Culture Wars and Dynamic Networks: A Hopfield Model of Emergent Structure

Michael W. Macya, James A. Kittsb, and Andreas Flache
c
aDepartment of Sociology, Cornell University, Ithaca, NY 14853; b Department of Sociology, University of Washington, Seattle, Washington 98195-3340; cThe Interuniversity Center for Social Science Theory and Methodology, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, NL

Why do populations often self-organize into antagonistic groups even in the absence of competition over scarce resources? Is there a tendency to demarcate groups of “us” and “them” that is inscribed in our cognitive architecture? We look for answers by exploring the dynamics of influence and attraction between computational agents. Our model is an extension of Hopfield’s attractor network. Agents are attracted to others with similar states (the principle of homophily) and tend to converge toward the states of agents to whom they are attracted, given the strength and valence of the social tie. Negative valence implies xenophobia (instead of homophily) and differentiation (instead of imitation). Consistent with earlier work on structural balance, we find that networks can self-organize into two antagonistic factions, without the knowledge or intent of the agents. We investigate this tendency as a function of network size, the number of potentially contentious issues, and agents’ openness and flexibility toward alternative positions. The surprising finding is that the global alignment of a multi-dimensional opinion space along a single polarizing definition of opposing ideologies is facilitated by the ideological flexibility and open-mindedness of local constituents.

INTRODUCTION: CULTURE WARS CONTROVERSY

A decade ago, James Davison Hunter’s (1) book, Culture Wars: The Struggle to Define America, raised the prospect of an increasingly polarized society. Hunter argues that societies have an underlying tendency to polarize, for the reason identified by Yeats (2) in The Second Coming: “The best lack all conviction, while the worst are full of passionate intensity.” When the center cannot hold, conflict shifts from divisible interests to divisive identities. Two centuries ago, concerns about identity politics motivated Madison to advocate “an equilibrium in the interests & passions of the Society itself” as the “best provision for a stable and free Gov’t” (3). However, two centuries later, conflicts over abortion laws, gay rights, school prayer, and other social issues have led commentators like Hunter to conclude that the equilibrium is tipping. For example, in The Restructuring of American Religion (1988), Robert Wuthnow sees “a deep hostility and misgiving” between “two opposing camps” in American society, with “fundamentalists,” ‘evangelicals,’ and ‘religious conservatives’ in one and “... ‘religious liberals,’ ‘humanists,’ and ‘secularists’ in the other” (4:371).

The “culture war” hypothesis has been challenged by more recent opinion research. DiMaggio, Evans, and Bryson (5) analyzed survey responses over a twenty year period, using National Election Studies and the General Social Survey, and found that “only attitudes toward abortion have become more polarized in the past twenty years” (p. 738), where polarization refers to the tendency to hold an extreme and uncompromising view. A follow-up study by Mouw and Sobel (6) challenges the DiMaggio team’s use of ordinal data to build interval measures. When they reanalyzed the NES data using a latent variable model, they found no evidence of polarization even on abortion.

While informative, these empirical challenges to the “culture wars” hypothesis have missed the theoretical target. The studies show no increase in the number of contentious issues or in the degree of contention on any given issue. However, they do not address the prediction that positions on one item are increasingly correlated with positions on others. Here, the critics seemed to implicitly accept Hunter’s basic thesis, that the formation of ideological “camps” is grounded in an underlying intransigence characterized by hard-line attitudes and narrow identification with a highly salient symbolic issue. Both sides in the debate seem to assume that polarization at the level of the group (attenuation of cross-cutting cleavages) reflects polarization at the level of the individual – reluctance to set aside predispositions, explore other viewpoints, or think about issues in other domains. The challengers only argue that there is little evidence of ideological hardening, which they assume contradicts the “culture war” hypothesis of polarization at the macro level.

