https://doi.org/10.1038/s41586-019-1507-6

Information gerrymandering and undemocratic decisions

Alexander J. Stewart[1]*, Mohsen Mosleh[2], Marina Diakonova[3], Antonio A. Arechar[2,4], David G. Rand[2] & Joshua B. Plotkin[5]

People must integrate disparate sources of information when making decisions, especially in social contexts. But information does not always flow freely. It can be constrained by social networks[1–3] and distorted by zealots and automated bots[4]. Here we develop a voter game as a model system to study information flow in collective decisions. Players are assigned to competing groups (parties) and placed on an 'influence network' that determines whose voting intentions each player can observe. Players are incentivized to vote according to partisan interest, but also to coordinate their vote with the entire group. Our mathematical analysis uncovers a phenomenon that we call information gerrymandering: the structure of the influence network can sway the vote outcome towards one party, even when both parties have equal sizes and each player has the same influence. A small number of zealots, when strategically placed on the influence network, can also induce information gerrymandering and thereby bias vote outcomes. We confirm the predicted effects of information gerrymandering in social network experiments with $n = 2{,}520$ human subjects. Furthermore, we identify extensive information gerrymandering in real-world influence networks, including online political discussions leading up to the US federal elections, and in historical patterns of bill co-sponsorship in the US Congress and European legislatures. Our analysis provides an account of the vulnerabilities of collective decision-making to systematic distortion by restricted information flow. Our analysis also highlights a group-level social dilemma: information gerrymandering can enable one party to sway decisions in its favour, but when multiple parties engage in gerrymandering the group loses its ability to reach consensus and remains trapped in deadlock.

Distorted and false information threaten to disrupt public discourse and democratic decision-making[5,6]. Social media platforms are particularly vulnerable, because they allow users to shut out dissenting voices[1–3], while providing adversarial actors with anonymity and opportunity to target messages for maximal effect[7]. The effect of information distortion is not limited to the online world, but filters out to traditional news media and voter behaviour[8]. Two distinct but intertwined threats have received considerable attention: information campaigns using fake news[5,9] and automated bots[4], and the growth of polarized political debate[10,11]. These issues pose a considerable social problem. Progress requires that we develop basic scientific methods to understand how networks that constrain the flow of information influence group decision-making.

Here we develop a voter game to study collective decisions under incomplete information. The game is simple enough to analyse mathematically and to use in controlled experiments with human subjects, yet it retains salient features of real-world collective decisions. Players are split into two parties of equal size and allowed to change their voting intention over time in response to continuously updated polling data. The aggregate polling information seen by a player is determined by their placement on a directed graph, called the influence network. Players are aware that polls represent a subset sampled from the entire population. At the end of the game, players receive the maximum payoff B if the final vote share for their assigned party exceeds a super-majority threshold V in which $V > 0.5$; they receive a lower payoff $b < B$ if the vote share of the opposing party exceeds V; and they receive no payoff if both parties fail to reach the threshold V, which we call 'deadlock' (Fig. 1). The possibility of deadlock forces players to consider both their personal preferences and the voting intentions of others in their decision-making.

We initially assume the payoff to the losing team is positive, $b > 0$, which reflects a 'compromise worldview': it is preferable to reach some decision than to end in deadlock. This payoff scheme captures the common practical value of broad consensus in collective decisions, even as individuals pursue partisan preferences[12,13]. There is ample evidence that large majorities of Americans, for example, adopt a compromise worldview in their attitudes towards political decisions[14]. However, others adopt a 'zero-sum worldview' in which they prefer deadlock to their party losing, that is, $b < 0$. The behaviour of people with a zero-sum worldview is simple: they act as zealots, meaning that they always vote for their preferred party regardless of the poll that they see. We begin by focusing on players with a compromise worldview, and then we study groups with a mixture of compromise and zero-sum players, including zero-sum bots.

We first considered how a player updates their intended vote over time as they integrate their partisan preference with noisy social information and the desire to avoid deadlock. We developed a simple mathematical model of voter behaviour based on calibration experiments that varied the super-majority threshold V, the payoff ratio B/b, the game duration and the poll size. In our model, a player expresses the intention to vote for their assigned party according to a probability that depends on the poll that they currently see and the time remaining in the game. In particular, we consider a six-parameter strategy space in which the voting probability of a player is conditioned on the state of their current poll (whether it projects their preferred party to win, the opposing party to win or deadlock) and the phase of the game (early or late) (Fig. 1c).

