Science AAAS

www.sciencemag.org/content/360/6393/1116/suppl/DC1

Supplementary Materials for

Experimental evidence for tipping points in social convention

Damon Centola*, Joshua Becker, Devon Brackbill, Andrea Baronchelli

*Corresponding author. Email: dcentola@asc.upenn.edu

Published 8 June 2018, *Science* **360**, 1116 (2018)
DOI: 10.1126/aas8827

This PDF file includes:

Materials and Methods

Experimental Design

Subjects were randomly assigned to a fully-connected network (i.e., homogeneously mixing population) containing a pre-determined number of experimental participants and confederates. The number of confederates as a fraction of the total population was varied between conditions. Our theoretical model predicts a tipping point with approximately 25% of the population. To test this effect, we ran 5 trials with fewer than 25% confederates ("below threshold") and five trials with greater than 25% confederates ("above threshold"). Because users had no way of distinguishing confederates from experimental participants, the user experience in below-threshold conditions was identical to the user experience in above-threshold conditions.

Experimental trials consisted of two phases. Once a trial began, the first phase of the experimental procedure (see "Subject Experience During the Experiment" below) was allowed subjects to interact until they established a shared social convention. This phase used the procedures from previous research on the emergence of conventions *(18)*. During this phase, confederate participants entered no response, except during trials 3 and 4, in which they played a single pre-determined strategy during phase one to speed up initial convergence. Phase one was considered complete when every experimental participant began using the same strategy, with at most two players deviating. This determination was based on previous experimental methods for identifying convergence on social conventions *(18)*, which indicated that even once a stable convention is established, not all players will always use the conventional strategy.

Once an endogenous convention was reached, the confederate participants simultaneously began entering a pre-determined response that differed from the established convention. All the confederates used the same response. In order to ensure that the alternative response was comparable to the established convention, we used responses selected from names that commonly emerged as conventions in previous studies (e.g., "Mary"). Confederate participants continued entering this alternative strategy until the game was complete.

The length of an entire experimental trial was determined prior to play beginning. Due to stochastic variation in the length of the first experimental phase, consistent with previous research on the emergence of conventions *(18)*, the number of interactions available for Phase 2 varied between trials. The number of interactions also varied because subjects entered responses at different rates, and the trials were run until each player participated in a minimum number of interactions.

Subject Recruitment

Participants in our study were recruited via the World Wide Web to be players in online games for research purposes. Players registered by providing their email address and selecting a username and an avatar. All players were required to provide informed consent in order to register. Once registered, users were placed into a recruitment pool for future experiments. Registered users were then sent an advertisement and a link to participate in a trial for this study. Trials were run over a 105 day period between July 15, 2015, and October 7, 2015.

Subject Experience During the Experiment

Upon arriving at the study website, participants viewed instructions on how to play the game (Figure S1). In the game, participants play a series of one-shot coordination games, as shown in Figure S2. The left hand column shows the other players in the game. However, they have no information about which player ("partner") they are matched with for a given interaction.

Participants cannot identify whether they are matched with another experimental participant or a confederate. They cannot identify if their partner for a given interaction is the same as their partner in a previous interaction.

For each one-shot game, participants are shown an image (a picture of a man or a woman) and prompted to enter a name in the response field. There is one picture per trial—i.e., the picture does not change across interactions and remains the same for all the interactions within the same experimental trial. Participants are instructed that if they enter the same name as their partner, they will be rewarded \$0.10. If they do not enter the same name as their partner, \$0.10 is deducted from their total winnings, with a minimum possible reward of \$0.00. On the right hand side, participants are shown their progress in the game. Completed interactions indicate whether the interaction was a match. After each interaction, participants are shown the name entered by their partner, regardless of whether or not it was a match. Centola & Baronchelli *(18)* showed that by this process, subjects will successfully establish a shared convention; i.e., this process leads subjects to converge on a single shared name for the person depicted in the image.

