Belief-Oriented Segregation in Policy Networks

Adam Douglas Henry*

West Virginia University, Division of Public Administration, Morgantown, WV 26506-6322, U.S.A.

Abstract

Many policy networks are characterized by belief-oriented segregation, where actors with shared belief systems are clustered together and few opportunities exist for communication across coalitions of like-minded stakeholders. This inhibits the ability of network actors to effectively learn about, and successfully manage, complex policy problems. Despite the importance of understanding why these structures emerge, the processes that generate belief polarization in networks are not well-studied. This paper derives a general agent-based model of network formation and belief change from the Advocacy Coalition Framework (ACF), a prominent theory of the policy process that has been widely applied to the study of belief conflict in contentious policy systems. Simulation results suggest that the ACF assumption of biased information processing plays a critical role in the emergence of belief-oriented segregation in networks. This model provides a starting point for re-thinking the role of cognitive bias in social and policy learning, as well as the relationships between belief change and the evolving structure of policy networks.

Keywords: Policy networks; agent-based simulation; network segregation; social learning; belief change; Advocacy Coalition Framework

1. Introduction

Networks are a crucial part of the machinery of political decision-making, and the structure of networks can yield insights into the rationales that drive cooperation within a particular policy domain (Henry et al., 2010). In addition, network structures have a direct influence on the ability of actors to collectively solve problems that are highly complex and involve large amounts of potentially value-laden scientific information (Hong and Page, 2004; Dietz & Henry, 2008). While some treatments of networks surrounding a particular policy issue—termed “policy networks” in this research—assume that network formation is an intrinsically desirable goal, the structure of many policy networks hinder problem solving processes by creating barriers between two or more groups of actors who must share information or otherwise interact in order to gain traction on a issue (Burt, 2004). One important structural barrier is network segregation, a term used to describe networks that exhibit two related properties: first, network...
actors are grouped into two or more “clusters” or “communities” with dense interactions within groups and relatively sparse connections between groups (Girvan & Newman, 2002), and second, actors have attributes that are strongly correlated with membership in a given cluster.1

Network segregation is of great applied interest, since this property may inhibit actors’ ability to deal effectively with shared problems. For example, research in sustainability frequently points to the challenges in learning and collective action that are posed by schisms within networks across vertical and horizontal levels of government (Schneider et al., 2003), between the scientific and policy-making communities (Cash et al., 2003; McNie, 2007), and (perhaps the most daunting barrier of all) across groups with divergent belief systems regarding the scope and severity of problems and appropriate solutions (Henry, 2009). The focus of this paper is to model the emergence of this latter type of segregation—termed belief-oriented network segregation—within policy networks.

Belief-oriented network segregation is a feature observed within many networks (e.g., Henry et al., 2010; Weible & Sabatier, 2005), and is thought to greatly exacerbate political conflict as it limits lines of communication and trust between actors with divergent viewpoints on policy. What is less well-known about this phenomenon, however, is the process by which it emerges. This research uses agent-based computer simulation to model hypothesized pathways to belief-oriented segregation in policy networks, based on theoretically-grounded expectations of how network actors engage in social learning (i.e., assimilate specific beliefs from others within their network neighborhood), and how actors’ networking decisions (i.e., with whom they form network ties) are in turn influenced by their belief systems.

The Advocacy Coalition Framework (ACF: Sabatier & Jenkins-Smith, 1993, 1999) is chosen as a theoretical starting point for defining a general class of agent-based models. The ACF is chosen for two reasons. First, the ACF is a prominent theory of the policy process that has been widely applied to the study of belief conflict in a wide variety of policy arenas, and the ACF focuses on the behavior of individuals embedded within segregated network structures. Second, the framework includes an explicit conceptual model of belief systems, networking, and social learning that is amenable to mathematical formalization and computer simulation. Given the lack of comprehensive theory in the social sciences regarding the interactions between learning processes, internal belief system structure, and networks (Henry, 2009), the ACF is a good starting point for the operationalization and testing of theoretical propositions.

