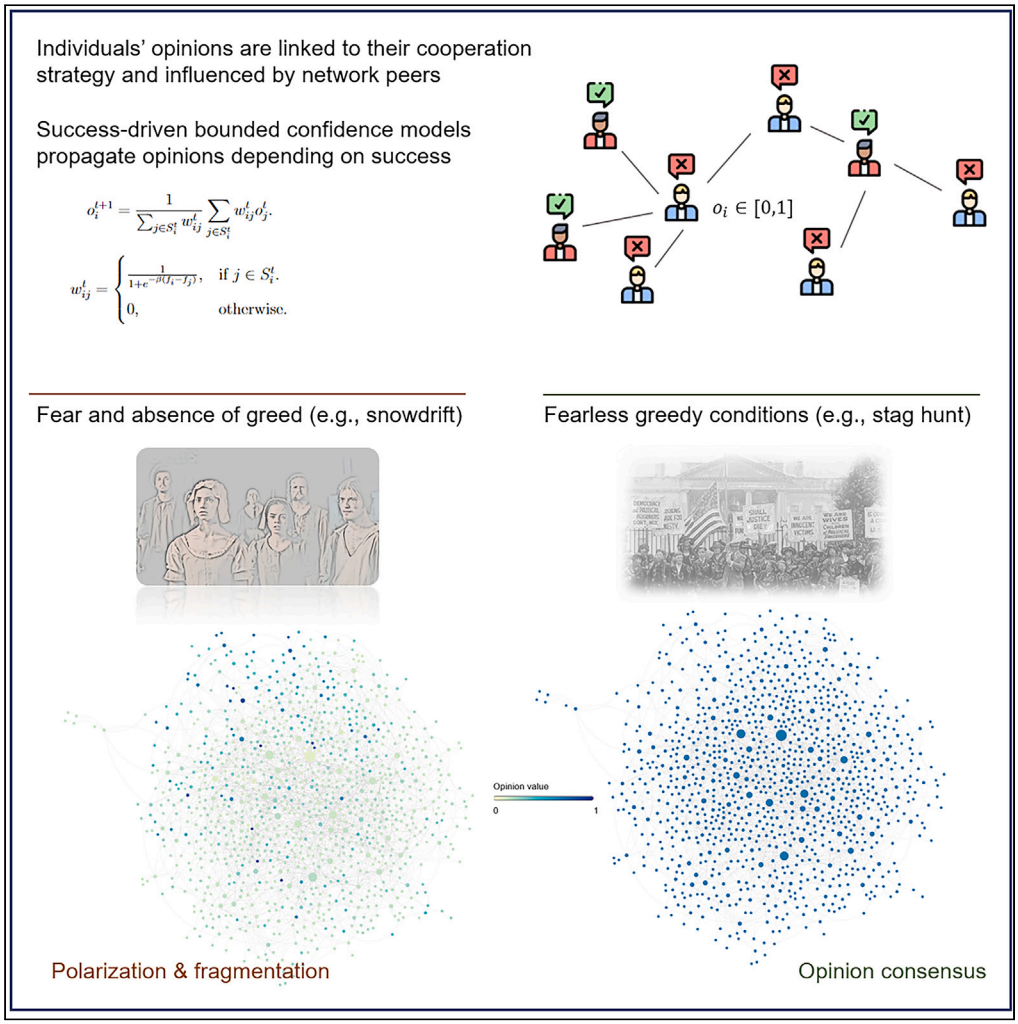
iScience

CellPress
OPEN ACCESS

Article

Success-driven opinion formation determines social tensions



Manuel Chica, Matjaž Perc, Francisco C. Santos

manuelchica@ugr.es

Highlights

Opinion formation is coupled with strategy selection in social dilemmas such as the PD

Under fear of punishment, opinion consensus is reached in the population

Fragmentation of opinions is high when individuals are greedy and defection is viable

Chica et al., iScience *27*, 109254
March 15, 2024
https://doi.org/10.1016/j.isci.2024.109254

iScience

CellPress
OPEN ACCESS

Article

Success-driven opinion formation determines social tensions

Manuel Chica,[1,2,8,*] Matjaž Perc,[4,5,6,7] and Francisco C. Santos[3]

SUMMARY

Polarization is common in politics and public opinion. It is believed to be shaped by media as well as ideologies, and often incited by misinformation. However, little is known about the microscopic dynamics behind polarization and the resulting social tensions. By coupling opinion formation with the strategy selection in different social dilemmas, we reveal how success at an individual level transforms to global consensus or lack thereof. When defection carries with it the fear of punishment in the absence of greed, as in the stag-hunt game, opinion fragmentation is the smallest. Conversely, if defection promises a higher payoff and also evokes greed, like in the prisoner's dilemma and snowdrift game, consensus is more difficult to attain. Our research thus challenges the top-down narrative of social tensions, showing they might originate from fundamental principles at individual level, like the desire to prevail in pairwise evolutionary comparisons.

INTRODUCTION

From partisanship and populism within and among nations[1] to social media exposure and the emergence of echo chambers,[2–5] a better understanding of how polarization and individual opinions are affected by others is of significant current interest. While mass media played a key role in the past, various online social networks and communication channels are taking over as the main drivers of these processes.[6–8] Notably, the problem of polarization is not restricted solely to the political arena or the lack of moderate views, but may also entail lifestyle preferences, consumer choices, and even morality.[9] Global challenges, such as climate inaction[10,11] or the COVID-19 pandemic,[12,13] also often generate massive cascades of polarized opinions.[9] What is more, the dynamics of information exchange on social media often encourages users to act rather aggressively when launching or protecting their opinions.[14] At the same time, and especially in more autocratically governed countries, the public expression of opinions is often restricted or tailored to meet a certain narrative,[15] which further facilitates an environment for social tensions.

A better understanding of how social tensions are related to opinion polarization can shed new light on our social interactions and how these translate back to moderate or escalate contentious dilemmas. Evolutionary game theory provides a comprehensive mathematical framework to study such social interactions, especially in the context of social dilemmas,[16,17] where individuals can choose whether or not to cooperate with their peers.[18] Prisoner's dilemma, snowdrift, and stag-hunt games are commonly used pairwise games to study the effects of individual interactions and cooperation. These games enable us to study how collective cooperation may survive in a world where individual selfish actions produce better short-term outcomes.[17,19–21] The set of applications of evolutionary games is immense, from collective disasters[22,23] to herding behavior,[24] sharing economy,[25] or tax fraud,[26] among others.

Social dilemmas also characterize social tensions in the population depending on the preferences of the players to unilaterally or mutually defect or cooperate.[21,27] These social tensions are greed and fear, which are two distinct motives that underlie non-cooperative behavior.[28] Greed corresponds to situations in which players prefer unilateral defection to mutual cooperation, while fear corresponds to situations in which players prefer mutual defection to unilateral cooperation.[27] Moreover, greed promises gains for exploiting cooperative peers, whereas fear warns against the cost of cooperation with exploitative peers.