Supplementary Text

<u>Model Definition</u>

Our theoretical model follows previous theoretical models of critical mass in studying asynchronous pairwise interaction *(14,25)*. However, while these prior models assume that agents randomly select from previously observed strategies, we follow game theoretic models of convention *(9)* in modeling strategic choice in which individuals attempt to choose the behavior most likely to generate successful coordination.

In each time step, two agents from a population of N agents are randomly selected to interact. One agent is randomly selected to be the "speaker" and the other agent is assigned the role of "hearer." The agent playing the role of speaker picks a best response strategy. The best response strategy is defined as the strategy most frequently observed in previous interactions in which that agent was the hearer. An agent's "memory" stores a record of the strategies observed in use by other players, and an agent only updates their memory during interactions in which they are the hearer. Agents do not respond to a complete history of past plays; rather, we assume that agents determine their best response strategy based only on the past M interactions. The agent decision rule is therefore defined by the single parameter M that determines the size of an agent's memory. This limit reflects both the assumption that agents have limited cognitive resources and also the assumption that recent interactions are more informative of population behavior *(9)*.

The formalization of memory as a sliding window within a vector of past plays was chosen in order both to be consistent with previous theoretical research in evolutionary game theory *(9)* and also to maintain a parsimonious model with minimal degrees of freedom. Other possible formalizations for memory include a continuous decay model. This model was not selected for this study because it would involve more degrees of freedom to specify a decay function and associated parameters. As a simple approximation, we find that a sliding window formalization correctly predicts approximately 80% of user plays. Thus far, we have not found that any alternative model provides a better fit with the empirical data.

Following a standard procedure from the literature *(14)*, we model a scenario in which a population has already converged on some convention, so that every agent uses some previously established strategy 'A', such that A is the only option contained in each agent's memory. The simulation is therefore initialized so that every non-committed agent has a memory that contains only A (i.e., each non-committed agent begins the simulation with a memory as if they had observed A for the previous M interactions). Our simulation studies the dynamics that occur when some fraction of that population begins to use an alternative strategy 'B' instead of following best-response dynamics. This committed minority always uses a single fixed strategy B when they are selected as speaker instead of using a best response strategy. Thus, we simulate the effect of a committed minority using strategy B in a population with an established convention A. With the exception of the decision heuristic used by non-committed agents, our model is identical to that studied by Xie et al. *(14)*, which is itself identical to Baronchelli et al. *(25)*. The best-response decision-rule for non-committed agents is identical to that studied by Young *(9)*.

We consider a population of N agents. Some fraction C of the total N agents are identified as committed agents who always play B. At time T=0, the agents playing best-response dynamics (i.e., non-committed agents) are initialized with a memory vector of length M, the entries of which are all A. The model is fully defined by the following parameters:

- N: the number of agents
- C: the fraction of the population belonging to the committed group
- M: the number of past interactions used in agent decisions

This model defines a Markov chain with only one absorbing state: the state in which the entire population has adopted the strategy promoted by the committed minority. However, this will only occur in infinite time when the committed minority is small *(14)*. In finite time, the model is characterized by two states. Above the tipping point, the alternative strategy promoted by the committed minority is very quickly adopted by the entire population. Below the tipping point, the model reaches a quasi-stationary state in which the initial convention is the dominant convention for very long periods of time *(14)*.

We therefore measure the tipping point in simulation by simulating T=1000*N interactions (i.e., to allow an average of 1000 interactions per agent) and then calculating the percentage of non-committed agents who have adopted the alternative strategy. Results are only minimally perturbed by larger values of T.

Estimating Memory Length

To develop our experimental hypotheses, we used data from previous experiments on the same web platform *(18)* to estimate the value for M by determining the value for which the model most accurately predicts user behavior. We then replicated this analysis for experimental data from this study, producing comparable results. Figure S3 shows the fraction of user choices which are accurately predicted by the model as a function of M. To generate this figure, each user's play was predicted for each interaction based on their M previous observations. We then plot the total fraction of interactions which were correctly predicted for each value of M. Using this parameter estimation, we predict a lower bound of M=12 for users in our experimental platform.

