Encouraging Moderation: Clues from a Simple Model of Ideological Conflict

Seth A. Marvel, Hyunsuk Hong, Anna Papush, and Steven H. Strogatz

Phys. Rev. Lett. **109**, 118702 — Published 11 September 2012

DOI: 10.1103/PhysRevLett.109.118702

Encouraging moderation: Clues from a simple model of ideological conflict

Seth A. Marvel,[1,*] Hyunsuk Hong,[2] Anna Papush,[3] and Steven H. Strogatz[3]

[1]*Department of Mathematics, University of Michigan, Michigan 48109, USA*
[2]*Department of Physics and Research Institute of Physics and Chemistry, Chonbuk National University, Jeonju 561-756, Korea*
[3]*Department of Mathematics, Cornell University, New York 14853, USA*

(Dated: August 22, 2012)

Some of the most pivotal moments in intellectual history occur when a new ideology sweeps through a society, supplanting an established system of beliefs in a rapid revolution of thought. Yet in many cases the new ideology is as extreme as the old. Why is it then that moderate positions so rarely prevail? Here, in the context of a simple model of opinion spreading, we test seven plausible strategies for deradicalizing a society and find that only one of them significantly expands the moderate subpopulation without risking its extinction in the process.

PACS numbers: 05.45.-a, 89.65.-s

The social history of ideas involves the frequent replay of a single story: there is a widely accepted and deeply ingrained dogma in the community. This dogma helps to justify the community's institutions and shape its common practices. Then, in the midst of this stable milieu, a new doctrine emerges. Backed by a small group of unwavering advocates, it challenges the status quo and steadily wins converts, eventually replacing the previous system to become the dominant ideology of the group.

In some cases, there is an enduring consensus that the new doctrine marks a tangible improvement on the old. This is the case for the American civil rights movement [1], women's suffrage [2, 3], and paradigm shifts in science [4–6]. However in many other situations, the newer doctrine is not clearly better. After some time as the dominant approach, it too is overtaken by a younger alternative, which in turn is itself replaced, and so on. This second situation is often seen in rapidly spreading political campaigns [7], the booms and busts of credit lending and consumer confidence [8, 9], cultural fashions and short-lived reforms (e.g. Prohibition in the United States) [10, 11], methodological or topical fads in academia, and various political revolutions [12].

A natural question is: why do communities, and sometimes entire societies, get caught in these swings from one ideological extreme to the other when neither delivers a sustainable solution? Why doesn't a majority of the population settle on an intermediate position that blends the best of the old and new?

There are several ways in which this question might be answered, but here we give one that is purely mathematical: the environment of successive ideological revolutions is not conducive to moderate-mindedness simply from a *dynamical* perspective. In particular, almost all of the intuitive ways of encouraging moderation either fail to expand the moderate subpopulation or make it vulnerable to collapse in the process of encouraging its growth.

In this Letter, we provide evidence for this claim by studying a minimal model of ideological revolution. Critically, this model only addresses large-scale ideological conversions and does not treat the many other common processes found in real communities, such as apparent conversions within the old paradigm and situations where there is no conversion at all but rather a splitting of opinions, or fragmentation.

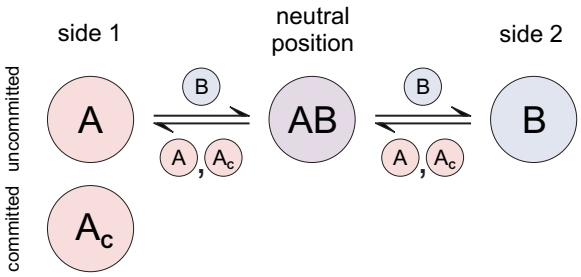

FIG. 1: Model structure (see text for definitions of A, B, AB and A_c). Labels on the arrows indicate the allowed affiliation(s) of a speaker that converts a listener from one subpopulation to another in the direction of the arrow.

The model (Fig. 1) starts from an assumption of a community consisting of four non-overlapping subpopulations: those that currently hold an extreme opinion A; those that currently hold the conceptually opposing opinion B (in the simplest case, just the negation of A); those that currently hold neither A nor B (the moderates); and those that hold A indefinitely and are immune to the influence of others (we call these committed believers or A zealots). We partially overload notation, using A, B, AB and A_c, respectively, to denote both the individuals in these four subpopulations and the subpopulations themselves. This model builds on earlier pioneering work in sociophysics [13–20] and is directly inspired by (but different from) a model examined in a recent study of opinion dynamics [21, 22].