Intuitively, this assumption seems entirely safe. Societies rent by civil war or religious crusade are not where we would expect to find individuals with flexible positions. However, it may be misleading to focus on static comparisons of individuals in polarized and pluralform societies. Consider instead the dynamics by which societies become polarized. While societal polarization may well harden individuals into narrow and uncompromising stances, this need not imply that individual inflexibility or single-mindedness aggregates into group polarization. Indeed, we cannot rule out the opposite possibility. Corporatist societies suggest the possibility that pluralist ideological differences can persist even when individual adherents to each position adamantly reject other points of view than their own.

Does single-mindedness and ideological rigidity underlie the polarization of social groups, causing a multidimensional opinion space to reduce to a rivalry of two opposing ideological “camps”? Or could it be that this global alignment of positions is actually facilitated by the ideological breadth, flexibility, and openness of local constituents?

We look for answers by exploring the dynamics of influence and attraction between computational agents. Our model draws heavily on recent work using agent-based models of cultural differentiation (9, 10). In Axelrod’s specification, neighboring agents on a square lattice interact with neighbors with a likelihood determined by the similarity of their cultural traits (given by a randomly assigned string of numbers). Interaction, in turn, reduces remaining differences. Equilibrium obtains when every agent differs from each of its neighbors either on every attribute (which precludes interaction) or on none (which precludes influence).

† Dimaggio et al. (5) define polarization in terms of “the extremity of and distance between responses” (p. 693), not the correlation between responses across issues.

* Address reprint requests to: mwm14@cornell.edu
Axelrod’s model incorporates one of the starkest regularities in the social world: “homophily,” or the tendency to interact with similar others (11, 12, 13). Several explanations for homophily have been proposed. Social psychologists posit a “Law of Attraction” based on an affective bias toward similar others (14; see also 15-18). Carley (19) and others (20, 21, 10) point to the greater reliability and facility of communication between individuals who share vocabulary and syntax.

Much less attention has been directed to an alternative explanation— that homophily is largely a byproduct of its antipole. It is not so much attraction to those who are similar that produces group homogeneity but repulsion from those who are different. Xenophobia leads to the same emergent outcome as attraction between similar actors: disproportionate homogeneity in relations. In fact, Rosenbaum (22) argues that many experimental findings of homophily in relations may have spuriously represented this effect of repulsion from those who are different.

Others counter that homophily is the spurious consequence of social influence, as argued in classic research on “pressures to uniformity” (23-25) and recent work in social networks (26, 27). Even if relations are held constant in an exogenously clustered social network, social influence from network neighbors will lead to local homogeneity, without the need to assume homophilous interaction.

Although Axelrod does not allow for negative influence or xenophobia, he does assume that both attraction and influence processes operate to produce the high levels of in-group homogeneity we see in the social world. The combination of social influence and homophilous interaction produces a feedback loop in which minimal initial similarity increases the probability of influence and homophilous interaction produces a feedback loop in processes operate to produce the high levels of in-group xenophobia, he does assume that both attraction and influence processes operate to produce the high levels of in-group homogeneity. 

Applying Heider’s Balance Theory (8) to social relations, Cartwright and Harary’s Structural Balance Theory assumes that dyadic social ties can be positive or negative, but that actors will seek “balance” in their relations with any two alters. This effectively means that the product of the signs in any cycle will be positive. Consider the simplest case—an ABC triad in which the AB and AC ties are positive but BC is negative. The triad is then out of balance (as the product AB×BC×AC is negative), creating pressure to change any one of the three relations, such that the product of the relations becomes positive. For example, either B and C will become “friends” or there will be a falling out between A and B or between A and C (but not both). This is the familiar idea that enemies of my enemies become my friends, while the friends of my enemies become my enemies. Extended to a larger network, the elemental triadic case suggests a perfect separation of two mutually antagonistic subgroups. Davis generalized the model to include “clustering” (31) by relaxing the assumption that enemies of enemies are always friends. This makes it possible for more than two antagonistic subgroups to persist, even while not formally balanced.