Tipping Point

Coordination dynamics show a tipping point similar to that observed by Xie et al. *(14)* even after accounting for strategic choice. Using the lower bound estimate M=12, we model short term and long term success of committed groups using strategy B in a population playing A, across range of committed minority sizes P and population sizes N. Figure S4 shows adoption by non-committed agents after T=45, T=100, and T=1000 "rounds" in a population of N=1,000 agents. To normalize time across variation in population size, each time step ("round") is measured as N interactions, or one interaction per agent. To generate this figure, we initialize the simulation as described above and run for N*T interactions (i.e., T rounds). We measure adoption as the usage of each strategy in the N interactions prior to measurement. We then plot the fraction of non-committed players adopting the alternative strategy after T interactions for each value of P.

Long term adoption at T=1000 rounds shows a sharp transition when committed groups reach approximately 24.2% of the population. Below this threshold, only a small number of non-committed agents use strategy B in the N interactions prior to measurement. Above this threshold, the population reaches an absorbing state in which all agents are consistently using strategy B.

For extremely short term dynamics, committed minorities above the tipping point do not always achieve full convergence on the alternative strategy. For N=1,000, for example, after 20 rounds (Fig S4, light blue line), full convergence is not reached even with P=50%. After 45 rounds (the average length of experimental trials), most above-threshold groups are successful, while committed groups between 25% and 30% are not yet guaranteed to reach convergence. After 100 rounds, nearly every above-threshold group has achieved widespread adoption.

Robustness to Population Size

The tipping point is robust to variation in population size. To generate figure 1B inset (main text) we run the same analysis shown in Figure S4, but with varying population size. For each population size and each value of C, we measure the percentage of simulations that reach complete convergence – i.e., the percentage of outcomes in which 100% of the non-committed population has adopted the alternative strategy after 1,000 interactions per person.

The grey area in Figure 1B inset (main text) shows values of C and N for which the committed minority is successful (i.e., they achieve full adoption of the alternative strategy) in greater than 1% of simulations, but in fewer than 99% of simulations. That is, for values of C and N that fall below this area, the minority group never succeeds in changing the social convention, while for values of C that fall above this area the minority group is successful more than 99% of the time.

For small population sizes, the tipping point has the appearance of non-monotonicity with N due to the fact that not all fractions can be converted into a discrete critical mass size: for example, when N=20, a committed group can comprise either 20% (4/20) or 25% (5/20) of the population. Thus, if the tipping point is between 20% and 25%, then it will take a minimum of 25% (5/20) committed individuals to overturn an established norm. For population sizes from 1000 to 100,000 the tipping point stabilizes at a value of 24.2%.

Effect of Memory Parameter

The existence of a tipping point is robust to variation in agent memory length, and a tipping point in long term adoption appears for all values of M. Figure 1B (main text) and Figure S5 show the tipping point as a function of M. To generate these figures, we identify the lowest value for P in which at least 99% simulations reach convergence (i.e., achieve full adoption of the alternative strategy) after N*1000 interactions. As shown in Fig S3 a group larger than 10% required even if agents choose their strategy only based on the previous 4 interactions. The tipping point remains below 50% even for very large values of M, with a critical mass of only 40% required when M=100.

Robustness to Network Density

Our experimental platform studies critical mass dynamics in a homogeneously mixing (fully connected) network, so that any two agents are equally likely to interact. Dynamics are qualitatively similar with sparse random networks, as shown in figure S6. To generate this figure, we generated random networks in which each node has an equal number of connections *(38)*. In sparse networks, the model is defined identically, but instead of uniformly selecting two agents for interaction, the model randomly selects an edge from the network for interaction. One node is then selected as speaker, and one node is then selected as hearer.

Consistent with findings in *(14)*, the tipping point is slightly lower in sparse networks. To generate Figure S6, we run analyses as shown in Figure 4 at varying network densities. At each point, we determine the smallest value for P at which 100% adoption of the alternative strategy is achieved in at least 99% of simulated outcomes. We hold network size constant, and therefore network density is determined by average degree (number of network neighbors) for each node.

