Convolution of individual and group identity: self-reliance increases polarisation in basic opinion model

Lennart Quante [a,b] *, Annika Stechemesser[a,c,d] *, Damian Hödtke [a,e] and Anders Levermann[a,d,f] †

*These authors contributed equally to this work.

[a]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, P.O. Box 6012 03, 14412 Potsdam, Germany

[b]Institute of Mathematics, University of Potsdam, Potsdam, Germany

[c]Mercator Research Institute on Global Commons and Climate Change (MCC), Berlin, Germany

[d]Institute of Physics, University of Potsdam, Potsdam, Germany

[e] Institute of Physics, Humboldt University of Berlin, Berlin, Germany

[f]Columbia University, LDEO, Palisades, New York, USA

† Correspondence to: anders.levermann@pik-potsdam.de

Abstract

Opinion formation within society follows complex dynamics. Towards its understanding, axiomatic theory can complement data analysis. To this end we propose an axiomatic model of opinion formation that aims to capture the interaction of individual conviction with social influence in a minimalistic fashion. Despite only representing that (1) an agent has an initial conviction with respect to a topic and is (2) being influenced by their neighbours, the model shows emergence of opinion clusters from an initially unstructured state. Here we show, that increasing individual self-reliance makes agents more likely to align their socially influenced opinion with their inner conviction which concomitantly leads to increased polarisation. The opinion drift observed with increasing self-reliance matches matches real world polarisation trends. Modelling the basic traits of striving for individual versus group identity, we find a trade-off between individual fulfilment and societal cohesion. This finding from fundamental assumptions can serve as a building block to explain societal opinion formation.

Humans make thousands of decisions every day. Most of them are small and largely inconsequential, or affect only their personal life, but some major decisions, like for example on election days, form the future of societies. These major decisions are preceded by an opinion formation process that does not take place in isolation but evolves dynamically in relation with others. Understanding the mechanics of opinion formation and decision making and its underlying mechanisms is of crucial relevance for social processes at all scales. This need is especially highlighted by the contrast between increasing political polarisation in many countries on the one hand and the urgent need for collective societal action to deal with major crises such as the Covid-19 pandemic or climate change.

The connection of complex systems theory with social sciences[1], which stresses the general importance of polarisation in society[2], have inspired a number of ways to incorporate polarisation and radicalisation in agent-based models, which are a popular tool to explore dynamic opinion formation and collective decision making[3;4;5;6]. Recent interdisciplinary studies have for example extended homophily based models[7] by introducing the concept of bias assimilation[8] which assumes a mechanism where agents are more likely to believe opinions that are similar to their own. Furthermore, it has been proposed to include a radicalisation parameter[9] which determines how agents perceive others' opinions. The effects of homophily have been explored with regard to polarisation regarding leisure activities[10] as well as with respect to multi-dimensional opinion modelling[11] or emergence of network structure[12]. An alternative are additions to attraction-repulsion based models of opinion formation, where Axelrod et al.[13] explore polarisation and possible intervention strategies. Recent studies also analyse the mechanisms of polarisation of elites[14], the emergence of political factions under increasingly partisan identities[15;16] or effects of political shocks on polarisation[17]. Further complex systems studies of specific conditions for polarisation in agent-based models explore the effects of coupled layers as a model of echo chambers[18] or combine polarisation and network evolution[19]. Recent empirical studies on social networks are divided between finding finding increasing effects of echo chambers[20;21] and recent high level publications finding no increase in polarisation due to dynamics of polarisation sampling data from Facebook[22;23].

Polarisation and its repercussions remain an important area of research about opinion formation[1;2] as it influences societal decisions which was exemplified during the recent Covid-19 pandemic[24;25;26]. Further effects of polarised society include the potential hindering of societal beneficial change processes after a tipping point like intervention changed circumstances[27], influencing election results election results in the USA[28], foreshadow right wing terrorism[29] or feedback into policies like climate mitigation[30].

Here, we add to this literature by proposing a simple agent-based model which captures dynamic opinion formation against the backdrop of two opposing but fundamental human social desires: Belonging to a group and at the same time pursuing individual goals, i.e. to some extent stick out of the group. These desires have been described for example in Brewer's optimal distinctiveness theory[31], that posits the need to balance assimilation and distinction from others, while Deci's and Ryans's self determination theory[32] also differentiates between individual and collective needs, labelled as "autonomy" and "relatedness". We explore whether this juxtaposition between the needs for similarity, belonging and likeness on the one hand and individuality and independence on the other hand is a plausible driving mechanism for opinion formation. Hence, we integrate the strive to be individual and similar into a dynamics simulation, thus bridging the gap between theories of self-expression and models of opinion formation.

We formalise this in our model (Fig. 1) by assigning every agent a continuous "self-reliance" parameter γ which describes how dependent on others the agent is in their opinion formation (continuous value between zero – "very dependent" to one "very self-reliant"). Further, every agent has an initial conviction A_i^*, which represents their intrinsic opinion on a topic (scaled continuously from zero - "full opposition" to one - "full agreement"). Both parameters are distributed uniformly in the basic version of the model. The agents are then randomly placed on a regular, periodic grid (100 x 100 agents by default). All qualitative results are obtained by averaging over model ensembles of 100 varying initial distributions. Every agent is equally influenced by the eight neighbours around them. At each time step, every agent updates their attitude A based on the self-reliance weighted influence of their neighbours and the disparity between their own opinion and the opinions in their neighbourhood (see Fig. 1). Once the model has completed the final time step and reached a stable state, every agent makes a final decision for zero or one, determined by a specified threshold of their final attitude (default is 0.5). We provide a mathematical description of the model in the methods. While simple, this model set-up has multiple advantages. First, it is easily adjustable as the agent number, size and weighting of neighbourhood influence and parameter distributions are modular. Second, it relies on very few input parameters. The genesis of opinions (going from initial to final attitude) is not influenced by externally specified thresholds and all polarisation observed is emergent. Finally, the interpretation is comprehensible, allowing an unobstructed view on the mechanism.

In the remainder of the paper, we first show the emergence of divided groups from the random initialisation described above (Fig. 2). Next, we explore how the share of agents with high self-reliance agents changes the strength of polarisation between the groups (Fig. 3, Fig. 4). Finally, we compute the alignment between the intrinsic attitude of the agents and the final decision and show the trade-off between this decision alignment and societal cohesion (figs. 5 and 6).

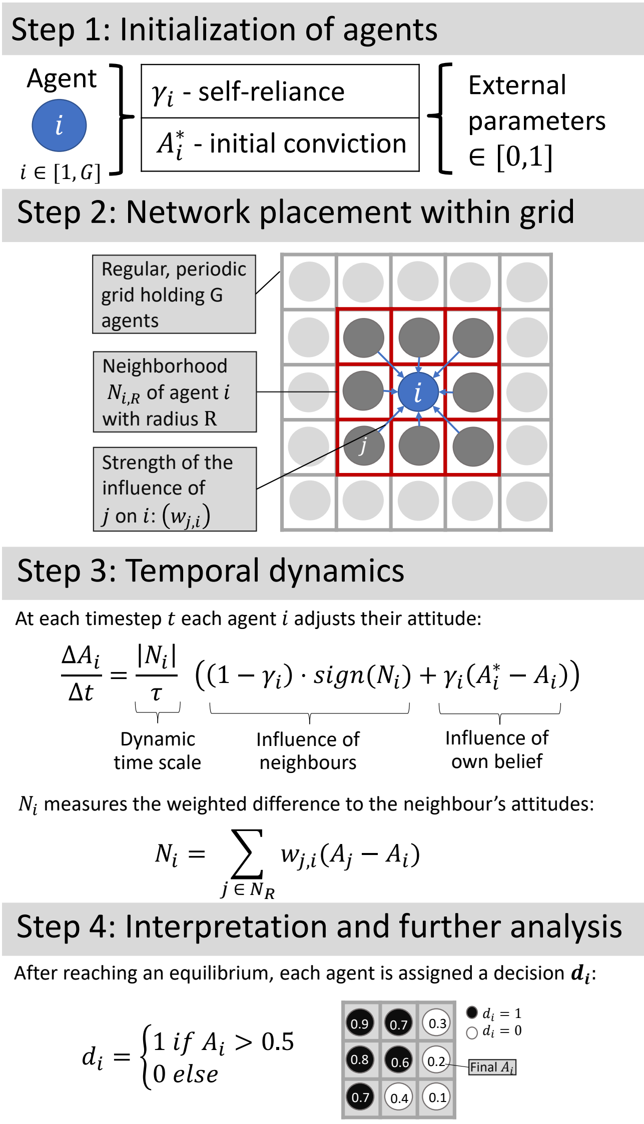


Figure 1: Model setup and dynamics. Agents are initialised with an degree of self-reliance γ_i, expressing individualism, and an initial conviction A_i^*, which is also the starting value for attitude A_i. They are placed on a regular periodic grid and assigned neighbourhoods based on a Chebyshev distance. At each time-step, agents form a attitude A_i with the dynamics factoring in the influence of the neighbourhood and the self-reliant strive towards the initial conviction. Once the equilibrium is reached, each agent is assigned a final decision and the final state is the basis for following analyses.

Results

Emergence of a stable and opinion-divided society

For uniform distributions of initial attitude and self-reliance, we consistently find an emergence of stable, polarised decision clusters. Fig. 2 shows the difference between the initial attitude (a) and initial decision (c) and the final attitude (b) and decision (d) exemplary on a 100x100 grid. In the initialisation, no patterns or structure are visible. After evolving the model for 1000 time steps, we see clusters of agents with final attitudes (b) that strongly diverge from the threshold (here 0.5) in either direction surrounded by agents with very moderate opinions. This suggests that neighbourhoods can be overly influenced by few self-reliant agents, which mirrors patterns observed in the structure of scale-free social networks, where few agents are influential (many connections) and many agents are influenced (few connections). Visualising the final decision (d) shows multiple distinct decision clusters. For both possible decisions (one or zero) there is one large cluster and multiple smaller clusters.

We expand on the uniform initialisation by considering normal distributions of agent self-reliance with varying means (Supplementary Fig. 1) as well as varying standard deviations (Supplementary Figs. 6, 11 and 16). Initial attitude as well as agent placement remain as before. Again we observe the formation of stable, opposing opinion clusters in all realisations. The size and balance of clusters changes with the distribution of individualistic agents. Smaller average self-reliance induces the formation of larger cluster, larger average self-reliance implies fine clustering. Small mean and standard deviation of self-reliance leads to very small differences in final attitude (Supplementary Fig. 6), showing that variation of self-reliance is necessary for strong clustering to emerge.

Overall, these results show that the proposed simple decision model leads to non-trivial decision patterns based on the trade-off between social influences and reliance on personal attitude. Few, strongly minded agents are sufficient to create large, stable clusters.

Societal polarisation increases with more individualistic agents

The exemplary results in Fig. 2 suggest that opinion clusters form around few agents with a very strong final attitude and many agents with a more moderate final attitude. We systematically explore this by considering an ensemble of societies, each with 10,000 agents, over 100 runs varying the uniform distribution of initial variables. To evaluate how the attitude evolved, we split agents according to their initial decision (threshold of 0.5, i.e. $A_i^* < 0.5$ and $A_i^* \geq 0.5$) and visualise the distribution of the final attitude separately for each group (Fig. 3 a).

We find that the final attitude is approximately normally distributed but with long tails. The majority of agents sticks with their inherent decision but their attitude becomes more moderate. In contrast to the initial uniform distribution, the median is much closer to 0.5 in the final distribution for both groups. A tail of agents with small or large attitudes remains. The joint near-normal distribution of final attitudes matches distributions of opinions observed in real-life[33;34]. To investigate the impact of changing distributions of self-reliance, we consider normally distributed self-reliance (γ) instead of a uniform distribution. All other initial values remain the same. We fix the standard deviation as $\sigma(\gamma) = 0.1$ and consider different means of the normal distribution ($\mu(\gamma) = 0.5$, $\mu(\gamma) = 0.7$, $\mu(\gamma) = 0.9$). We then evaluate how the distribution of the final attitude changes in response. With increasing mean of self-reliance, i.e. a higher share of individualistic agents, the variance of the final attitude distribution increases and the difference in the final mean attitude of the two groups grows (Fig. 4 b-d). This maps an opinion drift towards stronger polarisation. To assess which agents populate the opinion tails, we merge the runs with varying means ($\mu \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$) into a shared ensemble and visualise the distribution of the final attitude separately for different thresholds of self-reliance ($\gamma \leq \frac{1}{3}, \frac{1}{3} < \gamma \leq \frac{2}{3}, \frac{2}{3} < \gamma$). As shown in Fig. 4, agents that have a higher self-reliance populate the tails of the final attitude distribution, such that these highly self-reliant individuals drive the polarisation of society at large, an observation also made in previous studies of more complex models[35].

Varying the standard deviation of the normal distribution of self-reliance compared to Fig. 3 in Supplementary Figs. 7, 12 and 17 the characteristics of resulting distributions persist, with lower standard deviations leading to slightly more concentrated attitudes (Supplementary

Fig. 7) and higher standard deviation to a larger spread (Supplementary Fig. 17). This also holds for Fig. 4 as shown in Supplementary Figs. 8, 13 and 18.

We also relax the assumption that initial attitude A_i equals inherent conviction A_i^*, observing similar increasing opinion spread with increasing self-reliance (Supplementary Figs. 21–24).

The opinion drift and polarisation observed here for increasing numbers of self-reliant agents ties in with empirical observations of societies that experience a growing number of citizens with strong, opposing political opinions over time. An example is the political polarisation of the United States, where opinion polls show near-normal distributions that drift apart over time[34]. This suggests that the mechanism we model has potential to map real-world phenomena.

Trade-off between self fulfilment and social cohesion

Every agent is equipped with an initial attitude which represents their individual stance of a topic. Their final attitude, on the other hand, arises dynamically in the field of tension between the influence of the neighbours and the initial opinion, weighted according to self-reliance. We now assess if the decision the agent would have taken based on their initial attitude $(0$ if $A_i^* \leq 0.5$, else $1)$ aligns with their final decision, taken based on the final attitude. The societal level of self-fulfilment is computed as share of all agents where initial and final decision agree (Fig. 5a). This shows that societies with more self-reliant agent achieve a higher average decision alignment: the agents are more prone to follow their own beliefs independently of their neighbourhoods. This behaviour is approximated with a mean-field approximation (see methods, grey line in Fig. 5a), showing that the society-wide mean self reliance might be used as a proxy to estimate results emerging from simulations based on individual agents. We contrast the alignment with the intrinsic goals with a measure of societal cohesion, which has been defined along three core dimensions[36]: Quality of social relations, identification with society and orientation towards the common good. We compute the difference of 90th and 10th percentile of the final attitude (Fig. 5b). This opinion spread measures the distance of opinions between the radical ends of the opinion spectrum, proxying identification with society represented by the average opinion as well as the ability of society to agree on the common good. It also increases with self-reliance, indicating a larger spread of attitudes and thus a potential decrease of social cohesion, since opinions in society are drifting apart. This means we observe a trade-off: If the number of self-reliant agents is low the opinion spread is small and social cohesion is high. However, the societal decision alignment is also lower as agents have to compromise more. We show this in Fig. 6 combining decision alignment against opinion spread. Varying the percentile levels leads to qualitative similar results compared to Fig. 5 in Supplementary Figs. 2 and 4 as well as for Fig. 6 in Supplementary Figs. 3 and 5. Also variation of the standard deviation of self-reliance does not alter the main observations, as shown for Fig. 5 in Supplementary Figs. 9, 14 and 19 and Fig. 6 in Supplementary Figs. 10, 15 and 20.