Complex computational simulations and the field of "sociophysics",[29,30] which combines tools and methods from statistical physics to investigate social phenomena, have already provided policy-relevant insights into mechanisms for preventing extreme polarization such as in political contexts,[31] finding that, in some circumstances, repulsion from those of whom we are intolerant can reinforce a moderate majority. Social feedback was also injected into opinion dynamics (OD) models[32,33] for individuals to express their opinion about an issue, being

[1]Andalusian Research Institute DaSCI "Data Science and Computational Intelligence", University of Granada, 18071 Granada, Spain
[2]School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia
[3]INESC-ID & Instituto Superior Técnico, Universidade de Lisboa, 2744-016 Porto Salvo, Portugal
[4]Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia
[5]Community Healthcare Center Dr. Adolf Drolc Maribor, Vošnjakova ulica 2, 2000 Maribor, Slovenia
[6]Complexity Science Hub Vienna, Josefstädterstraße 39, Vienna 1080, Austria
[7]Department of Physics, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul, Republic of Korea
[8]Lead contact
*Correspondence: manuelchica@ugr.es
https://doi.org/10.1016/j.isci.2024.109254

more sensitive to approval and disapproval by their peers.[34] In general, OD consists of understanding the conditions under which consensus or diversity is reached from an initial population of individuals (agents) with different opinions.[35]

Some OD models were voter model extensions where agents can have three states ("leftists", "centrists", and "rightists") and extreme agents cannot interact because of their incompatibility in their opinions[36] or agents interact to time-fluctuating external influences such as new sources and media polarization.[37,38] An OD model in connection to the emergence of polarization to show three particular psychological traits in a population: floaters, contrarians, and stubborn agents.[39] Particularly, Galam found that floaters produce segregated polarization (zero entropy), contrarians produce a fluid polarization (high entropy), and stubbornness produces a frozen polarization (low entropy). Li et al.[40] explored opinion propagation with strategic interactions on social networks by using the game theoretical approach but without having the goal of studying social game dilemmas. Ding et al.[41] treated opinions discussions as a game theoretical approach, where heterogeneous agents adjust their behaviors to the environment during discussions, and their interacting strategies evolve together with opinions. Also, Li et al. proposed a network model where agents hold one of two different opinions and form dynamical links in the network.[42]

Other authors proposed an OD model applied to voting records of the US House of Representatives over a time span of decades where nodes' opinions jointly evolve with the network connections.[3] An agent-based model to represent the generation of eco chambers (i.e., situations in which one's opinion resonates with those of ones' social contacts) is presented in Wang et al. study.[4] The model, which is calibrated with Twitter political data, showed that polarized and segregated network structures are a function of ideological differences between the political campaigns and the open-mindedness of the agents. They also observed how confirmation and selection bias, generators of echo chambers, influence the co-evolution of political opinions and network structure.

In a recent and relevant study, Kawakatsu et al.[43] extended a cultural evolution model based on evolutionary game theory (concretely, the pairwise donation game with four strategies). Individuals have opinions on multi-dimensional issues and accumulate benefits through pairwise interactions and learn the successful strategies, showing that the diversity on political issues can promote both individual cooperation and societal cohesion but, extreme partisanship can introduce tension between the individual cooperation and social cohesion (i.e., inter-individual cooperation thrives at the cost of increased polarization). Following the social tensions concept between individuals and society depicted in the study of Kawakatsu et al.,[43] we primarily focus on the injection of success information into an OD model and the study of consensus for different social tensions represented by evolutionary game conditions. According to the state-of-the-art study, this phenomenon was not studied up to now.

Thus, our goal in this study is to computationally analyze the microscopic dynamics that is behind polarization and fragmentation in social networks, and to determine how these two processes affect the emergence of social tensions. In doing so, we aim to determine whether social tensions, such as greed and fear, are correlated with the final polarization and fragmentation states in a social network. But to be able to study the evolution of individual opinions and their success with respect to the social tensions in a population, novel computational methods are needed. To that effect, we first combine opinion formation with the strategy selection in social dilemmas in structured populations (specifically, we employ a heterogeneous social network with a power-law distribution, generated by the Barabasi-Albert algorithm[44]). The continuous opinion of a player is linked with her disposition to either cooperate or defect with other players in the game. In that way, and unlike previous research along similar lines,[40,45,46] we link opinions and evolutionary success and are thus able to study how social tensions condition the polarization and fragmentation of opinions.

First, we define, for each player or individual, her opinion o_i as a continuous value to either cooperate or not at a given time-step. In a pairwise interaction of two individuals i and j, their opinions to cooperate (i.e., o_i and o_j) will be used to determine the expected payoffs f_{ij} for each of them. Our proposal to couple the opinions of the individuals with their payoffs during game interactions is done by a success-driven OD model, derived from classical OD models with bounded confidence levels[33] such as the Hegselmann and Krause (HK)[47] and Deffuant–Weisbuch (DW)[48] models. Continuous OD model of bounded confidence were already biased to modify the selection rule of the discussion partners and thus mimicking the behavior of online media which suggest interaction with similar peers.[7] Also, Huang et al.[46] used the HK model in an evolutionary framework but with a different approach with respect to the one here, as cooperation was understood as a way of interchanging opinions and, when agents differ in their opinions, they suffer from a cost.

We use the HK model in all the experiments of the paper since we obtain similar results with the DW model, as already shown in the literature.[49] Both HK and DW models are defined in the STAR Methods section. Specifically in the success-driven HK model, opinions of the individuals o_i evolve in the following manner: $o_i^{t+1} = \frac{1}{\sum_{j \in S_i^t} w_{ij}^t} \sum_{j \in S_i^t} w_{ij}^t o_j^t$. S_i^t is a set of individuals within a confidence level: $\{j, \left|o_i^t - o_j^t\right| \leq \epsilon\}$, $|.|$ being the absolute value of a real number. Then, neighboring individuals with opinions not differing more than an ϵ confidence threshold are included in S_i^t. The weight of each neighbor's opinion to compute the new one is calculated by including success fitness information f_i and f_j as $w_{ij}^t = \frac{1}{1+e^{-\beta(f_i - f_j)}}$, only for those j individuals of the S_i^t set. The intensity of selection β plays here the role of controlling how success in the game influences the new opinion of the individual. When β equals to 0, the model evolves as in the traditional HK model, without any fitness or success information.

In what follows, we will show how success-driven opinions evolve over time subject to their evolutionary success in different social dilemmas and thus subject to fear or greed or both. Particularly, in the prisoner's dilemma game players are driven by greed for the expected gain from exploiting cooperative partners, as well as by fear of punishment for cooperating when partners defect.[27] Conversely, and along the same lines of reasoning, in the snowdrift game only greed is present, while in the stag-hunt game only fear is present. In our computational experiments, we determine the number of different opinion clusters and their evolution in success-driven settings for different selection intensities and adoption confidence levels, ultimately revealing that individual greed is the major factor in preventing opinion consensus, much