Our model integrates the attraction-influence feedback loop from Axelrod and the bivalent relations in Structural Balance Theory (7, 8). The model is an extension of Hopfield’s attractor network (32, 33). An important property of attractor networks is that individual nodes seek to minimize “energy” (or dissonance) across all relations with other nodes—a process that parallels but differs from the pursuit of balanced relations in Structural Balance Theory. These networks also posit self-reinforcing dynamics of attraction and influence as well as repulsion and differentiation.

More precisely, this class of models generally uses complete networks, with each node characterized by one or more binary or continuous states and linked to other nodes through endogenous weights. Like other neural networks, attractor networks learn stable configurations by iteratively adjusting the weights between individual nodes, without any global coordination. In this case, the weights change over time through a Hebbian learning rule (34): the weight $w_{ij}$ is a function of the correlation of states for nodes $i$ and $j$ over time. To the extent that $i$ and $j$ tend to occupy the same states at the same time, the tie between them will be increasingly positive. To the extent that $i$ and $j$ occupy discrepant states, the tie will become increasingly negative.

Following Nowak & Vallacher (35; see also 36, 37), we apply the Hopfield model of dynamic attraction to the study of polarization in social networks. In this application, observed similarity/difference between states determines the strength and valence of the tie to a given referent. This attraction and repulsion may be described anecdotally in terms of liking, respect, or credibility and their opposites.

**MODEL DESIGN**

In our application of the Hopfield model, each node has N-1 undirected ties to other nodes. These ties include weights, which determine the strength of influence between agents. Formally, social pressure on agent $i$ to adopt a binary state $s$ (where $s = \pm 1$) is the sum of the states of all other agents $j$, where influence from each agent is conditioned by the weight ($w_{ij}$) of the dyadic tie between $i$ and $j$ ($1.0 < w_{ij} < 1.0$):

$$P_s = \frac{\sum_{j \neq i} w_{ij}s_j}{N-1}, j \neq i$$

(1)

Thus, social pressure (-1 < $P_s$ < 1) to adopt $s$ becomes increasingly positive as $i$’s “friends” adopt $s$ ($s=1$) and $i$’s “enemies” reject $s$ ($s=-1$). The pressure can also become negative in the opposite circumstances. The model extends to multiple states in a
straightforward way, where [1] independently determines the pressure on agent \(i\) for each binary state \(s\).

Strong positive or negative social pressure does not guarantee that an agent will accommodate, however. It is effective only if \(i\) is willing and able to respond to peer influence. If \(i\) is closed-minded or \(i\) has a given trait is not under \(i\)'s control (e.g., ethnicity or gender), then no change to \(s\) will occur. Let \(v\) classify states as fixed \((v=0)\) or free \((v=1)\) and \(\chi\) represent an exogenous determination of \(s\). We then find probability \(\pi\) of \(s\) as a cumulative logistic function of social pressure:

\[
\pi_{is} = \frac{1}{1 + e^{-\beta \chi}} + (1 - \pi) \chi_i
\]

Agent \(i\) adopts state \(s\) if \(\pi_{i} > C + \chi\), where \(C\) is the inflection point of the sigmoid in [2], \(\chi\) is a random number drawn from a uniform distribution in the interval \([C-C, C+C]\), and \(\epsilon\) is an exogenous error parameter \((0 \leq \epsilon \leq 1)\). At one extreme, \(\epsilon = 0\) produces highly deterministic behavior, such that any social pressure above the trigger value always leads to conformity and pressures below the trigger value entail differentiation. Following Harsanyi (38), \(\epsilon > 0\) allows for a “smoothed best reply” in which pressure levels near the trigger point leave the agent relatively indiffident and thus likely to explore behaviors on either side of the threshold.