Discussion

Examining underlying drivers for human decision making is crucial for understanding societal processes. In this study, we use an agent-based approach to model opinion formation against the push-and-pull of individuality and group belonging. This approach is motivated by optimal distinctiveness theory[31] and the related concept of self-determination[32]. Even though there are individual studies exploring optimal distinctiveness theory in agent-based modelling[37], these studies do not discuss the broader implications for societal opinion formation. We pursue a strategy of model-driven exploration of behavioural patterns, positing that relatively simple rules for agents can reproduce emerging phenomena of society as demonstrated in recent work on emerging ostracism[38].

We find that even in a simple model, these opinion dynamics lead to polarisation with stable, opposing opinion clusters. Further, with an increasing number of independent agents which are difficult to influence, a stronger drift between opposing views emerges, culminating in a trade-off between the agent-level individual alignment with their personal opinion and societal cohesion.

Our findings align with empirical evidence for opinion drift that leads to stronger societal polarisation such as the political polarisation of the United States in the last decades[34]. The data the drifting apart of the mean distributions of political alignment between Republican and Democratic voters which maps the drift in attitude means we observe in Fig. 3. This suggests that the juxtaposition between belonging and self-reliance might be a mechanism that promotes social polarisation which complements other polarisation mechanisms that have been examined in the theoretical and empirical literature. The fact that more self-reliant agents lead the opinions of others can be connected to recent empirical work on political opinions[39]. Showing that complementing homophily by the preference to agree with more radical options, named acrophily, might contribute to political segregation. Thus our finding a trade-off between self-reliance and the higher frequency of more extreme opinions might be a harbinger of a divided opinion spectrum if self-reliance crosses a societal threshold. In contrast to theoretical work[9] or models[13] exploring polarisation of society, we do not model explicit repulsion from other opinions. Polarisation emerges from a stronger drive towards a personal inherent opinion. While this is a variation of similar mechanisms, the proposed interpretation as reliance on your own opinion aligns with psychological concepts like optimal distinction theory[31] or self-determination[32]. Introducing a bias towards a personal opinion, our model is related but distinct from the inclusion of biased-assimilation[8].

Since we aim to present a model as simple as possible, we omit additional mechanisms - but as Baldessari[2] argues it is hard to provide a granular general model and thus we present this new perspective on how polarisation might be driven by preference for inherent opinions as one potential part of the puzzle on the drivers and consequences societal polarisation. While our model is simple, it reproduces multiple previous findings of more complex models, especially we find polarisation emerging form a random initial state, the tendency to follow more radical agents and a connected trade-off between societal cohesion and individuality.

We have shown that by introducing an intuitive component of self-reliance into averaging neighbours model[7], polarisation emerges from an increasing spread of opinions and clustering occurs. Thus considering the individual reliance on inherent opinion complementing adjustment to opinions of social contacts may contribute to explaining multiple society wide phenomena of human decision making.

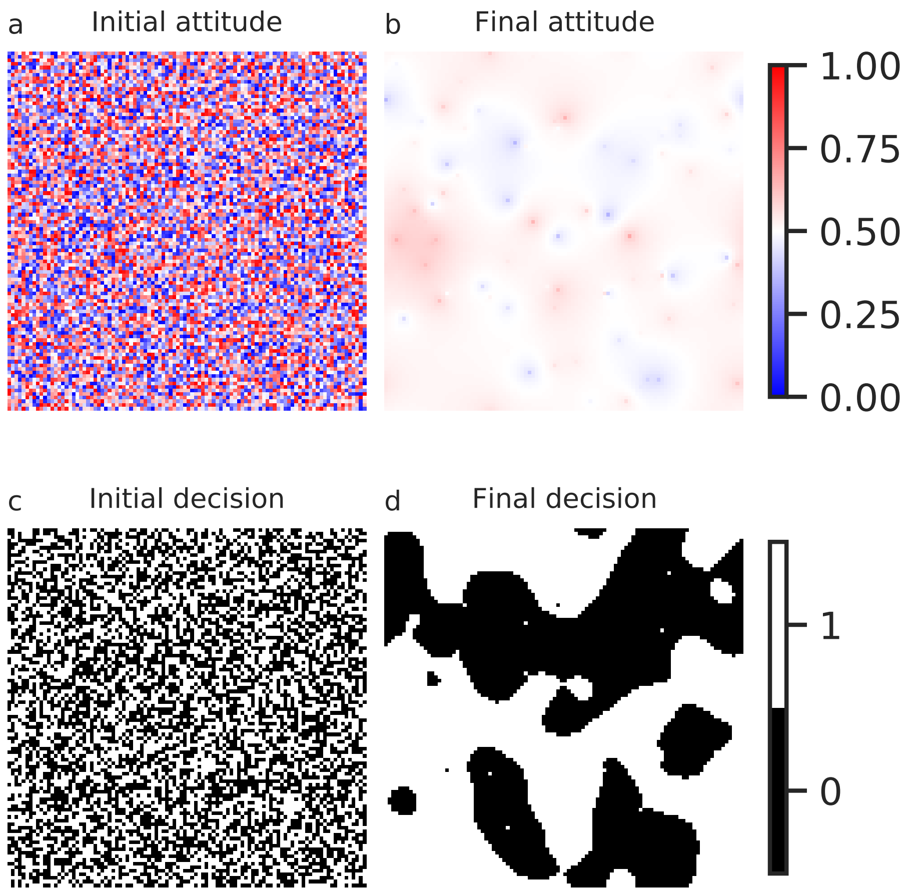


Figure 2: Trade-off between belonging and individualism leads to emergence of stable, opposing opinion clusters a shows the uniformly distributed initial attitude for each agent. Agents are placed randomly on the periodic grid. **b** shows the attitude after evolving the model for 1000 time steps. Opinion clusters emerge around few agents with strong opinions and many agents with more moderate views. **c** visualises the binary initial decision, which is based on the initial attitude (threshold of 0.5). No clusters are visible. In contrast, the final decision (**d**), which is based on the final attitude, shows clear opinion clusters.

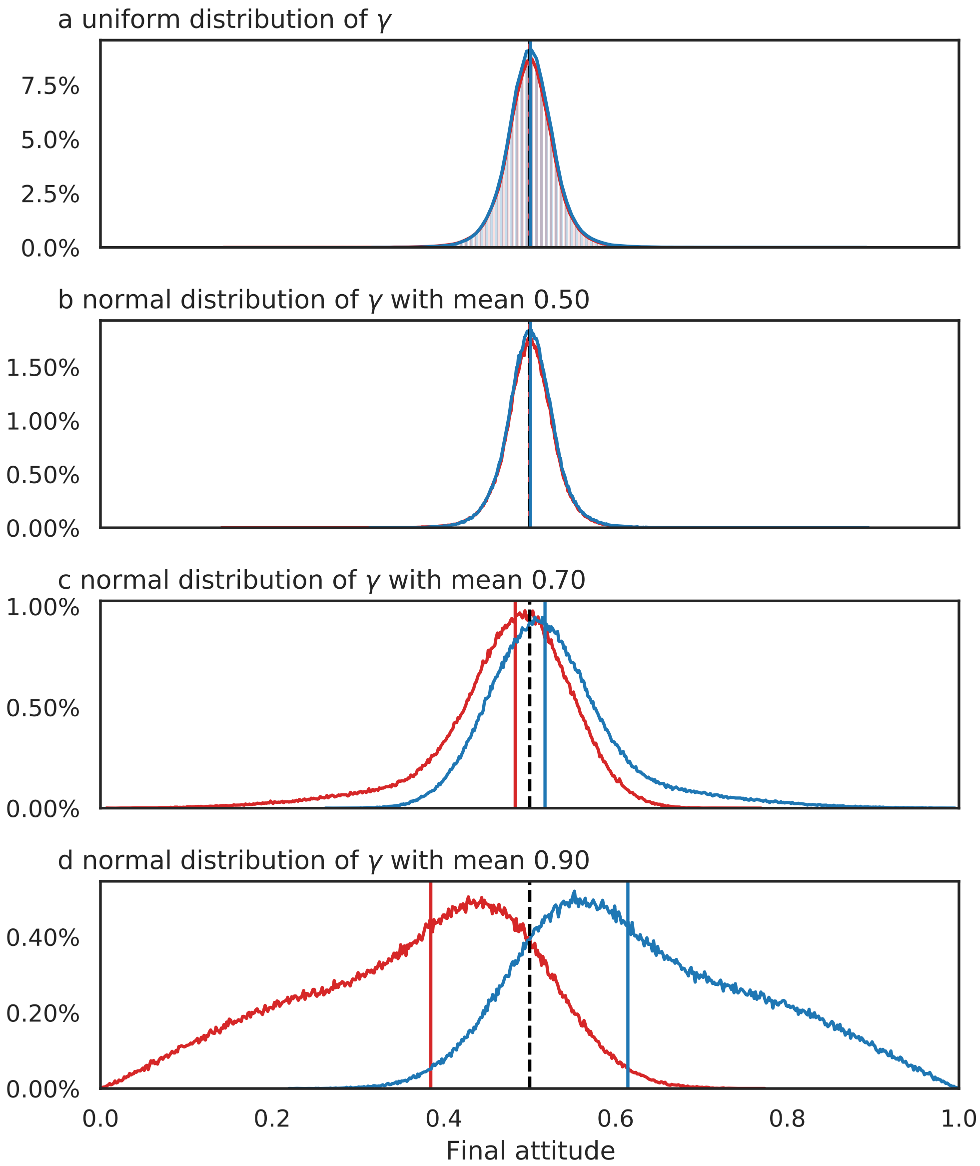


Figure 3: Opinion spread increases in more self-reliant societies. Panels show histograms of final attitude: **a** for uniform parameterised population, **b** for normal distributed γ with mean 0.5, **c** for normal distributed γ with mean 0.7 and **d** for normal distributed γ with mean 0.9. Red shows initial conviction < 0.5, blue shows initial attitude ≥ 0.5. Solid lines show medians, dashed black line 0.5.

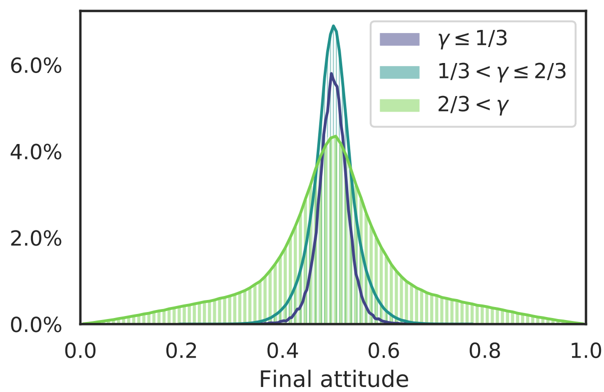


Figure 4: Highly self-reliant individuals have more polarised opinions. Histograms of final attitude for bins of individual self-reliance γ based on ensemble of normal distributions of γ with means between 0.5 and 0.9

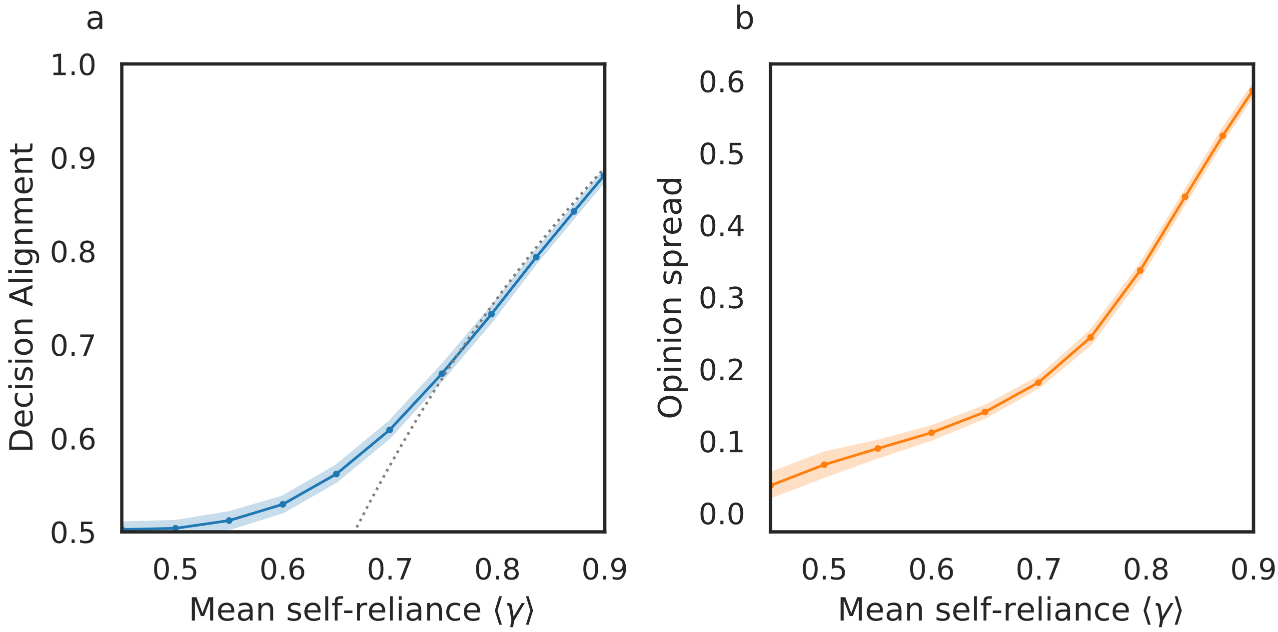


Figure 5: Trade-off between decision alignment and social cohesion. a shows the decision alignment in dependence of the average number of self-reliant agents. The decision alignment is computed as the average over the differences between the agents' initial decision and their final decision after evolving the model. The grey dots show an analytical approximation based on a mean-field approximation (Assuming 25% of agents are in equilibrium with their neighbours, i. e. $\mathbb{P}(N_i = 0) = 0.25$). **b** shows the opinion spread in dependence of the average number of self-reliant agents measured as the difference between the 90th and 10th percentiles of the distribution of the final attitude. (The initial opinion spread is about 0.8 due to the uniform distribution of the initial attitude.) If the opinion spread is large societal opinions are drifting apart and social cohesion lowers. Hence, there is a trade-off between higher personal decision alignment with more self-reliant agents and more social cohesion with less self-reliant agents. The average number of self-reliant agents corresponds to the mean of a normal distribution with mean γ and standard deviation σ = 0.1 from which the self-reliance was sampled. Confidence bands show the [5,95] Confidence interval based on 100 simulations with varying initial conditions.

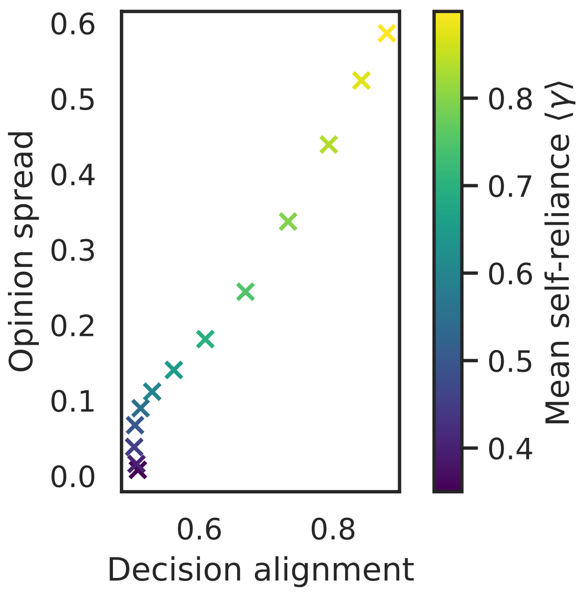


Figure 6: Trade-off between opinion spread and alignment with inherent decision. Decision alignment and opinion spread increase for higher mean self-reliance. Markers show mean values for colour-coded mean self-reliance.