The structure of the influence network has considerable, and surprising, effects on vote outcomes. In the simplest influence networks, every player has an equal number of players from each party represented in the polls that they see, so that the polls—although not identical—are representative samples of the entire group (Fig. 2a–c). The decision process on such a network is unbiased in the sense that the expected vote share for each party is equal to the frequency of its membership in the entire group. However, even when parties are equally matched in influence and representation, bias can arise when the parties differ in how their influence is assorted across the network (Fig. 2d). Even when all players have the same amount of influence—that is, they are seen by the same number of other players—the two parties can nonetheless distribute their influence in more- or less-effective ways.

A party is most effective when it influences the largest possible number of people just enough to flip their votes, without wasting influence on those who are already convinced. The phenomenon of

[1]Department of Biology and Biochemistry, University of Houston, Houston, TX, USA. [2]Sloan School of Management, MIT, Cambridge, MA, USA. [3]Environmental Change Institute, University of Oxford, Oxford, UK. [4]Center for Research and Teaching in Economics (CIDE), Aguascalientes, Mexico. [5]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA.
*e-mail: astewar6@central.uh.edu

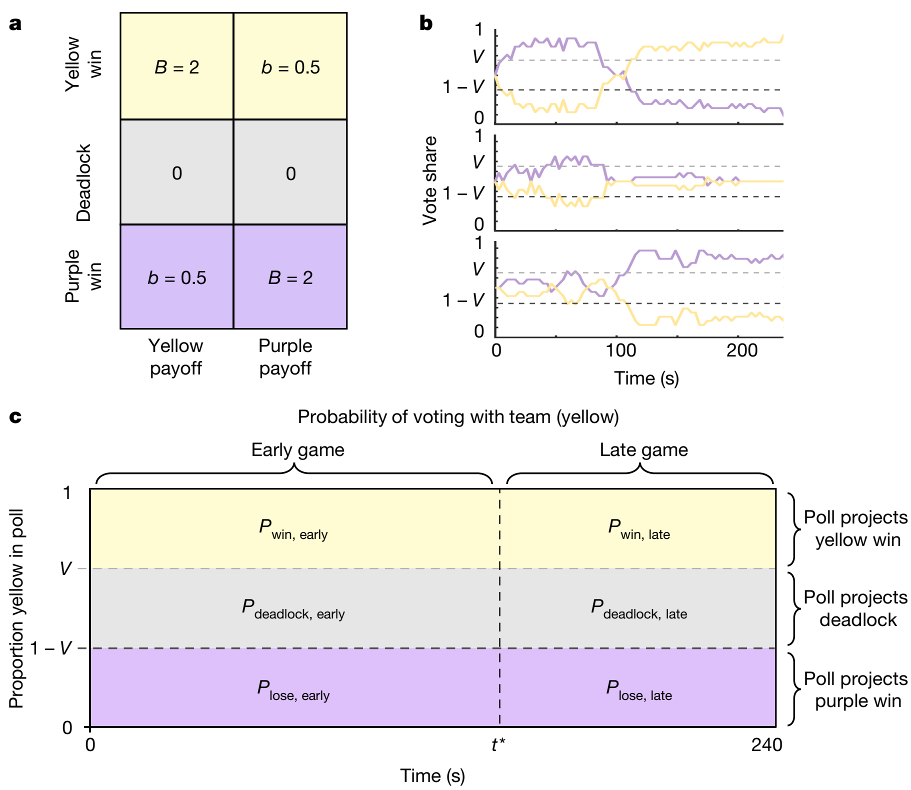

Fig. 1 | Strategies and payoffs in the voter game. Players are assigned to either the purple or yellow party and allowed to change their voting intention over time in response to continuously updated polling information. **a**, The voting intentions recorded at the end of the game (after 240 s) determine the payoffs to players. Players are incentivized primarily to vote for their party ($B > b$), but also to coordinate with the larger group to avoid deadlock ($b > 0$; compromise worldview). **b**, Example time series of the overall vote share in three experiments, which illustrate the yellow party winning a consensus (final vote share exceeding $V = 60\%$; top), neither party winning a consensus (deadlock; middle) and the purple party winning a consensus (bottom). **c**, A simple model of voter behaviour stipulates the probability a player will vote for their preferred party (yellow, in this example) at time t, given which of three possible outcomes is projected by the current polling information and whether the game is in the early ($t < t^*$) or late ($t > t^*$) phase. This six-parameter stochastic model of individual behaviour recapitulates the typical time series (Supplementary Fig. 4) and vote outcomes (Fig. 3) observed in experiments. **b**, **c**, Dashed lines indicate thresholds $V = 60\%$ and $1 - V = 40\%$.

information gerrymandering arises when one party punches above its weight by distributing its influence on a network so as to flip a disproportionate number of persuadable voters. To understand how a party can gain such an advantage, we adapt the principles of electoral gerrymandering—in which voting districts are drawn so that one party wins a disproportionate number of seats[15]—to construct influence networks in which one party has an advantage in persuading voters and the other party wastes much of its social influence (Fig. 2d, Supplementary Information section 4 and Supplementary Video 1).