In the Hopfield model, the path weight \(w_{ij}\) changes as a function of similarity in the states of node \(i\) and \(j\). Weights begin with uniformly distributed random values, subject to the constraints that weights are symmetric \((w_{ij} = w_{ji})\). Across a vector of \(K\) distinct states \(s_k\) (or the position of agent \(i\) on issue \(k\)), agent \(i\) compares its own states to the observed states of another agent \(j\) and adjusts the weight upward or downward corresponding to their aggregated level of agreement or disagreement. Based on the correspondence of states for agents \(i\) and \(j\), their weight will change at each discrete time point \(t\) in proportion to a parameter \(\lambda\), which defines the rate of structural learning \((0 < \lambda < 1)\):

\[
w_{ij,t+1} = w_{ij} + \lambda \left(1 - w_{ij}\right) + \frac{\lambda}{K} \sum_{k=1}^{K} s_{jk}s_{ik}, j \neq i
\]

As correspondence of states can be positive (agreement) or negative (disagreement), ties can grow positive or negative over time, with weights between any two agents always symmetric.

Note one significant departure from Structural Balance Theory. Although the agents in this model are clearly designed to maintain balance in their behaviors with both positive and negative referents, this assumption is not “wired in” to the relations themselves. That is, two agents \(i\) and \(j\) feel no direct need for consistency in their relations with a third agent \(h\). Indeed, \(i\) has no knowledge of the \(jh\) relationship and thus no ability to adjust the \(ij\) relation so as to balance the triad.

Given an initially random configuration of states and weights, these agents will search for a profile that minimizes dissonance across their relations. Structural Balance Theory predicts that system-level stability can only occur when the group either has become uniform or has polarized into two (Cartwright and Harary, 1956) or more (Davis 1967) internally cohesive and mutually antipathetic cliques. However, there is no guarantee in this model that they will achieve a globally optimal state in structural balance.

METHODS

We tested the effects of agent polarization on network polarization by manipulating three agent-level behavioral rules and observing the effects on system dynamics. Polarization at the agent level can be manifested as rigidity (unwillingness to explore less-preferred positions), narrow-mindedness, and closed-mindedness:

1. Rigid vs. flexible (a.k.a. “curious”): the extent to which the state adoption decision is random near indifference, ranging from \(\epsilon = 0\) (always pick the preferred position even when almost indifferent) to \(\epsilon = 1\) (a coin toss).
2. Narrow- vs. broad-minded: the multiplexity of the state space, that is, the number of salient cross-cutting dimensions that agents consider in evaluating each other.
3. Closed- vs. open-minded: the proportion of these salient dimensions that can be affected by social pressure, ranging from \(v=0\) (all states are fixed in [2]) to \(v=1\) (all states can be influenced). A fixed state can be a position on an issue about which an agent is strictly uncompromising. It can also be an attribute that is difficult to change, such as ethnicity or gender.

Thus, we consider a range of stylized character profiles for agents, ranging from “extremists” who focus on ascriptive differences and are rigidly narrow- and closed-minded, to “moderates” who are curious, broad-minded, open to influence, and/or focused on behavior rather than ascriptive traits.

We use two measures of network polarization—the number of mutually exclusive cohesive subgroups in the weight graph and the degree of segregation among these subgroups. Both measures are based on the graph theoretical LS-set (39). An LS-set is a subset of agents who have smaller social distance (i.e., larger weights) to members within the subset than they have to members outside the subset. For the cohesive subgroups thus identified, we calculated their segregation in the weight graph in terms of the normalized difference between the average internal tie strength and the average tie strength between the subgroup and its complement. To obtain an overall polarization score, we generated all maximal partitionings of the group into mutually exclusive cohesive subgroups and took from these the average segregation of all subgroups in the most segregated partitioning.

We allowed each experimental condition to repeat for 1000 iterations which was sufficient for almost all conditions to arrive at an equilibrium. Only equilibrium outcomes (converged solutions) were analyzed. For all the experiments, we set the structural learning rate \(\lambda\) at 0.5, which means that agents consider current states equally with previous impressions in determining the tie weight for the next (arbitrary) time interval. Other non-experimental parameter combinations were varied within experimental conditions as tests of robustness. Each parameter combination was repeated 20 times, yielding 44,000 observations.