The agent-based models developed here include parameters that represent assumptions regarding the endogenous formation of network linkages and social learning processes. The general model allows for various specifications that are consistent with the ACF, but that tweak the strength of the framework’s assumptions regarding network formation and belief change processes. In addition, competing models may also be specified. These models are then run using initial network structures that are well-mixed, in the sense that they are unsegregated and exhibit belief heterogeneity. Hypothetical agents are then allowed to learn from one another and rewire their network neighborhood over time, until equilibrium is reached in network actors’ belief systems. This allows for a direct investigation of whether models implied by the ACF are capable of producing belief-oriented network segregation, and in addition allows for an investigation of the relative importance of different theoretical assumptions.

2. Networks, social learning, and the Advocacy Coalition Framework

The ACF provides one of the more sophisticated treatments of belief systems, social learning, and network formation in the policy process literature (Sabatier & Jenkins-Smith, 1993, 1999). The ACF contains a clearly-explained set of theoretical foundations that imply a particular model of belief change and network formation, leading to the emergence of what the ACF calls “advocacy coalitions” (see, for example, Sabatier & Jenkins-Smith, 1999). Advocacy coalitions are defined in the ACF as groups of actors who both coordinate their political behavior and also share an integrated set of policy-relevant beliefs. Thus, networks that contain two or more advocacy

1 The concept of network segregation is related to community structure, however community structure indicates a property of the network structure only whereas network segregation imposes the additional requirement that actor attributes be homogenous within communities and heterogeneous across communities.
coalitions also exhibit belief-oriented segregation; network clusters correspond to advocacy coalitions, and belief systems tend to be homogenous within each cluster but quite different across clusters. While the ACF focuses on political elites, this model is also applicable to understanding ideological polarization within the general public—this is also an important area of research (Baldassarri & Gelman, 2008).

The ACF explains the emergence of advocacy coalitions and belief-oriented segregation through a process known as “biased assimilation.” Biased assimilation has been studied by social psychologists for a number of years, and assumes that policy actors tend to systematically filter information through their prior beliefs and discredit information that is not congruent with their belief system (Innes, 1978; Lord et al., 1979; Munro & Ditto, 1997; Munro et al., 2002). Since policy actors with shared beliefs have similar perceptual filters, information exchange, learning, and the development of common views occurs easily among them. Conversely, shared learning is exceedingly difficult between people with conflicting beliefs, since their perceptual filters will cause them to interpret the same piece of evidence differently. This breeds mistrust among people with conflicting beliefs, and causes networks to form and solidify among those with shared perceptual filters.

Thus, biased assimilation influences social learning and network evolution in two ways. The first is that the ease with which two actors are able to learn from each other is directly related to their similarity in beliefs. Second, biased assimilation further constrains learning processes by influencing where policy actors go to gather information and discuss policy issues. These effects are further compounded by a “devil shift” phenomenon that causes actors to emphasize the negative attributes of their competitors, which further inhibits the formation of ties between actors with divergent beliefs (Sabatier et al., 1987).

In modeling these processes, it is important to note that the architecture of a belief system will, in most cases, be quite complex. Not only are many different types of beliefs likely to influence networking decisions, but some beliefs are likely to be constrained or influenced by other beliefs. For example, beliefs that lack a strong empirical component (such as beliefs regarding the appropriate tradeoffs between environmental and economic sustainability in development decisions) may influence an actor’s interpretation of more specific, evidence-based beliefs (such as the reliability of certain economic indicators). The ACF provides a method of defining policy-relevant beliefs for a particular policy subsystem by categorizing beliefs into a three-tier hierarchy comprised of the “deep core,” the “policy core,” and “secondary aspects.” Deep-core beliefs are broad normative values that act as a general guide for political behavior, and are normally applicable to a wide variety of policy subsystems. The policy core includes basic strategies and beliefs concerning a specialized policy area, which are usually subsystem-wide in scope. Secondary aspects include specific policy rules and preferences that are constrained by the policy core, and are more narrow in geographic scope than either deep-core or policy-core beliefs.2

3. Overview of the model

The consequences of the ACF’s behavioural assumptions of belief change and networking are explored via an agent-based model of social learning in an evolving network. This model is built upon a prior model of belief change on a static social network described in Henry (2007). The model is written and implemented in the statistical computing package R (R Core Development Team, 2006), and visualizations make use of the “sna” package written for R (Butts, 2006).