The dynamics of the basic model are deterministic, continuous and mean-field, derived as the large-population limit of the following random process: time is discrete, and at each time step we select two individuals

TABLE I: Interactions that change the membership of subpopulations A, B and AB in the basic model. The A_c subpopulation is constant.

Speaker	Listener pre-interaction	Listener post-interaction
A, A_c	B	AB
	AB	A
B	A	AB
	AB	B

uniformly at random and randomly choose one of them to be the *speaker* and the other the *listener*. If the speaker is an A or B and the listener is a B or A, respectively, then the listener is dissuaded from his or her extremist position and becomes an AB. However if the listener is an AB, then the listener becomes an A if the speaker is an A and a B if the speaker is a B. In all other cases, there is no change in the state of the speaker and listener (Table I). Note that in this highly simplified model, moderate speakers do not produce a change of mind in either their listeners or themselves; only extremists successfully rally others to their cause.

Let n_A, n_B and n_{AB} denote the expected fractions of the total population of N individuals corresponding to the uncommitted A, B and AB subpopulations, respectively, and let p denote the constant fraction of the population in the committed A_c subpopulation. We will study how varying p, the proportion of zealots, affects the eventual state of the system. Using this new notation, we can consider the expected change to the subpopulation fractions in the limits of a large population and a vanishing time step (which we take to grow like N and shrink like $1/N$, respectively). This reduces our discrete dynamics to the rate equations

$$\begin{aligned} \dot{n}_A &= (p + n_A)n_{AB} - n_A n_B, \\ \dot{n}_B &= n_B n_{AB} - (p + n_A)n_B, \end{aligned} \tag{1}$$

where $n_{AB} = 1-p-n_A-n_B$ and the overdot denotes differentiation by time. Since we present no formal evidence that the dynamics of (1) do actually occur in practice, our work could alternatively be viewed as posing this model and its subsequent generalizations as interesting in their own right.

Now suppose we run the system (1) to equilibrium starting from a population initially composed of only A_c and B individuals. We will use this initial condition for all the systems considered in this Letter; the idea is that A represents the new doctrine and B the reigning view. If we then track the final fractions of n_A, n_B and n_{AB} as functions of p, we find (as in Ref. [21]) that the equilibrium state changes dramatically as we increase p through a critical value p_c (Fig. 2). For $p < p_c$ the system remains similar to how it started—most of the individuals maintain B. However as p is increased through p_c, the system

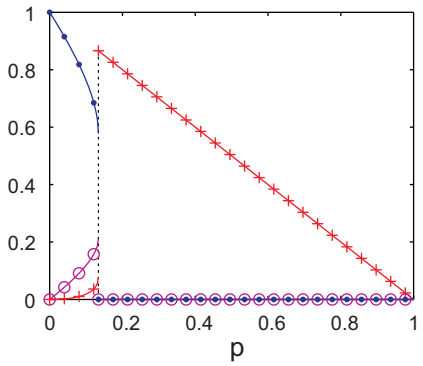

FIG. 2: (color online) The equilibrium values of n_A (red plus signs), n_B (blue dots) and n_{AB} (magenta open circles) for the basic model as functions of p, assuming an initial population with $(n_A, n_B) = (0, 1-p)$. The vertical dashed line marks the critical value $p_c = 1 - \sqrt{3}/2 \approx 0.134$. At values of p greater than p_c, n_B and n_{AB} are zero and $n_A = 1 - p$.

undergoes a discontinuous transition, and for $p > p_c$ the entire population quickly reaches a consensus on A. A bifurcation analysis shows that $p_c = 1-\sqrt{3}/2 \approx 0.134$ [23].

To test the robustness of these mean-field predictions, we simulate the model on a diverse set of real social networks. Figure 3 shows that in each case, the n_B vs. p curves resemble the mean-field result depicted in Fig. 2. The primary differences are a lower p_c value for the real networks and a small, stable fraction of peripherally located B believers for $p > p_c$.