To study information gerrymandering, we define the 'influence assortment' of an influence network. Positive influence assortment means that players are predominately exposed to the voting intentions of members from their own party; negative influence assortment means that players are predominately exposed to members of the opposing party. To be precise, the influence assortment of player i is defined as

$$a_i = \begin{cases} \Delta_i & \text{if} \Delta_i \geq 1/2 \\ -(1-\Delta_i) & \text{otherwise} \end{cases} \quad (1)$$

in which Δ_i denotes the proportion of the players who comprise the poll visible to player i who are assigned to the same party as player i. Notably, the influence assortment of a player depends in a nonlinear way on the proportion of their influencers who share their party. This definition is appropriate assuming players are strongly pulled towards the majority view that they observe (Supplementary Information section 3). The overall assortment of the influence network, denoted $\mathcal{A}_I$, is the average influence assortment of its nodes.

Information gerrymandering arises when parties have asymmetric influence assortment. We quantify information gerrymandering as the difference in assortment between a party $\mathcal{P}$ and its opposition, by defining the influence gap as

$$\mathcal{G}_I = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P} \bigcap \mathcal{H}} a_i - \frac{1}{N - |\mathcal{P}|} \sum_{i \in \mathcal{P}^c \bigcap \mathcal{H}} a_i \quad (2)$$

in which $\mathcal{P}$ are the nodes assigned to party $\mathcal{P}$, $\mathcal{H}$ are human nodes and N is the total number of nodes. Our model predicts that a party with a positive influence gap will benefit from information gerrymandering.

To test this prediction, we conducted experiments with human subjects ($n = 2{,}520$) playing the voter game, in which we varied only the structure of the influence network (Fig. 3). All games involved two parties of equal size (12 players each) with fixed payoffs ($B = 2$ and $b = 0.5$), super-majority threshold ($V = 60\%$) and duration (240 seconds) (for full details of experiments and pre-registrations, see Supplementary Information sections 1–3). In the baseline condition, each player sees a poll that consists of three players from their own party and three players from the opposing party, but the influence network is otherwise drawn randomly. Under this condition of no influence assortment, each party achieved a winning consensus in roughly one-quarter of experimental replicates; deadlock occurred in the remaining half of replicates (Fig. 3a). The time-series data from this condition were used to infer the probabilistic voting parameters of our behavioural model by maximum likelihood (Supplementary Information section 2.2). The distribution of inferred strategies is consistent with a Nash equilibrium for the voter

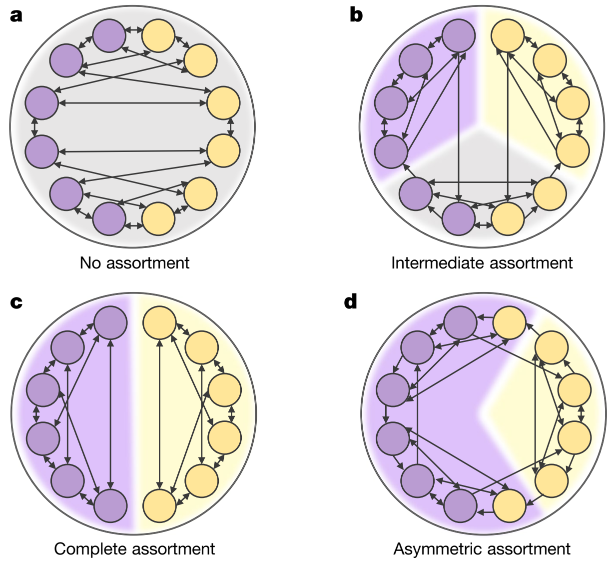

Fig. 2 | Influence assortment and information gerrymandering. **a**–**d**, The polling information available to a player in the voter game is determined by their placement on a directed graph, called the influence network. All of the example graphs here have nodes with identical indegree and outdegree equal to three; and background colours indicate the party with the majority of influence on each node (grey indicates no majority). Each individual may be influenced predominantly by their own party (positive influence assortment), predominately by the opposition party (negative influence assortment) or evenly split between parties (no influence assortment). **a**–**c**, When both parties have the same distribution of influence assortment across their members, assortment is symmetric and the decision outcome will be unbiased. **d**, An asymmetric distribution of assortment can distort the flow of information so that, even when all players have the same amount of influence, a majority of players are influenced primarily by one party's members—a phenomenon that we call information gerrymandering.

