Inequality, Identity, and Partisanship: How redistribution can stem the tide of mass polarization – Supplementary Information

Alexander J. Stewart[1*], Joshua B. Plotkin[2], and Nolan McCarty[3*]

[1]*School of Mathematics and Statistics, University of St Andrews, St Andrews, KY16 9SS, United Kingdom*

[2]*Department of Biology, University of Pennsylvania, Philadelphia, PA, USA*

[3]*School of Public and International Affairs, Princeton University, Princeton, NJ, USA*

[*]*ajs50@st-andrews.ac.uk, nmccarty@princeton.edu*

Contents

In this supplement we present additional numerical analysis and simulations to demonstrate the robustness of our findings to relaxation of model assumptions.

1 Decision Processes

In this section we describe two additional decision processes, beyond those described in Table 1 of the main text. Table S1 gives event probabilities for an AND logic decision process, in which an individual only regards another as part of their in-group if they belong to the same identity group and the same party:

AND Interaction type

	Same group, same party	Same group, other party	Other group, same party	Other group, other party
Probability of *i* choosing interaction	p_i^2	$p_i(1-p_i)$	$p_i(1-p_i)$	$(1-p_i)^2$
Probability of *j* agreeing to interaction	1	$1-p_j$	$1-p_j$	$(1-p_j)^2$
Probability of interaction success	q_I	q_I	q_O	q_O
Benefit from successful interaction	B_I	B_I	B_O	B_O

Table S1: **AND decision logic**.

Under this decision process the probability of an individual i, belonging to group 1 and party 1, choosing to interact with an individual with identity kl is λ_{kl}. We then have

$$
\begin{aligned}
\lambda_{II} &= \frac{(2p_i - p_i^2)x}{(2p_i - p_i^2)x + (p_i + (1-p_i)^2)(1-x) + (p_i + (1-p_i)^2)(1-x) + (1-p_i^2)x} \\
\lambda_{IO} &= \frac{(p_i + (1-p_i)^2)(1-x)}{(2p_i - p_i^2)x + (p_i + (1-p_i)^2)(1-x) + (p_i + (1-p_i)^2)(1-x) + (1-p_i^2)x} \\
\lambda_{OI} &= \frac{(p_i + (1-p_i)^2)(1-x)}{(2p_i - p_i^2)x + (p_i + (1-p_i)^2)(1-x) + (p_i + (1-p_i)^2)(1-x) + (1-p_i^2)x} \\
\lambda_{OO} &= \frac{(1-p_i^2)x}{(2p_i - p_i^2)x + (p_i + (1-p_i)^2)(1-x) + (p_i + (1-p_i)^2)(1-x) + (1-p_i^2)x}
\end{aligned}
\tag{1}
$$

The utility of a mutant when players attend to party AND group is thus

$$w_i(x) = [\lambda_{II}(x) + \lambda_{IO}(x)(1-p)]q_I F(B_I + \theta) + [(\lambda_{II}(x) + \lambda_{IO}(x))(1-q_I) + \lambda_{IO}(x)pq_I]F(\theta)$$
$$+[\lambda_{OI}(x)(1-p) + \lambda_{OO}(x)(1-p)^2]q_O F(B_O + \theta)$$
$$+[(\lambda_{OI}(x) + \lambda_{OO}(x))(1-q_O) + (\lambda_{OI}(x)p + \lambda_{OO}(x)(2p-p^2))q_O]F(\theta) \quad (2)$$

and the average selection gradient of a mutant can be calculated in the same way as described in Eqs 9-10 of the main text.

Table S2 gives event probabilities for an OR logic decision process, in which behavioral strategies have two components: p_p^i is the probability that a player i is willing to interact with a member of their party and p_g^i is the probability that a player i is willing to interact with a member of their identity group.

2D Strategy - OR Interaction type

	Same group, same party	Same group, other party	Other group, same party	Other group, other party
Probability of *i* choosing interaction	$p_g^i(1-p_p^i)+$ $p_p^i(1-p_g^i)+p_p^ip_g^i$	$p_g^i+(1-p_p^i)(1-p_g^i)$	$p_p^i+(1-p_p^i)(1-p_g^i)$	$1-p_p^ip_g^i$
Probability of *j* agreeing to interaction	1	$1-p_p^j$	$1-p_g^j$	$1-p_p^jp_g^j$
Probability of interaction success	q_I	q_I	q_O	q_O
Benefit from successful interaction	B_I	B_I	B_O	B_O

Table S2: **OR decision logic with a two-dimensional strategy**.

The fitness for this process is as described in Eq. 6 and Eq. 8, with the appropriate substitutions from Table S2.

2 Adaptive dynamics of polarization

In this section we present additional results for the "adaptive dynamics" analysis of the model. We perform invasion analyses under the assumption that that the population is infinitely large, and all members of the population adopt the same strategy. We then compute the selection gradient experienced by a rare mutant, in order to determine whether it will spread.

2.1 Divergent selection pressures

We first note that Eqs 9-10 of the main text describe the average selection gradient across both groups. However, in general when the distribution of identity group with respect to party is asymmetric, i.e. $0 < |\chi| < 1$, a member of group 1 belonging to party 1 will experience different selection pressures than a member of group 1 belonging to party 2, and so on. Under our assumption that identity groups and parties are of equal size and experience the same risk profiles, however, the only equilibria we find either occur when $\chi = 0$ (both parties are well mixed with respect to identity group) or $|\chi| = 1$ (identity groups align perfectly with parties). Under these conditions selection pressures are symmetric. Since any internal state in which selection pressures are asymmetric is unstable, we are able to ignore the complications that arise in such cases and focus on the stability of the symmetric equilibria. We note however that if different groups experience different risk profiles, or are of different size, this symmetry may not hold and new dynamics may arise.

2.2 Only party joint dynamics

Figure S1 shows the joint dynamics of sorting and polarization under the only party decision process (see main text Table 1, Eq. 5 and Eq. 7), in which individuals make decisions about who to interact with based only on party identity.