Methods

Model description

We include the effects of self-reliance in an agent-based model of opinion formation based on interaction with neighbours. We assume that each agent has two time-invariant attributes: An *initial conviction* $A_i^* \in [0,1]$ and a degree of *self-reliance* $\gamma_i \in [0,1]$, which expresses the agent's need for individualism with $\gamma_i = 1$ corresponding to individualistic opinion formation and $\gamma_i = 0$ implying opinion formation purely driven by the social neighbourhood. Further each agents current opinion is described by a time-varying *attitude* parameter $A_i \in [0,1]$, which evolves as described by eq. (1). To initialise agent i at $t = 0$, attitude is initialised equal to initial conviction $A_i^* = A_i(0)$. A_i^* and γ_i are drawn from independent random distributions. For model variation, we provide results with independent draws of initial conviction A_i^* and initial attitude $A_i(0)$ in the supplement.

Agent i interacts with a set of neighbours M_i. In our simple example on a grid with periodic boundaries, i.e. a torus, all agents j with Chebyshev distance $d_C := \max\left(x_j - x_i, y_j - y_i\right) \leq 1$ to agent i are part of the neighbourhood M_i. The strength of the influence of j on i is given by the weight $w_{j,i}$, which we assume to be constant as $w_{j,i} := \frac{1}{\#M_i}\ \forall\ i,j$ for the sake of simplicity. For applications of our modelling framework on a network structure, these assumptions can be relaxed in the open source model implementation.

At each time step t, agent i adjusts their *attitude* A_i according to its difference to the average attitude of its neighbourhood (defined via the *neighbourhood influence*, N_i) weighted by $1 - \gamma_i$ and to its initial conviction A_i^* weighted by $\gamma_i|N_i|$. Herein the self-reliance term is proportional to the attitude difference to the neighbourhood. Thus the stronger one's opinion differs from that of your neighbours, the stronger the influence of your initial conviction. If the neighbours already have the same attitude as oneself, the influence of the initial Including a time scale τ, the dynamics of A_i are given by

$$\frac{\Delta A_i}{\Delta t} = \frac{1}{\tau}\left((1-\gamma_i)N_i + \gamma_i|N_i|\left(A_i^* - A_i\right)\right) \tag{1}$$

with *neighbourhood influence* defined as $N_i = \sum_{j \in M_i} w_{j,i}(A_j - A_i)$. An alternative representation of the same dynamics offers a slightly different interpretation.

$$\begin{aligned}\frac{\Delta A_i}{\Delta t} &= \frac{|N_i|}{\tau}\left((1-\gamma_i)\frac{N_i}{|N_i|} + \gamma_i\left(A_i^* - A_i\right)\right)\\ &= \frac{|N_i|}{\tau}\left((1-\gamma_i)\mathrm{sign}\left(N_i\right) + \gamma_i\left(A_i^* - A_i\right)\right)\end{aligned} \tag{2}$$

In this representation the difference to the neighbourhood merely changes the time scale of the dynamics. The core of the dynamics (within the parenthesis) is now a competition between the influence of the individual initial conviction and the neighbourhood which now merely enters as a direction of change, since only the sign of the difference enters the dynamics.

To analyse the resulting equilibria in our simple framework, we run the model until it converges to a steady state. Once the model has reached its equilibrium, each agent is assigned a final decision d_i

$$d_i = \mathbf{1}_{A_i > 0.5} \tag{3}$$

The model structure and dynamics are visualised in Fig. 1.

Equilibrium conditions The model equilibrium is given by the concurrent individual agent equilibria defined by the condition

$$0 = \frac{1}{\tau}\left((1-\gamma_i)N_i + \gamma_i|N_i|(A_i^* - A_i)\right)\ . \tag{4}$$

This equation has two qualitatively different solutions:

$$N_i = 0 \text{ and } A_i^* - A_i = \frac{1-\gamma_i}{\gamma_i}\mathrm{sign}\left(N_i\right) \tag{5}$$

The first solution describes the situation when an agents attitude equals the average of their neighbours attitudes. The second equilibrium with $N_i \neq 0$ is reached if the dissonance between initial conviction and actual attitude is equal to the strength of individuality and direction of the disagreement with the neighbours. From the parameter ranges of γ_i, A_i^* and A_i follows that this equilibrium can only be met by agents with $\gamma_i \geq 0.5$, since

$$1 \geq |A_i^* - A_i| = \frac{1-\gamma_i}{\gamma_i} \Rightarrow \gamma_i \geq 0.5 \tag{6}$$

Analytical derivation of average decision alignment

To analyse the dependence of society wide decision alignment Δ on the distribution of self-reliance γ_i, we approximate the expected decision alignment of agent i with self-reliance γ_i based on the possible equilibria eq. (5). The expectation of δ_i in the final equilibrated state can be estimated as

$$\mathbb{E}\left[\delta_i)\right] \approx \mathbb{P}(N_i < 0)\mathbb{P}(\delta_i = 1|N_i < 0) + \mathbb{P}(N_i > 0)\mathbb{P}(\delta_i = 1|N_i > 0) + \mathbb{P}(N_i = 0)\mathbb{P}(\delta_i = 1|N_i = 0), \tag{7}$$

assuming independence of N_i from A_i, which is wrong on the individual level, but can be considered a mean-field approximation of societal behaviour at large.

Assuming equilibrium with $N_i \neq 0$ we derive in the next paragraph, that

$$\mathbb{P}(\delta_i = 1|N_i < 0) = \mathbb{P}(\delta_i = 1|N_i > 0) = \frac{1}{2} + \frac{1}{2}\mathbb{P}(\mathcal{U}(0, 0.5) > \frac{1-\gamma_i}{\gamma_i}). \tag{8}$$

For $N_i = 0$ it follows due to the uniform distribution of A_i and A_i^*, that

$$\mathbb{P}(\delta_i = 1|N_i = 0) = \mathbb{P}(\delta_i = 1|A_i < 0.5) + \mathbb{P}(\delta_i = 1|A_i \geq 0.5) = \frac{1}{4} + \frac{1}{4}.$$

Thus, with this simplifying mean-field analysis we find

$$\begin{aligned} \mathbb{E}\left[\delta_i)\right] = & \quad \left[\mathbb{P}(N_i < 0) + \mathbb{P}(N_i > 0)\right]\left[\frac{1}{2} + \mathbb{P}(\mathcal{U}(0, 0.5) > \frac{1-\gamma_i}{\gamma_i})\right] + \frac{1}{2}\mathbb{P}(N_i = 0) && (9) \\ = & \quad \frac{1}{2} + \left[\mathbb{P}(N_i \neq 0)\right]\mathbb{P}(\mathcal{U}(0, 0.5) > \frac{1-\gamma_i}{\gamma_i}) && (10) \end{aligned}$$

Considering $\mathbb{P}(\mathcal{U}(0, 0.5) > \frac{1-\gamma_i}{\gamma_i})$, it follows that only individuals with $\gamma_i \geq \frac{2}{3}$ can contribute to this term, since

$$0.5 \geq \frac{1-\gamma_i}{\gamma_i} \Leftrightarrow 0.5\gamma_i \geq 1-\gamma_i \Leftrightarrow \gamma_i \geq \frac{2}{3}$$

Detailed derivation of eq. (8) We consider the case $N_i < 0$ and $N_i > 0$ separately. We use the uniform distribution of $A_i^* \in [0, 1]$. We assume that N_i is independent of A_i or A_i^*. This assumption is simplifying and may be used to estimate the average over all agents, not for individual agents i. Decision alignment $\delta_i = 1$, if $A_i^* > \frac{1}{2}$ and $A_i > \frac{1}{2}$ or $A_i^* < \frac{1}{2}$ and $A_i < \frac{1}{2}$.

If $N_i < 0$, the equilibrium is given by

$$A_i = A_i^* + \frac{1-\gamma_i}{\gamma_i}.$$

For $A_i^* > \frac{1}{2}$, $A_i > \frac{1}{2}$ if $A_i^* + \frac{1-\gamma_i}{\gamma_i} > \frac{1}{2}$, which holds since $\frac{1-\gamma_i}{\gamma_i} \geq 0$. For $A_i^* < \frac{1}{2}$, $A_i < \frac{1}{2}$ if $A_i^* + \frac{1-\gamma_i}{\gamma_i} < \frac{1}{2}$, which holds if $A_i^* - \frac{1}{2} < -\frac{1-\gamma_i}{\gamma_i}$. Since $A_i^* < \frac{1}{2}$, this is equivalent to $\mathbb{P}(\mathcal{U}(-0.5, 0) < -\frac{1-\gamma_i}{\gamma_i})$.

If $N_i > 0$, the equilibrium is given by

$$A_i = A_i^* - \frac{1-\gamma_i}{\gamma_i}.$$

For $A_i^* > \frac{1}{2}$, $A_i > \frac{1}{2}$ if $A_i^* + \frac{1-\gamma_i}{\gamma_i} > \frac{1}{2}$, which holds if $A_i^* - \frac{1}{2} > \frac{1-\gamma_i}{\gamma_i}$. Since $A_i^* > \frac{1}{2}$, this is equivalent to $\mathbb{P}(\mathcal{U}(0, 0.5) > \frac{1-\gamma_i}{\gamma_i})$. For $A_i^* < \frac{1}{2}$, $A_i < \frac{1}{2}$ if $A_i^* - \frac{1-\gamma_i}{\gamma_i} < \frac{1}{2}$, which holds since $\frac{1-\gamma_i}{\gamma_i} \geq 0$.

With $\mathbb{P}(A_i^*) < 0.5) = \mathbb{P}(A_i^*) > 0.5) = \frac{1}{2}$ and $\mathbb{P}(\mathcal{U}(0, 0.5) > \frac{1-\gamma_i}{\gamma_i}) = \mathbb{P}(\mathcal{U}(-0.5, 0) < -\frac{1-\gamma_i}{\gamma_i})$, eq. (8) follows.

Supplementary Material

Supplementary material is available online.

Code availability

The model code and analysis code will be available open source on Github.

Data availability

The simulation data that support the findings of this study will be openly available at the public repository for this publication with identifier 10.5281/zenodo.8363819.

References

[1] Levin, S. A., Milner, H. V. & Perrings, C. The dynamics of political polarization. *Proceedings of the National Academy of Sciences* **118**, e2116950118 (2021). DOI 10.1073/pnas.2116950118.

[2] Baldassarri, D. & Page, S. E. The emergence and perils of polarization. *Proceedings of the National Academy of Sciences* **118**, e2116863118 (2021). DOI 10.1073/pnas.2116863118.

[3] Sakoda, J. M. The checkerboard model of social interaction. *The Journal of Mathematical Sociology* **1**, 119–132 (1971). DOI 10.1080/0022250X.1971.9989791.

[4] Schelling, T. C. Dynamic models of segregation. *The Journal of Mathematical Sociology* **1**, 143–186 (1971). DOI 10.1080/0022250X.1971.9989794.

[5] Bianchi, F. & Squazzoni, F. Agent-based models in sociology. *WIREs Computational Statistics* **7**, 284–306 (2015). DOI 10.1002/wics.1356.

[6] Sobkowicz, P. Whither Now, Opinion Modelers? *Frontiers in Physics* **8** (2020).

[7] DeGroot, M. H. Reaching a Consensus. *Journal of the American Statistical Association* **69**, 118–121 (1974). DOI 10.2307/2285509.

[8] Dandekar, P., Goel, A. & Lee, D. T. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences* **110**, 5791–5796 (2013). DOI 10.1073/pnas.1217220110.

[9] Chuang, Y.-L., D'Orsogna, M. & Chou, T. A bistable belief dynamics model for radicalization within sectarian conflict. *Quarterly of Applied Mathematics* **75**, 19–37 (2016). DOI 10.1090/qam/1446.

[10] DellaPosta, D., Shi, Y. & Macy, M. Why Do Liberals Drink Lattes? *AJS; American journal of sociology* **120**, 1473–1511 (2015). DOI 10.1086/681254.

[11] Korbel, J., Lindner, S. D., Pham, T. M., Hanel, R. & Thurner, S. Homophily-Based Social Group Formation in a Spin Glass Self-Assembly Framework. *Physical Review Letters* **130**, 057401 (2023). DOI 10.1103/PhysRevLett.130.057401.

[12] Pham, T. M., Korbel, J., Hanel, R. & Thurner, S. Empirical social triad statistics can be explained with dyadic homophylic interactions. *Proceedings of the National Academy of Sciences* **119**, e2121103119 (2022). DOI 10.1073/pnas.2121103119.

[13] Axelrod, R., Daymude, J. J. & Forrest, S. Preventing extreme polarization of political attitudes. *Proceedings of the National Academy of Sciences* **118**, e2102139118 (2021). DOI 10.1073/pnas.2102139118.

[14] Leonard, N. E., Lipsitz, K., Bizyaeva, A., Franci, A. & Lelkes, Y. The nonlinear feedback dynamics of asymmetric political polarization. *Proceedings of the National Academy of Sciences* **118**, e2102149118 (2021). DOI 10.1073/pnas.2102149118.

[15] Kawakatsu, M., Lelkes, Y., Levin, S. A. & Tarnita, C. E. Interindividual cooperation mediated by partisanship complicates Madison's cure for "mischiefs of faction". *Proceedings of the National Academy of Sciences* **118**, e2102148118 (2021). DOI 10.1073/pnas.2102148118.

[16] Kozlowski, A. C. & Murphy, J. P. Issue alignment and partisanship in the American public: Revisiting the 'partisans without constraint' thesis. *Social Science Research* **94**, 102498 (2021). DOI 10.1016/j.ssresearch.2020.102498.

[17] Chu, O. J., Donges, J. F., Robertson, G. B. & Pop-Eleches, G. The microdynamics of spatial polarization: A model and an application to survey data from Ukraine. *Proceedings of the National Academy of Sciences* **118**, e2104194118 (2021). DOI 10.1073/pnas.2104194118.

[18] Gajewski, L. G., Sienkiewicz, J. & Holyst, J. A. Transitions between polarization and radicalization in a temporal bilayer echo-chamber model. *Physical Review E* **105**, 024125 (2022). DOI 10.1103/PhysRevE.105.024125.

[19] Liu, J., Huang, S., Aden, N. M., Johnson, N. F. & Song, C. Emergence of Polarization in Coevolving Networks. *Physical Review Letters* **130**, 037401 (2023). DOI 10.1103/PhysRevLett.130.037401.

[20] Ferraz de Arruda, H. *et al.* Modelling how social network algorithms can influence opinion polarization. *Information Sciences* **588**, 265–278 (2022). DOI 10.1016/j.ins.2021.12.069.

[21] Cinus, F., Minici, M., Monti, C. & Bonchi, F. The Effect of People Recommenders on Echo Chambers and Polarization. *Proceedings of the International AAAI Conference on Web and Social Media* **16**, 90–101 (2022). DOI 10.1609/icwsm.v16i1.19275.

[22] Nyhan, B. *et al.* Like-minded sources on Facebook are prevalent but not polarizing. *Nature* 1–8 (2023). DOI 10.1038/s41586-023-06297-w.

[23] Guess, A. M. *et al.* How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* **381**, 398–404 (2023). DOI 10.1126/science.abp9364.

[24] Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M. & Ryan, J. B. Affective polarization, local contexts and public opinion in America. *Nature Human Behaviour* **5**, 28–38 (2021). DOI 10.1038/s41562-020-01012-5.

[25] Lobinska, G., Pauzner, A., Traulsen, A., Pilpel, Y. & Nowak, M. A. Evolution of resistance to COVID-19 vaccination with dynamic social distancing. *Nature Human Behaviour* **6**, 193–206 (2022). DOI 10.1038/s41562-021-01281-8.