In the model developed here, hypothetical policy actors are allowed to interact and learn within a dynamically-changing policy network. The term “learning” is here used interchangeably with the terms “social learning” and “social influence,” which (as noted above) is defined as change in an agent’s beliefs based on the beliefs of surrounding agents. The assumption here is that policy actors are engaged in a continual process of discussion, persuasion, and adjustment of their own beliefs based on information received from their network partners (i.e., those they are connected to within the network). The most important feature of this simulation is that learning

2 This research uses the ACF convention of labeling these cognitive elements “beliefs,” although it should be noted that the usage of terms varies considerably across literatures. For example, cognitions such as values, attitudes, and certain types of norms also fit well into the ACF hierarchy of beliefs.
processes are constrained by the underlying network structure, which governs opportunities of agents to communicate and learn from one another. This network, in turn, evolves dynamically as beliefs are updated and agents reassess their cognitive similarities and differences with others in the system.

The following discussion provides an overview of the two main components of the model: how agents’ belief systems are structured, and how agents update their network ties.

3.1. Social influence: A simple model of belief systems

The model begins by populating a policy system with a fixed number of agents, \( N \), each of whom is assigned a static “core” belief that may be used to filter evidence or arguments received from the surrounding social environment. This core belief is represented by the variable \( c \), where \( c_i \) represents the core belief of the \( i \)th agent. Core beliefs are assumed to be discrete, nominal variables that take on one of two values, say –1 or 1 (the model dynamics are independent of how \( c \) is represented numerically). In the models presented here, the system is assumed to contain a total of 50 interacting agents (\( N = 50 \)), with a uniform distribution of competing values of \( c_i \) in the population (thus, 25 agents have one level of \( c_i \) and the other 25 agents have a different level of \( c_i \)). Agents are represented graphically as nodes, where the shape of the node (circle or triangle) represents their corresponding value of \( c_i \).

Dynamically-changing beliefs are represented by a continuous variable \( b \), ranging in the interval \([-1, 1]\), where \( b_i \) represents the policy belief of the \( i \)th agent. Representing \( b_i \) as a continuous variable allows for the modelling of beliefs with varying intensity—in this model, \( b_i \) values close to zero are meant to denote more moderate beliefs whereas beliefs further from zero are more extreme. Moreover, the valence of \( b_i \) also represents competing belief orientations in the population. In the initialization of agent beliefs, \( c_i \) is causally prior to \( b_i \), where negative values of \( c_i \) will tend to correspond with negative values of \( b_i \) and vice versa. Thus, \( b_i \) takes on a bimodal distribution in the population, where negative and positive values tend to reflect the competing deep- and policy-core belief orientations that exist in the policy system.

According to the ACF, belief-oriented segregation usually occurs on the basis of policy-core beliefs, and so it is perhaps most appropriate to think of the dynamic belief \( b \) as the policy core, and the core belief \( c \) as the belief type that filters and influences the policy core; namely, the deep core. In this case, the model is useful for illustrating the circumstances under which policy core beliefs solidify within advocacy coalitions. On the other hand, one could also investigate whether (and under what conditions) network segregation takes place on secondary aspects (\( b \)) by viewing the policy core beliefs (\( c \)) as static over time and the primary filter on more specific secondary beliefs. Both approaches are consistent with the ACF’s theoretical expectations, and illustrate how this model may be used to test (and more concretely operationalize) the ACF’s hypothesized dynamics of learning and coalition formation.

Agent beliefs are visualized through the shape, colour, and size of nodes, as represented in Figure 1. Node shapes represent values of \( c_i \)—in this case, agents are either represented as circles (\( c_i = +1 \)) or triangles (\( c_i = -1 \)). Note that, since \( c_i \) is static, node shapes will never change throughout the course of a simulation. The valence of \( b_i \) is represented by node colour: red is used to denote positive values of \( b_i \) and green is used to denote negative values of \( b_i \). Since \( b_i \) is distributed as a function of \( c_i \), in the initial conditions circles will tend to be red and triangles will tend to be green, although this will not always be the case. Finally, node size reflects the intensity of the belief \( b_i \), where smaller nodes have \( b_i \) close to zero and larger nodes have correspondingly larger absolute values of \( b_i \).

