

# Cultural Polarization and the Role of Extremist Agents: A Simple Simulation Model

Shade T. Shutters

Center for Social Dynamics and Complexity and School of Sustainability,  
Arizona State University, PO Box 875402, Tempe, AZ 85287-5402, USA  
shade.shutters@asu.edu  
www.public.asu.edu/~sshutte

**Abstract.** Cultural dynamics can be heavily influenced by extremists. To better understand this influence, temporal dynamics of an arbitrary cultural belief are simulated in a simple computational model. Extremist agents, holding an immutable and extreme belief, are used to examine the process of polarization – adoption of the extremist belief by the entire population. Two possible methods of counteracting polarization are examined, removal of the extremist agent and introducing a counter-extremist which holds an immutable belief at the opposite extreme. Eliminating the extremist agent is only effective at the onset of cultural transition, while introducing a counter-extremist is effective at any time and will lead to a dynamic intermediate belief. Finally, a parameter governing the society's willingness to adopt new beliefs is varied. As it decreases, extremist agents are unable to polarize a society. Instead the population breaks permanently into two or more belief groups. The study closes with a possible pathway for extremists to nevertheless polarize a society not open to new beliefs.

**Keywords:** extremism, cultural transitions, consensus, networks, social simulation.

## 1 Introduction

Though conflict is an inextricable component of social organisms, humans are unique in that culture plays a central role in many conflicts [1]. To address conflict and to better understand social dilemmas more generally it is important to understand the dynamics of cultural beliefs and how those dynamics may be influenced. Previous studies have examined social diffusion of ideas [2], norms [3], innovations [4], social values [5], and diseases [6]. Others have focused on how the topology of the network governing a society influences the rate, dynamics, and efficacy of diffusion [7-11].

This study is largely a continuation of work presented in [11], which addresses cultural consensus and sources of perpetuated conflict. In the current paper, the effects of extremist agents are examined. Given a continuum of values representing an arbitrary cultural belief or norm, extremist agents hold a belief at one endpoint of the continuum and the belief cannot be changed. Because extremist agents may be sources of incitation to violence or other socially disruptive behavior it is important to

understand how they affect cultural dynamics and how they might respond to counter measures.

This study uses very simple and highly abstract social simulations to better understand how extremist agents affect the timing and ability of a population to become polarized. It further tests and contrasts two intuitive methods of preventing polarization: removing extremist agents from a population and introducing a counter-extremist.

## 2 Simulation Description

The base-case simulation initiates by embedding  $N$  agents in one of four social network structures. Each agent holds a single arbitrary belief that is assigned an initial random value, with uniform probability, on  $[0, 1]$ . The model then proceeds through a number of pairwise interactions until the population either converges to a single, universal belief or the simulation reaches the maximum allowable number of interactions.

During a single interaction, a member of the population is selected at random and paired randomly with one its immediate neighbors as defined by the network type. Let  $a_0$  and  $b_0$  represent the initial belief values of two interacting agents so that the initial difference between their beliefs is

$$T = |a_0 - b_0| \quad (1)$$

The interacting agents influence each other's beliefs so that they are updated to  $a_1$  and  $b_1$  respectively. In this study, the new values are equal to each other and to the mean of their original beliefs

$$a_1 = b_1 = (a_0 + b_0) / 2 \quad (2)$$

Given enough interactions, the population will converge to a single belief equal to the mean value of the initial population [11, 12].

A population level parameter  $D$ , determines whether the beliefs of two interacting agents are sufficiently similar for the agents to influence each other. This threshold represents the willingness of agents to adjust their beliefs towards others. One might also consider this parameter the degree to which a society is "open-minded" or dogmatic. During a pairwise interaction, the difference between agent beliefs  $T$  is compared to the threshold  $D$ . If the difference is too great, the interaction ends without any changes in beliefs

$$\begin{aligned} T \leq D: a_1 = b_1 = (a_0 + b_0) / 2 \\ T > D: a_1 = a_0 \text{ and } b_1 = b_0 \end{aligned} \quad (3)$$

In the first series of simulations,  $D = 1$  so that all agents change their beliefs when interacting with an agent that holds a different belief. In later treatments  $D$  is varied to understand how and when polarization is affected by a population's willingness to adopt new beliefs.