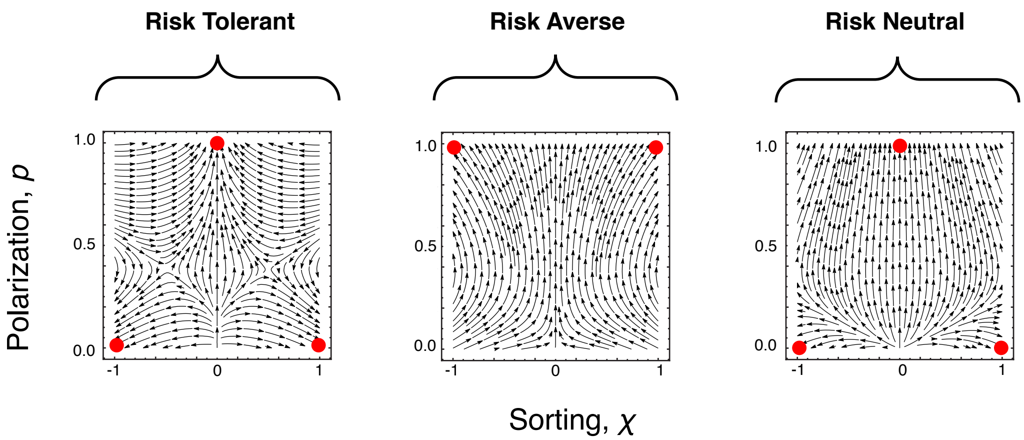

Figure S1: **Only party – Polarization and sorting**. Phase portraits illustrate the dynamics of polarization p and degree of sorting χ under our model of economic interactions and party switching, with fixed identity groups. Arrows indicate the average selection gradient experienced by a local mutant against a monomorphic background (see Methods). Red dots indicate stable equilibria. (left) When the decision process for social interactions considers only party identity, both high and low polarization states are stable. When polarization is low, sorting is always high $|\chi| = 1$, but when polarization is high, low sorting ($\chi = 0$) becomes stable. (center) However when the environment is risk averse, only high polarization and high sorting are stable. (right) And finally when the environment is risk neutral, the system returns to bistable polarization with high sorting when polarization is low, and low sorting when polarization is high. These plots show dynamics for $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$, $h = 10$ and $a = 0.02$. The phase portraits here show the selection gradient experienced by a monomorphic population in which parties and groups are of equal size.

2.2.1 AND logic joint dynamics

Figure S2 shows the joint dynamics of sorting and polarization under the AND party decision process (see Table S1), in which individuals make decisions about who to interact with based on party and group identity using AND logic.

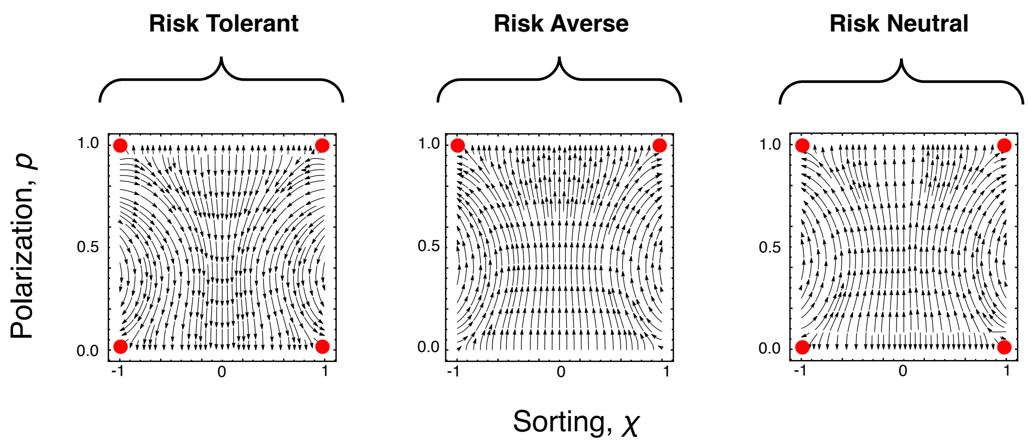

Figure S2: **Party AND group – Polarization and sorting**. Phase portraits illustrate the dynamics of polarization p and degree of sorting χ under our model of economic interactions and party switching with fixed identity groups. Arrows indicate the average selection gradient experienced by a local mutant against a monomorphic background (see Methods). Red dots indicate stable equilibria. (left) When the decision process for social interactions considers group AND party identity, both high and low polarization states are stable, but sorting is always high $|\chi| = 1$. (center) However when the environment is risk averse, only high polarization and high sorting are stable. (right) And finally when the environment is risk neutral, the system returns to bistable polarization with high sorting. These plots show dynamics for $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$, $h = 10$ and $a = 0.02$. The phase portraits here show the selection gradient experienced by a monomorphic population in which parties and groups are of equal size.

2.3 Joint dynamics of group and party polarization

Figure S3 shows the joint dynamics of group and identity polarization under the OR party decision process (see Table S2), in which individuals make decisions about who to interact with based on party and group identity using OR logic, and strategies are two-dimensional, meaning that an individual can weight the two dimensions of identity differently. We see that across all environments, there are two equilibria, with either high group and low party polarization, or vice versa.

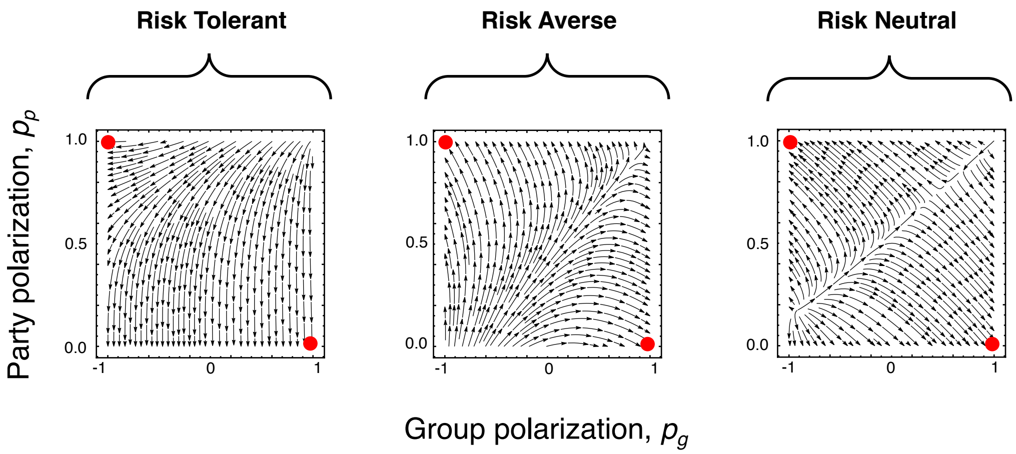

Figure S3: **Party OR group – Group and identity polarization**. Phase portraits illustrate the dynamics of group polarization p_g and party polarization p_p under our model of economic interactions, with high sorting $\chi = 1$. Arrows indicate the average selection gradient experienced by a local mutant against a monomorphic background (see Methods). Red dots indicate stable equilibria. When the decision process for social interactions considers group and party identity using OR logic, either high group polarization $p_g = 1$ and low polarization $p_p = 0$ or $p_p = 1$ and low polarization $p_g = 0$ are stable. These plots show dynamics for $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$, $h = 10$ and $a = 0.02$. The phase portraits here show the selection gradient experienced by a monomorphic population in which parties and groups are of equal size.