[26] Charron, N., Lapuente, V. & Rodríguez-Pose, A. Uncooperative society, uncooperative politics or both? Trust, polarization, populism and COVID-19 deaths across European regions. *European Journal of Political Research* **62**, 781–805 (2023). DOI 10.1111/1475-6765.12529.

[27] Ehret, S., Constantino, S. M., Weber, E. U., Efferson, C. & Vogt, S. Group identities can undermine social tipping after intervention. *Nature Human Behaviour* 1–11 (2022). DOI 10.1038/s41562-022-01440-5.

[28] Abramowitz, A. & McCoy, J. United States: Racial Resentment, Negative Partisanship, and Polarization in Trump's America. *The ANNALS of the American Academy of Political and Social Science* **681**, 137–156 (2019). DOI 10.1177/0002716218811309.

[29] Ravndal, J. A. Explaining right-wing terrorism and violence in Western Europe: Grievances, opportunities and polarisation. *European Journal of Political Research* **57**, 845–866 (2018). DOI 10.1111/1475-6765.12254.

[30] Moore, F. C. *et al.* Determinants of emissions pathways in the coupled climate–social system. *Nature* **603**, 103–111 (2022). DOI 10.1038/s41586-022-04423-8.

[31] Brewer, M. B. The Social Self: On Being the Same and Different at the Same Time. *Personality and Social Psychology Bulletin* **17**, 475–482 (1991). DOI 10.1177/0146167291175001.

[32] Deci, E. L. & Ryan, R. M. Self-Determination Theory. In *Handbook of Theories of Social Psychology: Volume 1*, 416–437 (SAGE Publications Ltd, London, 2012).

[33] Baldassarri, D. & Gelman, A. Partisans without Constraint: Political Polarization and Trends in American Public Opinion. *AJS; American journal of sociology* **114**, 408–446 (2008). DOI 10.2139/ssrn.1010098.

[34] Geiger, A. Political Polarization in the American Public (2014).

[35] Turner, M. A. & Smaldino, P. E. Paths to Polarization: How Extreme Views, Miscommunication, and Random Chance Drive Opinion Dynamics. *Complexity* **2018**, e2740959 (2018). DOI 10.1155/2018/2740959.

[36] Schiefer, D. & van der Noll, J. The Essentials of Social Cohesion: A Literature Review. *Social Indicators Research* **132**, 579–603 (2017). DOI 10.1007/s11205-016-1314-5.

[37] Smaldino, P., Pickett, C., Sherman, J. & Schank, J. An Agent-Based Model of Social Identity Dynamics. *Journal of Artificial Societies and Social Simulation* **15**, 7 (2012).

[38] Lindström, B. & Tobler, P. N. Incidental ostracism emerges from simple learning mechanisms. *Nature Human Behaviour* **2**, 405–414 (2018). DOI 10.1038/s41562-018-0355-y.

[39] Goldenberg, A. *et al.* Homophily and acrophily as drivers of political segregation. *Nature Human Behaviour* **7**, 219–230 (2023). DOI 10.1038/s41562-022-01474-9.

[40] Datseris, G., Vahdati, A. R. & DuBois, T. C. Agents.jl: a performant and feature-full agent-based modeling software of minimal code complexity. *SIMULATION* **0**, 003754972110688 (2022). DOI 10.1177/00375497211068820.

Acknowledgements

We thank the developers of the Agents.jl[40] package for providing the basis for our model development and Maximilian Kotz for helpful discussions.

This research has received funding from the German Federal Ministry of Education and Research (BMBF) under the research projects QUIDIC (01LP1907A), from the Volkswagen foundation and the Siemens foundation.

The authors gratefully acknowledge the European Regional Development Fund (ERDF), the German Federal Ministry of Education and Research, and the Land Brandenburg for supporting this project by providing resources on the high performance computer system at the Potsdam Institute for Climate Impact Research.

Author Contribution

L.Q, A.S. and A.L. designed the study. L.Q. and A.S. developed the model code. D.H. conducted exploratory analysis and extended the model following discussions with all authors. D.H. analysed the supplementary two-agent case. L.Q. and A.S. wrote the manuscript. All authors discussed the results and approved the final manuscript.

Competing Interests

The authors declare that they have no competing interests.

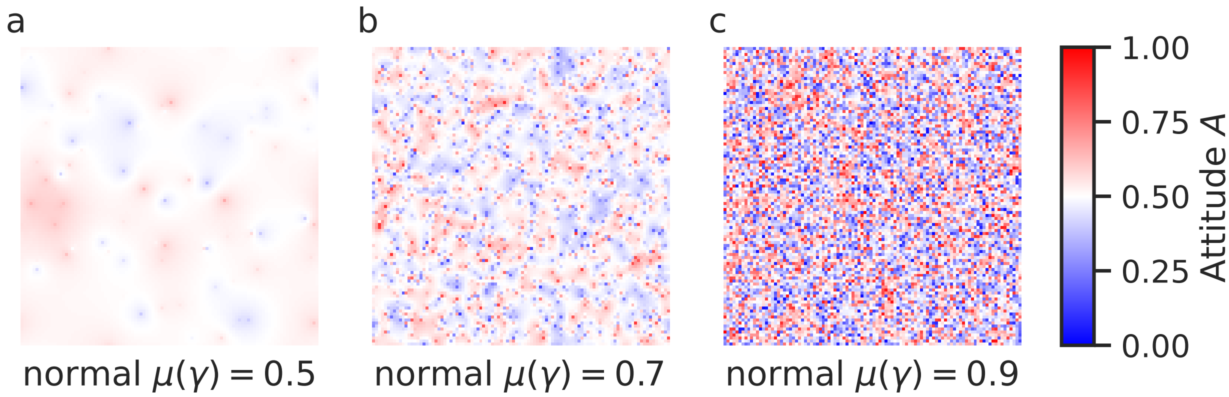
Supplementary information for

Convolution of individual and group identity: self-reliance increases polarisation in basic opinion model

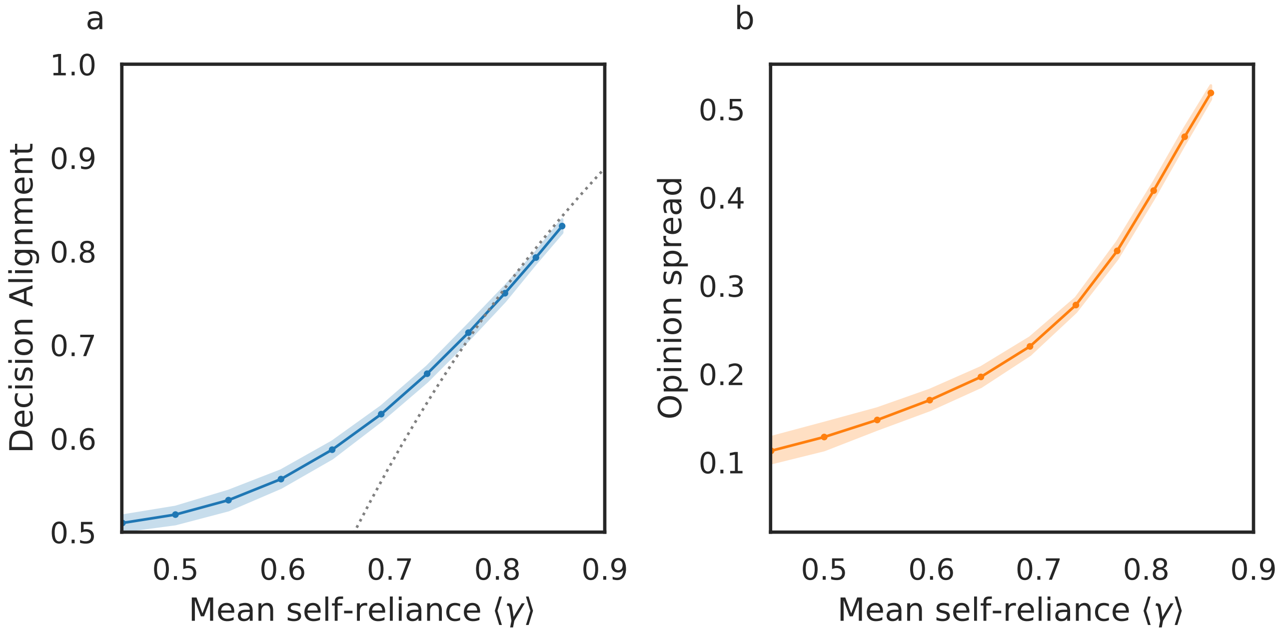
L. Quante*, A. Stechemesser*, D. Hödtke, A. Levermann

*These two authors contributed equally.

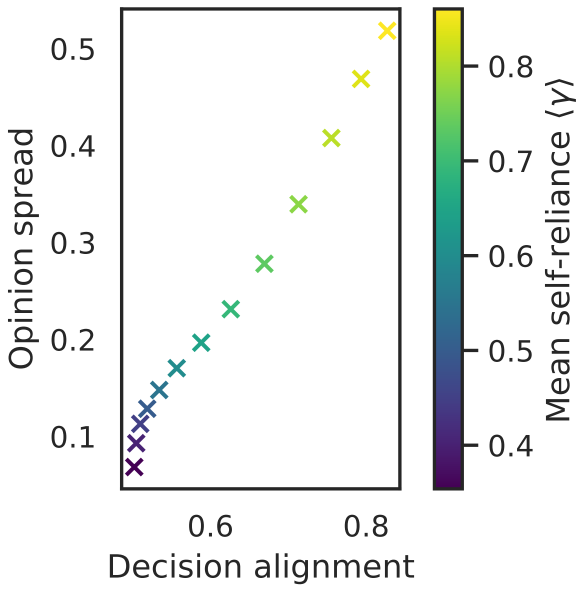
Supplementary figures



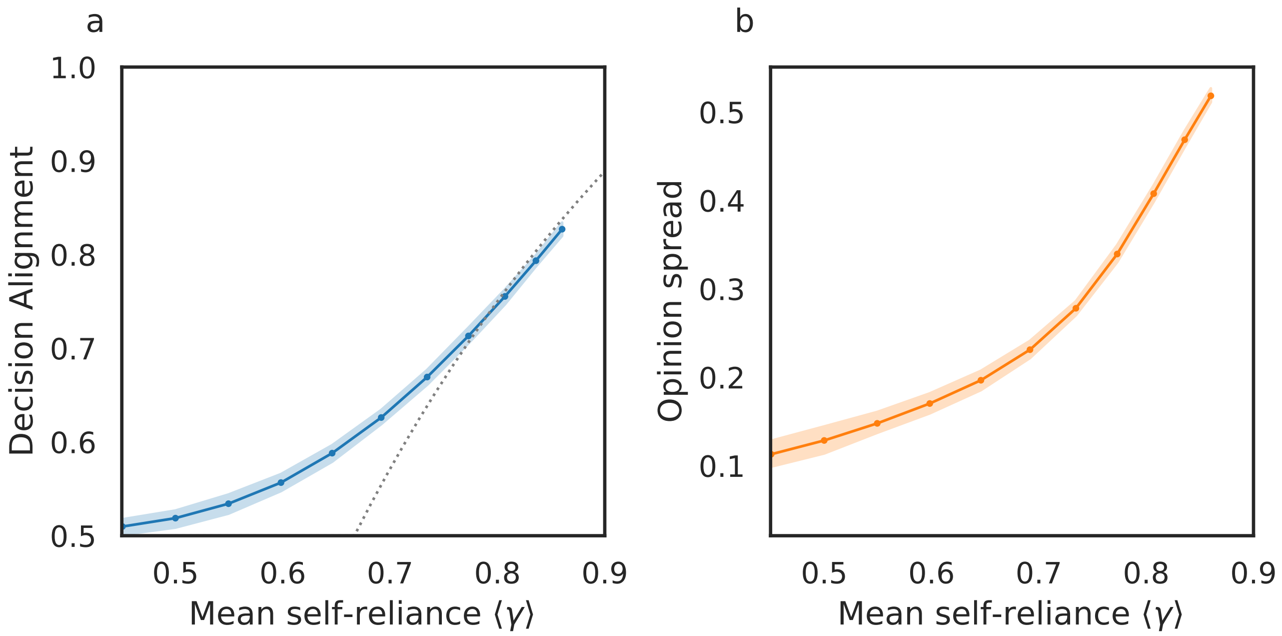
Supplementary Figure 1 Emergence of opinion clusters changes with distribution of γ **– standard deviation** $\sigma = 0.1$ **as in the main** Panels shows the attitude after evolving the model for 1000 time steps **a** for normal distributed γ_i with mean 0.5, **b** for normal distributed γ_i with mean 0.7 and **c** for normal distributed γ_i with mean 0.9



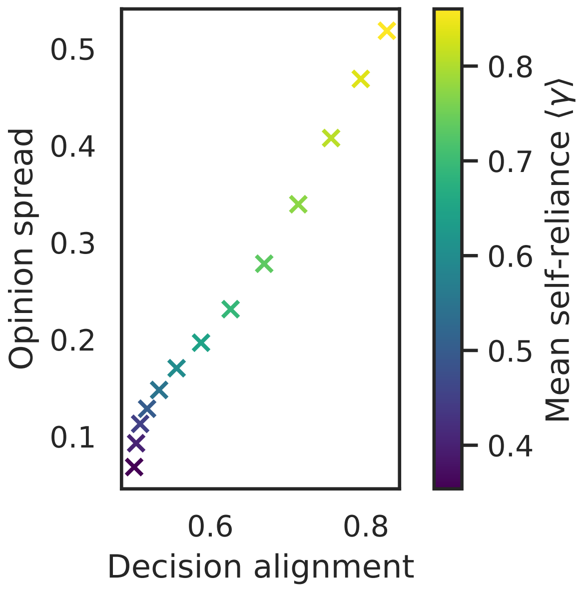
Supplementary Figure 2 Trade-off between decision alignment and social cohesion – standard deviation $\sigma = 0.10$. a shows the decision alignment in dependence of the average number of self-reliant agents. The decision alignment is computed as the average over the differences between the agents' initial decision and their final decision after evolving the model. The grey dots show an analytical approximation based on a mean-field approximation (Assuming 25% of agents are in equilibrium with their neighbours, i. e. $\mathbb{P}(N_i = 0) = 0.25$). **b** shows the opinion spread in dependence of the average number of self-reliant agents measured as the difference between the 95th and 5th percentiles of the distribution of the final attitude. (The initial opinion spread is about 0.9 due to the uniform distribution of the initial attitude.) If the opinion spread is large societal opinions are drifting apart and social cohesion lowers. Hence, there is a trade-off between higher personal decision alignment with more self-reliant agents and more social cohesion with less self-reliant agents. The average number of self-reliant agents corresponds to the mean of a normal distribution with mean γ and standard deviation σ = 0.1 from which the self-reliance was sampled. Confidence bands show the [5,95] Confidence interval based on 100 simulations with varying initial conditions.



Supplementary Figure 3 Trade-off between opinion spread and alignment with inherent decision – standard deviation $\sigma = 0.10$. Decision alignment and opinion spread increase for higher mean self-reliance. Opinion spread between the 95th and 5th percentiles of final attitudes. Markers show mean values for colour-coded mean self-reliance.