## 2.1 Extremist Agents

In some simulations a single agent is chosen randomly from the initialized population and converted to an extremist agent. Its belief value is set to 0 and it cannot be changed for the duration of the simulation. Given that an extremist agent is a participant in an interaction, let  $a_0 = 0$  be the belief of the extremist agent and let  $b_0$  be the initial belief of the other participant. The interaction results in  $a_1 = a_0 = 0$  while  $b_1 = b_0 / 2$ . Thus the belief of the normal agent moves toward that held by the extremist agent, but the belief of the extremist agent remains unchanged.

In this case, given enough interactions, the belief of every agent in the population will converge to the belief of the extremist agent, a process referred to in this study as polarization.

## 2.2 Counter-Extremist Agents

In additional simulations a single extremist agent is again included in the initial population. After a number of interactions, which can be varied, an agent (other than the extremist agent) is selected randomly from the population and converted to what is referred to here as a counter-extremist agent. Its belief value is set to 1 and cannot be changed for the duration of the simulation. Thus its behavior during a pairwise interaction is identical to that of an extremist agent but causes beliefs to move towards the opposite extreme.

## 2.3 Network Structures

Four network topologies – complete, scale-free, small-world, and regular – are used to examine the role of social structure on polarization. All networks are unweighted and undirected. Scale-free networks are generated using a Barabási-Albert algorithm of preferential growth [13] with no nodal limit on links and in which each new node links to the existing network at two nodes. Small-world networks are generated using the Watts-Strogatz algorithm [14] in which each node in a ring substrate is linked to the two neighbors on either side and edges are randomly rewired with a probability  $p = 0.05$ . Regular networks are torroidal lattices in which each node has four adjacent neighbors – up, down, left, and right. For all networks other than complete mean degree  $k = 4$ , meaning differences in results among those networks are due to topological attributes other than mean degree or network density.

## 3 Results and Discussion: Time Until Polarization

Under all network topologies, the introduction of an extremist agent eventually led to polarization of the population (Table 1). However, the number of interactions required for polarization differed significantly among the four network types, both with population  $N = 64$  (ANOVA,  $F = 2,332$ ,  $p < 0.001$ ) and  $N = 400$  (ANOVA,  $F = 516$ ,

$p < 0.001$ ). Compared to small-world networks, populations on complete, scale-free, and regular networks required relatively few interactions to become polarized. However, populations on small-world networks required up to 20 times more interactions to reach polarization than populations on a complete network.

In [11] we determined, for several network topologies, the number of interactions required for a population to converge to a single belief value. In comparison to these results, the addition of an extremist agent increased convergence time by as much as 160 times (Table 1). This suggests that polarization is a much slower phenomenon than forming a consensus at some intermediate belief value.

**Table 1.** Mean interactions (in thousands) until belief convergence, with and without extremist agents. Mean calculated from 100 runs.

Network type	$N = 400$		$N = 64$	
	With	Without	With	Without
Regular ( $k = 4$ )	3,975	176	77	5
Scale-free	3,330	37	87	5
Small-world	38,083	526	271	36
Complete	1,739	10	47	2