RESULTS

Experiment 1 tests the effects on network polarization of agent broad-mindedness. Network polarization is measured as the degree of segregation in the most segregated equilibrium partitioning of cohesive subgroups. Maximum segregation (1.0) occurs in a perfectly bifurcated network, composed of exactly two internally cohesive and mutually antagonistic groups. Broad-mindedness is operationalized as multiplexity, or the number of dimensions along
which agents can differentiate by changing their positions. Intuitively, we would expect polarization to be most likely in a world where agents focus single-mindedly on one highly salient issue, and indeed, Figure 1 confirms the intuition. What is surprising, however, is that network polarization can also be caused by too many salient issues. With 100 agents scattered over a state space larger than 2^5, the dynamics quickly approach the bifurcation we would expect in a single-issue population. As N declines, the non-monotonic effect becomes less pronounced. For N=100, the U-shape function is little changed but the tipping point increases, from 5 dimensions with N<100 to 11 dimensions with N=300.

Experiment 1 assumes rigidly deterministic agents (ε=0) who never explore alternative positions, even when they approach indifference. It also assumes that agents are willing and able to change all positions in response to social influence (v=1), which would not be the case, for example, if race or ethnicity were a salient dimension of social differentiation. Experiment 2 relaxes these assumptions under conditions that make polarization relatively unlikely, namely, that there are at least five dimensions of differentiation (see Figure 1). Intuitively, we might expect less network polarization in a population that is less rigid and more polarizing in a population that is less open-minded. The surprising result is that the effect is quite the opposite.

Figure 2 displays the polarizing effects of flexibility and open-mindedness in greater detail. As in Experiment 1, polarization is measured as the overall degree of segregation in the weight graph. Flexibility is simply the error rate, ranging from ε=0 (always pick the preferred position even when almost indifferent) to ε=1 (toss a fair coin). Openness is the proportion of all states K that are affected by social pressure (Σv/K). Results were averaged over 44000 observations in which N ranged from 10 to 100 and K ranged from 5 to 10.

The results show positive main effects of flexibility and openness on polarization, as well as a strong interaction, such that openness and flexibility have stronger polarizing effects when combined.

We also found that increased flexibility washes out the non-monotonic effect of multiplexity evident in Figure. Flexibility leads to polarization even when agents attend to multiple dimensions of differentiation that would otherwise produce the cross-cutting cleavages characteristic of a pluri-form society.

DISCUSSION

With binary agent states, the dynamics of homophilous influence and xenophobic differentiation create an energy landscape in which there is only one basin of attraction, polarization. Any configuration with more than two cohesive subgroups represents global tension for a reason that is readily apparent: A binary state precludes the ability to hold a position that differs from each of two opposing positions. As the number of dimensions increases, many more than two combinations are possible at equilibrium, but for each, there must always be some similarity with all other combinations except one. Surprisingly, as these combinations explode, they lead not to pluralism but to polarization. The mechanism is similar to that identified by Axelrod. The more opportunities for neighborhoods to be linked via overlapping positions, the higher the probability they will find a way to coalesce.

The key parameter in this mechanism is the density of the population distribution across the state space. When there are many agents but few dimensions, then, from a random start, every possible combination of states (corresponding to “ideologies” if the states are variable and to “identities” if the states are fixed) will find multiple incumbents. The presence of identical neighbors (as well as dissimilar antipodes who provide negative referents) creates strong pressures to remain loyal to a shared set of states. This resistance to change can then support a pluri-form equilibrium. Increasing the number of dimensions in this region of the parameter space expands the set of possible combinations without unraveling the balance of influences underlying the pluri-form solution. Accordingly, polarization initially declines with increasing multiplexity.

However, if population density in the state space falls below a critical level (either due to low population or high multiplexity), it becomes impossible for every distinct vector of states to find an incumbent. The sparseness of the population relative to the number of possible ideologies/identities requires that some agents begin with neither allies (who overlap perfectly) nor enemies (who overlap not at all). These agents lack sufficient pressure to retain their own unique ideologies or identities and will thus be pulled into coalitions with agents who overlap their positions. As the multiplexity of the state space increases, equilibria increasingly depend on a delicate balance, where a distribution of agents on ideologies and identities needs to be found that contradicts the distances between positions, such that they are not too close to each other to be pulled together and not too far apart to be pushed further.
differentiation. In this higher range, increasing multiplexity makes it more difficult for pluriform configurations to persist, such that collapse into a simple bifurcation becomes the most likely alternative.