Supplementary Figure 4 Trade-off between decision alignment and social cohesion – standard deviation $\sigma = $ 0.10. a shows the decision alignment in dependence of the average number of self-reliant agents. The decision alignment is computed as the average over the differences between the agents' initial decision and their final decision after evolving the model. The grey dots show an analytical approximation based on a mean-field approximation (Assuming 25% of agents are in equilibrium with their neighbours, i. e. $\mathbb{P}(N_i = 0) = 0.25$). **b** shows the opinion spread in dependence of the average number of self-reliant agents measured as the difference between the 75th and 25th percentiles of the distribution of the final attitude. (The initial opinion spread is about 0.5 due to the uniform distribution of the initial attitude.) If the opinion spread is large societal opinions are drifting apart and social cohesion lowers. Hence, there is a trade-off between higher personal decision alignment with more self-reliant agents and more social cohesion with less self-reliant agents. The average number of self-reliant agents corresponds to the mean of a normal distribution with mean γ and standard deviation σ = 0.1 from which the self-reliance was sampled. Confidence bands show the [5,95] Confidence interval based on 100 simulations with varying initial conditions.

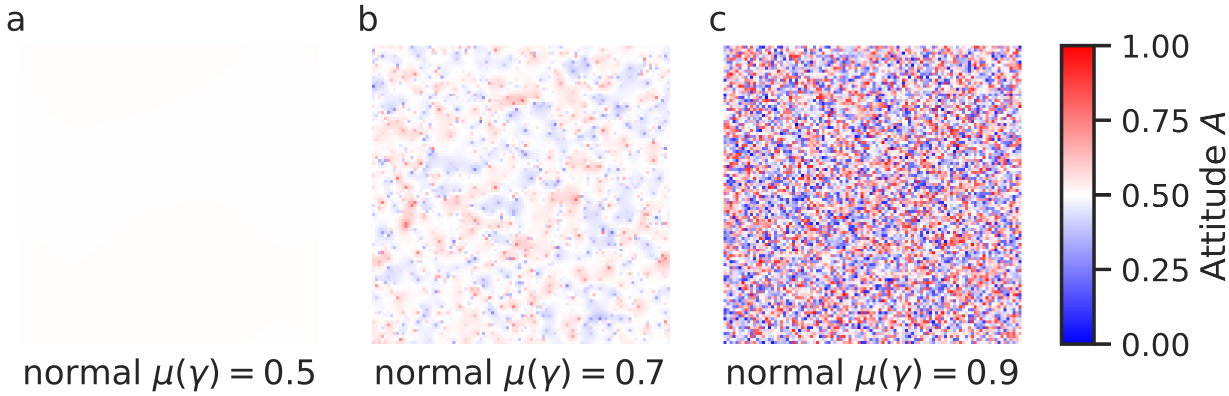


Supplementary Figure 5 Trade-off between opinion spread and alignment with inherent decision – standard deviation $\sigma = 0.10$. Decision alignment and opinion spread increase for higher mean self-reliance. Opinion spread between the 75th and 25th percentiles of final attitudes. Markers show mean values for colour-coded mean self-reliance.

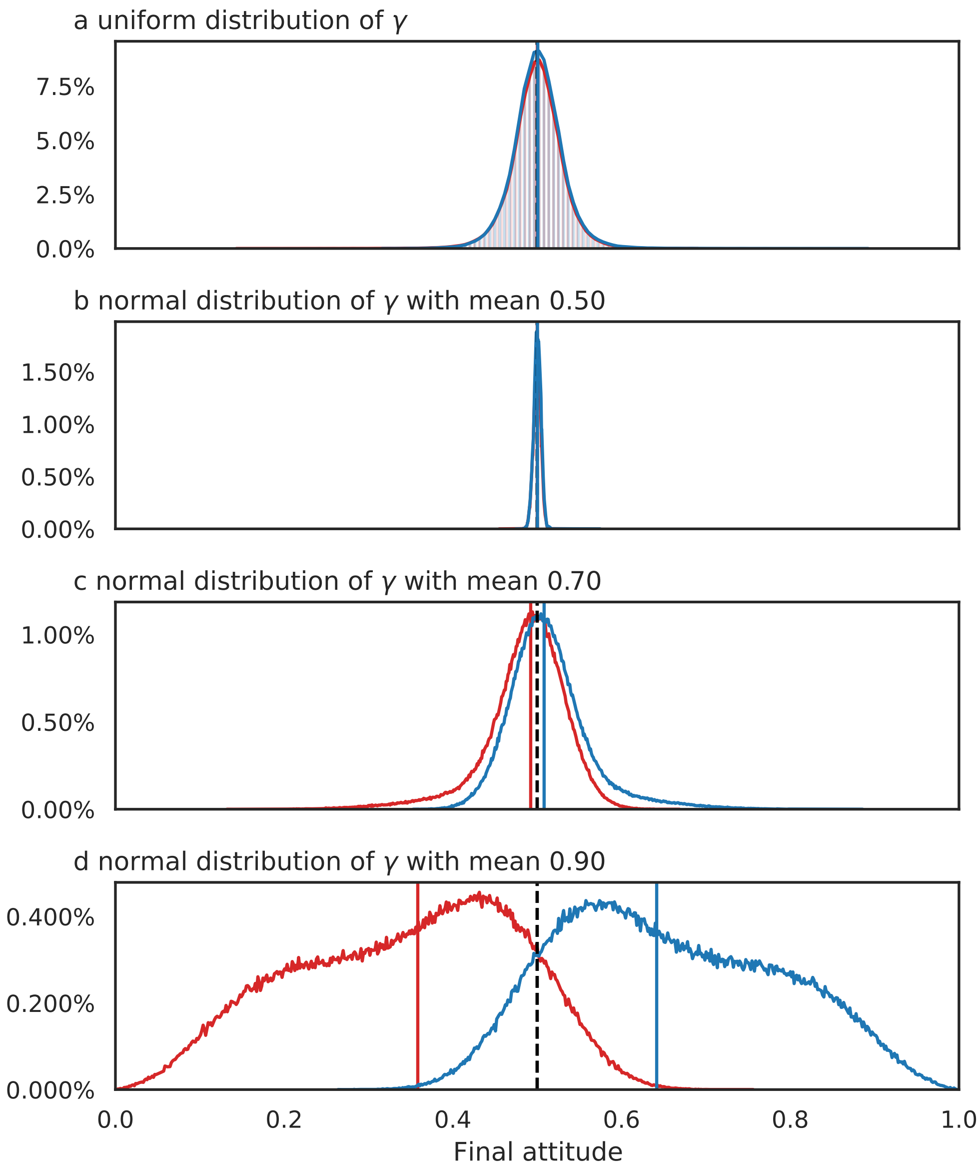
Sensitivity to standard deviation for the normal distributions of self-reliance

To check sensitivity of the main results using a standard deviation of 0.1 for the normal distribution of self-reliance γ, we show results for alternative parameters:

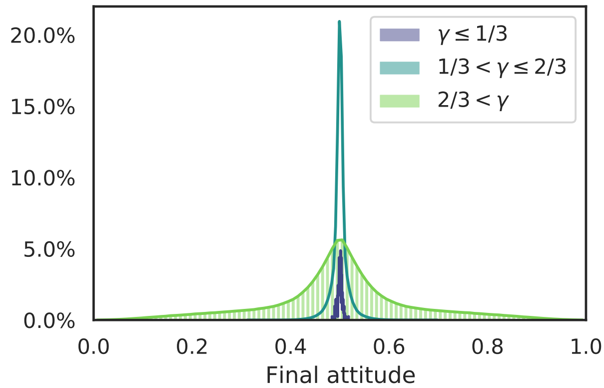
Standard deviation $\sigma = 0.05$



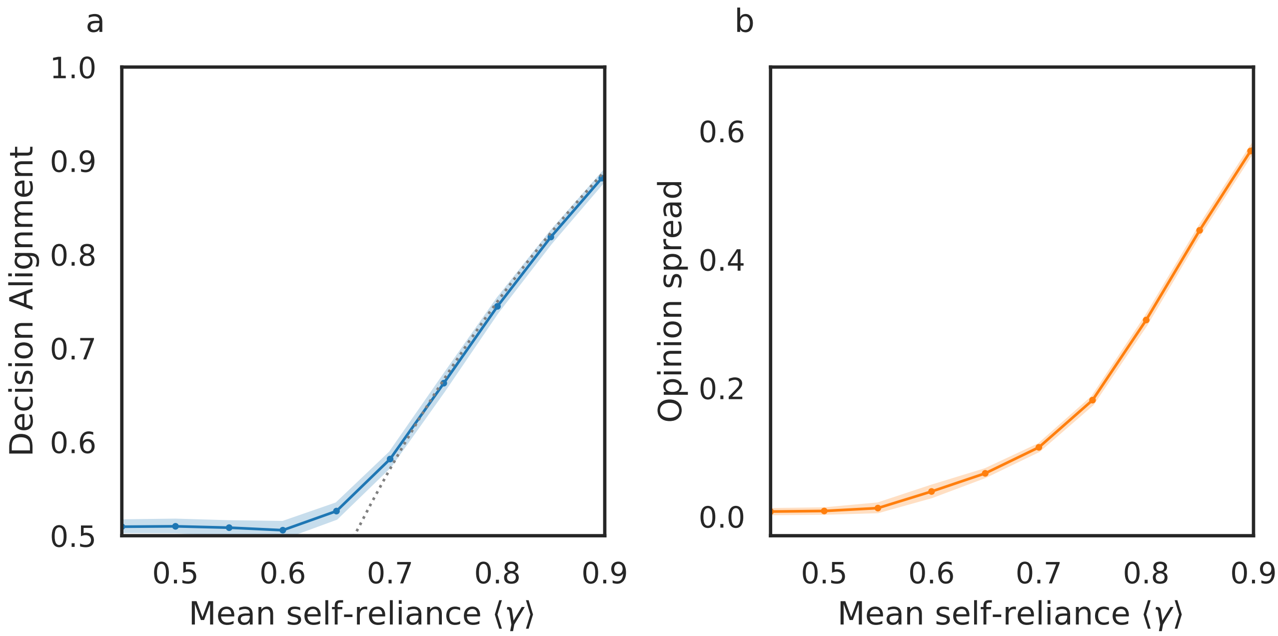
Supplementary Figure 6 Emergence of opinion clusters changes with distribution of γ – standard deviation $\sigma = 0.05$. Panels shows the attitude after evolving the model for 1000 time steps **a** for normal distributed γ_i with mean 0.5, **b** for normal distributed γ_i with mean 0.7 and **c** for normal distributed γ_i with mean 0.9



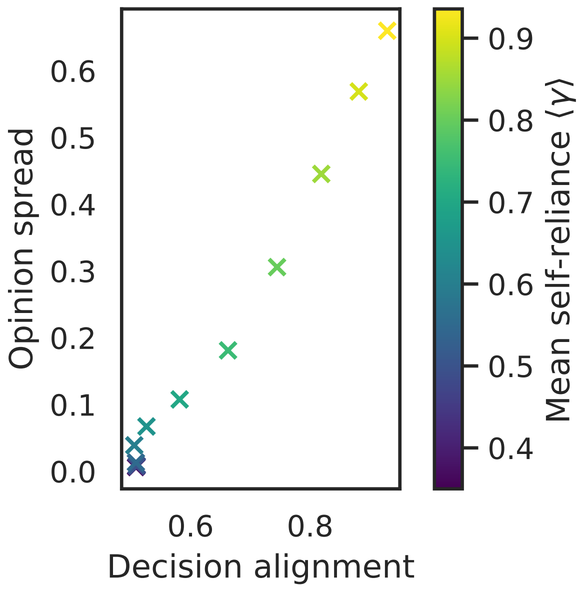
Supplementary Figure 7 Opinion spread increases in more self-reliant societies – standard deviation $\sigma = 0.05$. Panels show histograms of final attitude: **a** for uniform parameterised population, **b** for normal distributed γ with mean 0.5, **c** for normal distributed γ with mean 0.7 and **d** for normal distributed γ with mean 0.9. Red shows initial conviction < 0.5, blue shows initial attitude ≥ 0.5. Solid lines show medians, dashed black line 0.5.



Supplementary Figure 8 Highly self-reliant individuals have more polarised opinions – standard deviation $\sigma = 0.05$. Histograms of final attitude for bins of individual self-reliance γ based on ensemble of normal distributions of γ with means between 0.5 and 0.9

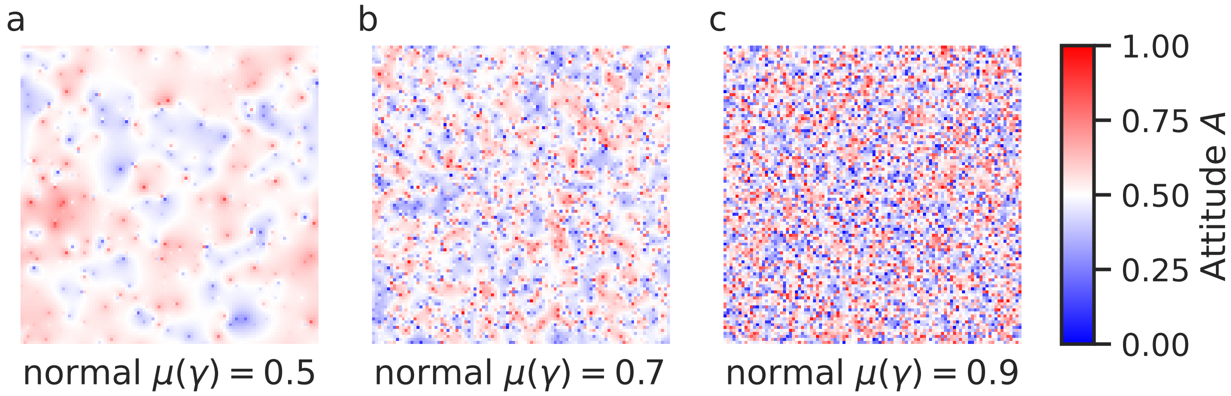


Supplementary Figure 9 Trade-off between decision alignment and social cohesion – standard deviation $\sigma = 0.05$. a shows the decision alignment in dependence of the average number of self-reliant agents. The decision alignment is computed as the average over the differences between the agents' initial decision and their final decision after evolving the model. The grey dots show an analytical approximation based on a mean-field approximation (Assuming 25% of agents are in equilibrium with their neighbours, i. e. $\mathbb{P}(N_i = 0) = 0.25$). **b** shows the opinion spread in dependence of the average number of self-reliant agents measured as the difference between the 90th and 10th percentiles of the distribution of the final attitude. (The initial opinion spread is about 0.8 due to the uniform distribution of the initial attitude.) If the opinion spread is large societal opinions are drifting apart and social cohesion lowers. Hence, there is a trade-off between higher personal decision alignment with more self-reliant agents and more social cohesion with less self-reliant agents. The average number of self-reliant agents corresponds to the mean of a normal distribution with mean γ and standard deviation σ = 0.1 from which the self-reliance was sampled. Confidence bands show the [5,95] Confidence interval based on 100 simulations with varying initial conditions.