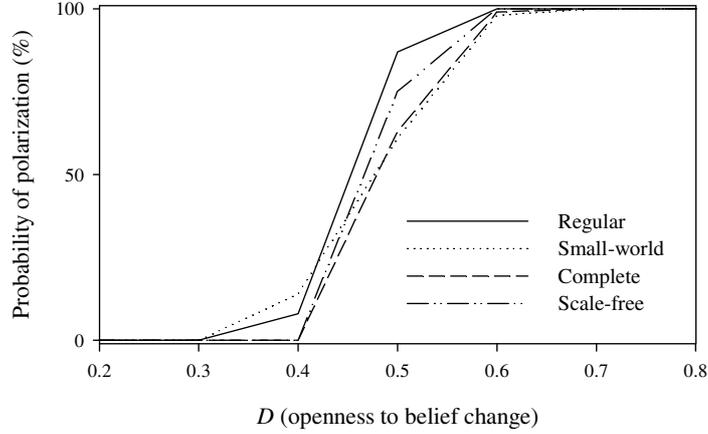
### 3.1 Societies Less Open to New Beliefs

Results shown in Table 1 were collected from simulations with  $D = 1.0$ . In additional simulations  $D$  was varied in increments of 0.1 to determine how a society's openness to new beliefs affects the ability of an extremist agent to polarize a population. Results (Fig. 1) show that at high values of  $D$ , an extremist agent will always result in a polarized society. At intermediate values of  $D$  ( $\sim 0.4$  to  $0.6$ ), the probability of polarization drops rapidly until when  $D \leq 0.3$ , polarization never takes place.

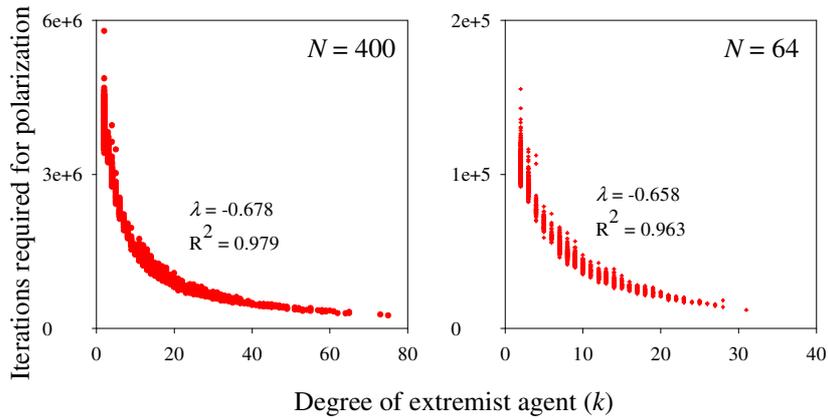
### 3.2 Further Considerations of Scale-Free Networks

Because scale-free networks are ubiquitous among physical systems as well as communication networks in social systems [15, 16], it is important to understand how cultural dynamics might be influenced by a scale-free topology. Thus, effects of scale-free networks on the ability of extremist agents to polarize a society are examined in more detail.

Results show that as the nodal degree  $k$  of the extremist agent increases the mean number of interactions required to achieve polarization decreases according to a power law (Fig. 2). This concurs with the intuitive notion that highly connected actors have a disproportionately strong influence on a society's cultural trajectory.



**Fig. 1.** Probability of polarization vs.  $D$ . Population  $N = 64$ . Probability calculated for 100 simulation runs per  $D$  value on each network type. As threshold  $D$  decreases below  $\sim 0.35$ , the probability of polarization drops to 0 on all networks.



