback of the participants of the workshop "Political Polarization Meeting." S.M.C. acknowledges funding from the Center for Policy Research on Energy and the Environment at the School of Public and International Affairs at Princeton University. V.V.V. acknowledges funding from the Princeton Institute for International and Regional Studies, Rapid Switch Community. V.V.V. and S.L. acknowledge funding by the NSF Grant GEO-1211972 and the Army Research Office Grant W911NF-18-1-0325. V.V.V., E.W., and S.L. acknowledge funding from the Princeton Data-Driven Social Science Initiative Small-Scale grant on the project "Social norms dynamics as a complex adaptive system." A.D. and M.L. acknowledge funding from the University of Kassel through the project "The Relationship between Environmentally-relevant Behavior and the Development of Values and Norms (ZumWert)."

1. A. I. Abramowitz, K. L. Saunders, Is polarization a myth? *J. Polit.* **70**, 542–555 (2008).
2. Y. Lelkes, Mass polarization: Manifestations and measurements. *Public Opin. Q.* **80**, 392–410 (2016).
3. L. Boxell, M. Gentzkow, J. Shapiro, *Cross-Country Trends in Affective Polarization* (National Bureau of Economic Research, 2020).
4. N. Gidron, J. Adams, W. Horne, *American Affective Polarization in Comparative Perspective* (Cambridge University Press, Elements in American Politics, 2020).
5. M. Granovetter, Threshold models of collective behavior. *Am. J. Sociol.* **83**, 1420–1443 (1978).
6. S. Gavrilets, P. J. Richerson, Collective action and the evolution of social norm internalization. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6068–6073 (2017).
7. H. P. Young, The evolution of social norms. *Annu. Rev. Econom.* **7**, 359–387 (2015).
8. P. Sharkey, The US has a collective action problem that's larger than the coronavirus crisis, *Vox*, 10 April, 2020. https://www.vox.com/2020/4/10/21216216/coronavirus-social-distancing-texas-unacast-climate-change. Accessed 10 September 2021.
9. K. Nyborg *et al.*, Social norms as solutions. *Science* **354**, 42–43 (2016).
10. C. Efferson, S. Vogt, E. Fehr, The promise and the peril of using social influence to reverse harmful traditions. *Nat. Hum. Behav.* **4**, 55–68 (2020).
11. D. McAdam, Social movement theory and the prospects for climate change activism in the United States. *Annu. Rev. Polit. Sci.* **20**, 189–208 (2017).
12. R. Croson, M. Marks, Step returns in threshold public goods: A meta- and experimental analysis. *Exp. Econ.* **2**, 239–259 (2000).
13. E. van Dijk, H. Wilke, M. Wilke, L. Metman, What information do we use in social dilemmas? Environmental uncertainty and the employment of coordination rules. *J. Exp. Soc. Psychol.* **35**, 109–135 (1999).
14. C. Schill, T. Lindahl, A.-S. Crépin, Collective action and the risk of ecosystem regime shifts: Insights from a laboratory experiment. *Ecol. Soc.* **20**, 48 (2015).
15. T. Lindahl, A.-S. Crépin, C. Schill, Potential disasters can turn the tragedy into success. *Environ. Resour. Econ.* **65**, 657–676 (2016).
16. S. Barrett, A. Dannenberg, Climate negotiations under scientific uncertainty. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17372–17376 (2012).
17. S. Barrett, A. Dannenberg, Sensitivity of collective action to uncertainty about climate tipping points. *Nat. Clim. Chang.* **4**, 36–39 (2014).
18. M. Milinski, R. D. Sommerfeld, H.-J. Krambeck, F. A. Reed, J. Marotzke, The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2291–2294 (2008).
19. A. Tavoni, A. Dannenberg, G. Kallis, A. Löschel, Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11825–11829 (2011).
20. E. F. Domingos *et al.*, Timing uncertainty in collective risk dilemmas encourages group reciprocation and polarization. *iScience* **23**, 101752 (2020).
21. F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, M. Starnini, Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.* **124**, 048301 (2020).
22. C. R. Sunstein, *Republic: Divided Democracy in the Age of Social Media* (Princeton University Press, 2018).
23. S. Iyengar, G. Sood, Y. Lelkes, Affect, not ideology: A social identity perspective on polarization. *Public Opin. Q.* **76**, 405–431 (2012).