2.3.1 Bistability

We now describe the evolution of polarization under fixed levels of sorting, χ, for each of the three different decision processes described in Table 1 of the main text, across different economic environments, θ (Figure S4). We see that high polarization strategies are the only stable outcome in risk-adverse economic environments ($\theta \approx 0$), whereas the availability of stable low polarization strategies, and the associated basin of attraction for such strategies, varies with the economic environment, the decision process, and the degree of sorting in the population.

Most strikingly, when only party identity is used to make decisions, low levels of sorting result in high polarization regardless of the economic environment; whereas a strategy that uses party *or* group identity to make decisions tends to increase the basin of attraction to low-polarization strategies. In risk-tolerant environments ($\theta < 0$), the low-polarization outcome may even be the only stable strategy.

The intuition for this is simple: when sorting is low, a decision process that only accounts for party identity provides little information about the likely success of interactions. In contrast, a decision process that uses party identity or group identity to define the in-group widens the pool of potential out-group interaction partners and the associated increased benefits of those interactions. However, such a widening of the pool is of no use when the environment favors risk aversion ($\theta \approx 0$, Figure S4).

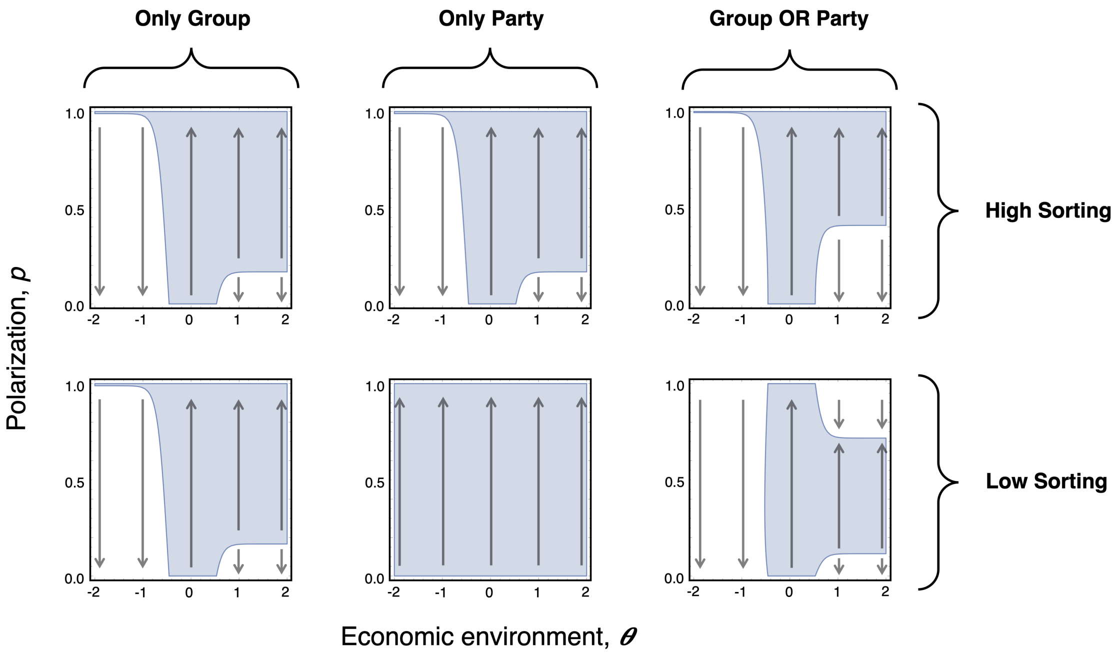

Figure S4 **Polarization across economic environments**. Shown are the evolutionary dynamics of polarization p, i.e. the probability of choosing in-group or in-party interactions, for different underlying economic environments θ. Arrows indicate the selection gradient (see Methods) experienced by a monomorphic population employing strategy p in environment θ. Blue regions indicate a positive gradient (increasing polarization) while white regions indicate a negative gradient (declining polarization). Six cases are presented. (top, left) When only group identity is attended to, polarization is bistable provided the environment is not risk averse. However under risk aversion ($\theta \approx 0$), only high polarization is stable. (bottom, left) The dynamics do not depend on sorting when only group identity is attended to. (top, center) When only party identity is attended to and sorting is high ($|\chi| = 1$), the dynamics and basins of attraction for group and party identity are identical. (bottom center) However when sorting is low ($\chi = 0$), high polarization always evolves in a population only attends to party identity. (top, right) When interaction decisions are based on party or group identity, and sorting is high, the basin of attraction for low polarization is large compared to the case when decisions are based on party or group alone, meaning that polarization is comparatively stable. (bottom, right) When sorting is low, low polarization becomes the only stable equilibrium under risk tolerance, while an intermediate polarization state becomes stable under risk aversion. These plots show dynamics for $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$, $h = 10$ and $a = 0.02$. Dynamics describe the selection gradient experienced by a monomorphic population in which parties and groups are of equal size.

2.4 Evolution of attention

We now consider whether a population can be incentivized to switch from a group-only decision process to a group- or party-decision process (see Table 1 of the main text and Table S2). We calculated the fitness of each decision process as a function of sorting χ and group polarization strategy p_g. We assume that initially $p_p = 0.5$ for the OR logic decision strategy, indicating indifference to party identity. Below we also consider individual-based simulations in which attention is parameterized to vary smoothly between 0 and 1 (Figure S9).