![Figure 1: Visualization of agents’ belief systems. Numbers in the center of nodes denote values of \( b_i \).](image-url)
At each time step, all agents are given an opportunity to update their beliefs based on information received from agents with whom they are connected in the network (more details on how this network is structured are given below). Agents learn sequentially, where agents are selected at random to learn from their neighbours until all actors have updated their beliefs. Thus, learning occurs at the level of egocentric networks. When an agent is selected to learn, they become Ego while adjacent agents (actors who have a link to Ego) become persuading agents. Ego has \( d \) persuading agents, where \( d \) is the degree of Ego’s vertex, that is, the number of other agents with whom Ego shares a direct connection.

Before Ego interacts with her \( d \) persuading agents, she has belief \( b_i = B_E \). Each persuading agent simultaneously attempts to convince Ego to update her beliefs to be in closer agreement with that of the persuader. Ego then learns from all of these persuading agents by balancing the competing forces of persuasion and choosing an actual belief change, \( \Delta B_E \).

In the example illustrated in Figure 2, Ego is being persuaded by 5 actors, each of whom wants Ego to adopt their belief. While none of the persuading agents agree entirely with Ego, two of them (PA\(_2\) and PA\(_4\)) share Ego’s underlying belief orientation, since all of their belief scores are positive. The belief orientations of the other three persuading agents are in competition with Ego’s orientation. In Figure 2, note that node size does not represent belief intensity.

In choosing \( \Delta B_E \), Ego first determines the belief change that would best balance out the competing forces of the ideologically-diverse persuading agents. If Ego is a rational information processor, then this is done by simply averaging out the belief change that each persuading agent is pressuring Ego to undertake. Thus, in the rational case,

\[
\Delta B_E (\text{max}) = \frac{\sum_{i=1}^{d} (b_i - B_E)}{d}.
\]

Ego does not, however, have a diffuse prior about the correct belief to adopt. Rather, these persuasive forces must also be balanced by Ego’s prior belief \( (B_E) \). Thus, this “optimal” belief change (given the persuasive forces of Ego’s adjacent agents) represents an upper bound of change rather than an actual change to be undertaken. The next step taken by Ego is to choose, uniformly and at random, an actual belief change \( \Delta B_E \) that lies between zero and \( \Delta B_E (\text{max}) \), inclusive. In the above example,

\[
\Delta B_E (\text{max}) = \frac{(-0.5 - 0.2) + (0.5 - 0.2) + (-0.7 - 0.2) + (1 - 0.2) + (-1 - 0.2)}{5}, \text{ or}
\]
\[
\Delta B_E^{(\text{max})} = \frac{(-0.7) + (0.3) + (-0.9) + (0.8) + (-1.2)}{5}, \text{ or } \\
\Delta B_E^{(\text{max})} = -0.34.
\]

Thus, Ego chooses a random actual belief change to adopt such that her new belief, \(B_E'\), lies somewhere between her initial belief \(B_E\) and the belief that averages out the competing persuasive forces, \(B_E + \Delta B_E^{(\text{max})}\):

\[
B_E' \in [B_E, B_E + \Delta B_E^{(\text{max})}] = [0.2, -0.14].
\]

In this case, there is nearly a 50% chance that ego will shift her underlying belief orientation, from red to green. This is not only because a majority of her persuading agents have competing belief orientations, but also because she began with a fairly moderate belief score.

### 3.3. Social influence: Incorporating biased assimilation

The above specification of learning processes assumes that each agent will give equal weight to the competing beliefs of all other adjacent agents. But what if agents are susceptible to biased assimilation, as the ACF suggests? In this case, Ego will give disproportionate weight—or even a polarizing weight—to persuading agents who share her core beliefs. Conversely, the arguments of persuading agents with competing core beliefs will be perceived as less influential, and will have a comparatively weak effect in Ego’s learning process.

To capture the effect of biased assimilation, consider a new model parameter called the \textit{biased assimilation index}, denoted \(\beta\). This index may take on any value between –1 and 1 (inclusive), and depending on the value of \(\beta\) Ego may be a rational learner as described above (where all agents are given equal weight), Ego may give diminishing weights to forces exerted by agents with different values of \(c_i\), or forces from persuading agents with different levels of \(c_i\) may cause Ego’s beliefs to polarize away from the persuading agents’ \(b_i\) values.