<u>Robustness to Agent Strategy Preference</u>

Our experimental design and theoretical model both reflect coordination dynamics in which two agents must decide between two possible strategies, both of which are equally preferred by every non-committed individual in the population. To model the possibility that agents may prefer one strategy over another strategy, we adopt the standard game theoretic formalization of coordination games in which each strategy is assigned a numeric payoff. In our base model, both strategies would be assigned the same payoff since they are equally preferred.

To choose a strategy based on numeric payoff, agents select the strategy with the greatest "expected payoff," which is calculated as the probability of success multiplied by the numeric payoff of a successful interaction. In our computational model of coordination, probability of success is determined by the percentage of recent interactions (where recent is defined by memory length M) in which a particular strategy was observed. For example, if strategy A provides a payoff of 1 and was observed in 60% of the previous M interactions, and strategy B provides a payoff of 2 and was observed in 40% of the previous M actions, then the payoff for each strategy is calculated as follows:

$$\text{payoff of strategy B} = 0.6 \times 1 = 0.6$$

$$\text{payoff of strategy A} = 0.4 \times 2 = 0.8$$

Thus, strategy A offers a higher expected payoff despite being less frequently observed and offering a lower probability of success.

We use this model of agent preference to test whether we observe critical mass dynamics in situations where agents have a bias towards either the entrenched convention or the alternative strategy. Figure S7 shows the effect of agent preference on adoption dynamics as a function of the relative preference for the established convention, which is measured as the ratio between the payoff for the established convention and the alternative strategy. This figure indicates that even when agents prefer the established convention twice as much as the alternative strategy (i.e., the payoff ratio between the established convention and the alternative strategy equals 2:1) a committed minority can still establish critical mass and achieve widespread adoption of the alternative strategy.

Data Availability

The complete dataset is publicly available for download from the following URL: http://ndg.asc.upenn.edu/experiments/creating-critical-mass/

This dataset contains a time series for each trial starting at phase 2 as described in Materials and Methods. Each row indicates the percentage of responses which are the established convention, the percentage of responses which are the alternative strategy, and the percentage of responses which are any other strategy.

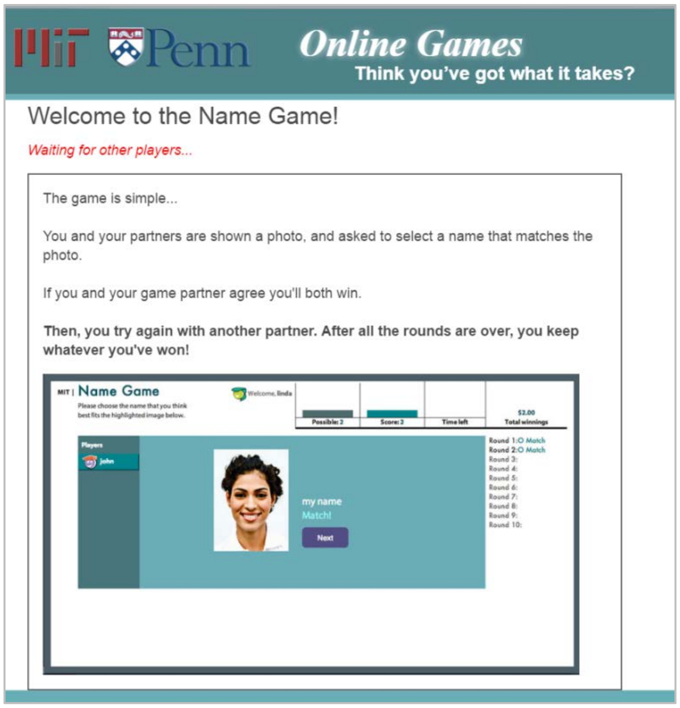

Figure S1. Screenshot of the waiting page showing instructions.

Figure S2. Screenshot of game interface. Note that what is called a "round" in the user interface is a single "interaction" as discussed throughout this text.