Flexibility and openness promote polarization by the same process. Network self-organization can become trapped in high-energy pluriform equilibria that are nevertheless a local minimum of the energy landscape, as depicted in Figure 3. Volatile agents can disturb these local solutions, allowing the system to continue searching for the global attractors.

Figure 3. How agent volatility allows the network to find a lower energy minimum—hence greater polarization.

Homogeneity is also a stable equilibrium of the model, but this equilibrium cannot obtain unless we preclude negative weights and negative influence (as in the Axelrod model). If we initialize the model at homogeneity, it will remain there indefinitely. However, from a random start, the system cannot find the single-group solution without falling into an inescapable trap of in-group/out-group hostility. Further, a sufficient exogenous perturbation will disturb the unitary equilibrium and yield a two-group solution, but the reverse shift is much more difficult to obtain and is extremely unlikely.

To explore this interpretation further, we measured the energy level of equilibria by summing over the product of agreement and tie strength across all dyads in the weight graph. As expected, we found that networks with perfect polarization in two opposing camps had an energy level of zero, while energy levels increased with the number of different positions in the state space that co-existed at equilibrium. That analysis examined tension in dyads for the variety of converged configurations. A continuous index of graph balance (40), assessing balance in triads and cycles of any higher length, yielded the same conclusion. Equilibria with either one or two internally cohesive groups represent maximum balance, while convergence at any higher number of groups implies imbalanced relations in the social network.

CONCLUSIONS AND IMPLICATIONS

This study has explored the effects of agent-level polarization on the polarization of a self-organizing network, using a Hopfield model of dynamic attraction. The project integrates Axelrod’s positive feedback model of influence-interaction with a bivalent model of social relations and cognitive balance. The Hopfield model emulates processes of homophily, xenophobia, and influence from positive and negative referents that have been observed in natural settings. Thus, we may account for emergent social phenomena with a rigorous and parsimonious model, including only basic behavioral principles that have been broadly supported in experimental work. The results also have interesting implications for the notion of structural balance in social relations. Although the agents in this model are clearly designed to maintain balance in their behaviors with both positive and negative referents, this assumption is not “wired in” to the relations themselves, as it is in Structural Balance Theory. That is, two agents feel no direct need for consistency in their relations with a third agent. However, the results show that triads (or cycles of any length) tend to become balanced over time as agents seek balance in their dyadic relations.

The model also has interesting implications for Social Identity Theory (41), which posits an in-group bias toward those who share a salient trait, prejudice against the out-group, and a tendency to ignore or change discrepant traits. The Hopfield model produces dynamic networks that self-organize a similar pattern, but without a higher-order cognitive framework of social categories. In fact, these agents are not even aware that they belong to “groups” at all. This demonstrates the possibility that in-group/out-group differentiation and antagonism are emergent properties of network self-organization and are not inscribed in agents’ cognitive architectures as assumed by identity theorists. In effect, the model demonstrates a distributed representation of the formation of social identities, which are not reducible to local representations in the minds of individual agents.

Previous theoretical work has emphasized the global stability of social homogeneity, where convergence to unanimity is an almost irresistible force in closely interacting populations. Our bivalent model suggests instead that a social structure is most stable when the network self-organizes along a single dimension of differentiation which thus determines a clear “right” and “wrong” choice on all behavioral dimensions. This resonates with our sociological intuitions that social conflict in fixed populations does not diminish monotonically over time, but instead consolidates into larger coalitions with larger conflicts. The model can thus generate a variety of empirically testable propositions.

ACKNOWLEDGEMENT

The research was supported by a grant to the first author from the U.S. National Science Foundation (SES-0079381). The research of the third author was made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

REFERENCES