24. E. Knudsen, Affective polarization in multiparty systems? Comparing affective polarization towards voters and parties in Norway and the United States. *Scand. Polit. Stud.* **44**, 34–44 (2021).
25. A. Lauka, J. McCoy, R. Firat, Mass partisan polarization: Measuring a relational concept. *Am. Behav. Sci.* **62**, 107–126 (2018).
26. A. Reiljan, 'Fear and loathing across party lines' (also) in Europe: Affective polarisation in European party systems. *Eur. J. Polit. Res.* **59**, 376–396 (2020).

27. M. Wagner, Affective polarization in multiparty systems. *Elect. Stud.* **69**, 102199 (2021).
28. C. McConnell, Y. Margalit, N. Malhotra, M. Levendusky, The economic consequences of partisanship in a polarized era. *Am. J. Pol. Sci.* **62**, 5–18 (2018).
29. S. Iyengar, Y. Lelkes, M. Levendusky, N. Malhotra, S. J. Westwood, The origins and consequences of affective polarization in the United States. *Annu. Rev. Polit. Sci.* **22**, 129–146 (2019).
30. R. Axelrod, J. J. Daymude, S. Forrest, Preventing extreme polarization of political attitudes. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102139118 (2021).
31. O. J. Chu, J. F. Donges, G. B. Robertson, G. Pop-Eleches, The microdynamics of spatial polarization: A model and an application to survey data from Ukraine. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104194118 (2021).
32. C. K. Tokita, A. M. Guess, C. E. Tarnita, Polarized information ecosystems can reorganize social networks via information cascades. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102147118 (2021).
33. R. Durrett, S. A. Levin, Can stable social groups be maintained by homophilous imitation alone? *J. Econ. Behav. Organ.* **57**, 267–286 (2005).
34. M. Mildenberger, D. Tingley, Beliefs about climate beliefs: The importance of second-order opinions for climate politics. *Br. J. Polit. Sci.* **49**, 1279–1307 (2019).
35. S. Cowan, D. Baldassarri, It could turn ugly: Selective disclosure of political views and biased network perception. *Am. Sociol. Rev.* **52**, 1–17 (2016).
36. A. J. Stewart *et al.*, Information gerrymandering and undemocratic decisions. *Nature* **573**, 117–121 (2019).
37. P. M. Blau, J. E. Schwartz, P. M. Blau, *Crosscutting Social Circles* (Routledge, 2018).
38. D. Centola, The social origins of networks and diffusion. *AJS* **120**, 1295–1338 (2015).
39. V. V. Vasconcelos, P. M. Hannam, S. A. Levin, J. M. Pacheco, Coalition-structured governance improves cooperation to provide public goods. *Sci. Rep.* **10**, 9194 (2020).
40. P. R. Ehrlich, S. A. Levin, The evolution of norms. *PLoS Biol.* **3**, e194 (2005).
41. V. V. Vasconcelos, S. A. Levin, F. L. Pinheiro, Consensus and polarization in competing complex contagion processes. *J. R. Soc. Interface* **16**, 20190196 (2019).
42. D. Centola, *Change: The Surprising Science of How New Ideas, Behaviors and Innovations Take Off and Take Hold* (Little, Brown & Co., 2021).
43. B. Ryan, N. C. Gross, The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociol.* **8**, 15 (1943).
44. T. W. Valente, Social network thresholds in the diffusion of innovations. *Soc. Netw.* **18**, 69–89 (1996).
45. E. Peer, L. Brandimarte, S. Samat, A. Acquisti, Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *J. Exp. Soc. Psychol.* **70**, 153–163 (2017).
46. E. Peer, D. M. Rothschild, Z. Evernden, A. Gordon, E. Damer, MTurk, Data quality of platforms and panels for online behavioral research. *J. Behav. Res. Methods*, 1–20 (10 January 2021).
47. M. P. Fiorina, S. J. Abrams, Political polarization in the American public. *Annu. Rev. Polit. Sci.* **11**, 563–588 (2008).
48. L. Mason, "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *Am. J. Pol. Sci.* **59**, 128–145 (2015).
49. L. Mason, A cross-cutting calm: How social sorting drives affective polarization. *Public Opin. Q.* **80**, 351–377 (2016).
50. Y. Lelkes, Affective polarization and ideological sorting: A reciprocal, albeit weak, relationship. *Forum Fam. Plan. West. Hemisph.* **16**, 67–79 (2018).