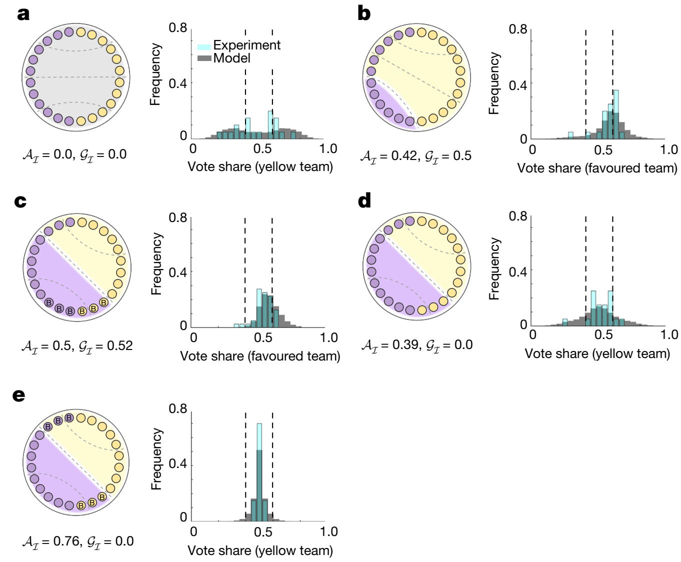

Fig. 3 | Undemocratic outcomes and polarization in the voter game. We conducted experimental voter games on human subjects ($n = 2{,}560$), varying only the structure of the influence network. Each game involved 24 players, including any bots. **a**, We inferred the parameters of our behavioural model (Fig. 1c) from experiments in a baseline condition: networks with no influence assortment and no influence gap ($\mathcal{A}_I = 0$ and $\mathcal{G}_I = 0$). The model recapitulates the observed, bimodal distribution of voting outcomes. **b**–**e**, We used the model to predict the distribution of voting outcomes in 4 additional conditions each with ≥20 replicates (model predictions are shown in grey and experimental results are overlaid in light blue). **b**, Information gerrymandering ($\mathcal{G}_I = 0.5$) produced vote shares as large as 67% for the more assorted party, which received a mean vote share of 57% across experimental replicates, consistent with the model prediction (Table 1). **c**, Asymmetric placement of 6 zealot bots also favoured the party with a positive influence gap, resulting in vote shares as large as 63% and a mean vote share of 53%. **d**, Symmetric influence assortment gave neither party an advantage, and the frequency of a consensus (15%) was markedly reduced compared to networks without assortment (55%). **e**, Symmetrically placed bots gave neither party an advantage and resulted in deadlock for all replicates. Dashed lines indicate thresholds $V = 60\%$ and $1 - V = 40\%$. The party favoured by information gerrymandering is depicted as yellow in the example graphs, but was in fact assigned to yellow and purple evenly across experimental replicates.

game (Supplementary Information section 5), which includes a portion of players who behave as zealots.

We used our behavioural model to predict the quantitative effects of influence assortment and information gerrymandering on voting outcomes in four other experimental conditions. All treatments retained the constraint that players have fixed and equal indegrees and outdegrees, and thus the same amount of influence. Our model predicts that information gerrymandering will skew the final vote towards the party with a positive influence gap ($\mathcal{G}_I > 0$) and that this party will achieve a winning consensus more often than its opposition. Both of these predictions were validated experimentally (one-sided Wilcoxon signed-rank test, $P = 0.003$ and one-sided binomial test, $P = 0.02$; Fig. 3b), demonstrating that a party does indeed gain a considerable advantage by information gerrymandering.

If one party benefits from information gerrymandering then, understanding this, the opposing party will naturally seek to do the same. The party that has a disadvantage ($\mathcal{G}_I < 0$) can redress the imbalance only by increasing the influence assortment of its members (equation (2)). But when both parties have equally high levels of influence assortment ($\mathcal{A}_I > 0$ and $\mathcal{G}_I = 0$), neither party will have an advantage. In fact, our model predicts that both parties will suffer from their self-constructed echo chambers, resulting in deadlock more often than in the case of no influence assortment. This prediction was also validated experimentally (one-sided t-test, $t = 2.5$, $P = 0.006$; Fig. 3), demonstrating that increasing the influence assortment of your party—although a rational response to information gerrymandering by your opponent—decreases the rate of consensus and therefore decreases payoffs for both parties.