With (1) as our starting point, we now ask how we might alter the model to encourage moderation. Specifically, we would like to (i) increase the equilibrium size of the moderate subpopulation, and (ii) decrease the chance that this equilibrium size could drop substantially if the parameter values (just p for the basic model) were to vary a little. In search of a strategy that does both, we explore seven different generalizations of the basic model. Three generalizations are discussed here in the main text; the rest are treated in the Supplemental Ma-

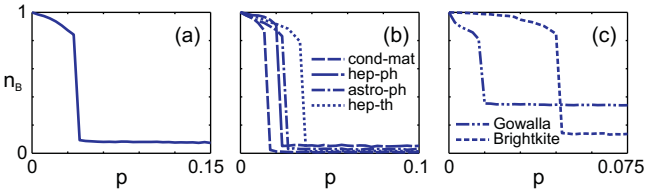

FIG. 3: The equilibrium fraction of B believers remaining after the basic model is run on the giant connected component of (a) the U.S. network of corporate board memberships in 1994 [24], (b) four coauthorship networks in the physics division of arXiv.org [25], and (c) the friendship networks of the location-based social networking websites Gowalla and Brightkite [26]. Note the abrupt transitions in n_B; compare with the corresponding curve in Fig. 2.

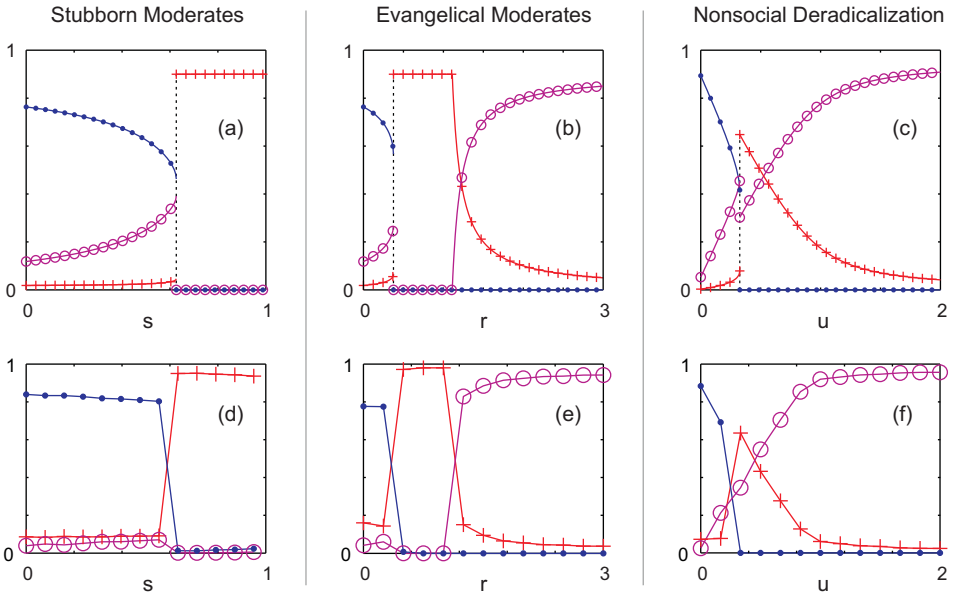

FIG. 4: (color online) (a)-(c): Mean-field results obtained analytically for generalizations of the basic model with (a) stubborn moderates, (b) evangelical moderates and (c) nonsocial deradicalization (see the text for details). The final equilibrium values of n_A (red plus signs), n_B (blue dots) and n_{AB} (magenta open circles) for the initial condition $(n_A, n_B) = (0, 1-p)$ are plotted as a function of the new parameter (s, r or u) in the corresponding generalized model. Of the three strategies shown—and in fact for all seven considered in the Supplemental Material [23]—only nonsocial deradicalization allows for growth of the moderate fraction up to $1-p$ without risking its extinction. (d)-(f): Representative simulation results for the discrete-time versions of the models with (d) stubborn moderates, (e) evangelical moderates and (f) nonsocial deradicalization when these models are run on the arXiv coauthorship network for high energy physics theory (hep-th) [25]. The plots show the equilibrium fractions of n_A (red plus signs), n_B (blue dots) and n_{AB} (magenta open circles) obtained. Each simulation is started from the state in which a random but highly interconnected fraction p of the population is committed to a belief in A and the rest believe B. The simulation is then run for 10^8 time steps after which the values of n_A, n_B and n_{AB} are tabulated. The constant fractions of zealots in the six panels of this figure are (a) 0.1, (b) 0.1, (c) 0.05, (d) 0.035, (e) 0.02 and (f) 0.02.