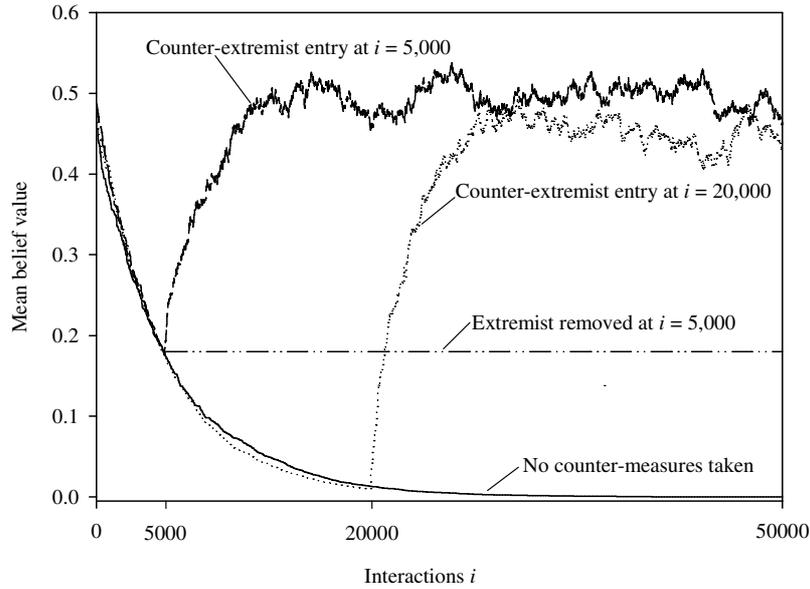
**Fig. 2.** Degree of extremist agent vs. time required for polarization on scale-free networks. Results are shown for 2,000 runs each at population sizes  $N = 400$  and  $N = 64$ . Power law exponent  $\lambda$  and simple correlation coefficient  $R^2$  are shown.

### 3.3 Counteracting Polarization

I now consider two intuitive methods for preventing polarization in scale-free networks. In both cases  $D = 1$ . The first method is removal of the extremist agent and the second method is introduction of a counter-extremist agent.

Removal of the extremist agent immediately halts further movement of the populations mean belief toward the extreme value. However, the future value to which the population will converge is fixed at time of removal and equal to the population's

mean at that time. As shown in Fig. 3, this value drops rapidly after introduction of the extremist agent and so simply removing the extremist agent is only effective if done very early after it has begun to influence a society. Fig. 3 shows that when the extremist agent is removed after only 5,000 interactions, or about 6% of the interactions required for polarization, the population's mean belief has already fallen to 0.18, a value relatively close to the extremist belief.



**Fig. 3.** Temporal dynamics of the population's mean belief. The first 50,000 interactions of four simulations are presented. Results are for a scale-free network with population size  $N = 64$  and with the degree of both the extremist and counter-extremist agents  $k = 5$ .

On the other hand, the introduction of a counter-extremist agent will lead to a certain dynamic equilibrium belief value  $b_{\text{eq}}$  regardless of when the counter-extremist is introduced. In networks other than scale-free networks  $b_{\text{eq}} \approx 0.50$ . In scale-free networks  $b_{\text{eq}}$  is a function of the nodal degrees of the extremist  $k_e$  and counter-extremist  $k_c$  approximated by the belief-weighted relative probability of each being randomly chosen for an interaction

$$b_{\text{eq}} \approx (k_c + 1) / [(k_e + 1) + (k_c + 1)] \quad (4)$$

Within the constraints of this abstract model, the introduction of a counter-extremist agent offers a better counter-measure to polarization than simple removal of the extremist agent. This occurs when  $D = 1$  and the effectiveness of introducing a counter-extremist would likely decrease over time in real world situations as the population slowly moves out of the counter-extremist's range of influence.

## 4 How Extremist Agents Can Overcome Low Willingness to Change

The concept of  $D$  is highly abstract and it is likely that any analog in real human societies is not only highly heterogeneous among individuals, but also relatively low. This presents the extremist agent with an obstacle to polarization. I close by discussing how an extremist agent may overcome this obstacle by introducing an additional agent type, the dogmatic agent. Like extremists and counter-extremists, dogmatic agents hold an immutable belief. However, the value of that belief lies at some intermediate point between extremes.

In the presence of such dogmatic agents, extremist agents have an opportunity to polarize a population even when  $D$  is low. The extremist must rely on dogmatic agents to convert the beliefs of those in the population that are far out of the extremists range of influence ( $T \gg D$ ). The extremist then must go through a stepwise process of eliminating dogmatic agents, starting first with those that differ most with respect to belief value. The extremist should also allow sufficient time between the elimination of dogmatic agents so that followers of recently-eliminated dogmatic agents will be fully drawn to the next closest dogmatic agent.