To incorporate the role of biased assimilation in the model, Equation 1 is reformulated into a more general form:

\[
\Delta B_E^{(\text{max})} = \frac{\sum_{i} (\beta_i \cdot (b_i - B_E))}{d},
\]

where \(\beta_i = \beta\) if \(c_{\text{ego}} \neq c_i\), and

where \(\beta_i = 1\) if \(c_{\text{ego}} = c_i\).

Thus, when \(\beta = 1\) we have the rational learning scenario described above. When \(\beta\) lies in the interval \([0,1)\), Ego will give smaller weight to persuading agents with differing levels of \(c_i\), where values of \(\beta\) closer to zero indicate diminishing weights. When \(\beta\) is negative, then ideological forces from persuading agents with different underlying \(c_i\) orientations will cause Ego to choose beliefs that move away from her persuading agents.

It should be noted that these \(\beta_i\) weights are a function of similarity in core beliefs, and can be viewed as the degree to which Ego devalues information received from a persuading alter with a competing core belief. It is plausible, of course, that agents’ valuation of information from certain sources is also a function of other actor attributes, such as qualifications, history of providing useful policy information, or disciplinary identity. It is possible that some of these influences could be modelled within the current framework. For instance, by viewing the core variable \(c\) as an indicator of whether an agent is a scientist or a political decision-maker, \(\beta\) could then be used to model communication difficulties between scientific and decision-making communities, or the degree to which members of each community devalue information received from outside their community. Other types of influences could be directly specified, but at the cost of increasing model complexity.

To illustrate how these weights influence the learning process, consider once again the Ego from our above example (see Figure 2). How would we revise our estimates of her potential belief change if biased assimilation were at work? Suppose, for example, that \(\beta = 0.8\). Then,
Thus, under biased assimilation, Ego will choose:

$$B_E' \in [0.2, -0.028].$$

The likelihood that Ego will shift her underlying orientation from positive (red) to negative (green) is now very small, despite the fact that a majority of her persuading agents are green. Disproportionate weight has been given to the arguments of the red minority.

3.4. Network formation: Initial conditions

At this point it is clear that the dynamic process of belief change depends largely on the underlying structure of the network. Many agent-based models assume very simple types of lattice networks (e.g., Axelrod, 1997, Johnson & Huckfeldt, 2005), and in many cases these structures do not change during the course of a simulation. However, more recent work demonstrates that equilibrium conditions can change radically when different assumptions are made about underlying network structures (Ohtsuki et al., 2006). Thus, a key point of departure is to make a decision about what the underlying network should look like.

This model assumes that, in the initial conditions, agents are connected with each other in an Erdős-Renyi random graph, as depicted in Figure 3. Specifying an initial network requires a single parameter $p$; to create the initial network, each dyad (possible actor pair) is independently populated with a single network link with probability $p$, and not populated with a link with probability $1 - p$. Density of the network is proportional to $p$. Note also that this method of specifying an initial network assumes no correlation between agents’ belief systems and the existence of linkages between them; thus the simulations are able to explore whether agents will begin to sort themselves into advocacy coalitions as they learn and rewire their surrounding network.

Using a random graph in the initial condition allows us to consider many types of network structures as a starting point. Even though many realistic structures, such as power-law graphs, will be realized with extremely low probability, the use of random graphs allows for a stochastic starting point that controls for initial network topology over many distinct simulation runs. This is an advantage over approaches that always use the same network (such as a lattice structure) without introducing any random perturbations of that initial structure. Nonetheless, future elaborations of this model should consider alternative types of random initial networks.
3.5. Network formation: The rewiring process

After agents are given an opportunity to update their beliefs, they are given the opportunity to rewire their local network. According to the ACF, this occurs primarily through a process of aversion from dissimilar alters due to biased assimilation and the devil shift—the process of link formation, on the other hand, is not clearly explained within the framework. Although ACF scholars generally assume that tie formation occurs through an attraction to others with shared belief systems, the model outlined here assumes that new partners are chosen uniformly at random. This yields a more conservative view on how coalitions are potentially formed through the network rewiring process, and restricts our focus only to these theoretical processes that are made explicit within the ACF.