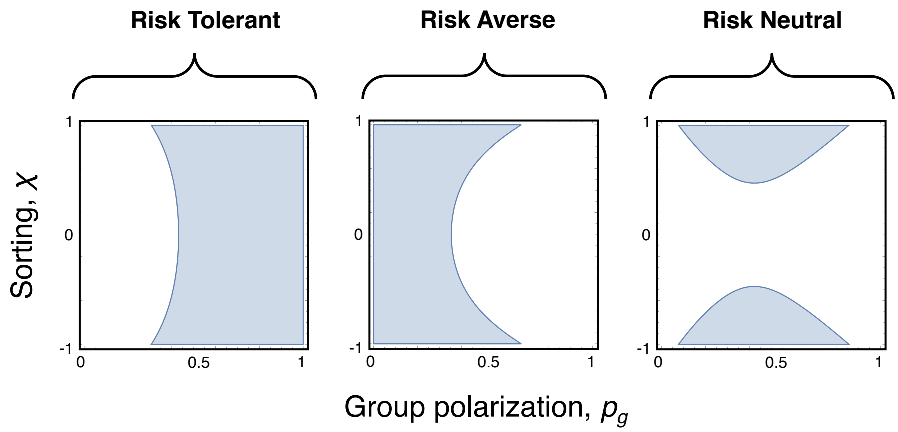

Figure S5: **Switching decision logic**. Shown in blue are the conditions under which an OR decision logic (Table S2) produces higher fitness than a group-only decision logic, as a function of sorting, χ and group polarization strategy p_g in a monomorphic population. We assume that the OR decision logic uses $p_p = 0.5$ indicating initial indifference to party identity (i.e neither in-party nor out-party favoring). (left) When the environment is risk tolerant, high polarization group only decision strategies can be invaded by group OR polarization decision strategies. (center) When the environment is risk averse, low polarization group-only decision strategies can be invaded by group OR polarization decision strategies. (right) When the environment is risk neutral, highly sorted populations with intermediate levels of polarization under a group only decision strategy can be invaded by group OR party decision strategies. These plots show dynamics for $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$, $h = 10$ and $a = 0.02$. Dynamics describe the selection gradient experienced by a monomorphic population in which parties and groups are of equal size.

2.5 Varying model parameters

Here we vary the parameters for the payoffs and for the utility function, under the group OR party decision logic. We show that the bi-stability of the system is highly robust to parameter variation, but if the utility function becomes too shallow, risk aversion does not stimulate polarization.

2.5.1 Varying payoffs

We varied the value of B_O, keeping other parameters fixed, and retaining the risk profile in which in-group interactions are less risky but less advantageous.

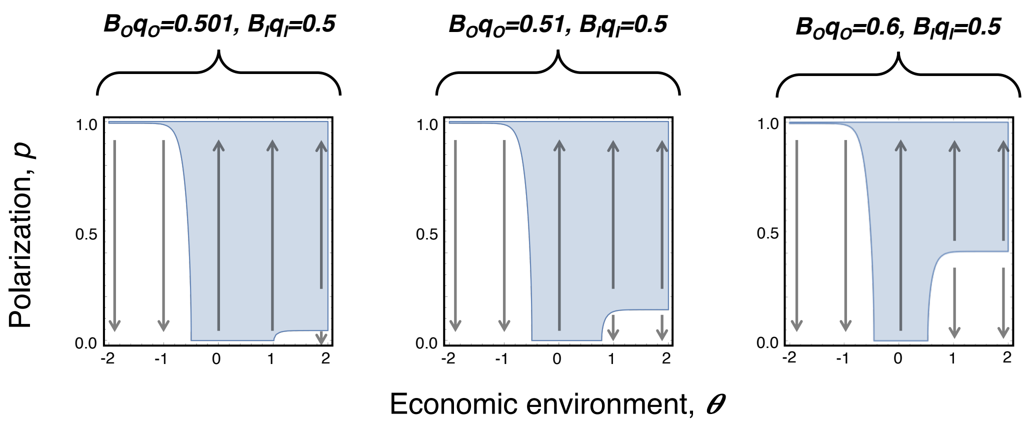

Figure S6: **Polarization under different payoffs**. Shown are the evolutionary dynamics of polarization p, i.e. the probability of choosing in-group or in-party interactions, for different underlying economic environments θ. Arrows indicate the selection gradient (see Methods) experienced by a monomorphic population employing strategy p in environment θ. Blue regions indicate a positive gradient (increasing polarization) while white regions indicate a negative gradient (declining polarization). In all cases we show the group OR party decision logic as described in main text Table 1. (left) When the expected benefit of out-group interactions is only slightly advantageous ($B_o q_o = 0.501$, $B_i q_i = 0.5$), polarization is bistable provided the environment is not risk averse. Under risk aversion ($\theta \approx 0$), only high polarization is stable. However when the environment is risk neutral, the basin of attraction for the low polarization equilibrium is small. (center) When the expected benefit of out-group interactions is moderately advantageous ($B_o q_o = 0.51$, $B_i q_i = 0.5$), polarization is bistable provided the environment is not risk averse. Under risk aversion ($\theta \approx 0$), only high polarization is stable. When the environment is risk neutral, the basin of attraction for the low polarization increases. (right) When the expected benefit of out-group interactions is more advantageous ($B_o q_o = 0.6$, $B_i q_i = 0.5$), polarization is bistable provided the environment is not risk averse. Under risk aversion ($\theta \approx 0$), only high polarization is stable. When the environment is risk neutral, the basin of attraction for the low polarization increases further. These plots show dynamics for $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$, $h = 10$ and $a = 0.02$. Dynamics describe the selection gradient experienced by a monomorphic population in which parties and groups are of equal size.

2.5.2 Varying utility function non-linear component

We varied the sharpness of the non-linear component of the utility function, h. We see that risk aversion is not sufficient to ensure the evolution of polarization when the h is small and the slope of the S-curve is shallow.

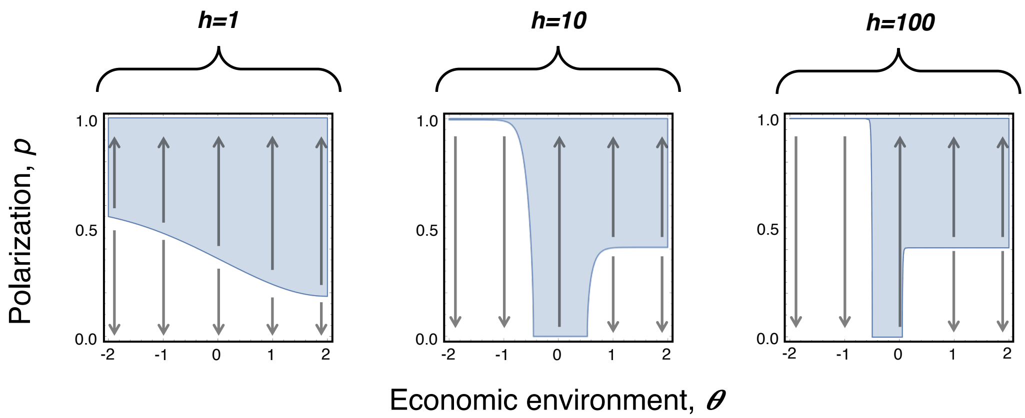

Figure S7: **Polarization under different utility function sharpness,** h. Shown are the evolutionary dynamics of polarization p, i.e. the probability of choosing in-group or in-party interactions, for different underlying economic environments θ. Arrows indicate the selection gradient (see Methods) experienced by a monomorphic population employing strategy p in environment θ. Blue regions indicate a positive gradient (increasing polarization) while white regions indicate a negative gradient (declining polarization). In all cases we show the group OR party decision logic as described in main text Table 1. (left) When the sharpness of the utility function threshold is shallow, $h = 1$, polarization is always bistable. (center) When the sharpness of the utility function threshold is intermediate, $h = 10$, polarization is bistable provided the environment is not risk averse. Under risk aversion ($\theta \approx 0$), only high polarization is stable. (right) When the sharpness of the utility function threshold is sharp, $h = 100$, polarization is bistable provided the environment is not risk averse. Under risk aversion ($\theta \approx 0$), only high polarization is stable. These plots show dynamics for $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$, $h = 10$ and $a = 0.02$. Dynamics describe the selection gradient experienced by a monomorphic population in which parties and groups are of equal size.