Information gerrymandering by differential influence assortment requires a degree of coordination among party members that may be impractical in some settings. Another way to achieve the same advantage is to encourage players to adopt a zero-sum worldview and act as zealots. Or, in online interactions, bots can be deployed in place of actual human zealots. In the context of the voter game, zealot bots always project the intention to vote for their party regardless of the polls. Placed in strategic locations, zealot bots can increase the influence assortment of their party and decrease the influence assortment of the opposing party, generating a positive influence gap. When one party's zealot bots are so deployed ($\mathcal{G}_I > 0$), our model predicts that the vote will be skewed in its favour and the party will win a consensus more often than its opposition. The first of these predictions was validated experimentally (one-sided Wilcoxon sign-rank test, $P = 0.002$; Fig. 3) and the second was not statistically significant (one-sided binomial test, $P = 0.2$; Fig. 3). Thus a party receives some advantage from information gerrymandering by zealot bots. However, if both parties seek to use bots in the same way, then overall influence assortment increases, neither party receives an advantage and deadlock occurs in all experimental replicates (Fig. 3e).

Collective decisions often involve more individuals, with greater heterogeneity in influence, than used in our experiments. To study information gerrymandering on complex networks, we simulated our experimentally derived behavioural model on large influence networks with long-tailed degree distributions (Supplementary Information section 6). Information gerrymandering arises easily in these networks, and the influence gap $\mathcal{G}_I$ continues to be predictive of the resulting vote skew (Fig. 4a). Information gerrymandering induces vote skews that

Table 1 | Vote share and consensus

	Vote share model	Vote share experiment	Consensus model	Consensus experiment
No assortment	0.50 (0.41–0.59)	0.48 (0.42–0.54)	0.63 (0.35–0.85)	0.55 (0.25–0.70)
Asymmetric assortment	0.60 (0.56–0.63)	0.57 (0.53–0.60)	0.47 (0.20–0.70)	0.45 (0.20–0.65)
Asymmetric bots	0.56 (0.54–0.59)	0.53 (0.51–0.55)	0.19 (0.05–0.33)	0.18 (0.03–0.28)
Symmetric assortment	0.50 (0.46–0.55)	0.51 (0.47–0.55)	0.18 (0–0.35)	0.15 (0–0.30)
Symmetric bots	0.50 (0.49–0.51)	0.50 (0.49–0.51)	0.01 (0–0.05)	0 (0)

Mean vote share and frequency of a winning consensus observed in experiments and predicted by the behavioural model, for the five experimental conditions shown in Fig. 3. Parentheses show 95% confidence intervals derived by Efron bootstrap (experiments) or by Monte Carlo simulations (model).

are as large as 40% when two parties have equal size. Gerrymandering can also reverse a 2:1 difference in party size, allowing the minority party to win a majority of votes under our model (Fig. 4a).

To study groups with heterogeneous incentives, we simulated the voter game on complex networks in which 20% of randomly placed players hold a zero-sum worldview, resulting in 20% more zealots in each game. Information gerrymandering continues to arise in this setting, and the influence gap continues to predict vote outcomes (Supplementary Figs. 12–14). Across diverse networks—in which party representation ranged from 50:50 to 80:20—the influence gap accounts for at least 40% of the variance in vote share under our behavioural model (Supplementary Table 2).

Our results on the voter game raise the question of whether real-world collective decisions bear the hallmarks of information gerrymandering. To investigate this, we constructed networks from a variety of empirical influence relationships, in which assortment may either arise by strategic design or emerge spontaneously by self-assembly. We measured influence relationships among lawmakers, using data on bill co-sponsorship and among participants in online political discussions (Supplementary Information section 7). We found significant differences ($P < 0.01$, t-test) in influence assortment between the governing and non-governing parties in six out of eight European legislative bodies; and between Democrats and Republicans in online political discussions preceding US federal elections (Fig. 4c). The influence gaps observed in these diverse forms of political discourse are comparable in magnitude to those that induce large (>10%) vote skews in our model and experiments. Influence gaps also occur in the networks of bill co-sponsorship of the US Congress (Fig. 4b), which show a pattern of increasing influence assortment over time, consistent with previous accounts of increasing polarization in Congress based on established metrics of political ideology[16] (Supplementary Fig. 17).

Political polarization and echo chambers are the focus of intense research and public discussion[1–3,10]. Unravelling the psychological mechanisms at play when people interact with different identity groups[17,18], opposing viewpoints[19], hot-button topics such as climate change[20], fake or misleading news[9], trusted versus distrusted sources[21] and bots[4,22] is vital for understanding decision making in real-world settings. Furthermore, affective polarization[10]—negative attitudes to members of the other party, rather than to specific policies—is of great importance as it may cause people to adopt a zero-sum worldview. Nevertheless, our study on the voter game highlights how sensitive collective decisions are to information gerrymandering on an influence network, how easily gerrymandering can arise in realistic networks and how widespread it is in real-world networks of political discourse and legislative process. Our analysis provides a new perspective and a quantitative measure to study public discourse and collective decisions across diverse contexts.