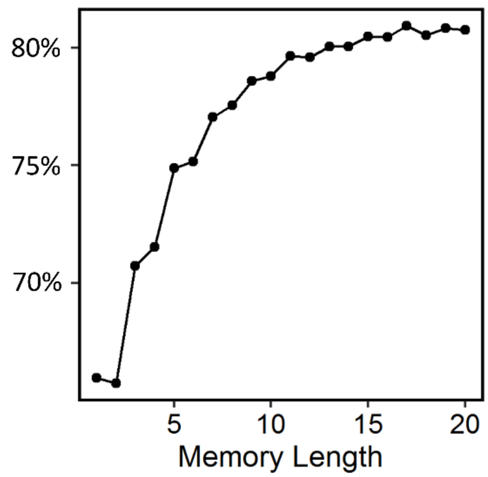

Figure S3. Estimating Memory Length from Empirical Data. Each panel shows the fraction of interactions successfully predicted in the data from the current experiment. When memory length is greater than 10, our model correctly predicts 80% of the choices by our experimental subjects.

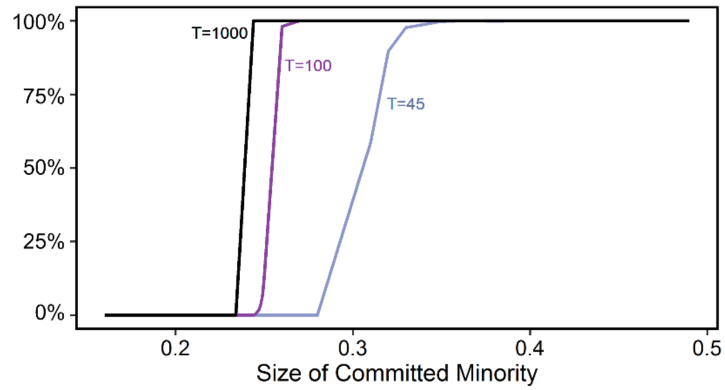

Figure S4. Estimating the Tipping Point. Each line shows the proportion of simulations in which every agent adopted the alternative strategy after T interactions per agent for a network of N=1,000 agents as a function of P. When T=1000, there is a sharp threshold between P=0.241 and P=0.242 indicating that a small change in the size of a committed minority can generate a dramatic shift in the adoption of an alternative convention. It is worth noting that the threshold is well defined only over long time-scales. For shorter time periods, committed minorities that are sufficient in long term dynamics may have not yet achieved convergence.

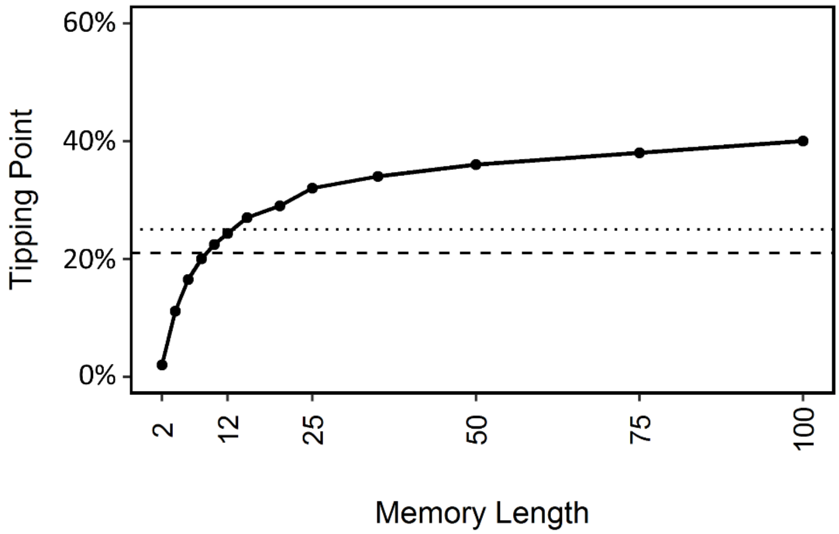

Figure S5. Critical mass dynamics are robust to variation in agent memory length. Each point in this figure shows the tipping point as a function of agent memory length (M). Even when M=100, the tipping point is well below 50%. Horizontal lines indicate the largest value for P which failed in experimental trials (P=21%, dashed line) and the smallest value for P which succeeded in experimental trials (P=25%, dotted line) suggesting that M is between 9% and 13%.