2.5.3 Varying utility function linear component

We varied the slope of the linear component of the utility function, a. We see that the qualitative dynamics of the system are robust to the choice of a over two orders of magnitude.

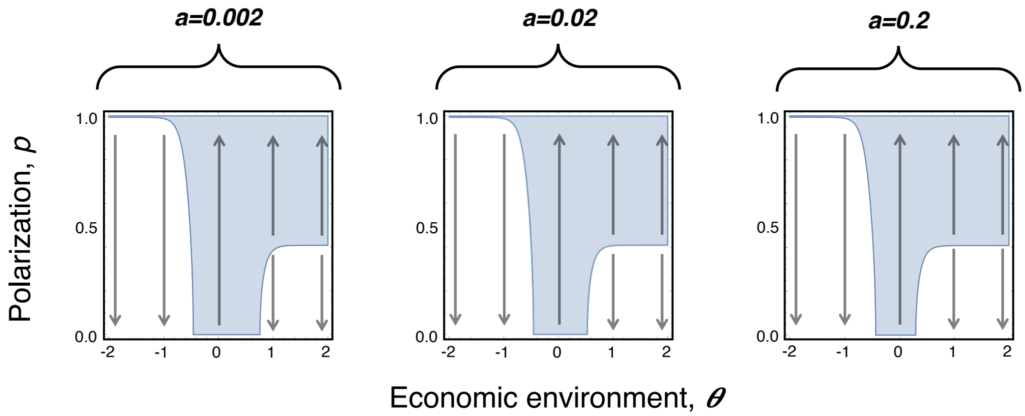

Figure S8: **Polarization under different utility function linear component slope,** a. Shown are the evolutionary dynamics of polarization p, i.e. the probability of choosing in-group or in-party interactions, for different underlying economic environments θ. Arrows indicate the selection gradient (see Methods) experienced by a monomorphic population employing strategy p in environment θ. Blue regions indicate a positive gradient (increasing polarization) while white regions indicate a negative gradient (declining polarization). In all cases we show the group OR party decision logic as described in main text Table 1. Regardless of the the slope of the linear part of utility function, a, polarization is bistable provided the environment is not risk averse. Under risk aversion ($\theta \approx 0$), only high polarization is stable. These plots show dynamics for $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$, $h = 10$ and $a = 0.02$. Dynamics describe the selection gradient experienced by a monomorphic population in which parties and groups are of equal size.

3 Simulations with heterogeneous group structure

We now discuss additional simulations in which we allow variation in the number and size of identity groups (keeping the number of parties fixed at 2).

3.1 Attention and Sorting

In order to allow changes to attention level and party makeup (i.e. sorting), we must make several additional assumptions to those described in the main text. First, to model attention we assume that, with probability p_a, an individual uses OR logic when assessing another person's identity, and with probability $(1 - p_a)$ they consider ONLY party identity. Thus, in the notation of main text equations 5-6, an individual encountering someone with identity kl attempts to interact with probability $p_a\phi_{kl} + (1 - p_a)\pi_{kl}$.

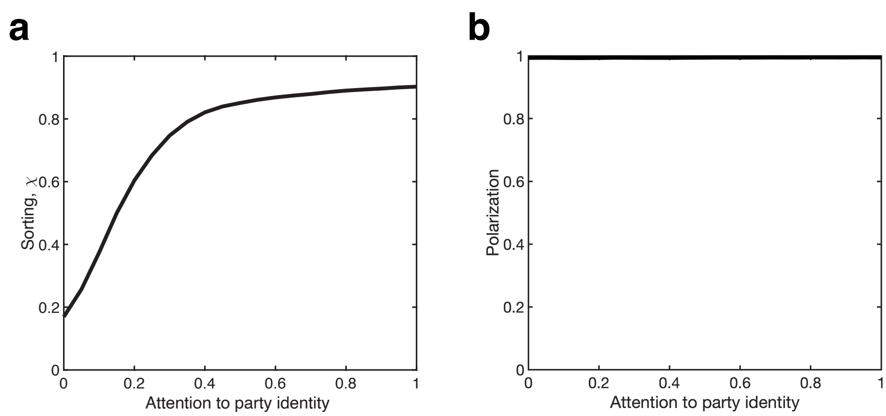

Figure S9: **Attention and sorting.** Ensemble mean equilibrium from individual-based simulations for a population with homogeneous incentives, without redistribution, $\alpha = 0$, under a risk averse environment, $\theta_0 = -0.5$. Attention, which functions as described in the text, level is varied endogenously, while sorting and polarization are allowed to co-evolve (a) As attention to party identity increases, sorting of group identities along party lines increases from initially low levels to almost complete sorting. b) Polarization remains high throughout under a risk averse environment. Plots show ensemble mean values across 10^4 replicate simulations, for a population of 2000 individuals. Success probabilities and benefits are fixed at $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$ with $h = 10$ and $a = 0.02$, while $\gamma = 0$ and $b_p = 0.01$. Evolution occurs via the copying process (see methods) with selection strength $\sigma = 10$, mutation rate $\mu = 10^{-3}$ and mutation size $\Delta = 0.01$. Mutations in which players flip party identity occur at rate $\mu_p = 10^-3$.

In order to allow the evolution of sorting, we assume that individuals are able to copy the party identity of in-group members, using the same copying process as described in the main text.

However, we also assume that there is pressure to keep parties of roughly equal size, e.g. through party elites adjusting their policies to increase their vote share if they are losing. From the point of view of a voter, this means there tends to be an incentive to switch to the smaller of the two main parties (since we assume this party will adjust policies to closely align with their interests on average). This assumption is consistent with (**?**) who notes that polarization in the US tends to track with partisan competition and parity.

To model this effect we assume that the utility of individual i is increased by a factor $b_p|\rho_i - \bar{\rho}|$ where $\rho_i = 0$ if they identify with the first party and $\rho_i = 1$ if they identify with the second party, where $\bar{\rho}$ is the average party affiliation in the population and b_p is the benefit of misalignment with the average. This model tends to keep parties of similar size.