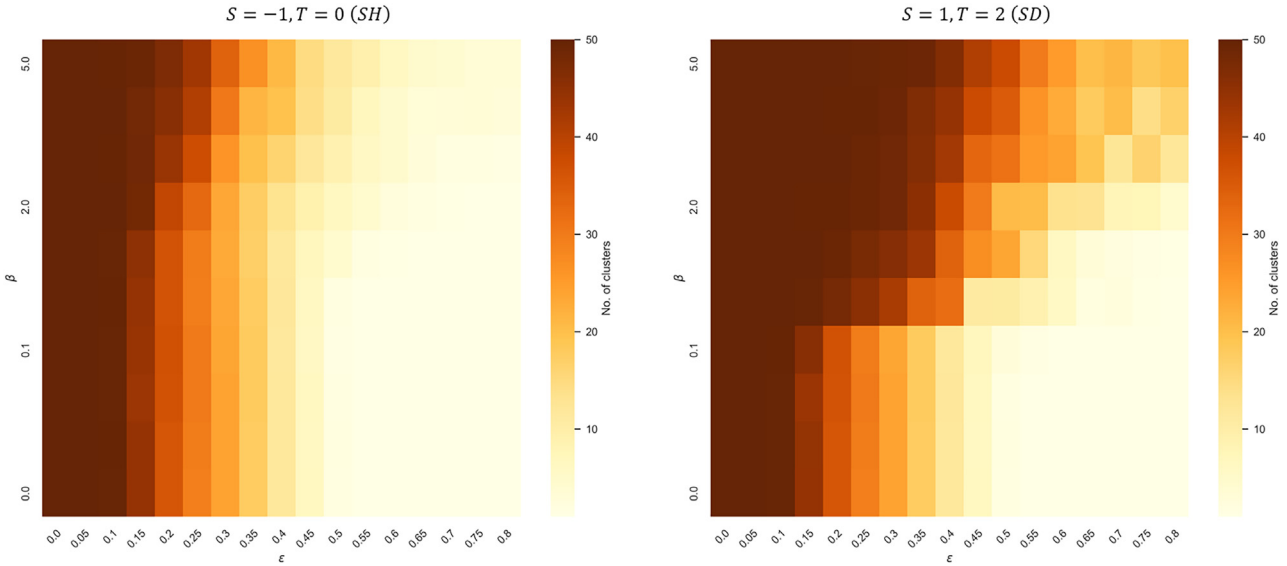


Figure 1. Number of clusters of opinions for two extreme social tensions (fear in SH and greed in SD)

Panel of sensitivity analysis on (ϵ, β) for two different games (SH and SD). A higher fragmentation is shown on the right heatmap (SD) when β and ϵ are sufficiently high to inject success-driven dynamics in the model.

more so than fear. This challenges the top-down narrative of polarization, showing that it may well have completely individual origins, as simple as wanting to attain the best fitness or payoff in an evolutionary standoff.

RESULTS

Impact of social tensions on fragmentation

The results of this study first show how, for different values of S and T that define the type of the game, the final opinions of the population change dramatically. Mainly, we will utilize the number of clusters of opinions at the end of the simulation to show the final state of the population's opinion. When consensus is not reached by the population, the opinion evolves into clusters which are opinions of agents separated by a given distance above an opinion threshold.[50] These clusters of opinions are processed by defining bins within the opinion's interval and later computing the frequency of opinions falling into each bin (note that the process is more precisely defined in the STAR Methods details).

Figure 1 shows a sensitivity analysis on intensity of selection β and confidence ϵ for two combinations of S, T values, $(S = -1, T = 0)$ and $(S = 1, T = 2)$, which define opposite tensions in the dilemmas. These tensions define well-known games such as prisoner's dilemma (PD), stag hunt (SH), and snowdrift (SD). First, and as it occurs in traditional OD models, higher values of ϵ, which defined the confidence level of each agent, tend to consensus; while lower values of ϵ make the population more fragmented in terms of opinions. When $\beta = 0$, the success-driven model converts into a traditional OD model and the payoffs of the game have no effect. We see in the heat-maps how the number of clusters are identical for both games when intensity $\beta \leq 0.1$.

More importantly, the heat-maps of the figures show the number of clusters of opinions and how the fragmentation is increased when increasing S and T values. Thus, the highest fragmentation is when an SD game is defined (i.e., by parameters $S = 1$ and $T = 2$). Changes are more relevant when β is increasing (x axis of the heat-maps). This change has an expected outcome as β is regulating the success-driven impact of the opinions.

To deeply show these variations in the final state of the population's opinions, panel of Figure 2 represents the increase in opinions' clusters (i.e., increase in fragmentation) when moving to PD $(S = -0.1, T = 1.1)$ from SH $(S = -1, T = 0)$ in the left plot; and to SD $(S = 1, T = 2)$ from PD. The increase is clear as we can notice increments between 200% and 350%. Mainly, this phenomenon takes place when the intensity of selection β is relevant (>0.1) and ϵ is greater to 0.2 (as lower values always incur in very fragmented opinions[32]). Within this wide are of relevant impact of the success in the game and high confidence level, the increase in the fragmentation is up to 350%.

Opinions' evolution under different tensions

Panel of Figure 3 shows the evolution of opinions for three games, SD, PD, and SH, given by three combinations of S, T. Additionally, Figure 4 shows the opinions' evolution on a social network at the beginning of the simulation (same opinions and structure for both configurations) and at $t = 500$ for SH and SD. Results for both figures were obtained by the same confidence level for opinions $(\epsilon = 0.7)$ and same intensity of selection $(\beta = 5)$. All the plots point out to the same direction as previous heat-maps. The social tensions represented by the different games change the landscape of the population opinions. The highest fragmentation is obtained when having an SD game (i.e., unilateral defection is preferred to mutual cooperation and a greedy sentiment is in a fearless population). The lowest fragmentation and therefore, a facilitation of

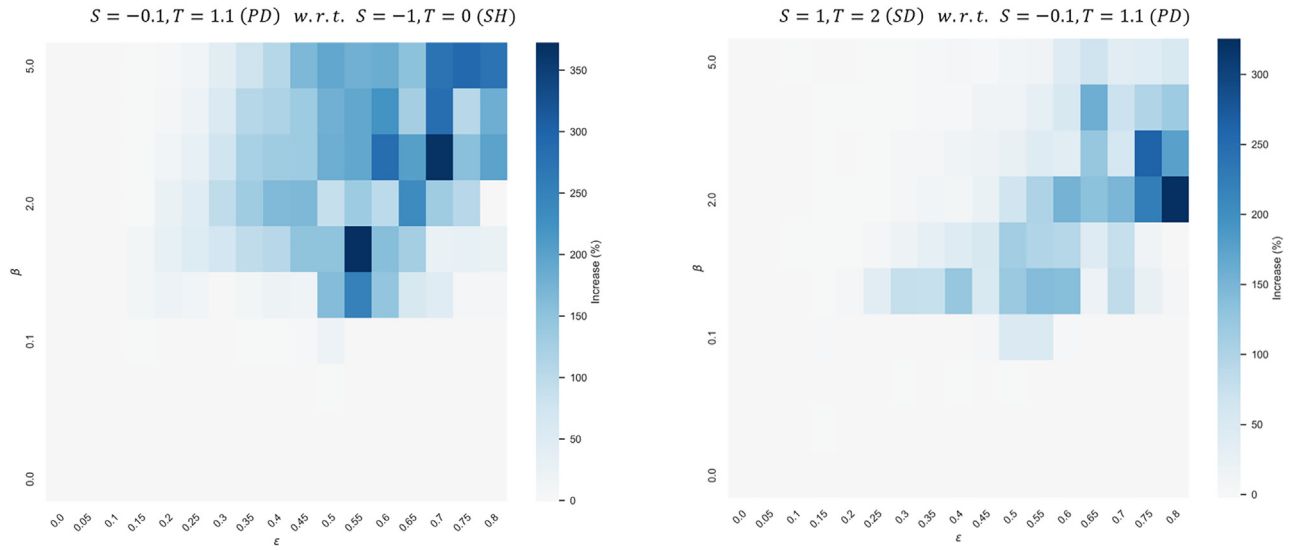


Figure 2. Increase in number of clusters of opinions when comparing different social tensions (S and T values of the game)