terial [23]. Figure 4(a)-(c) summarize mean-field results for these three generalizations, and for comparison, the corresponding simulation results on a real social network are shown in the panels beneath them [Fig. 4(d)-(f)]. Importantly, these do not constitute full empirical validations of the model and its generalizations (which would require dynamical data that is hard to obtain). Rather we include these simulations only as an indication of where the results of such tests might lie. Furthermore, we only consider the equilibrium values reached from the pre-revolution initial condition $(n_A, n_B) = (0, 1-p)$; as dynamical systems, the basic model and its generalizations are capable of a wider range of behaviors [23].

As a first attempt at achieving (i) and (ii) above, suppose we could somehow make the moderates less likely to convert to either of the two radical positions. We can represent this by generalizing the basic model to

$$\begin{aligned} \dot{n}_A &= (1-s)(p+n_A)n_{AB} - n_A n_B, \\ \dot{n}_B &= (1-s)n_B n_{AB} - (p+n_A)n_B, \end{aligned} \tag{2}$$

where the stubbornness parameter s indicates how likely a moderate is to remain moderate after listening to an extremist. When s is zero, we recover the basic model.

Intuitively, one might expect that increasing s should increase the size of the moderate subpopulation. Indeed, when s is small enough, the moderate subpopulation does grow slightly with increasing s [Fig. 4(a),(d)]. But remarkably, if s increases past a certain threshold, the moderates are driven to extinction; the size of their subpopulation drops to zero.

We can examine this surprising behavior in another way by calculating how s affects p_c (the critical fraction of zealots needed for the revolution to succeed). Intuition would suggest that p_c should increase with s; the more stubborn the moderates are, the more zealots are needed to persuade them and everyone else. In fact the opposite is true: p_c decreases with s, dropping monotonically from $1-\sqrt{3}/2$ at $s=0$ to zero at $s=1$ [23]. Thus, increasing the stubbornness of the moderates makes the population *more vulnerable* to takeover by the zealots.

To make sense of why p_c should decrease with s, it helps to realize that increasing s not only reduces the flow of AB individuals to opinion A but also to opinion B, thereby depleting *both* the uncommitted A and B subpopulations. With competition from B extremists over the AB subpopulation weakened as a result, it takes fewer A zealots (and hence a lower p_c) to convert the moderates to the A camp.

This explanation suggests that evangelism is an important force in the dynamics. So as a second strategy, we might try having the moderates actively promote moderation via the generalization:

$$\begin{aligned} \dot{n}_A &= (p + n_A)n_{AB} - n_A n_B - r n_A n_{AB}, \\ \dot{n}_B &= n_B n_{AB} - (p + n_A)n_B - r n_B n_{AB}, \end{aligned} \tag{3}$$

where the new parameter r is a nonnegative real number that reflects the intensity of the moderates' evangelism.

Again it may seem intuitively clear that the size of the moderate fraction should increase if the moderates start actively deradicalizing the population. For r up to unity however, the outcome is similar to that of making the moderates more stubborn. Figure 4(b),(e) shows that at a certain value of r, the size of the moderate subpopulation snaps discontinuously to zero. If the moderates' campaign of persuasion is sufficiently successful from the start—that is, if r starts and stays large enough—then the moderates do in fact maintain a large, robust equilibrium population. However if they fail to sustain this level of persuasiveness indefinitely, their evangelistic efforts can instigate their own extinction.

Finally, let us consider a third strategy: suppose that the fanatics are deradicalized by a pro-moderation media campaign or other environmental stimulus rather than through social interaction with moderates. We could then expect the dynamics to take the form:

$$\begin{aligned} \dot{n}_A &= (p + n_A)n_{AB} - n_A n_B - u n_A, \\ \dot{n}_B &= n_B n_{AB} - (p + n_A)n_B - u n_B, \end{aligned} \tag{4}$$

where u is a nonnegative parameter representing the rate at which the radicals abandon their radical position in response to the nonsocial stimulus.