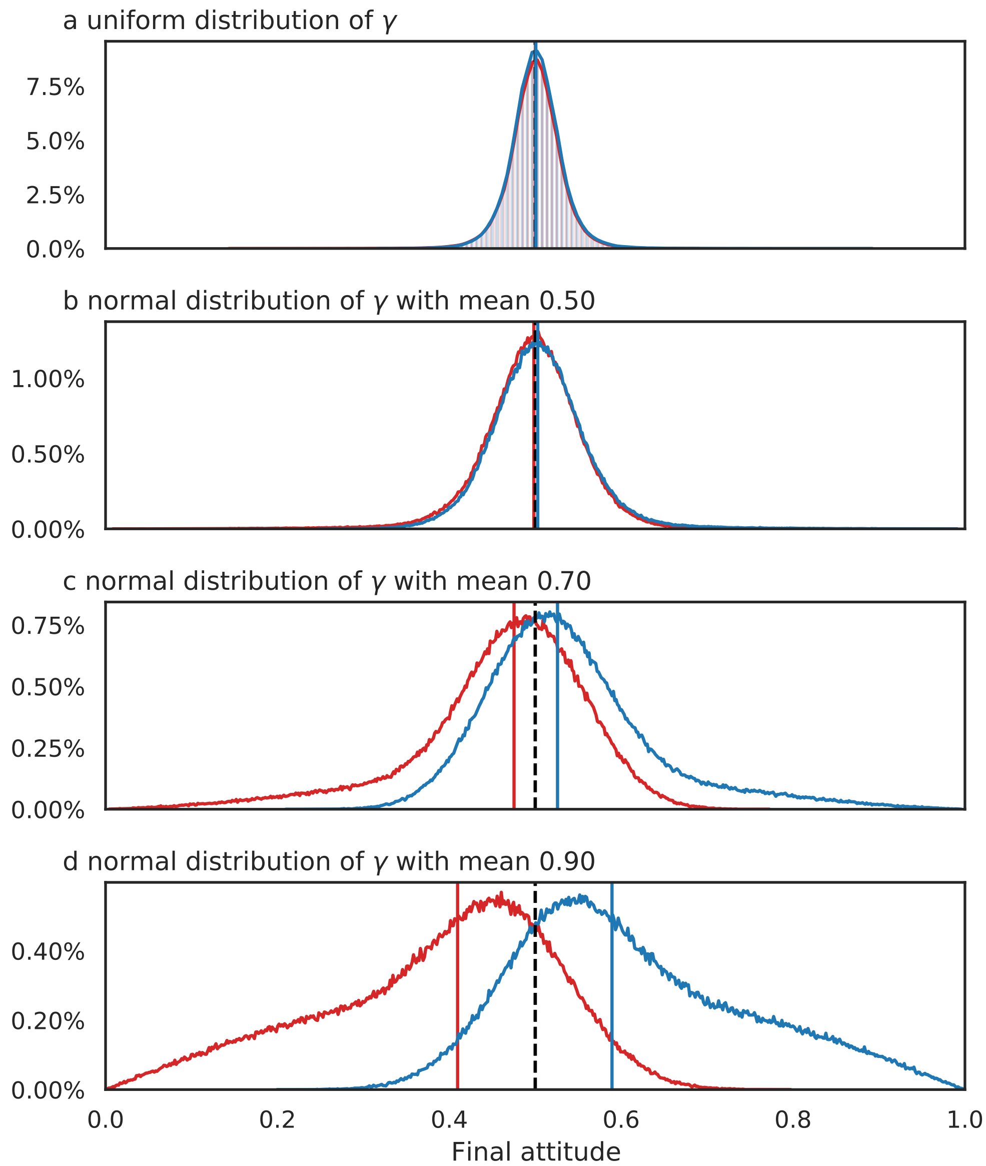


Supplementary Figure 10 Trade-off between opinion spread and alignment with inherent decision – standard deviation $\sigma = 0.05$. Decision alignment and opinion spread increase for higher mean self-reliance. Markers show mean values for colour-coded mean self-reliance.

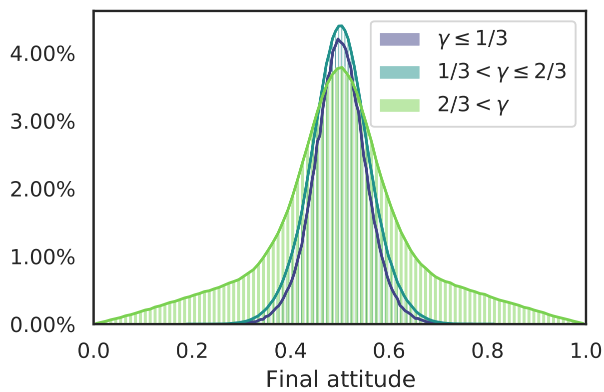
Standard deviation $\sigma = 0.15$



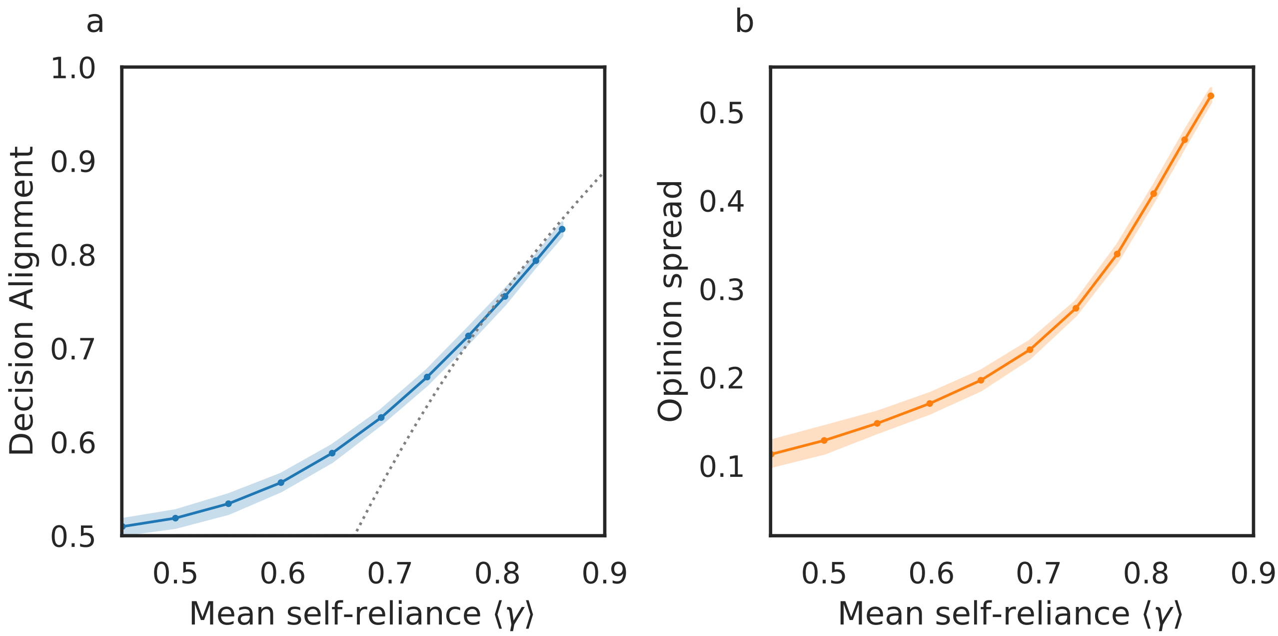
Supplementary Figure 11 Emergence of opinion clusters changes with distribution of γ – standard deviation $\sigma = 0.15$. Panels shows the attitude after evolving the model for 1000 time steps **a** for normal distributed γ_i with mean 0.5, **b** for normal distributed γ_i with mean 0.7 and **c** for normal distributed γ_i with mean 0.9



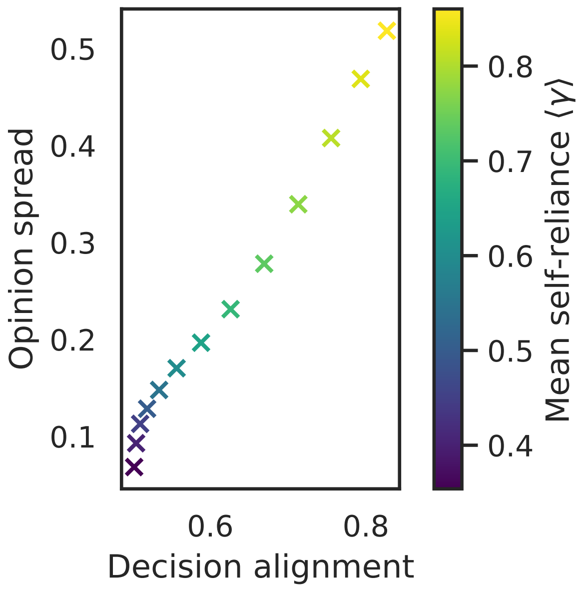
Supplementary Figure 12 Opinion spread increases in more self-reliant societies – standard deviation $\sigma = 0.15$. Panels show histograms of final attitude: **a** for uniform parameterised population, **b** for normal distributed γ with mean 0.5, **c** for normal distributed γ with mean 0.7 and **d** for normal distributed γ with mean 0.9. Red shows initial conviction < 0.5, blue shows initial attitude ≥ 0.5. Solid lines show medians, dashed black line 0.5.



Supplementary Figure 13 Highly self-reliant individuals have more polarised opinions – standard deviation $\sigma = 0.15$. Histograms of final attitude for bins of individual self-reliance γ based on ensemble of normal distributions of γ with means between 0.5 and 0.9

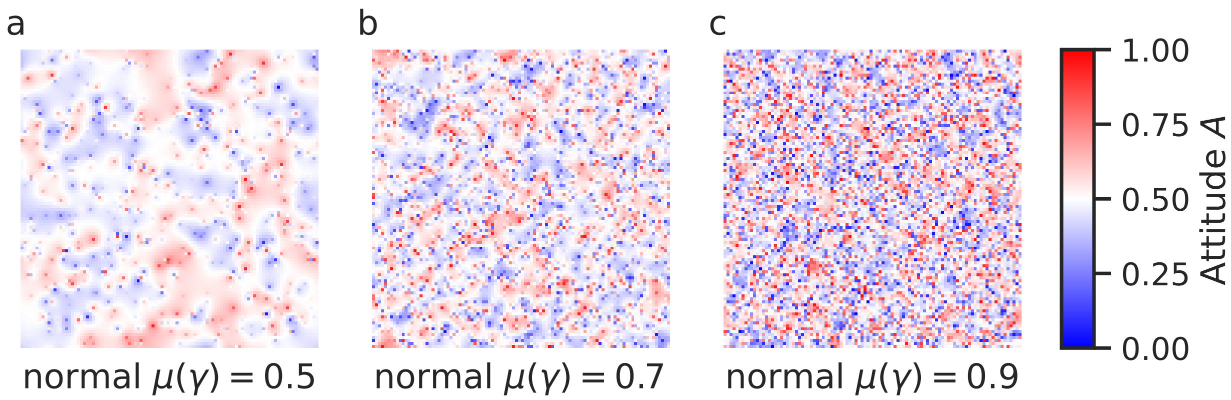


Supplementary Figure 14 Trade-off between decision alignment and social cohesion – standard deviation $\sigma = 0.15$. a shows the decision alignment in dependence of the average number of self-reliant agents. The decision alignment is computed as the average over the differences between the agents' initial decision and their final decision after evolving the model. The grey dots show an analytical approximation based on a mean-field approximation (Assuming 25% of agents are in equilibrium with their neighbours, i. e. $\mathbb{P}(N_i = 0) = 0.25$). **b** shows the opinion spread in dependence of the average number of self-reliant agents measured as the difference between the 90th and 10th percentiles of the distribution of the final attitude. (The initial opinion spread is about 0.8 due to the uniform distribution of the initial attitude.) If the opinion spread is large societal opinions are drifting apart and social cohesion lowers. Hence, there is a trade-off between higher personal decision alignment with more self-reliant agents and more social cohesion with less self-reliant agents. The average number of self-reliant agents corresponds to the mean of a normal distribution with mean γ and standard deviation σ = 0.1 from which the self-reliance was sampled. Confidence bands show the [5,95] Confidence interval based on 100 simulations with varying initial conditions.

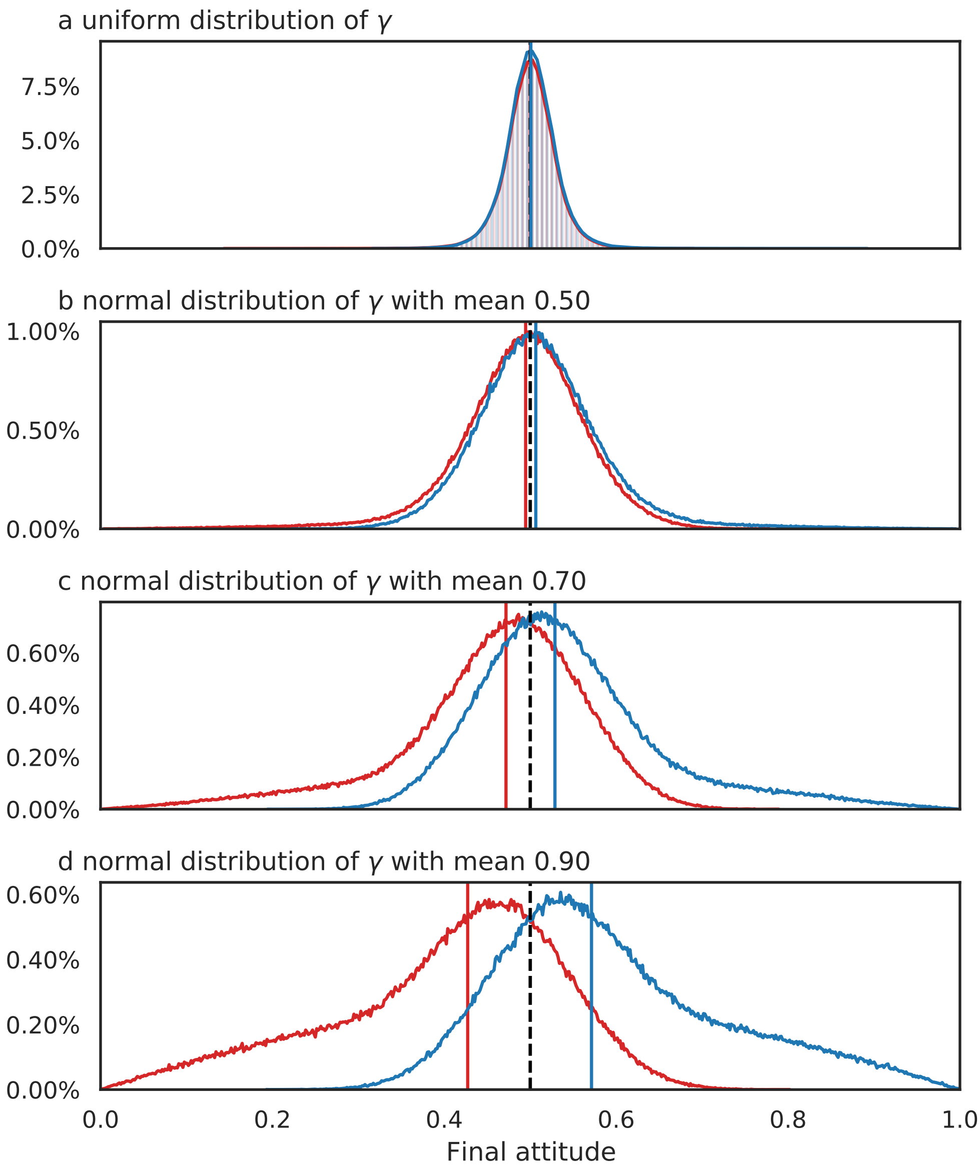


Supplementary Figure 15 Trade-off between opinion spread and alignment with inherent decision – standard deviation $\sigma = 0.15$. Decision alignment and opinion spread increase for higher mean self-reliance. Markers show mean values for colour-coded mean self-reliance.