Central to this perspective is the understanding that influence assortment presents a group-level social dilemma[23]. Symmetric influence assortment allows for democratic outcomes, in which the expected vote share of a party is equal to its representation among voters; and low influence assortment allows decisions to be reached with broad consensus despite different partisan goals. A party that increases its own influence assortment relative to that of the other party by coordination, strategic use of bots or encouraging a zero-sum worldview benefits from information gerrymandering and wins a disproportionate share

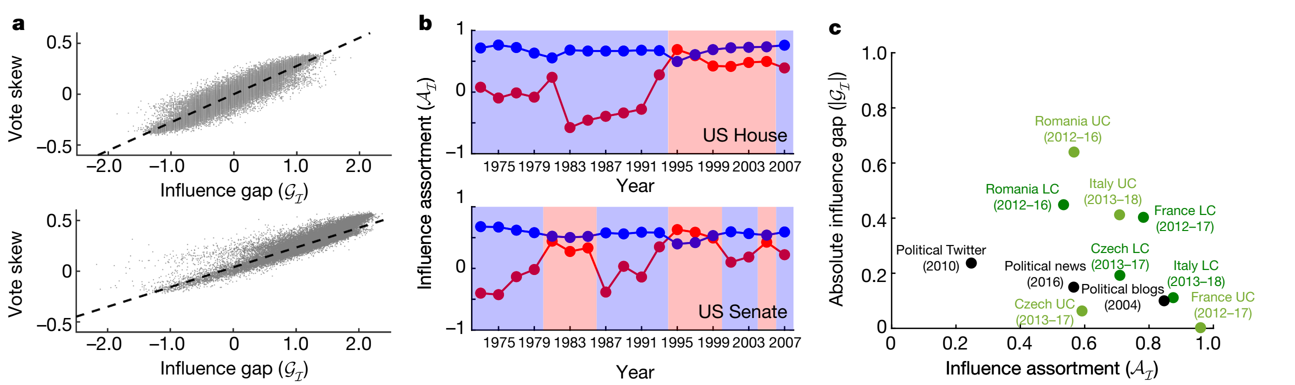

Fig. 4 | Information gerrymandering on simulated influence networks and empirical networks of political discourse. a, We simulated our behavioural model of the voter game on 10^5 networks with power-law degree distributions that reflect broad variation in individual influence. The influence gap $\mathcal{G}_I$ between two parties assigned randomly to nodes correlates strongly with voting outcomes on these networks. Information gerrymandering ($\mathcal{G}_I > 0$) induces large vote skews (top; equal-sized parties), allowing even a minority party to win a majority of votes (bottom; one party is twice as large). **b**, We constructed networks of influence among lawmakers in the US Congress based on records of bill co-sponsorship[25] (Supplementary Information section 7). The Democrats (blue) in the House and Senate exhibit consistent positive influence assortment; whereas the Republicans (red) have historically been less assorted, suffering from a negative influence gap. Starting in 1994, Republican influence assortment increased considerably and has since remained nearly as high as among Democrats, consistent with increased polarization[16] in Congress since the mid-1990s. **c**, We analysed bill co-sponsorship in the upper (UC; light green) and lower (LC; dark green) chambers of the Czech, French, Italian and Romanian legislatures during their most recent sessions[26]. We observe significant influence gaps in all except the Czech and French upper chambers (Supplementary Table 5). We analysed datasets of online political discussion[27–29] (black) during the 2004, 2010 and 2016 US elections. We observe a significant Republican-leaning influence gap in the 2004 blog network and in the 2010 Twitter-mention network and a significant Democratic-leaning influence gap in the 2016 news intake network (Supplementary Information section 7).

of the vote—that is, an undemocratic outcome. However, other parties are then incentivized to increase their own influence assortment, which leaves everyone trapped in deadlock.

This dilemma is reminiscent of the two-player Prisoner's Dilemma. Although mutual defection is the only Nash equilibrium for two individuals, a large literature has established mechanisms to avoid defection in the Prisoner's Dilemma, such as reciprocity, punishment and reputation[24]. Future research must seek to resolve the group-level dilemma that arises from the presence of information gerrymandering in collective decisions.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data necessary to reproduce the results are available at https://github.com/jplotkin/InformationGerrymandering.