Panel of sensitivity analysis on (ϵ, β) showing increase in fragmentation when comparing two pairs of games (PD w.r.t. SH, and SD w.r.t. PD). Both heat-maps show a clear increase in the number of opinion clusters when values of S and T are higher and, therefore, a greedy and fearless population is defined.

opinion consensus, is achieved for the SH game (mutual defection is preferred, having a fear sentiment). These differences in opinions for SD and SH are easily observed when comparing the nodes' colors of network in Figure 4. A midway between the highest fragmentation and highest consensus is done in a PD game, which reflects a greedy and but brave society[27]).

The effects of the intensity of selection

In our last experiment we explore how the intensity of selection β of the success-driven rules of OD change the number of clusters of opinions and fragmentation. Panel of Figure 5 show heat-maps for S, T when having different values of $\epsilon = \{0.2, 0.3\}$ and $\beta = \{0.5,1,2,5\}$, respectively. First and as a confirmation of previous analysis, there is a gradient of higher fragmentation toward the top-right corner of the (S, T) space. This can be observed for all the β and ϵ values. In fact, differences are significant if we compare two pairwise games defined by S, T from a top-right area and a bottom-left area.

Another insight from the panel is the relevance of the intensity of selection. A higher implication of success and thus, the evolutionary game dynamics, a higher fragmentation. This increase in the intensity of selection makes that, for high values of β such as 5 (last row of the panel), the fragmentation is high in all the S, T values of the space and then, differences among games is reduced.

DISCUSSION

Our goal in performing this research was to find relationships between the fragmentation of opinions, polarization, and the social tensions of individuals at a microscopic level. And in doing so, to go beyond the top-down narrative of macroscopic or even planetary-scale determinants of these processes. To that effect, we have developed a novel success-driven computational model where opinion formation is coupled with individual evolutionary success in social dilemmas. We have performed large-scale simulations to obtain new evolutionary insights into what drives social tensions and how this translates to opinion fragmentation and polarization. Due to the importance of these processes in mitigating the spread of misinformation and intolerance in social networks, public forums, as well as in the mass media and the political arena, our insights have important consequences for fostering social cohesion, opinion unison, and inclusivity.

In our model, the opinion of individuals is understood as a probability to cooperate with others on a given topic. Then, by considering different social dilemmas in which individuals can either cooperate or defect, we have observed distinguishable effects that are generated by the different tensions in these dilemmas. We have found that greed is the main driver of polarization, such that in a population dominated by greed and unilateral defection, as in the prisoner's dilemma and the snowdrift game, the fragmentation of opinions is maximal. On the contrary, in a population where mutual defection instills the fear of punishment in the absence of greed, as in the stag-hunt game, opinion fragmentation is smallest.

One of our main conclusions is thus that the conditions of the governing game and its corresponding social tensions[27] can dramatically influence the opinion landscape. Importantly, this is not a top-down phenomenon, but a microscopically driven process that essentially starts at an individual level and then percolates through the social network. In fact, percolation has often been associated with information spreading across social networks,[51,52] and this traditionally physical process has even been linked rigorously to social phenomena in terms of the governing universality class.[53] But despite the microscopic drivers behind these processes, our results can nevertheless explain also why conditions imposed on a society top down can lead to a more fragmented state, in terms of the imposed governing social dilemma, which then translates to social tensions on an individual level.

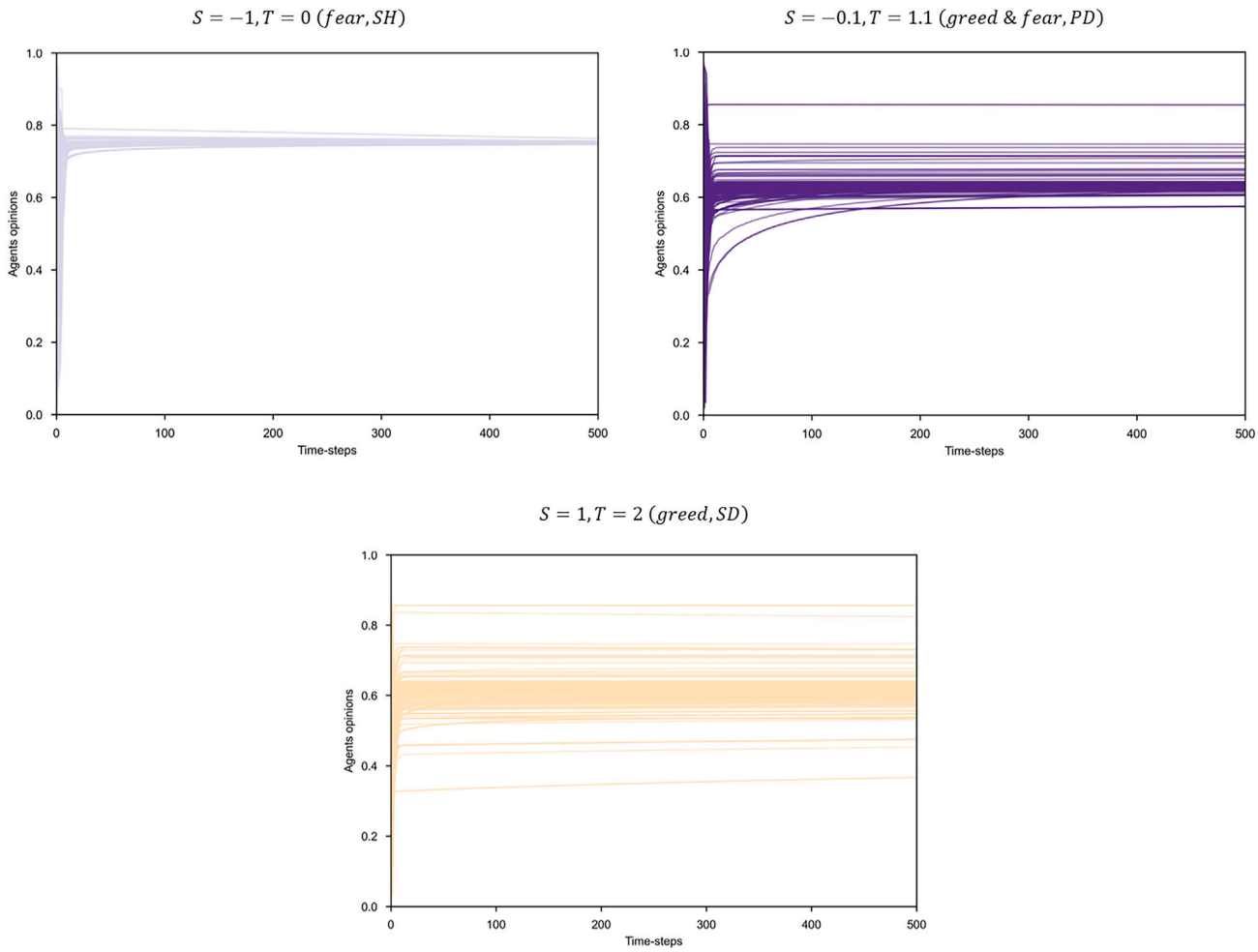


Figure 3. Evolution of opinions for different social tensions

Opinions' evolution for three parameters combinations of the game (SH with $S = -1, T = 0$, PD with $S = -0.1, T = 1.1$, and SD with $S = 1, T = 2$) until time-step 500 (although simulations are run until 5,000 steps). Values for success-driven OD model are $\epsilon = 0.7, \beta = 5$. In a fearful population (SH), agents achieve a consensus; but in a fearless and greedy population (SD), fragmentation of opinions is high.