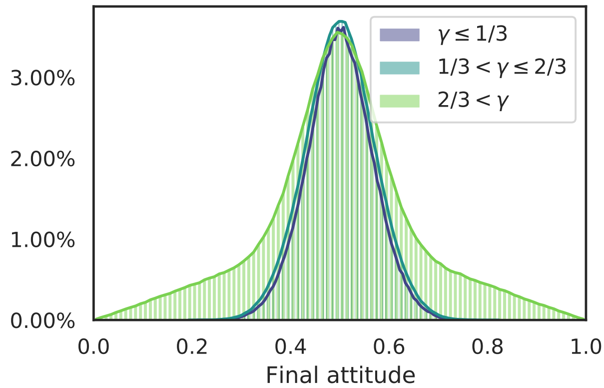
Standard deviation $\sigma = 0.20$



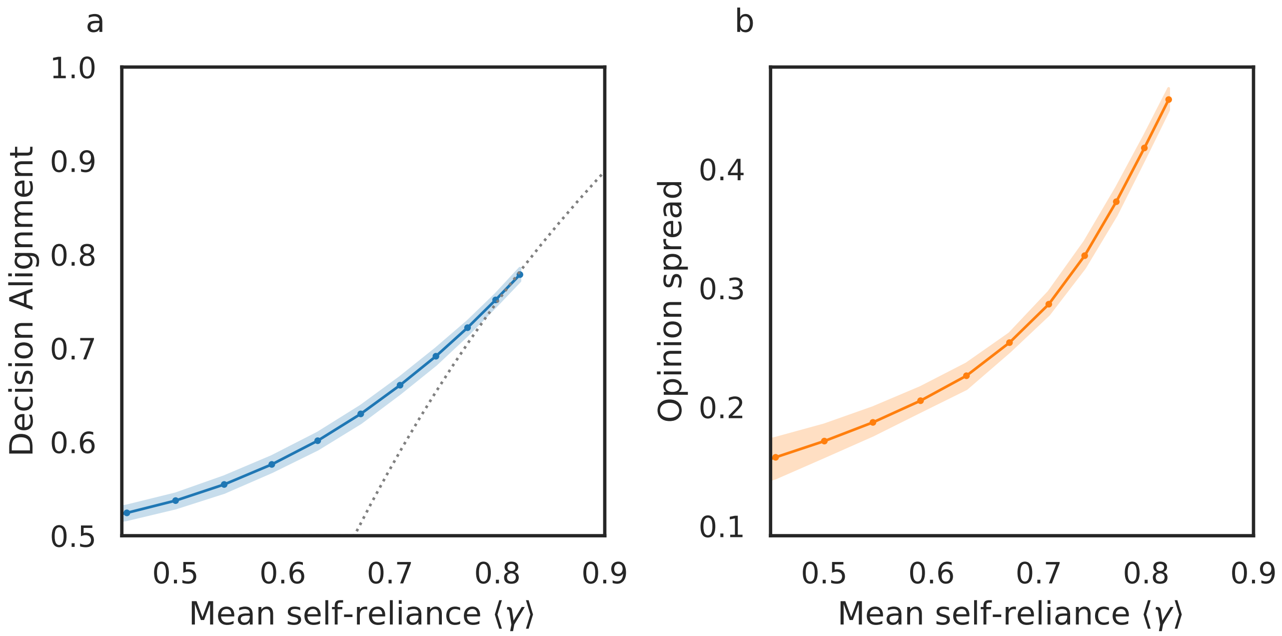
Supplementary Figure 16 Emergence of opinion clusters changes with distribution of γ – standard deviation $\sigma = 0.20$. Panels shows the attitude after evolving the model for 1000 time steps **a** for normal distributed γ_i with mean 0.5, **b** for normal distributed γ_i with mean 0.7 and **c** for normal distributed γ_i with mean 0.9



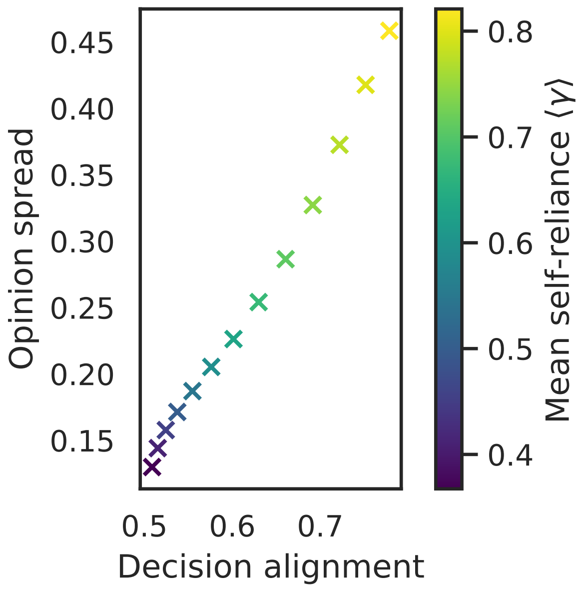
Supplementary Figure 17 Opinion spread increases in more self-reliant societies – standard deviation $\sigma = 0.20$. Panels show histograms of final attitude: **a** for uniform parameterised population, **b** for normal distributed γ with mean 0.5, **c** for normal distributed γ with mean 0.7 and **d** for normal distributed γ with mean 0.9. Red shows initial conviction < 0.5, blue shows initial attitude ≥ 0.5. Solid lines show medians, dashed black line 0.5.



Supplementary Figure 18 Highly self-reliant individuals have more polarised opinions – standard deviation $\sigma = $ 0.20. Histograms of final attitude for bins of individual self-reliance γ based on ensemble of normal distributions of γ with means between 0.5 and 0.9



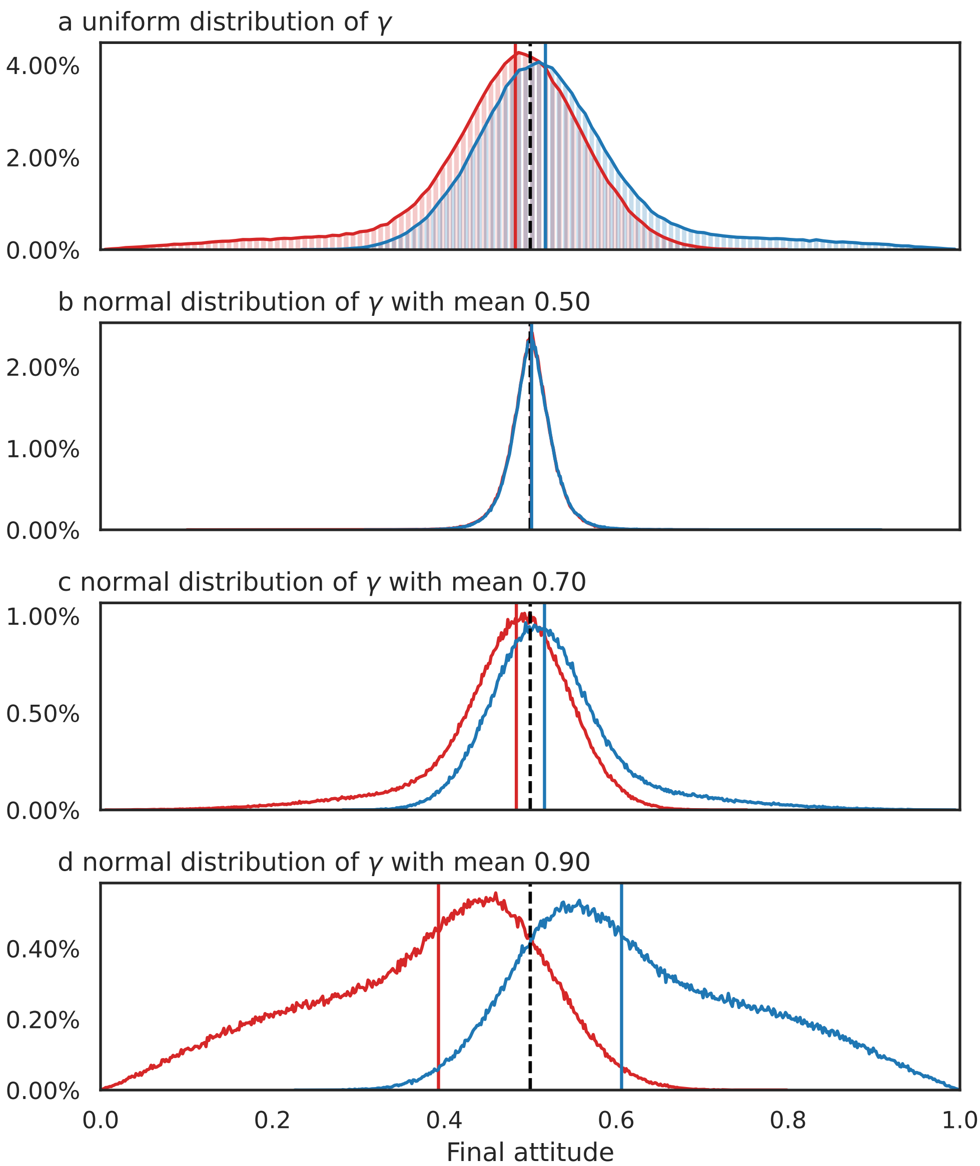
Supplementary Figure 19 Trade-off between decision alignment and social cohesion – standard deviation $\sigma = 0.20$. a shows the decision alignment in dependence of the average number of self-reliant agents. The decision alignment is computed as the average over the differences between the agents' initial decision and their final decision after evolving the model. The grey dots show an analytical approximation based on a mean-field approximation (Assuming 25% of agents are in equilibrium with their neighbours, i. e. $\mathbb{P}(N_i = 0) = 0.25$). **b** shows the opinion spread in dependence of the average number of self-reliant agents measured as the difference between the 90th and 10th percentiles of the distribution of the final attitude. (The initial opinion spread is about 0.8 due to the uniform distribution of the initial attitude.) If the opinion spread is large societal opinions are drifting apart and social cohesion lowers. Hence, there is a trade-off between higher personal decision alignment with more self-reliant agents and more social cohesion with less self-reliant agents. The average number of self-reliant agents corresponds to the mean of a normal distribution with mean γ and standard deviation σ = 0.1 from which the self-reliance was sampled. Confidence bands show the [5,95] Confidence interval based on 100 simulations with varying initial conditions.



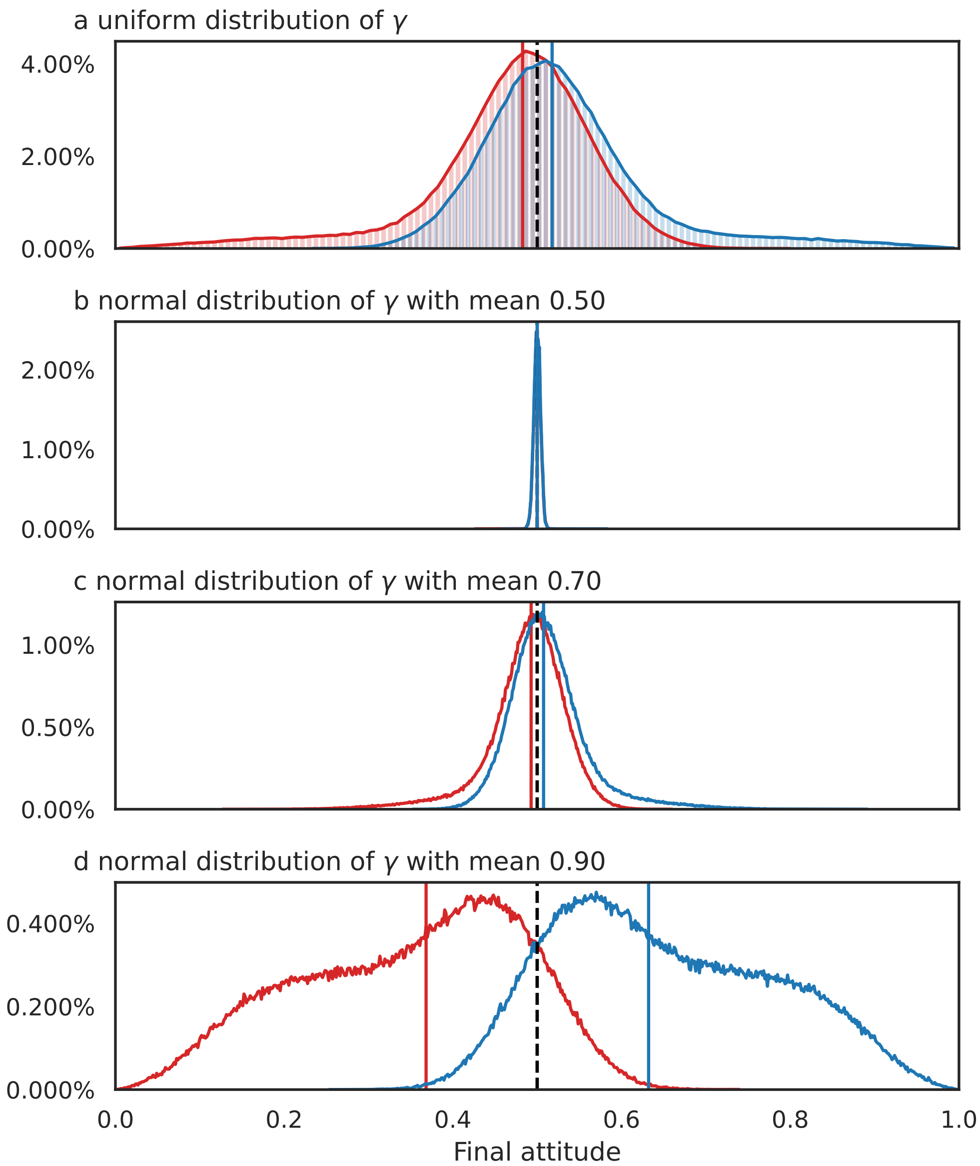
Supplementary Figure 20 Trade-off between opinion spread and alignment with inherent decision – standard deviation $\sigma = 0.20$. Decision alignment and opinion spread increase for higher mean self-reliance. Markers show mean values for colour-coded mean self-reliance.

Assuming independence of initial conviction and initial attitude

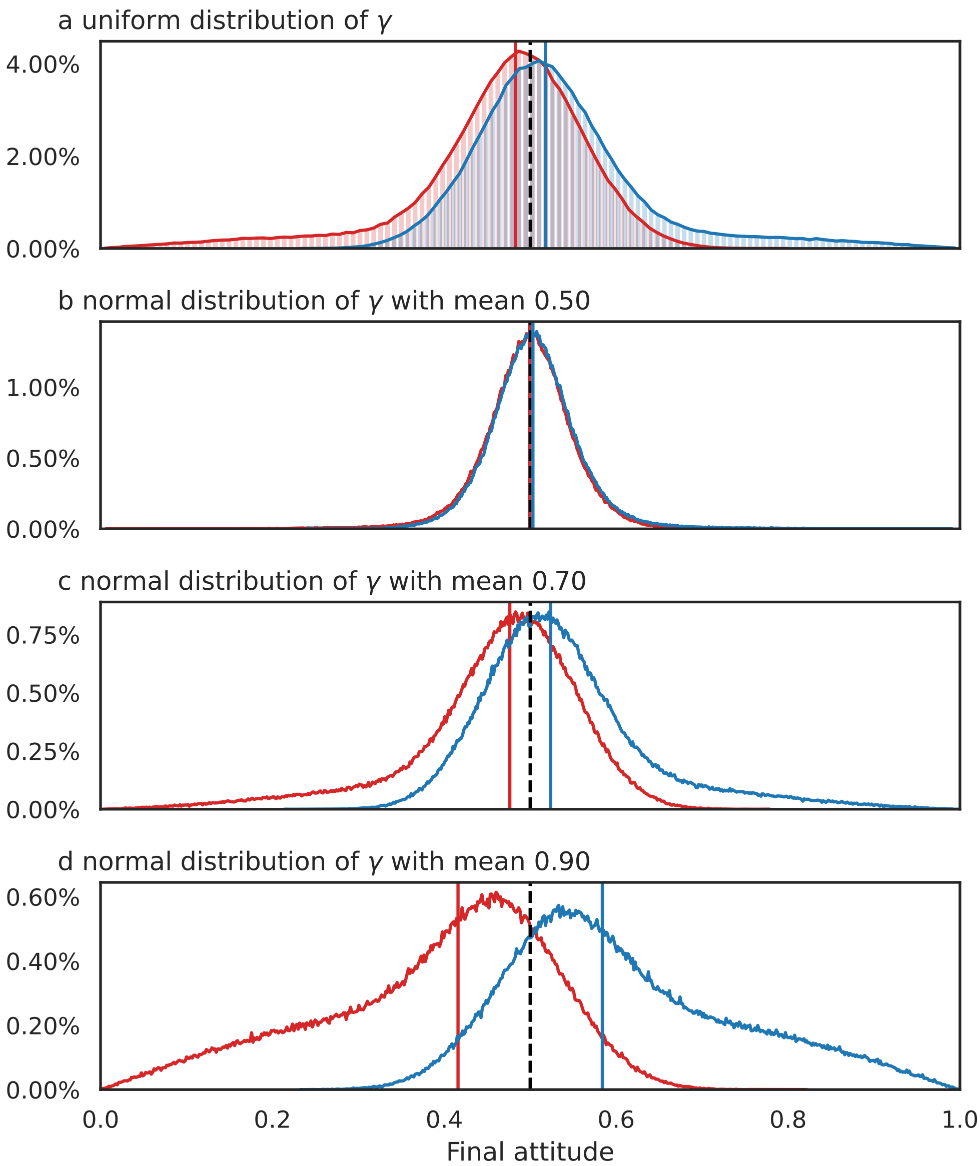
Relaxing the assumption that initial conviction equals initial attitude, we draw both parameters independently and find similar distributions of final attitude as shown in Supplementary Figs. 21–24.