Code availability

All scripts necessary to reproduce the results are available at https://github.com/jplotkin/InformationGerrymandering.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; and details of author contributions and competing interests are available at https://doi.org/10.1038/s41586-019-1507-6.

Received: 12 November 2018; Accepted: 30 July 2019;
Published online 4 September 2019.

1. Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A. & Bonneau, R. Tweeting from left to right: is online political communication more than an echo chamber? *Psychol. Sci*. **26**, 1531–1542 (2015).
2. Del Vicario, M. et al. The spreading of misinformation online. *Proc. Natl Acad. Sci. USA* **113**, 554–559 (2016).
3. Bakshy, E., Messing, S. & Adamic, L. A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**, 1130–1132 (2015).
4. Woolley, S. C. Automating power: social bot interference in global politics. *First Monday* **21**, 4 (2016).
5. Lazer, D. M. J. et al. The science of fake news. *Science* **359**, 1094–1096 (2018).
6. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
7. Matz, S. C., Kosinski, M., Nave, G. & Stillwell, D. J. Psychological targeting as an effective approach to digital mass persuasion. *Proc. Natl Acad. Sci. USA* **114**, 12714–12719 (2017).
8. Bond, R. M. et al. A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–298 (2012).
9. Pennycook, G. & Rand, D. G. Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* **188**, 39–50 (2019).
10. Iyengar, S., Sood, G. & Lelkes, Y. Affect, not ideology: a social identity perspective on polarization. *Public Opin. Q*. **76**, 405–431 (2012).
11. Crockett, M. J. Moral outrage in the digital age. *Nat. Hum. Behav*. **1**, 769–771 (2017).
12. Kearns, M., Judd, S., Tan, J. & Wortman, J. Behavioral experiments on biased voting in networks. *Proc. Natl Acad. Sci. USA* **106**, 1347–1352 (2009).
13. Couzin, I. D. et al. Uninformed individuals promote democratic consensus in animal groups. *Science* **334**, 1578–1580 (2011).
14. Newport, F. Americans favor compromise to get things done in Washington. *Gallup* https://news.gallup.com/poll/220265/americans-favor-compromise-things-done-washington.aspx (2017).
15. Stephanopoulos, N. O. & McGhee, E. Partisan gerrymandering and the efficiency gap. *Univ. Chic. Law Rev*. **82**, 831–900 (2015).
16. Lewis, J. B. & Poole, K. T. Measuring bias and uncertainty in ideal point estimates via the parametric bootstrap. *Polit. Anal*. **12**, 105–127 (2004).
17. Guilbeault, D., Becker, J. & Centola, D. Social learning and partisan bias in the interpretation of climate trends. *Proc. Natl Acad. Sci. USA* **115**, 9714–9719 (2018).
18. Gavrilets, S. & Richerson, P. J. Collective action and the evolution of social norm internalization. *Proc. Natl Acad. Sci. USA* **114**, 6068–6073 (2017).
19. Bail, C. A. et al. Exposure to opposing views on social media can increase political polarization. *Proc. Natl Acad. Sci. USA* **115**, 9216–9221 (2018).
20. Weber, E. U. & Stern, P. C. Public understanding of climate change in the United States. *Am. Psychol*. **66**, 315–328 (2011).
21. Pennycook, G. & Rand, D. G. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc. Natl Acad. Sci. USA* **116**, 2521–2526 (2019).
22. Shirado, H. & Christakis, N. A. Locally noisy autonomous agents improve global human coordination in network experiments. *Nature* **545**, 370–374 (2017).
23. Hardin, G. The tragedy of the commons. The population problem has no technical solution; it requires a fundamental extension in morality. *Science* **162**, 1243–1248 (1968).
24. Rand, D. G. & Nowak, M. A. Human cooperation. *Trends Cogn. Sci*. **17**, 413–425 (2013).
25. Fowler, J. H. Legislative cosponsorship networks in the US House and Senate. *Soc. Networks* **28**, 454–465 (2006).
26. Briatte, F. Network patterns of legislative collaboration in twenty parliaments. *Netw. Sci*. **4**, 266–271 (2016).
27. Adamic, L. A. & Glance, N. The political blogosphere and the 2004 U.S. election: divided they blog. In *Proc. 3rd International Workshop on Link Discovery* 36–43 (ACM, 2005).
28. Conover, M. et al. Political polarization on Twitter. In *ICWSM* 89–96 (2011).
29. Faris, R. et al. *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*. Berkman Klein Center Research Publication 2017-6 https://ssrn.com/abstract=3019414 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements We acknowledge funding from the Defense Advanced Research Projects Agency NGS2 program (grant D17AC00005; to A.J.S., J.B.P., M.M., A.A.A. and D.G.R.), the Ethics and Governance of Artificial Intelligence Initiative of the Miami Foundation (to D.G.R.), the Templeton World Charity Foundation and the John Templeton Foundation (to D.G.R.), the Army Research Office (grant W911NF-17-1-0083; to J.B.P.) and the David & Lucile Packard Foundation (to J.B.P.). The views, opinions, and findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the US Government.