All network rewiring is performed simultaneously by all agents. The probability that any given tie is deleted is governed by a new model parameter $\alpha$, which takes on a value in the interval $[0,1]$. Smaller values of $\alpha$ indicate that agents are subject to only a small bias towards cutting ties with ideologically-dissimilar network partners; larger values correspond with larger probabilities of tie deletion when two agents have divergent beliefs.

Let $D(i,j)$ denote the probability that the link connecting agents $i$ and $j$ will be deleted. Then,

$$D(i, j) = \alpha \left(1 - \frac{1}{e^{b+1}}\right).$$

Thus, large differences in the $b$ values of two connected agents corresponds to a high probability that the tie will be deleted—however, this effect is attenuated or even negated if alpha is close to zero.

Each agent simultaneously assesses their existing links with other agents in the network, and terminates each link with probability $D(i,j)$. Each link is therefore subject to termination twice—once by each agent connected in the relationship. Deleted relationships are then rewired by the agent that terminated the link, who selects a new network partner uniformly at random from the set of all agents who lie outside of their network neighbourhood. Note that it is possible in this process for a tie to be deleted and subsequently reformed; this is a rare occurrence, however, and will not influence the main results presented here.

4. Simulation runs

Approximately 1,000 simulations were run under different initial values of $p$, yielding initial networks with expected densities ranging from about 0.05 (very sparse networks) to about 0.25 (fairly dense networks). For each level of $p$, simulations were run across the full range of possible $\alpha$ and $\beta$ values, which allows for an exploratory test of the ACF model of learning and coalition formation. For any given combination of $\alpha$ and $\beta$, several separate simulations were run to check the stability of equilibrium conditions.

Each simulation was run to convergence. Model convergence is assumed to occur at time $t$ if the maximum belief change (in absolute value terms) for any agent between time steps $t-1$ and $t$ does not exceed 0.005. This ensures that no substantial amount of belief change is likely to occur, although the simulation will never stabilize completely due to the randomness built into the learning and network rewiring processes.

4.1. Summary statistics

There are a number of key features of interest that may be used as a starting point to summarize and assess simulation results. What follows is a brief description of these features, and how they will be empirically identified in networks that have run to convergence.

4.1.1. Network segregation and ideological rift

The primary result of interest is whether or not model runs with a given set of input parameters tend to produce belief-oriented network segregation as described within the ACF. We are thus interested in the extent to which agents’ beliefs have converged to a state where beliefs are relatively similar within communities of shared levels of the core belief, $c$, and different across communities of actors with competing values of $c$. This ideological polarization across densely connected network clusters is captured by a statistic called “ideological rift.” In order
for ideological rift to be positive, we must have a convergence state where mean belief scores in each community are different, and there exists little or no overlap between agent belief scores across differing core beliefs.

Suppose that \( b_1 \) is the mean belief score for agents with core belief \( c_i = 1 \), and \( b_2 \) is the mean belief score for agents with core belief \( c_i = -1 \). The standard deviation of beliefs within each group is given by \( \sigma_1 \) and \( \sigma_2 \), respectively. Then ideological rift, denoted \( R \), is given by:

\[
R = |b_1 - b_2| - (\sigma_1 + \sigma_2).
\] (4)

Ideological rift is a continuous measure that identifies the presence or absence of belief-oriented segregation, and reflects the intensity of the aggregated ideological conflict between advocacy coalitions.\(^3\)

4.1.2. Ideological consensus

An alternative to ideological polarization is a convergence state where roughly all agents share the same belief. The degree of ideological consensus that exists in the entire network is measured simply by the variance of all agents’ beliefs. Variance scores approaching zero indicate ideological consensus, whereas increasing variance in beliefs indicate decreasing consensus.

This parameter may be summarized by ideological rift: when ideological rift is close to zero, the degree of ideological consensus in the network is high.