Limitations of the study

We propose that the research done here can be upgraded and tested directly with data from social media, open repositories on topics that reflect controversy, or with human experiments, as done recently in the work of Li et al.[54] This is a limitation of our study which is mainly methodological and does not include the use of real data about opinions' evolution. works. As stated in the study of F. Vazquez,[29] there is a need for collecting empirical data from social experiments that would allow to test assumptions about social interactions at the microscopic level and to test predictions at the macroscopic level. Additionally, it should be possible to employ powerful generative artificial intelligence methods such as large language models (e.g., Chat-GPT)[55] to emulate conversations about contentious topics as an additional input to the success-driven opinion model. Concretely, the use of climate-related opinions could facilitate the use of decision-making tools to trigger climate change actions.[56] Of course, our model also lends itself well to alterations of the governing OD model, for example, different from the bounded confidence level group of models.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability

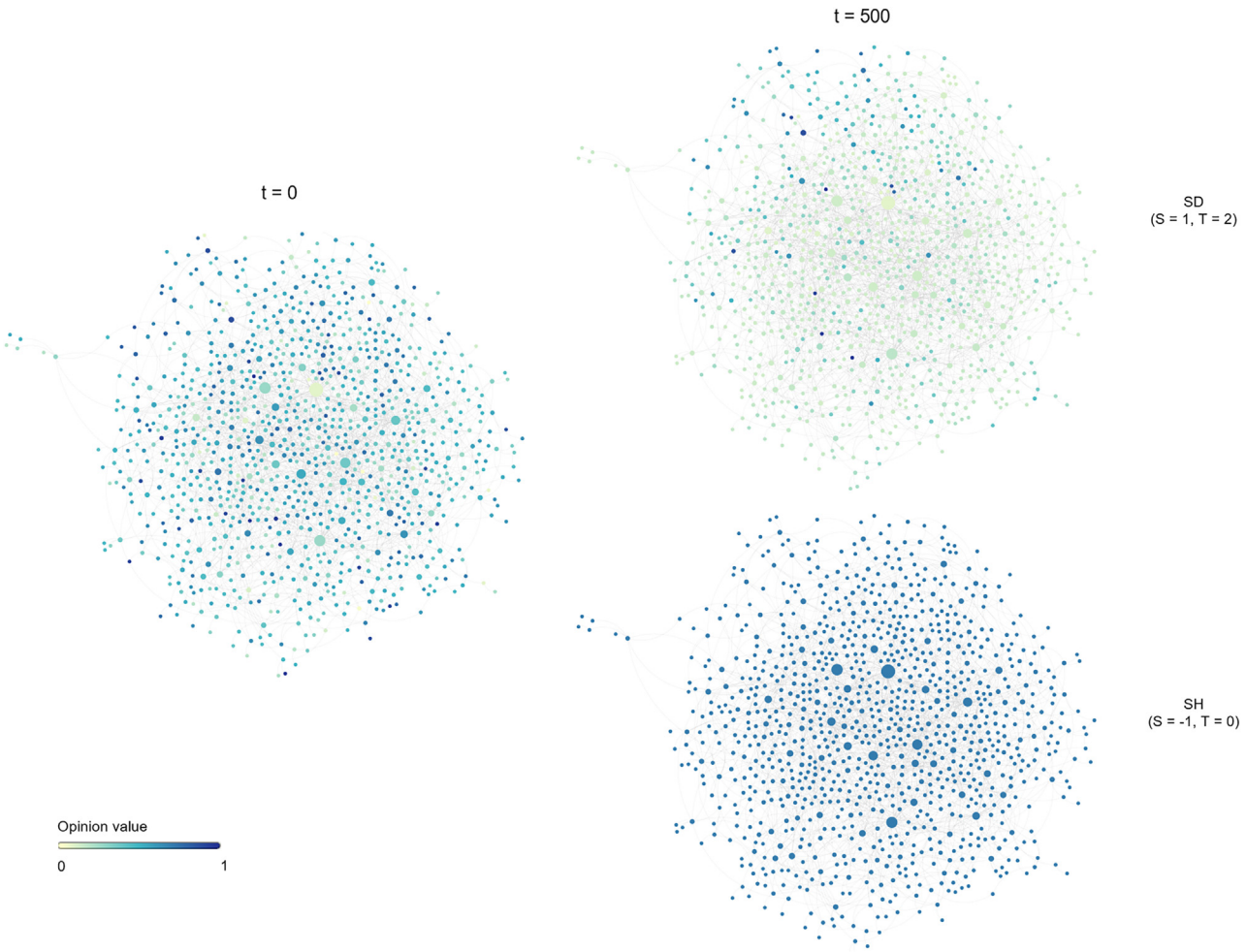


Figure 4. Snapshots of the same social network structure at time-steps $t = 0$ and $t = 500$ showing the evolution of opinions for each node for SD (top) and SH (bottom)

Nodes' diameter is related to their degree while nodes' color is the opinion value o_i. Main parameters for OD model are $\epsilon = 0.7, \beta = 5$. At $t = 500$, diversity in opinions (i.e., fragmentation) is shown for the nodes when playing SD ($S = 1, T = 2$). In contrast, all the nodes reach consensus when game is SH ($S = -1$, $T = 0$).

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109254.

ACKNOWLEDGMENTS

M.C. was supported by EMERGIA21_00139, funded by Consejería de Universidad, Investigación e Innovación of the Andalusian Government and by "ERDF A way of making Europe"; and CONFIA (PID2021-122916NB-I00), granted by the Spanish Ministry of Science. M.P. was supported by the Slovenian Research and Innovation Agency (Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije) (Grant Nos. P1-0403and N1-0232).

AUTHOR CONTRIBUTIONS

M.C.: Conceptualization, Methodology, Software, Validation, Writing – Original Draft, Funding acquisition. F.S.: Conceptualization, Methodology, Investigation, Writing – review editing. M.P.: Validation, Methodology, Writing – review editing.