51. N. Geiger, J. K. Swim, Climate of silence: Pluralistic ignorance as a barrier to climate change discussion. *J. Environ. Psychol.* **47**, 79–90 (2016).
52. E. Noelle-Neumann, *The Spiral of Silence: Public Opinion–Our Social Skin* (University of Chicago Press, 1993).
53. D. A. Prentice, D. T. Miller, Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *J. Pers. Soc. Psychol.* **64**, 243–256 (1993).
54. M. E. Tankard, E. L. Paluck, Norm perception as a vehicle for social change. *Soc. Issues Policy Rev.* **10**, 181–211 (2016).
55. T. Bolsen, J. N. Druckman, F. L. Cook, The influence of partisan motivated reasoning on public opinion. *Polit. Behav.* **36**, 235–262 (2014).
56. S. M. Constantino, A. Rinscheid, S. Pianta, R. Frey, E. U. Weber, The source is the message: The impact of institutional signals on climate change-related norm perceptions and behaviors. *Clim. Change* **166**, 35 (2021).
57. A. Coppock, A. Guess, J. Ternovski, When treatments are tweets: A network mobilization experiment over Twitter. *Polit. Behav.* **38**, 105–128 (2016).
58. D. Guilbeault, J. Becker, D. Centola, Social learning and partisan bias in the interpretation of climate trends. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9714–9719 (2018).
59. F. P. Santos, Y. Lelkes, S. A. Levin, Link recommendation algorithms and dynamics of polarization in online social networks. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102141118 (2021).
60. S. Gächter, B. Herrmann, C. Thöni, Culture and cooperation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 2651–2661 (2010).
61. A. Falk *et al.*, Global evidence on economic preferences. *Q. J. Econ.* **133**, 1645–1692 (2018).
62. H. Taijfel, Experiments in intergroup discrimination. *Sci. Am.* **223**, 96–102 (1970).
63. H. Tajfel, J. C. Turner, W. G. Austin, S. Worchel, An integrative theory of intergroup conflict. *Organ. Identity Read.* **56**, 9780203505984–16 (1979).
64. E. Dimant, *Hate Trumps Love: The Impact of Political Polarization on Social Preferences* (Social Science Research Network, 2020).
65. B. Shell-Duncan, Y. Hernlund, K. Wander, A. Moreau, Legislating change? Responses to criminalizing female genital cutting in Senegal. *Law Soc. Rev.* **47**, 803–835 (2013).
66. J. W. Brehm, *A Theory of Psychological Reactance* (Academic Press, 1966).
67. Z. Kunda, The case for motivated reasoning. *Psychol. Bull.* **108**, 480–498 (1990).
68. A. Guess, A. Coppock, Does counter-attitudinal information cause backlash? Results from three large survey experiments. *Br. J. Polit. Sci.* **50**, 1497–1515 (2020).
69. S. Jasanoff, Just transitions: A humble approach to global energy futures. *Energy Res. Soc. Sci.* **35**, 11–14 (2018).
70. R. Pollin, B. Callaci, The economics of just transition: A framework for supporting fossil fuel–dependent workers and communities in the United States. *Labor Stud. J.* **44**, 93–138 (2019).
71. A. J. Stewart, N. McCarty, J. J. Bryson, Polarization under rising inequality and economic decline. *Sci. Adv.* **6**, eabd4201 (2020).
72. M. Schaub, J. Gereke, D. Baldassarri, Does poverty undermine cooperation in multiethnic settings? Evidence from a cooperative investment experiment. *J. Exp. Political Sci.* **7**, 27–40 (2020).
73. M. Abascal, D. Baldassarri, Love thy neighbor? Ethnoracial diversity and trust reexamined. *AJS* **121**, 722–782 (2015).
74. P. Dal Bó, G. R. Fréchette, On the determinants of cooperation in infinitely repeated games: A survey. *J. Econ. Lit.* **56**, 60–114 (2018).
75. A. Dannenberg, C. Gallier, The choice of institutions to solve cooperation problems: A survey of experimental research. *Exp. Econ.* **23**, 716–749 (2019).
76. V. Vasconcelos *et al.*, "Replication Data for: Segregation and Clustering of Preferences Erode Socially Beneficial Coordination." Harvard Dataverse. 10.7910/DVN/PMN61R. Deposited 17 September 2021.