Consider a simple illustration. Let  $D = 0.3$  for a certain population, meaning it is largely averse to change. As shown in Fig. 1, an extremist agent with belief 0 will be unable to polarize the society. Let dogmatic agents exist with beliefs at 0.75, 0.50, and 0.25. Given sufficient time, all agents originally holding beliefs on  $(0.75, 1]$  will be adjusted to 0.75 (or less). At that time, the extremist agent should take measures to eliminate the dogmatic agent with belief 0.75. This will lead to all agents holding a belief on  $(0.50, 0.75]$  to eventually hold beliefs of 0.50 or less, since that entire interval is within the range of influence of the next closest dogmatic agent (belief 0.50). At this point, the dogmatic agent with belief 0.50 is targeted for elimination and so on until the entire population is moved to the extremist belief.

Anecdotal evidence for such a strategy exists throughout history in cases where agents, once considered relatively extreme, come to be viewed as moderate in their beliefs and are eliminated. Well known examples include the 1922 assassination of Irish militant/politician Michael Collins by more extreme Irish nationalists and the 1917 Russian Revolution, in which dogmatic agents first overthrew the Tsarist regime but were then expelled some months later by the more extreme Bolsheviks. Contemporary examples might include the frequent execution of moderate Muslim clerics in Dagestan by more radicalized Islamic militants.

## 5 Future Directions

This study has used a simplistic binary method to determine the degree to which an agent may be influenced by another. An agent is either completely influenced or not at all. A more realistic assumption is that the ability of one agent to influence the beliefs of another decays as a function of the difference of their current beliefs.

Let  $i$  equal the degree to which agents may influence each other so that equation (2) becomes

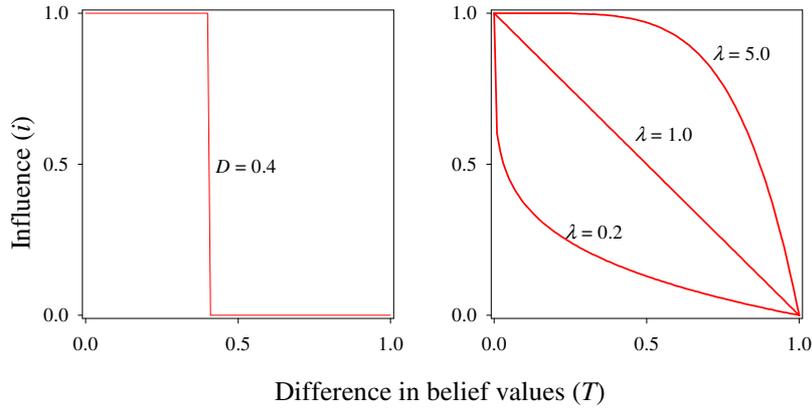
$$\begin{aligned} a_0 \geq b_0; & \quad a_1 = a_0 - iT/2 \quad \text{and} \quad b_1 = b_0 + iT/2 \\ a_0 < b_0; & \quad a_1 = a_0 + iT/2 \quad \text{and} \quad b_1 = b_0 - iT/2 \end{aligned} \quad (5)$$

As with (2), this equation does not apply to extremist or dogmatic agents. In addition, implicitly  $D = 1$  (though it may be explicitly set to values less than 1 to explore additional parameter space).

It is important then to understand how  $i$  decays as a function of the initial beliefs of two agents. A simple decay function might take the form

$$i = 1 - T^\lambda \quad (6)$$

Fig. 4 compares this decay function at three values of  $\lambda$  to the binary influence used above. However, evidence suggests that  $\lambda = 0.2$  is the most realistic of the decay functions presented [17] and future empirical work should seek to refine this value. More importantly, the efficacy of extremists and extremist counter-measures discussed above should be reassessed under different decay models of influence.