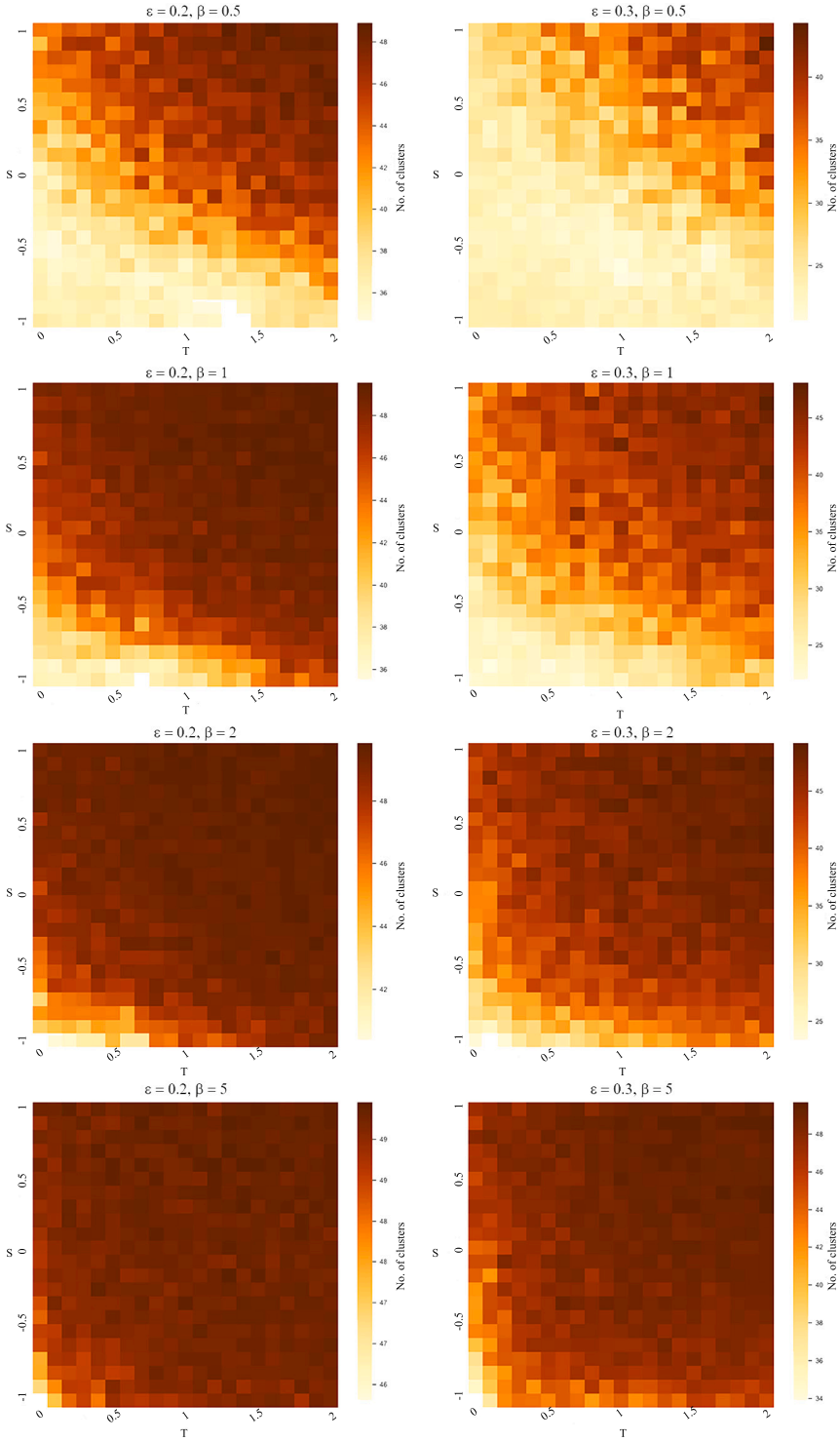


Figure 5. Number of clusters of opinions when S and T parameters change

Sensitivity analysis on (S, T) for different confidence levels $\epsilon = \{0.2, 0.3\}$ and intensity of selection $\beta = \{0.5,1,2,5\}$. One can see a higher opinion fragmentation when S and T increases (toward top-right corner of the heat-maps). When success-driven importance increases (higher β), opinion is also more fragmented.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 3, 2023
Revised: January 19, 2024
Accepted: February 13, 2024
Published: February 16, 2024