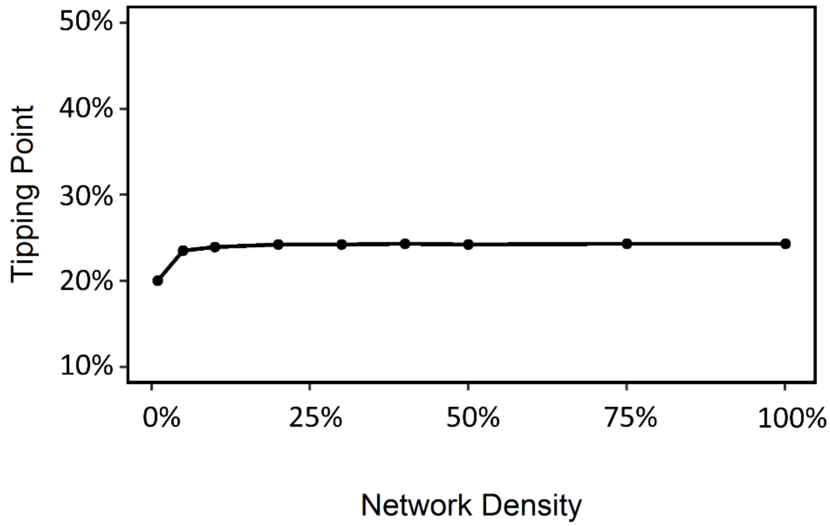

Figure S6. The tipping point is robust to changes in network density. Each point in this figure shows the tipping point as a function of network density for simulations with M=12, T=1000*N, N=1000. In very sparse networks where agents only have a few network neighbors, the tipping point drops slightly but remains above 20%. The point furthest to the left shows simulations for populations where agents have 10 network neighbors, producing a network density of 1%. Network density is defined to be the number of connected edges in a network as a percent of all possible edges that could be connected.