**Fig. 4.** Comparison of binary influence (left) and influence as a decaying function belief differences (right). In the binary model,  $D = 0.4$  so that when  $T > 0.4$ , influence = 0. The decay model is presented for three values of  $\lambda$ . When  $\lambda = 0.2$ , an agent's beliefs can only be significantly influenced by another agent holding very similar beliefs.

## 6 Conclusion

This brief study has described a very simple and abstract model of the effects of extremist agents. It shows that, given enough time, extremists can polarize a population and that either removing the extremist agent or introducing counter-extremists can

mitigate the extremist's polarizing effect. However, caution should be taken before drawing broad conclusions from such an abstract model. Future research should incrementally introduce more realistic parameters that are empirically grounded. In this manner, this and similar models will continue to make small but valuable steps toward a better understanding of cultural dynamics.

## References

1. Bowles, S.: Warriors, Levelers, and the Role of Conflict in Human Social Evolution. *Science* 336, 876–879 (2012)
2. Buskens, V., Yamaguchi, K.: A new model for information diffusion in heterogeneous social networks. *Sociol. Methodol.* 29, 281–325 (1999)
3. Axelrod, R.: An Evolutionary Approach to Norms. *Am. Polit. Sci. Rev.* 80, 1095–1111 (1986)
4. Mahajan, V., Muller, E.: Innovation Diffusion and New Product Growth-Models in Marketing. *J. Marketing* 43, 55–68 (1979)
5. Bikhchandani, S., Hirshleifer, D., Welch, I.: A Theory of Fads, Fashion, Custom, and Cultural-Change as Informational Cascades. *J. Polit. Econ.* 100, 992–1026 (1992)
6. Santos, F.C., Rodrigues, J.F., Pacheco, J.M.: Epidemic spreading and cooperation dynamics on homogeneous small-world networks. *Phys. Rev. E* 72, 56128 (2005)
7. Cowan, R., Jonard, N.: Network structure and the diffusion of knowledge. *J. Econ. Dyn. Control* 28, 1557–1575 (2004)
8. Abrahamson, E., Rosenkopf, L.: Social network effects on the extent of innovation diffusion: A computer simulation. *Organ. Sci.* 8, 289–309 (1997)
9. Delre, S.A.: The Effects of Social Networks on Innovation Diffusion and Market Dynamics. Department of Economics and Business, vol. PhD, p. 148. University of Groningen, Groningen, Netherlands (2007)
10. Santos, F.C., Rodrigues, J.F., Pacheco, J.M.: Graph topology plays a determinant role in the evolution of cooperation. *Proc. R. Soc. Lond. B. Biol. Sci.* 273, 51–55 (2006)
11. Shutters, S.T., Cutts, B.B.: A Simulation Model of Cultural Consensus and Persistent Conflict. In: Subrahmanian, V.S., Kruglanski, A. (eds.) *Proceedings of the Second International Conference on Computational Cultural Dynamics*, pp. 71–78. AAAI Press, Menlo Park (2008)
12. Boyd, S., Ghosh, A., Prabhakar, B., Shah, D.: Randomized Gossip Algorithms. *IEEE T. Inform. Theory* 52, 2508–2530 (2006)
13. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. *Science* 286, 509–512 (1999)
14. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. *Nature* 393, 440–442 (1998)
15. Barabási, A.L.: Scale-Free Networks: A Decade and Beyond. *Science* 325, 412–413 (2009)
16. Mendes, J.F., Dorogtsev, S.N., Povolotsky, A., Abreu, F.V., Oliveira, J.G. (eds.): *Science of Complex Networks: From Biology to the Internet and WWW; CNET 2004*. American Institute of Physics, Melville, New York (2005)
17. Cialdini, R.B., Trost, M.R.: Social influence: social norms, conformity, and compliance. In: Gilbert, D., Fiske, S., Lindzey, G. (eds.) *The Handbook of Social Psychology*, 4th edn., vol. 2, pp. 151–192. McGraw-Hill, New York (1998)