REFERENCES

1. Levin, S.A., Milner, H.V., and Perrings, C. (2021). The Dynamics of Political Polarization.
2. Bail, C.A., Argyle, L.P., Brown, T.W., Bumpus, J.P., Chen, H., Hunzaker, M.B.F., Lee, J., Mann, M., Merhout, F., and Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. Proc. Natl. Acad. Sci. USA *115*, 9216–9221.
3. Evans, T., and Fu, F. (2018). Opinion formation on dynamic networks: identifying conditions for the emergence of partisan echo chambers. R. Soc. Open Sci. *5*, 181122.
4. Wang, X., Sirianni, A.D., Tang, S., Zheng, Z., and Fu, F. (2020a). Public discourse and social network echo chambers driven by socio-cognitive biases. Phys. Rev. X *10*, 041042.
5. Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., and Starnini, M. (2021). The echo chamber effect on social media. Proc. Natl. Acad. Sci. USA *118*. e2023301118.
6. Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., and Quattrociocchi, W. (2016). The spreading of misinformation online. Proc. Natl. Acad. Sci. USA *113*, 554–559.
7. Sîrbu, A., Pedreschi, D., Giannotti, F., and Kertész, J. (2019). Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. PLoS One *14*, e0213246.
8. Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A.A., Eckles, D., and Rand, D.G. (2021). Shifting attention to accuracy can reduce misinformation online. Nature *592*, 590–595.
9. Macy, M., Deri, S., Ruch, A., and Tong, N. (2019). Opinion cascades and the unpredictability of partisan polarization. Sci. Adv. *5*, eaax0754.
10. Cook, J., and Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using bayesian networks. Top. Cogn. Sci. *8*, 160–179.
11. Wang, Z., Jusup, M., Guo, H., Shi, L., Geček, S., Anand, M., Perc, M., Bauch, C.T., Kurths, J., Boccaletti, S., and Schellnhuber, H.J. (2020b). Communicating sentiment and outlook reverses inaction against collective risks. Proc. Natl. Acad. Sci. USA *117*, 17650–17655.
12. Green, J., Edgerton, J., Naftel, D., Shoub, K., and Cranmer, S.J. (2020). Elusive consensus: Polarization in elite communication on the covid-19 pandemic. Sci. Adv. *6*, eabc2717.
13. Pennycook, G., McPhetres, J., Zhang, Y., Lu, J.G., and Rand, D.G. (2020). Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychol. Sci. *31*, 770–780.
14. Rost, K., Stahel, L., and Frey, B.S. (2016). Digital social norm enforcement: Online firestorms in social media. PLoS One *11*, e0155923.
15. Yilmaz, I., and Shipoli, E. (2022). Use of past collective traumas, fear and conspiracy theories for securitization of the opposition and authoritarianisation: the turkish case. Democratization *29*, 320–336.
16. Weibull, J.W. (1997). Evolutionary Game Theory (MIT press).
17. Nowak, M.A. (2006). Five rules for the evolution of cooperation. Science *314*, 1560–1563.
18. Rand, D.G., and Nowak, M.A. (2013). Human cooperation. Trends Cogn. Sci. *17*, 413–425.
19. Nowak, M.A., Sasaki, A., Taylor, C., and Fudenberg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. Nature *428*, 646–650.
20. Axelrod, R.M. (1997). The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration (Princeton University Press).
21. Santos, F.C., Pacheco, J.M., and Lenaerts, T. (2006). Evolutionary dynamics of social dilemmas in structured heterogeneous populations. Proc. Natl. Acad. Sci. USA *103*, 3490–3494.
22. Sun, W., Liu, L., Chen, X., Szolnoki, A., and Vasconcelos, V.V. (2021). Combination of institutional incentives for cooperative governance of risky commons. iScience *24*, 102844.
23. Domingos, E.F., Grujić, J., Burguillo, J.C., Kirchsteiger, G., Santos, F.C., and Lenaerts, T. (2020). Timing uncertainty in collective risk dilemmas encourages group reciprocation and polarization. iScience *23*, 101752.
24. Chica, M., Rand, W., and Santos, F.C. (2023). The evolution and social cost of herding mentality promote cooperation. iScience *26*, 107927.
25. Chica, M., Chiong, R., Adam, M.T.P., and Teubner, T. (2019). An evolutionary game model with punishment and protection to promote trust in the sharing economy. Sci. Rep. *9*, 19789.
26. Chica, M., Hernández, J.M., Manrique-de Lara-Peñate, C., and Chiong, R. (2021). An evolutionary game model for understanding fraud in consumption taxes [research frontier]. IEEE Comput. Intell. Mag. *16*, 62–76.
27. Macy, M.W., and Flache, A. (2002). Learning dynamics in social dilemmas. Proc. Natl. Acad. Sci. USA *99* (suppl_3), 7229–7236.
28. Płatkowski, T. (2017). Greed and fear in multiperson social dilemmas. Appl. Math. Comput. *308*, 157–160.
29. Vazquez, F. (2022). Modeling and analysis of social phenomena: challenges and possible research directions. Entropy *24*, 491.
30. Galam, S. (2012). What Is Sociophysics about? (Springer US), pp. 3–19.
31. Axelrod, R., Daymude, J.J., and Forrest, S. (2021). Preventing extreme polarization of political attitudes. Proc. Natl. Acad. Sci. USA *118*. e2102139118.
32. Sîrbu, A., Loreto, V., Servedio, V.D., and Tria, F. (2017). Opinion Dynamics: Models, Extensions and External Effects (Participatory sensing, opinions and collective awareness), pp. 363–401.
33. Dong, Y., Zhan, M., Kou, G., Ding, Z., and Liang, H. (2018). A survey on the fusion process in opinion dynamics. Inf. Fusion *43*, 57–65.
34. Banisch, S., and Olbrich, E. (2019). Opinion polarization by learning from social feedback. J. Math. Sociol. *43*, 76–103.
35. Vazquez, F., Krapivsky, P.L., and Redner, S. (2003). Constrained opinion dynamics: Freezing and slow evolution. J. Phys. A: Math. Gen. *36*, L61–L68.
36. Mobilia, M. (2011). Fixation and polarization in a three-species opinion dynamics model. Europhys. Lett. *95*, 50002.
37. Bhat, D., and Redner, S. (2020). Polarization and consensus by opposing external sources. J. Stat. Mech. *2020*, 013402.
38. Mobilia, M. (2023). Polarization and consensus in a voter model under time-fluctuating influences. Physics *5*, 517–536.
39. Galam, S. (2023). Unanimity, coexistence, and rigidity: Three sides of polarization. Entropy *25*, 622.
40. Li, Z., Chen, X., Yang, H.-X., and Szolnoki, A. (2022b). Game-theoretical approach for opinion dynamics on social networks. Chaos *32*, 073117.
41. Ding, F., Liu, Y., and Li, Y. (2009). Co-evolution of opinion and strategy in persuasion dynamics: An evolutionary game theoretical approach. Int. J. Mod. Phys. C *20*, 479–490.
42. Li, X., Mobilia, M., Rucklidge, A.M., and Zia, R.K.P. (2021). How does homophily shape the topology of a dynamic network? Phys. Rev. E *104*, 044311.
43. Kawakatsu, M., Lelkes, Y., Levin, S.A., and Tarnita, C.E. (2021). Interindividual cooperation mediated by partisanship complicates madison's cure for "mischiefs of faction". Proc. Natl. Acad. Sci. USA *118*. e2102148118.
44. Albert, R., and Barabási, A.L. (2002). Statistical mechanics of complex networks. Rev. Mod. Phys. *74*, 47.
45. Yang, H.-X. (2016). A consensus opinion model based on the evolutionary game. Europhys. Lett. *115*, 40007.
46. Huang, C., Hou, Y., and Han, W. (2023). Coevolution of consensus and cooperation in evolutionary hegselmann–krause dilemma with the cooperation cost. Chaos, Solit. Fractals *168*, 113215.
47. Hegselmann, R., and Krause, U. (2002). Opinion dynamics and bounded confidence models, analysis, and simulation. J. Artif. Soc. Soc. Simulat. *5*.
48. Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000). Mixing beliefs among interacting agents. Adv. Complex Syst. *3*, 87–98.

49. Lorenz, J. (2010). Heterogeneous bounds of confidence: meet, discuss and find consensus. Complexity *15*, 43–52.

50. Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H.E., and Quattrociocchi, W. (2017). Modeling confirmation bias and polarization. Sci. Rep. *7*, 40391.

51. Ji, P., Ye, J., Mu, Y., Lin, W., Tian, Y., Hens, C., Perc, M., Tang, Y., Sun, J., and Kurths, J. (2023). Signal propagation in complex networks. Phys. Rep. *1017*, 1–96.

52. Jusup, M., Holme, P., Kanazawa, K., Takayasu, M., Romić, I., Wang, Z., Geček, S., Lipić, T., Podobnik, B., Wang, L., et al. (2022). Social physics. Phys. Rep. *948*, 1–148.

53. Hinrichsen, H. (2000). Non-equilibrium critical phenomena and phase transitions into absorbing states. Adv. Phys. X. *49*, 815–958.

54. Li, J., Zhao, X., Li, B., Rossetti, C.S.L., Hilbe, C., and Xia, H. (2022a). Evolution of cooperation through cumulative reciprocity. Nat. Comput. Sci. *2*, 677–686.

55. Dwivedi, Y.K., Kshetri, N., Hughes, L., Slade, E.L., Jeyaraj, A., Kar, A.K., Baabdullah, A.M., Koohang, A., Raghavan, V., Ahuja, M., et al. (2023). "so what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy. Int. J. Inf. Manag. *71*, 102642.

56. Molina-Perez, E. (2023). Harnessing the power of decision-support tools to trigger climate action. Nat. Comput. Sci. *3*, 461–463.

57. Traulsen, A., Nowak, M.A., and Pacheco, J.M. (2006). Stochastic dynamics of invasion and fixation. Phys. Rev. E *74*, 011909.

58. Macal, C.M., and North, M.J. (2005). Tutorial on agent-based modeling and simulation. In Proceedings of the 37th conference on Winter simulation (ACM), pp. 2–15.

59. Adami, C., Schossau, J., and Hintze, A. (2016). Evolutionary game theory using agent-based methods. Phys. Life Rev. *19*, 1–26.