In contrast to the first two strategies (as well as four others treated in the Supplemental Material [23]), increasing the new parameter (u) in this system generally increases the equilibrium n_{AB} toward a limit of $1 - p$, with the one exception of a discontinuous drop partway through the ascent in Fig. 4(c). However the drop is not to zero as it was for the other strategies, and it vanishes in the limit of small p. Furthermore, following the drop, regrowth of n_{AB} is rapid. Hence this mechanism of promoting moderation, which we might call *nonsocial deradicalization*, provides the first acceptable means that we have found for expanding the moderate population in the midst of an ideological revolution. This holds for the three strategies in this Letter, and also for the four others in the Supplemental Material [23].

By itself, this final assessment should be regarded with caution. We suggest a greater emphasis be placed on our general approach as a framework for testing possible strategies as part of a continuing research program, which through further study might well uncover other means of fostering moderation more sophisticated than those considered here.

H. Hong acknowledges the hospitality of Cornell University during her visit for a sabbatical year. This research was also supported in part by NSF grants CCF-0835706 and CCF-0832782 (S. H. S.).

* Electronic address: smarvel@umich.edu

[1] A. D. Morris, *Origins of the Civil Rights Movement* (The Free Press of Simon & Schuster, New York, NY, 1984).
[2] E. Flexner and E. Fitzpatrick, *Century of Struggle* (Harvard University Press, Cambridge, MA, 1996), 3rd ed.
[3] F. O. Ramirez, Y. Soysal, and S. Shanahan, Am. Sociol. Rev. **62**, 735 (1997).
[4] T. S. Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, Chicago, IL, 1996).
[5] C. Chen, Proc. Natl. Acad. Sci. USA **110**, 1009 (2004).
[6] L. M. A. Bettencourt, A. Cintrón-Arias, D. I. Kaiser, and C. Castillo-Chávez, Physica A **364**, 513 (2006).
[7] J. Gerring, *Party Ideologies in America, 1828-1996* (Cambridge University Press, New York, NY, 2001).
[8] G. Dell'Ariccia and R. Marquez, J. Finance **61**, 2511 (2006).
[9] S. C. Ludvigson, J. Econ. Perspect. **18**, 29 (2004).
[10] G. B. Sproles, J. Marketing **45**, 116 (1981).
[11] S. Bikhchandani, D. Hirshleifer, and I. Welch, J. Polit. Economy **100**, 992 (1992).
[12] J. A. Goldstone, Ann. Rev. Pol. Sci. **4**, 139 (2001).
[13] F. Vazquez, P. L. Krapivsky, and S. Redner, J. Phys. A: Math. Gen. **36**, L61 (2003).
[14] F. Vazquez and S. Redner, J. Phys. A **37**, 8479 (2004).
[15] D. Centola, R. Willer, and M. Macy, Am. J. Sociol. **110**, 1009 (2005).
[16] S. Gekle, L. Peliti, and S. Galam, Eur. Phys. J. B **45**, 569 (2005).
[17] M. S. de la Lama, I. G. Szendro, J. R. Iglesias, and H. S. Wio, Eur. Phys. J. B **51**, 435 (2006).
[18] X. Castelló, V. M. Eguíluz, and M. S. Miguel, New J. Phys. **8**, 308 (2006).
[19] S. Galam and F. Jacobs, Physica A **381**, 366 (2007).
[20] C. Castellano, S. Fortunato, and V. Loreto, Rev. Mod. Phys. **81**, 591 (2009).
[21] J. Xie, S. Sreenivasan, G. Korniss, W. Zhang, C. Lim, and B. K. Szymanski, Phys. Rev. E **84**, 011130 (2011).
[22] To our knowledge, the only other model in literature with a committed fraction and three opinion states is in Ref. [21]. This model is based on the Naming Game [20], and so has a substantially different list of interactions than our basic model. For example, it assumes that if both the speaker and listener are moderate, then they will spontaneously persuade each other to adopt one or the other extreme position. By contrast, our basic model regards this outcome as infrequent relative to radicalization when the speaker is extremist rather than moderate. For other models involving committed individuals (or similar entities), see Refs. [15, 19]. Additional models with three opinion states are proposed in Refs. [13, 14, 16–18].
[23] See Supplemental Material at http://link.aps.org/supplemental/ for derivations of results in this paper.
[24] G. F. Davis, Corp. Gov. **4**, 154 (1996).
[25] J. Leskovec, J. Kleinberg, and C. Faloutsos, ACM TKDD **1**, 1 (2007).
[26] E. Cho, S. A. Myers, and J. Leskovec, in *ACM SIGKDD* (San Diego, CA, 2011).