Under these assumptions we explored the effect of attention on the evolution of sorting and polarization (Figure S9), when the environment favors risk aversion ($\theta_0 = -0.5$).

3.2 Number of identity groups

We explored the evolution of sorting and polarization as a function of the number of identity groups, assuming homogeneous incentives across groups (Figure S10).

We see that as number of identity groups increases, equilibrium sorting and polarization decline, under a decision process that attends to group OR party identity. This occurs because it becomes increasingly difficult to find successful in-group interactions as the number of different identity groups increases.

3.3 Asymmetrical group size

We also explored the effect of asymmetry in identity group size on the evolution of sorting and polarization. We considered both the case of two identity groups of unequal size (one large, one small), and the case of multiple identity groups, with many small groups and one large group.

We find that, when there is one large group and many small groups, sorting and polarization remain high, in contrast to the scenario when there are many equally sized small groups (Figure S10). High sorting in this case reflects a situation in which one party is made up of a single large

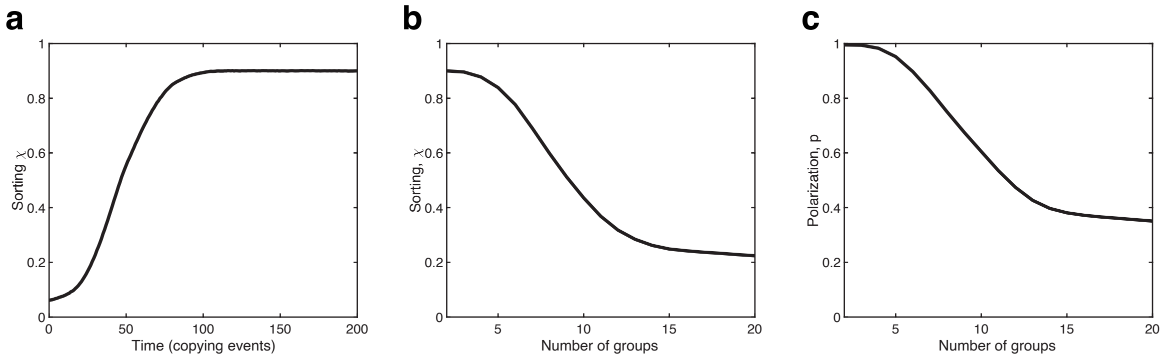

Figure S10: **Number of groups.** Individual-based simulations for a population with homogeneous incentives, without redistribution, $\alpha = 0$, under a risk averse environment, $\theta_0 = -0.5$. We fixed the number of party identities at 2, but varied the number of group identities, assuming that individuals treat anyone who is not part of their in-group as part of the out-group. (a) Time series of sorting evolution with two parties. Sorting, like polarization, evolves to high levels when individuals attend to party OR group identity, as predicted (main text Figure 2) b) Ensemble mean equilibrium for sorting as a function of group number, with number of identity groups varied from 2 to 20. c) Ensemble mean equilibrium for polarization as a function of group number, with number of identity groups varied from 2 to 20. Plots show ensemble mean values across 10^4 replicate simulations, for a population of 2000 individuals, with each individual randomly assigned to an identity group. Success probabilities and benefits are fixed at $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$ with $h = 10$ and $a = 0.02$, while $\gamma = 0$ and $b_p = 0.01$. Evolution occurs via the copying process (see methods) with selection strength $\sigma = 10$, mutation rate $\mu = 10^{-3}$ and mutation size $\Delta = 0.01$. Mutations in which players flip party identity occur at rate $\mu_p = 10^-3$.

identity group and the other party is made up of many small identity groups.

When there are only two identity groups of unequal size, sorting and polarization decline as asymmetry increases, with the larger party less sorted than the smaller party. This occurs because the large identity group necessarily becomes spread across multiple parties when it becomes sufficiently large.