60. Liu, L., Wang, X., Chen, X., Tang, S., and Zheng, Z. (2021). Modeling confirmation bias and peer pressure in opinion dynamics. Front. Phys. *9*, 649852.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Java 11	Oracle Java	https://jdk.java.net/11/
Apache Math v3.2	Apache Foundation	commons.apache.org/math
Mason v20	George Mason Univ.	cs.gmu.edu/~eclab/projects/mason
GraphStream v2.0	GraphStream Project	graphstream-project.org
NumPy v1.25.1	Github NumPy	github.com/numpy/numpy
Matplotlib v3.5.1	Matplotlib	matplotlib.org
Pandas v1.4.3	Scipy	scipy.org
Seaborn v0.11.2	PyData	seaborn.pydata.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Manuel Chica (manuelchica@ugr.es).

Materials availability

No materials were newly generated for this paper.

Data and code availability

- Data generated by the simulations is available upon request. No specific data-sets were used to feed the simulations.
- The code used to generate the simulations and the figures is available from the corresponding authors upon request.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

A success-driven opinion model

The methods needed to define the evolutionary success-driven OD model we use for the experiments are defined here. First, we explain an approximation method to compute the fitness of a pairwise interaction among two players given their opinions. Second, the proposed success-driven OD methods that include the fitness coming from two players interaction in the opinions evolution. Experimental setup and simulation specifications to obtain the results are defined at the end of the section.

Fitness computation based on players' opinions

First, our proposal includes a way to define the probability of either cooperate or defect for a player by a continuous opinion variable which evolves over-time. The opinion of a player $o_i \in [0,1]$ defines the probability to cooperate with another player j in a pairwise interaction. This opinion is the same for all the possible pairwise interactions the player i is having in the same time-step t. The closer the value of o_i to 0, the most probable for player i to defect. If the opinion of a player is close to 1, cooperating with others is the most probable option. In that way, we define the game strategy of a player as a non-binary opinion.

In order to computationally calculate the fitness output for every pairwise interaction of players i and j, we estimate and compute their fitness outcome by considering their opinions o_i and o_j with respect to the expected fitness or payoffs defined in the game and their possible strategies. Then, if having a traditional payoff matrix with parameters P, R, S, T, the expected fitness or payoffs f_{ij} for player i in its interaction with player j is as in Equation 1:

$$f_{ij} = T(1 - o_i)o_j + Ro_io_j + P(1 - o_i)(1 - o_j) + So_i(1 - o_j). \quad \text{(Equation 1)}$$

In the case of considering a generalized pairwise game of two parameters $S \in [-1, 1]$ and $T \in [0, 2]$, which are sufficient to characterize the main pairwise games[21] (see supplemental information for more details), the estimated fitness function is simplified as follows in Equation 2, as P is set to 0 and R to 1:

$$f_{ij} = T(1 - o_i)o_j + o_i o_j + So_i(1 - o_j). \quad \text{(Equation 2)}$$

Update rule using opinion dynamics models

Opinions of the players are influencing others and this can be seen as an imitation mechanism in an evolutionary game. Then, this imitation mechanism modifies the opinion to cooperate or not by the opinions of other players at every time-step t. A common imitation process is using Fermi's function where a player i imitates another player j of the same population Z with a probability $p_{i \to j}$ that increases with their payoff difference ($f_j - f_i$, being f_i and f_j the accumulated fitness or payoffs of i and j in $t - 1$, respectively):[57]

$$p_{i \to j} = \frac{1}{1 + e^{-\beta(f_j - f_i)}}, \quad \text{(Equation 3)}$$

where the amplitude of noise β controls the intensity of selection.

When players have a real-valued opinion, the imitation process must follow an approach similar to the one of OD models. Thus, our proposal facilitates using the well-known DW and HK OD models (described in the supplemental information document) together with fitness information of the players obtained after their pairwise interactions using their opinions. The proposed mechanism bias the propagation of opinions depending on their success rate.

Success-driven HK model. This model extends the traditional HK model by including, for each weight of the agent's opinion w_{ij}, the comparison in the fitness obtained by both players with their opinions. Then, o_i is calculated as in Equation 4, where weights w_{ij}^t for individuals' opinions are given by Equation 5 that includes a fitness bias.

$$o_i^{t+1} = \frac{1}{\sum_{j \in S_i^t} w_{ij}^t} \sum_{j \in S_i^t} w_{ij}^t o_j^t. \quad \text{(Equation 4)}$$

$$w_{ij}^t = \begin{cases} \dfrac{1}{1 + e^{-\beta(f_i - f_j)}}, & \text{if } j \in S_i^t. \\ 0, & \text{otherwise.} \end{cases} \quad \text{(Equation 5)}$$

As in the HK original model, S_i^t are those players bounded by a confidence level of agent i ($S_i^t = \{j, \left| o_i^t - o_j^t \right| \leq \epsilon\}$) and ϵ the confidence level, $|.|$ denotes the absolute value of a real number. Take into account that the agent itself is included in this set S_i. In this proposed success-driven OD model, if β equals to 0, then the intensity of selection is null and we have the original HK model.

Success-driven DW model. In a similar way to the HK model, the success-driven DW model is dervied from the traditional DW model by modifying the μ function of each pair of opinions, defined by Equation 6.

$$o_i^{t+1} = o_i^t + \mu(f_i, f_j)\left(o_j^t - o_i^t\right). \quad \text{(Equation 6)}$$

In this model, μ is a function defined by Equation 7 and μ_0 is the original μ parameter of the DW model. The function includes the comparison of the fitness values of both agents and thus, when intensity of selection β equals to 0, the success-driven DW model is converted to the original DW model.

$$\mu(f_i, f_j) = \frac{2\mu_0}{1 + e^{-\beta(f_i - f_j)}}. \quad \text{(Equation 7)}$$

Experimental setup and simulation methods

We perform Monte Carlo simulations of the success-driven OD model via agent-based modeling,[58,59] performed on computer clusters to obtain the stationary states of the model specifications. Agents of the simulations represent the individuals and their opinions as well as the interactions among them to share their opinions and obtain their payoffs. The initial opinions of all agents are set at random using a uniform random distribution ∈ [0,1]. We perform the simulations with a population size $Z = 1{,}024$. Players are connected through a heterogeneous social network with a power-law distribution, generated by the Barabasi-Albert algorithm[44] (with parameter $m =$ to obtain a network with average degree $\langle k \rangle \approx 4$ and density of 0.004). The system is run for 5,000 time-steps to achieve a stable state. Results were averaged for 30 Monte-Carlo realizations.

Values for parameters S and T of the pairwise game as well as ϵ of the HK model and β intensity factor are specified for each experiment. In general, we mainly explore the number of opinion clusters when the system reaches the stable state (by averaging the last 25% time-steps of the simulation of the 30 Monte-Carlo realizations). Complying with previous studies, we utilize the number of peaks in

distribution of opinions to represent the number of opinion clusters.[50,60] The calculation of the number of clusters in the opinions' spectrum, which denotes the consensus or fragmentation state of the opinions, is done as follows. We divide the opinion interval $[0, 1]$ into 50 bins and compute the frequency of opinion values falling into each bin. By also following the method performed in previous studies, we merge clusters if their distance among middle points of the clusters are below a threshold τ. This threshold is set to 0.01. Therefore, two opinions belong to different clusters if they are separated by a difference higher than $\tau = 0.01$.