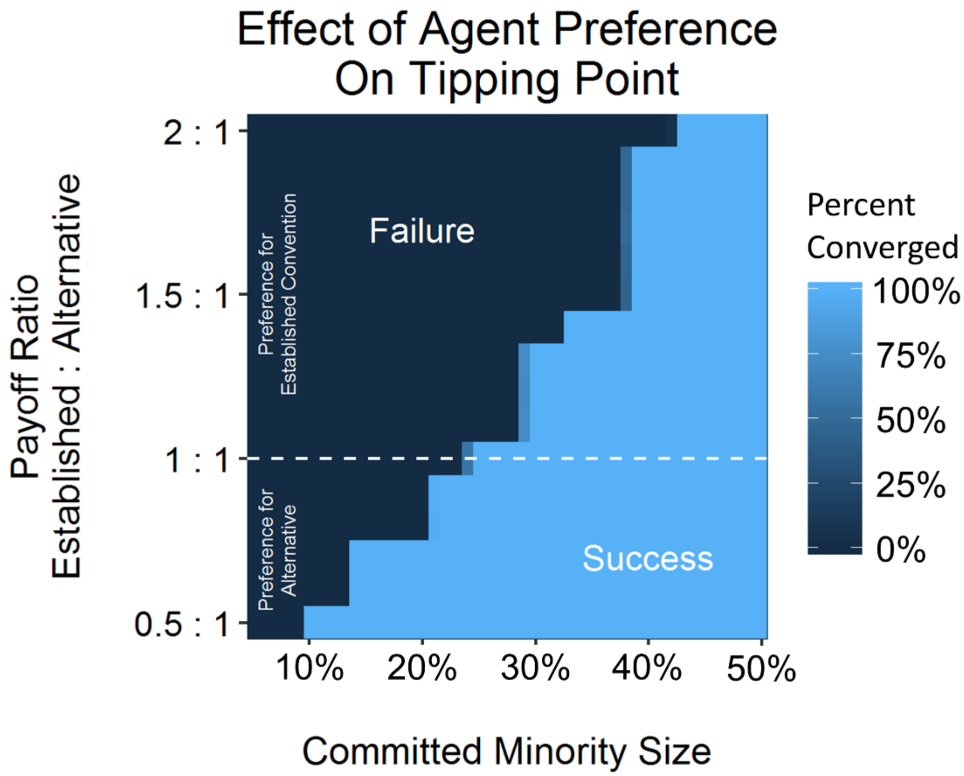

Figure S7. Critical mass threshold as a function of relative preference. This figure shows the proportion of simulations in which the alternative norm is adopted, as a function of minority size and relative preference for each convention for simulations with M=12, N=1000, T=1000*N. The Y axis of this figure indicates the relative payoff of the established convention as compared with the alternative strategy. When the payoff ratio is equal to 1, both strategies are equally desirable (i.e., agents simply adopt the most popularly used strategy) and this model is equivalent to our general model of conventions, showing a tipping point of approximately 25%. When the payoff ratio is equal to 2, agents prefer the established convention twice as much as the alternative strategy, but a committed minority can nonetheless overturn the established convention. When the payoff ratio is equal to 0.5, agents prefer the established convention half as much as the alternative strategy (i.e., they prefer the alternative strategy twice as much as the established convention) and the critical mass can be reached with a very small committed minority comprising less than 10% of the population.

References

1. T. Kuran, The inevitability of future revolutionary surprises. *Am. J. Sociol.* **100**, 1528–1551 (1995). doi:10.1086/230671

2. K. D. Opp, C. Gern, Dissident groups, personal networks, and spontaneous cooperation: The East German revolution of 1989. *Am. Sociol. Rev.* **58**, 659–680 (1993). doi:10.2307/2096280

3. R. M. Kanter, Some effects of proportions on group life: Skewed sex ratios and responses to token women. *Am. J. Sociol.* **82**, 965–990 (1977). doi:10.1086/226425

4. D. Dahlerup, L. Freidenvall, Quotas as a 'fast track'to equal representation for women: Why Scandinavia is no longer the model. *Int. Fem. J. Polit.* **7**, 26–48 (2005). doi:10.1080/1461674042000324673

5. S. Bikhchandani, D. Hirshleifer, I. Welch, A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* **100**, 992–1026 (1992). doi:10.1086/261849

6. T. Kuran, *Private Truths, Public Lies: The Social Consequences of Preference Falsification* (Harvard Univ. Press, 1997).

7. J. C. Harsanyi, R. Selten, *A General Theory of Equilibrium Selection in Games* (MIT Press, Cambridge, 1988).

8. J. F. Nash, Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. U.S.A.* **36**, 48–49 (1950). doi:10.1073/pnas.36.1.48 Medline

9. H. P. Young, The evolution of conventions. *Econometrica J. Econom. Soc.* **61**, 57–84 (1993). doi:10.2307/2951778

10. M. Kandori, G. J. Mailath, R. Rob, Learning, mutation, and long run equilibria in games. *Econometrica J. Econom. Soc.* **61**, 29–56 (1993). doi:10.2307/2951777

11. G. Ellison, Learning, local interaction, and coordination. *Econometrica J. Econom. Soc.* **61**, 1047–1071 (1993). doi:10.2307/2951493

12. T. Schelling, *Micromotives and Macrobehavior* (Norton, New York, 1978).

13. M. Granovetter, Threshold models of collective behavior. *Am. J. Sociol.* **83**, 1420–1443 (1978). doi:10.1086/226707

14. J. Xie, S. Sreenivasan, G. Korniss, W. Zhang, C. Lim, B. K. Szymanski, Social consensus through the influence of committed minorities. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **84**, 011130 (2011). doi:10.1103/PhysRevE.84.011130 Medline

15. D. M. Centola, Homophily, networks, and critical mass: Solving the start-up problem in large group collective action. *Ration. Soc.* **25**, 3–40 (2013). doi:10.1177/1043463112473734

16. G. Marwell, P. Oliver, *The Critical Mass in Collective Action* (Cambridge Univ. Press, 1993).

17. K. Nyborg, J. M. Anderies, A. Dannenberg, T. Lindahl, C. Schill, M. Schlüter, W. N. Adger, K. J. Arrow, S. Barrett, S. Carpenter, F. S. Chapin 3rd, A.-S. Crépin, G. Daily, P. Ehrlich, C. Folke, W. Jager, N. Kautsky, S. A. Levin, O. J. Madsen, S. Polasky, M. Scheffer, B.