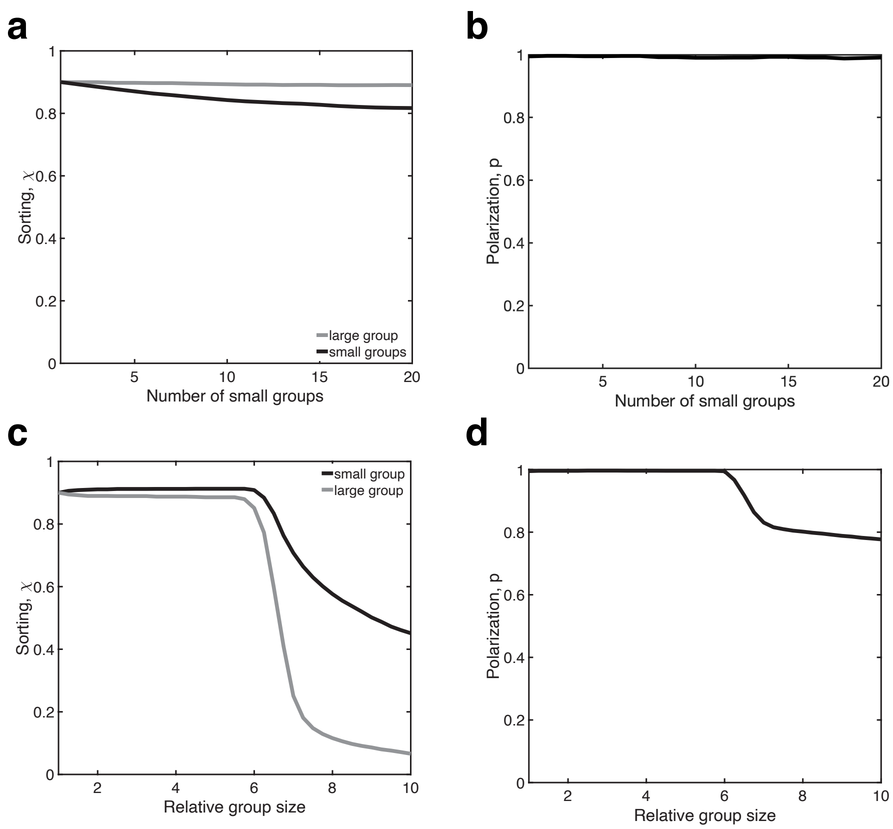

Figure S11: **Asymmetrical groups.** Ensemble mean equillibria from individual-based simulations for a population with homogeneous incentives, without redistribution, $\alpha = 0$, under a risk averse environment, $\theta_0 = -0.5$. We fixed the number of party identities at 2, but varied the relative size and the number of group identities, assuming that individuals treat anyone who is not part of their in-group as part of the out-group. (a) When there is one large group of 1000 individuals, and between 1 and 20, equally sized small groups making up the rest of the population, we see that the degree of sorting in the large group (gray line) and the small groups (black line) remains high even when the number of small groups is large. Similarly (b) polarization remains high for all group sizes. c) When only two identity groups are present, but differ in size, we see that sorting for the large group (gray) declines as the size of the large group increases, as does sorting in the smaller group (black line) to a lesser degree. d) Polarization also undergoes a moderate decline as relative group size increases. Plots show ensemble mean values across 10^4 replicate simulations, for a population of 2000 individuals, with each individual randomly assigned to an identity group. Success probabilities and benefits are fixed at $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$ with $h = 10$ and $a = 0.02$, while $\gamma = 0$ and $b_p = 0.01$. Evolution occurs via the copying process (see methods) with selection strength $\sigma = 10$, mutation rate $\mu = 10^{-3}$ and mutation size $\Delta = 0.01$. Mutations in which players flip party identity occur at rate $\mu_p = 10^{-3}$.

4 Simulations with heterogeneous wealth

We now discuss additional simulations, conducted under the same assumptions as described in the main text, in which we allow different groups to have different levels of wealth (see main text).

4.1 Sorting

We exogenously varied the amount of sorting, χ in the presence of inequality. We see that sorting tends to increase polarization, but it can have complex effects on levels of inequality and population average utility. This is because intermediate levels of polarization tend to result in lower levels of utility. Where reducing sorting can also reduce polarization to low levels, it has a beneficial effect in reducing inequality and increasing population average utility.

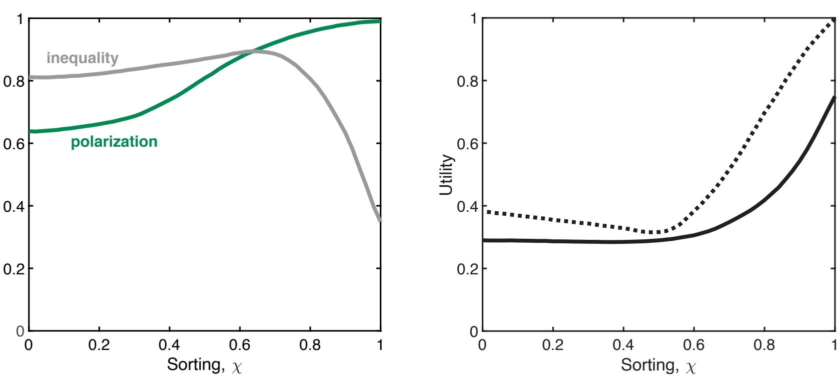

Figure S12: **Sorting and inequality.** Ensemble mean equillibria from individual-based simulations for a population initialized in a low polarization state in the presence of wealth redistribution (Eq. 4). We show results in the case of no underlying economic inequality, $\beta = 0.5$ (dashed lines), as well the case of high underlying inequality, $\beta = 0.01$ (solid lines). Results shown here arise from a decision process that attends to group OR party identity, and redistribution is fixed exogenously at $\alpha = 0.5$. (a) When public goods are not multiplicative ($r = 1$ and $\theta_0 = 0.5$), as sorting increases, polarization increases, but inequality changes non-linearly, achieving its lowest value at maximum sorting. b) Increasing sorting tends to increases overall utility. Plots show ensemble mean values across 10^4 replicate simulations, for groups of 1000 individuals each. Success probabilities and benefits are fixed at $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$ with $h = 10$ and $a = 0.02$, while $\gamma = 0$. Evolution occurs via the copying process (see methods) with selection strength $\sigma = 10$, mutation rate $\mu = 10^{-3}$ and mutation size $\Delta = 0.01$.

4.2 Public goods multiplication factor

We explored the effect of redistribution on inequality when public goods have a multiplicative effect, $r > 1$.