Author contributions A.J.S., D.G.R., M.D. and J.B.P. conceived the project. A.J.S., D.G.R., M.D. and J.B.P. developed the model. A.J.S. ran the simulations with input from J.B.P. and D.G.R. A.J.S., M.M., D.G.R. and J.B.P. designed the experiments. M.M. and A.A.A. ran the experiments. A.J.S., M.M. and J.B.P. analysed the experimental data with input from D.G.R. and A.A.A. A.J.S. analysed the empirical networks with input from D.G.R. and J.B.P. A.J.S. and J.B.P. wrote the paper with input from D.G.R., M.M., A.A.A. and M.D.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1507-6.

Correspondence and requests for materials should be addressed to A.J.S.

Peer review information *Nature* thanks Carl T. Bergstrom, Wei Chen and Brian Uzzi for their contribution to the peer review of this work.

nature research

Corresponding author(s): Alexander Stewart

Last updated by author(s): 2019-7-3

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Breadboard (https://github.com/human-nature-lab/breadboard) was used to code Mturk experiments. All other data was from published literature
Data analysis	Matlab and R were used for data analysis. All analysis scripts are provided.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Experimental data is included with the Supplementary Information
Publicly available datasets are available at the following locations:
Senate and House co-sponsorship data - http://jhfowler.ucsd.edu/cosponsorship.htm
European legislatures bill co-sponsorship data - https://f.briatte.org/parlviz/
Political blog data - available as supplement to Adamic, L. A. & Glance, N. The political blogosphere and the 2004 us election: divided they blog. In Proceedings of the 3rd international workshop on Link discovery, 36–43 (ACM, 2005)
Political Twitter data - available as supplement to Conover, M.et al. Political polarization on twitter. In ICWSM
Political news consumption - Available as supplement to Faris, R.et al.Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election. Berkman Klein Center Research Publication 2017-6

Figure 3 and Figure 4 are associated with raw data.
Simulation data and simulation code are available as supplements to this paper

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We study a "Voter Coordination Game" in which a pool of N=24 players continuously update their voting intention from which we generate "polls" leading up to a final "vote", in which they must choose between two "parties" (purple, P, and yellow, Y). Each voter is assigned an intrinsic preference for one or the other party, such that they receive a higher payoff if that party "wins" the vote by receiving a "supermajority" of V votes, relative to if the other party wins the supermajority. If neither party receives a supermajority, however, no players receive any payoff. Thus players have two potentially conflicting shared goals - achieving their individual preference (having their party achieve a supermajority) and reaching successful coordination across the population (having either party achieve a supermajority)
Research sample	We recruited at a total of 2,520 unique subjects from Amazon Mechanical Turk, over the course of 120 experimental sessions in total. Once joining a session, subjects were randomly assigned to one of the two teams (purple or yellow), and were asked to take a tutorial, followed by four multiple-choice questions about the game rules and the payment structure (the questions were the same across conditions, to avoid potential selection issues). Those who could not provide a correct answer to all questions were not allowed to continue to the game.
Sampling strategy	Sample sizes were calculated based on pilot experiments. Monte Carlo simulations were performed using inferred strategies and number of replicates were determined based on our simulated power to detect differences in vote outcome. Once data collection began we increased our sample size and updated our pre-registration (included as supplement) to account for lower than expected rates of vote consensus in one condition.
Data collection	Participants were recruited from Amazon Mechanical Turk (MTurk), paying $1 show-up with the opportunity to earn additional bonus based on the outcome of the game. MTurk workers with MTurk approval ratings of greater than 85% were eligible.
Timing	Data was collected between March 13th 2018 and March 22nd 2018
Data exclusions	No data was excluded
Non-participation	No participants dropped out
Randomization	Each participant was randomly assigned to a "party" (pruple or yellow) and to a position on the voting network

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above

Recruitment	Participants were recruited from Amazon Mechanical Turk (MTurk), paying $1 show-up with the opportunity to earn additional bonus based on the outcome of the game. MTurk workers with MTurk approval ratings of greater than 85% were eligible.
Ethics oversight	The human-subject research was approved by the Yale University Committee of the Use of Human Subjects IRB protocol #1307012383

Note that full information on the approval of the study protocol must also be provided in the manuscript.