Walker, E. U. Weber, J. Wilen, A. Xepapadeas, A. de Zeeuw, Social norms as solutions. *Science* **354**, 42–43 (2016). doi:10.1126/science.aaf8317 Medline

18. D. Centola, A. Baronchelli, The spontaneous emergence of conventions: An experimental study of cultural evolution. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1989–1994 (2015). doi:10.1073/pnas.1418838112 Medline
19. D. Lewis, *Convention: A Philosophical Study* (Harvard Univ. Press, Cambridge, 1969).
20. C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics. *Rev. Mod. Phys.* **81**, 591–646 (2009). doi:10.1103/RevModPhys.81.591
21. S. Garrod, G. Doherty, Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition* **53**, 181–215 (1994). doi:10.1016/0010-0277(94)90048-5 Medline
22. L. Wittgenstein, *Philosophical Investigations* (John Wiley & Sons, 2009).
23. S. Grey, Numbers and beyond: The relevance of critical mass in gender research. *Polit. Gend.* **2**, 492–502 (2006). doi:10.1017/S1743923X06221147
24. Materials and methods are available as supplementary materials.
25. A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, L. Steels, Sharp transition towards shared vocabularies in multi-agent systems. *J. Stat. Mech. Theory Exp.* **2006**, P06014 (2006). doi:10.1088/1742-5468/2006/06/P06014
26. L. Backstrom, D. Huttenlocher, J. Kleinberg, X. Lan, Group formation in large social networks: membership, growth, and evolution. *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 44–54 (2006).
27. F. Kooti, H. Yang, M. Cha, P. K. Gummadi, W. A. Mason, The Emergence of Conventions in Online Social Networks, *Proc. Sixth Int. AAAI Conf. Weblogs Soc. Media ICWSM 2012* (2012).
28. D. Dahlerup, From a small to a large minority: Women in Scandinavian politics. *Scand. Polit. Stud.* **11**, 275–298 (1988). doi:10.1111/j.1467-9477.1988.tb00372.x
29. C. Tilly, S. Tarrow, *Contentious Politics* (Oxford Univ. Press, New York, ed. 2nd, 2015).
30. G. King, J. Pan, M. E. Roberts, How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. *Am. Polit. Sci. Rev.* **111**, 484–501 (2017). doi:10.1017/S0003055417000144
31. G. King, J. Pan, M. E. Roberts, Reverse-engineering censorship in China: Randomized experimentation and participant observation. *Science* **345**, 1251722 (2014). doi:10.1126/science.1251722 Medline
32. A. Antoci, A. Delfino, F. Paglieri, F. Panebianco, F. Sabatini, Civility vs. Incivility in Online Social Interactions: An Evolutionary Approach. *PLOS ONE* **11**, e0164286 (2016). doi:10.1371/journal.pone.0164286 Medline
33. E. L. Paluck, H. Shepherd, P. M. Aronow, Changing climates of conflict: A social network experiment in 56 schools. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 566–571 (2016). doi:10.1073/pnas.1514483113 Medline

34. C. Shih, *The Facebook Era: Tapping Online Social Networks to Market, Sell, and Innovate* (Pearson Education, 2010).

35. F. Polletta, *Freedom is An Endless Meeting: Democracy in American Social Movements* (Univ. of Chicago Press, 2002).

36. R. Dodge, *Andrea and Sylvester: Challenging Marriage Taboos and Paving the Road to Same-Sex Marriage* (Algora Publishing, 2015).

37. L. Trotsky, *1905* (Haymarket Books, 2017).

38. S. Maslov, K. Sneppen, U. Alon, S. Bornholdt, H. G. Schuster, *Handbook of Graphs and Networks* (Wiley-VCH New York, 2003).