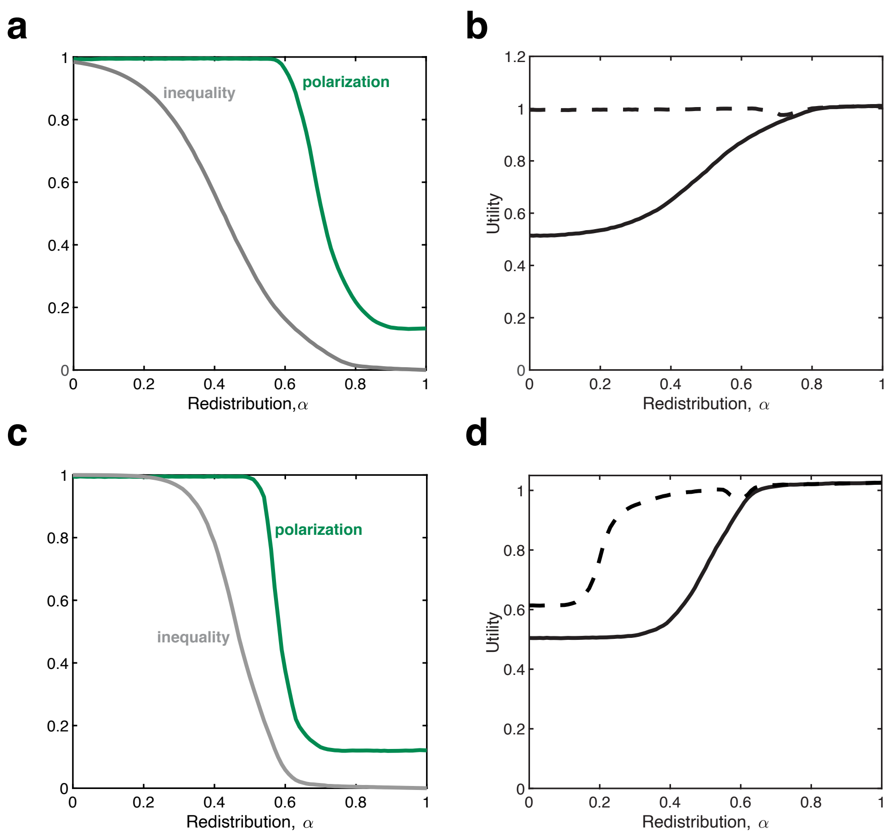

Figure S13: **Multiplicative public goods.** Ensemble mean equilibria from individual-based simulations for a population initialized in a low polarization state in the presence of wealth redistribution (Eq. 4). We show results in the case no underlying economic inequality, $\beta = 0.5$ (dashed lines), as well the case of high underlying inequality, $\beta = 0.01$ (solid lines). Results shown here arise from a decision process that attends to group or party identity, and sorting is fixed exogenously at $\chi = 1$. When public goods are not multiplicative ($r = 1$ and $\theta_0 = 0.5$), and redistribution is absent ($\alpha = 0$) overall inequality (gray line, measured as the relative difference in utility – see SI) and polarization (green line) are high. With increasing rates of redistribution, first overall inequality and then polarization decline to zero. b) Increasing redistribution increases overall utility towards the level achieved when underlying inequality is absent. b) When public goods are multiplicative ($r = 2$ and $\theta_0 = 1$), we see the same dynamics, but lower levels of redistribution are required to reduce inequality and polarization. d) In this case increasing redistribution also increases the utility of the population when inequality is absent (dashed line). Plots show ensemble mean values across 10^4 replicate simulations, for groups of 1000 individuals each. Success probabilities and benefits are fixed at $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$ with $h = 10$ and $a = 0.02$, while $\gamma = 0$. Evolution occurs via the copying process (see methods) with selection strength $\sigma = 10$, mutation rate $\mu = 10^{-3}$ and mutation size $\Delta = 0.01$.

4.3 Deadweight loss of taxation

We explored the effect of deadweight losses of taxation, $\gamma > 0$, on the effectiveness of redistribution on reducing polarization and inequality. We see that such losses reduce the effectiveness of redistribution and make it harder to mitigate both polarization and inequality.

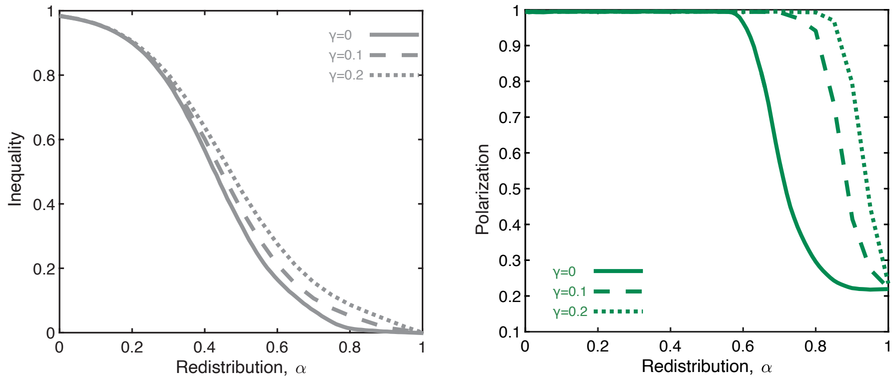

Figure S14: **Loss due to taxation.** Ensemble mean equillibria from individual-based simulations for a population initialized in a low polarization state in the presence of wealth redistribution (Eq. 4). We show results for the case of no underlying economic inequality, $\beta = 0.5$ (dashed lines), as well the case of high underlying inequality, $\beta = 0.01$ (solid lines). Results shown here arise from a decision process that attends to group OR party identity and where sorting is fixed exogenously at $\chi = 1$. (left) When public goods are not multiplicative ($r = 1$ and $\theta_0 = 0.5$), and redistribution is absent ($\alpha = 0$) overall inequality is high. With increasing rates of redistribution, polarization declines, but the decline is slower when deadweight losses due to taxation increase. (right) Increasing redistribution increases overall utility towards the level achieved when underlying inequality is absent. b) Similarly, redistribution decreases polarization, but becomes less effective as γ increases. Plots show ensemble mean values across 10^4 replicate simulations, for groups of 1000 individuals each. Success probabilities and benefits are fixed at $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$ with $h = 10$ and $a = 0.02$. Evolution occurs via the copying process (see methods) with selection strength $\sigma = 10$, mutation rate $\mu = 10^{-3}$ and mutation size $\Delta = 0.01$.

4.4 Economic shocks

As shown above (Figure S4) the basin of attraction for the high polarization equilibrium declines to the point of almost vanishing in a risk tolerant environment. Therefore, we considered a scenario in which a high-polarization population enters a very poor economic environment due to an economic shock. We see that, when this occurs, and we fix $\theta = -1.5$, a population will evolve from a state of high to a state of low polarization, whereas it will remain in a state of high polarization in a risk averse, $\theta = 0$ or risk neutral $\theta = 1.5$ environment.

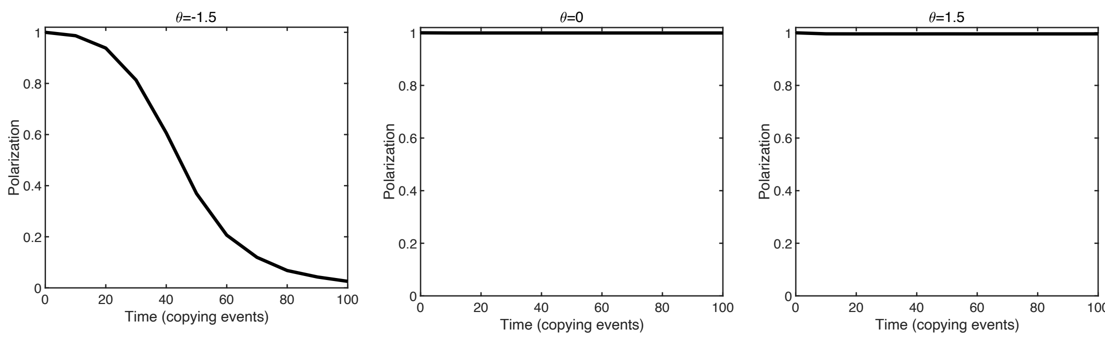

Figure S15: **Economic shocks.** Ensemble mean time trajectories for a population initialized in a high polarization state, from individual-based simulations. We show results in the case no underlying economic inequality, $\beta = 0.5$ for different (fixed) environments (θ). Results shown here arise from a decision process that attends to group or party identity, with sorting is fixed exogenously at $\chi = 1$ and redistribution at $\alpha = 0$. (left) When the environment is very bad, $\theta = -0.5$, corresponding to an economic shock, the basin of attraction for the high polarization equilibrium shrinks sufficiently that even small mutations are enough to escape. (center and right) When the environment is not risk tolerant, the population remains stuck in the high polarization equilibrium. Plots show ensemble mean values across 10^4 replicate simulations, for groups of 1000 individuals each. Success probabilities and benefits are fixed at $B_I = 0.5$, $B_O = 1$, $q_I = 1.0$, $q_O = 0.6$ with $h = 10$ and $a = 0.02$. Evolution occurs via the copying process (see methods) with selection strength $\sigma = 10$, mutation rate $\mu = 10^{-3}$ and mutation size $\Delta = 0.01$.