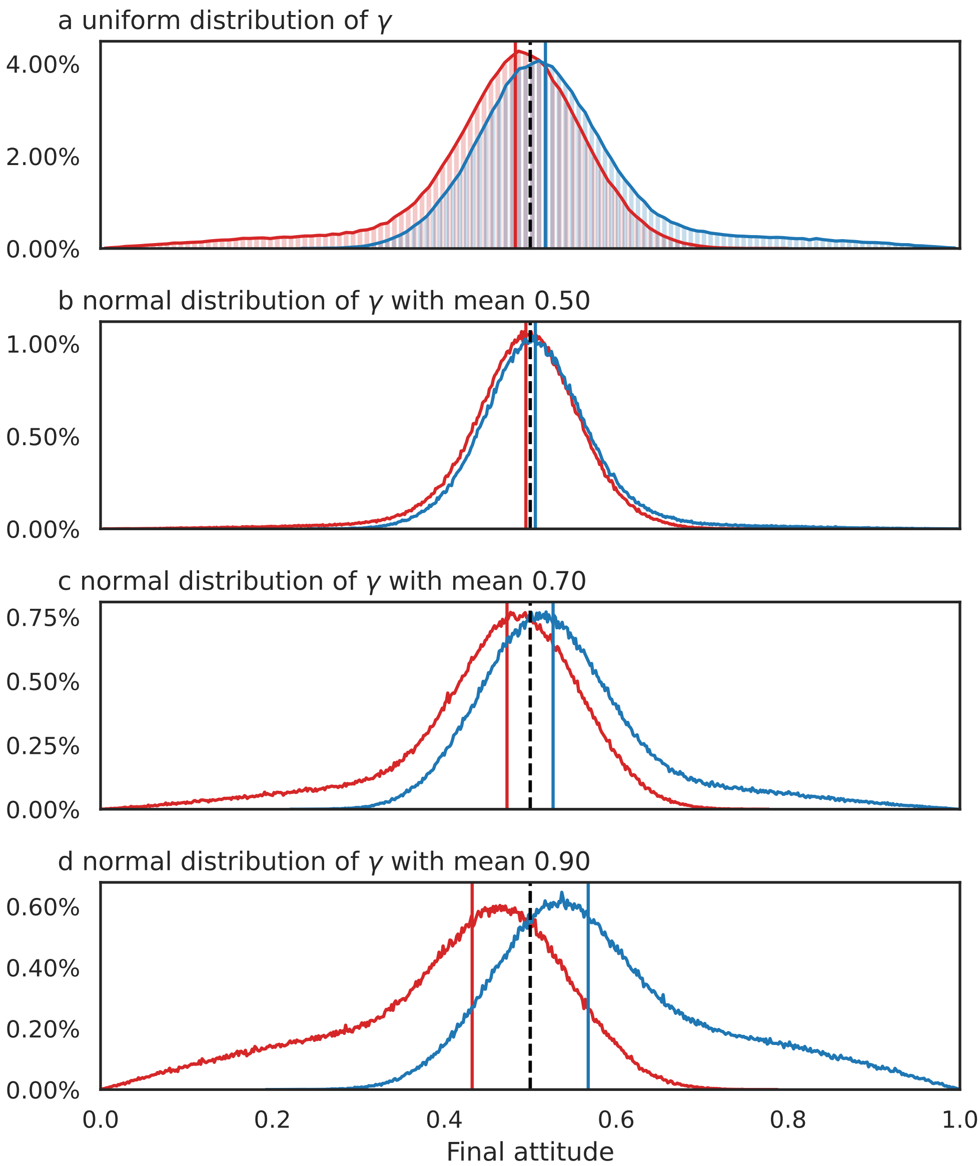
Supplementary Figure 21 Opinion spread increases in more self-reliant societies – standard deviation $\sigma = 0.10$. Panels show histograms of final attitude: **a** for uniform parameterised population, **b** for normal distributed γ with mean 0.5, **c** for normal distributed γ with mean 0.7 and **d** for normal distributed γ with mean 0.9. Red shows initial conviction < 0.5, blue shows initial attitude ≥ 0.5. Solid lines show medians, dashed black line 0.5.



Supplementary Figure 22 Opinion spread increases in more self-reliant societies – standard deviation $\sigma = 0.05$. Panels show histograms of final attitude: **a** for uniform parameterised population, **b** for normal distributed γ with mean 0.5, **c** for normal distributed γ with mean 0.7 and **d** for normal distributed γ with mean 0.9. Red shows initial conviction < 0.5, blue shows initial attitude ≥ 0.5. Solid lines show medians, dashed black line 0.5.



Supplementary Figure 23 Opinion spread increases in more self-reliant societies – standard deviation $\sigma = 0.15$. Panels show histograms of final attitude: **a** for uniform parameterised population, **b** for normal distributed γ with mean 0.5, **c** for normal distributed γ with mean 0.7 and **d** for normal distributed γ with mean 0.9. Red shows initial conviction < 0.5, blue shows initial attitude ≥ 0.5. Solid lines show medians, dashed black line 0.5.



Supplementary Figure 24 Opinion spread increases in more self-reliant societies – standard deviation $\sigma = 0.20$. Panels show histograms of final attitude: **a** for uniform parameterised population, **b** for normal distributed γ with mean 0.5, **c** for normal distributed γ with mean 0.7 and **d** for normal distributed γ with mean 0.9. Red shows initial conviction < 0.5, blue shows initial attitude ≥ 0.5. Solid lines show medians, dashed black line 0.5.

Example: Analytical considerations for a two-agent model

For the most general two-agents model with $\gamma_1 \neq \gamma_2$ we analyse conditions for the existence of a disagreeing equilibrium. Assuming the time-scaling factor τ to be 1, the dynamic is described by the map

$$\mathbf{f}(x,y) = \begin{pmatrix} \gamma_1 x + (1-\gamma_1)y + \gamma_1(\hat{x}-x)|y-x| \\ \gamma_2 y + (1-\gamma_2)x + \gamma_2(\hat{y}-y)|y-x| \end{pmatrix} \tag{11}$$

and has the fixed points

$$x = y \qquad \text{and} \qquad x = \hat{x} + \sigma\frac{1-\gamma_1}{\gamma_1}, \quad y = \hat{y} - \sigma\frac{1-\gamma_2}{\gamma_2}\,. \tag{12}$$

For the sign of the attitude difference $\sigma \in \{-1, 1\}$ we can assume w.l.o.g. $\sigma = 1$ due to symmetry. Thus $2^3 \cdot \frac{1}{2} = 4$ qualitatively different parameter configurations are possible.

Case 1	Case 2	Case 3	Case 4
$\hat{y} > \hat{x}$, $\gamma_2 > \gamma_1$	$\hat{y} > \hat{x}$, $\gamma_2 < \gamma_1$	$\hat{y} < \hat{x}$, $\gamma_2 > \gamma_1$	$\hat{y} < \hat{x}$, $\gamma_2 < \gamma_1$

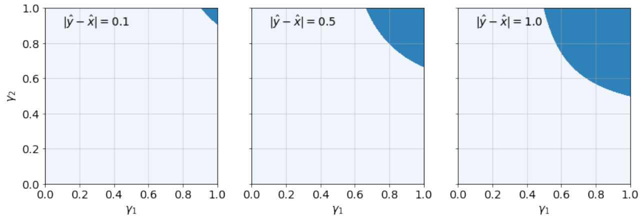
The second fixed point exists (and is stable as shown in the following) if

$$y > x \quad \leftrightarrow \quad \hat{y} - \frac{1-\gamma_2}{\gamma_2} \geq \hat{x} + \frac{1-\gamma_1}{\gamma_1} \quad \rightarrow \quad 2 + \hat{y} - \hat{x} \geq \frac{\gamma_1+\gamma_2}{\gamma_1\gamma_2}\,. \tag{13}$$

The area of the parameter space spanned by γ_1 and γ_2 where this condition is fulfilled is shown for different values of the initial convictions difference in Supplementary Fig. 25. For $\hat{x} > \hat{y}$ this is never the case so that case 3 and 4 of the configurations only allow the stable equilibrium $x = y$.

The condition for the initial parameters that allow a change of σ ($\hat{y} > \hat{x}$) becomes:

$$(\gamma_2 + \gamma_1 - 1)\,\sigma + \gamma_2\hat{y} - \gamma_1\hat{x} > \gamma_2 y - \gamma_1 x\,. \tag{14}$$



Supplementary Figure 25 Parameter space for γ of the two-agents model. Points inside the dark area meet condition eq. (13) which is required for the existence of a disagreeing equilibrium. Each panel shows the parameter space for a different gap of initial conviction in the system.

The Jacobian matrix is given by eq. (15).

$$\mathbf{Df}(x,y) = \begin{pmatrix} \gamma_1(1-\hat{x}-y+2x) & 1-\gamma_1(1-\hat{x}+x) \\ 1-\gamma_2(1+\hat{y}-y) & \gamma_2(1+\hat{y}-2y+x) \end{pmatrix} \tag{15}$$

Evaluated at the fixed point defined by $x = y \equiv \eta$ this becomes:

$$\mathbf{Df}(\eta,\eta) = \begin{pmatrix} q_1 & 1-q_1 \\ 1-q_2 & q_2 \end{pmatrix} \quad , \quad q_1 \equiv \gamma_1(1-\hat{x}+\eta) \quad q_2 \equiv \gamma_2(1+\hat{y}-\eta) \ . \tag{16}$$

The Eigenvalues are given by

$$0 = (q_1-\lambda)(q_2-\lambda) - (1-q_1)(1-q_2) \tag{17}$$
$$[\hat{q} \equiv q_1 + q_2] \tag{18}$$
$$= \lambda^2 - \hat{q}\lambda + \hat{q} - 1 \quad \rightarrow \quad \lambda_+ = \hat{q} - 1 \quad \lambda_- = 1 \ . \tag{19}$$

Since every point along the line $x = y$ is a fixed point of the dynamics λ_- is equal to 1.

$$\mathbf{0} = \begin{pmatrix} (q_1-\hat{q}+1)v_+^1 + (1-q_1)v_+^2 \\ (1-q_2)v_+^1 + (q_2-\hat{q}+1)v_+^2 \end{pmatrix} \quad \rightarrow \quad \mathbf{v}_+ = \begin{pmatrix} \frac{q_1-1}{1-q_2}\nu \\ \nu \end{pmatrix} \tag{20}$$

The fixed points are attracting along $\mathbf{v}_+$ if $|\lambda_+| < 1$. If $\gamma_1 > 0 \vee \gamma_2 > 0$ it holds

$$\begin{aligned} & |\gamma_1(1+\eta-\hat{x}) + \gamma_2(1+\hat{y}-\eta) - 1| < 1 \\ \leftrightarrow \quad & 0 < \gamma_1(1+\eta-\hat{x}) + \gamma_2(1+\hat{y}-\eta) < 2. \end{aligned} \tag{21}$$

In general, it is not trivial to determine if the condition is satisfied as it depends strongly on the specific values of the parameters and η. The left side of the inequality is always true for the parameters lying inside the open unit interval since $\min(1-a-b) = \epsilon > 0$ for $a, b \in [0, 1]$.

$$\gamma_1(1+\eta-\hat{x}) + \gamma_2(1+\hat{y}-\eta) > \gamma_1\epsilon_1 + \gamma_2\epsilon_2 > 0 \tag{22}$$

For $\hat{x} > \hat{y}$ the right side is also met because

$$\gamma_1(1+\eta-\hat{x}) + \gamma_2(1+\hat{y}-\eta) \leq 2+\eta-\eta+\hat{y}-\hat{x} < 2\,. \tag{23}$$

It follows that the fixed point is semi-stable. For $\hat{y} > \hat{x}$ it is possible to verify numerically using the appropriate long-term limit for η that the inequality is satisfied if and only if eq. (13) is not satisfied. In fact, it can be shown that the second fixed point given in eq. (12) is stable if it exists. Therefore the Jacobian matrix evaluates to

$$\mathbf{Df}(\eta_1, \eta_2) = \begin{pmatrix} q_1 & q_2 \\ q_3 & q_4 \end{pmatrix} \tag{24}$$

$$\begin{aligned} q_1 &\equiv \gamma_1(1-\hat{x}-y+2x) = \gamma_1(1-\hat{y}+\hat{x}) + 2\sigma(1-\gamma_1) + \sigma\gamma_1\frac{1-\gamma_2}{\gamma_2} \\ q_2 &\equiv 1-\gamma_1(1-\hat{x}+x) = 1-\gamma_1-\sigma(1-\gamma_1) \\ q_3 &\equiv 1-\gamma_2(1+\hat{y}-y) = 1-\gamma_2-\sigma(1-\gamma_2) \\ q_4 &\equiv \gamma(1+\hat{y}-2y+x) = \gamma_2(1-\hat{y}+\hat{x}) + 2\sigma(1-\gamma) + \sigma\gamma_2\frac{1-\gamma_1}{\gamma_1} \end{aligned} \tag{25}$$

Using $\sigma = 1$ it follows that $q_2 = 0$ and $q_3 = 0$. The resulting matrix is diagonal with Eigenvalues q_1, q_4. Stability is guaranteed if $|\lambda| < 1$.

$$\begin{aligned}
&1 > |\lambda| = |q_1| \\
&\Leftrightarrow -1 > \gamma_1(1 - \hat{y} + \hat{x}) - 2\gamma_1 + \gamma_1 \frac{1-\gamma_2}{\gamma_2} \\
&\Leftrightarrow 1 + \hat{y} - \hat{x} > \frac{\gamma_2}{\gamma_1\gamma_2} + \gamma_1 \frac{1-\gamma_2}{\gamma_1\gamma_2} \\
&\Leftrightarrow 1 + \hat{y} - \hat{x} > \frac{\gamma_1 + \gamma_2 - \gamma_1\gamma_2}{\gamma_1\gamma_2} \\
&\Leftrightarrow 2 + \hat{y} - \hat{x} > \frac{\gamma_1 + \gamma_2}{\gamma_1\gamma_2} \qquad \text{(for } q_2 \text{ analogously)}
\end{aligned} \tag{26}$$

The resulting equation matches the condition for the existence of the fixed point, sucht that if the fixed point exists it is stable. Thus the disagreeing equilibrium is stable if it can be reached, which depends on the parameters of the model.