4.1.3. Ideological heterogeneity

When ideological rift is large, we have network segregation. When ideological rift is close to zero, we have consensus. When ideological rift is strongly negative, we have “ideological heterogeneity.” Ideological heterogeneity suggests that consensus does not exist within the network, however the network also does not exhibit belief segregation. As noted above, this situation represents a state where there is significant overlap in beliefs across communities, but variance in beliefs across the network as a whole remains large.

![](image)

Figure 4: Ideological rift for selected \( \beta > 0 \) and varying levels of \( \alpha \). The x-axes represent increasing levels of \( \alpha \) (from 0 to 1); y-axes represents levels of ideological rift in the convergent state.

\(^3\) Ideological rift is positive when there is no overlap within one standard deviation of the mean belief score in each partition. To see where the equation comes from, suppose that \( b_H \) is the larger mean belief score for the two communities, and \( b_L \) is the smaller mean belief score. The standard deviation of beliefs within each community is given by \( \sigma_H \) and \( \sigma_L \), respectively, then ideological rift is given by:

\[
R = (b_H - \sigma_H) - (b_L + \sigma_L) = (b_H - b_L) - (\sigma_H + \sigma_L).
\]
5. Simulation results

Simulations reveal that the emergence of belief-oriented segregation is very sensitive to assumptions regarding the degree to which policy actors are rational information processors. Generally speaking, it is difficult to model the emergence of ideological polarization when agents are not biased assimilators; on the other hand, biased assimilation seems to be an important driver of both coalition formation and the persistence of existing coalitions.

5.1. The role of alpha

The degree of belief homophily at work (parameter $\alpha$) seems to have a negligible influence on ideological rift when actors are assumed to give even small weight to persuading agents with competing core beliefs. On the other hand, when agents with competing core beliefs are discounted (when $\beta$ is small), higher values of $\alpha$ may exacerbate belief-oriented segregation, yielding higher ideological rift values. Figure 4 illustrates this result; it displays a scatter plot of $\alpha$ values against ideological rift, for $\beta$ levels from 0.2 to 1.0, in 0.2-point increments. Recall that the learning processes represented in these scatter plots are rational, in the sense that individual actors will use information from all persuading agents to update their beliefs, although for smaller levels of $\beta$ smaller weights are given to ideological forces from agents with differing core beliefs ($c_i$). However, the key here is that competing beliefs will still converge towards global consensus, as illustrated by ideological rift values that are close to zero in almost all cases.

This is a surprising result. It demonstrates the importance of diffusion processes in balancing, even dominating, network-rewiring forces that move a network towards clustering around segregated communities. In the case of small $\beta$ parameters, diffusion may be slow but nonetheless still leads to states of ideological convergence states in the long run. Indeed, the primary difference between model runs with small but positive $\beta$ values, and models without biased assimilation ($\beta = 1$), is that smaller values of $\beta$ lead to longer convergence times.

5.2. Beta and the emergence of belief-oriented segregation

In order to model the emergence of networks where agents sort themselves into belief-oriented coalitions, it is essential to have at least some degree of belief polarization at work (i.e., $\beta < 0$). In nearly all simulation runs using negative $\beta$ parameters, convergent networks were characterized by positive ideological rift values. Figure 5 displays two representative networks obtained when negative $\beta$ values are assumed.

This pattern of clear belief-oriented network segregation is observed consistently across simulation runs, when ideological forces from agents with divergent core beliefs cause polarization rather than some degree of consensus. These patterns lend strong support for the ACF view that biased assimilation is a necessary condition for the emergence of advocacy coalitions.

The far right panel of Figure 5 provides further evidence of the importance of biased assimilation in driving ideological rift. In this scatterplot, positive $\beta$ values (no biased assimilation) tend to yield ideological consensus in convergent networks, whereas negative $\beta$ values consistently yield positive ideological rift and, in many cases, extreme belief polarization within networks. While some of the variance in ideological rift for negative values of $\beta$ may be explained by $\alpha$, the interactions between $\alpha$ and $\beta$ are clearly overshadowed by the direct influence that biased assimilation has in structuring polarized networks.

6. Conclusion

This paper presents a simplified first analysis of a simulation results from an agent-based model of learning that incorporates a model of co-evolving beliefs and policy network structures. This paper operationalizes a general learning model that is consistent with the ACF. While the ACF model of learning has been subject to extensive theoretical development and some empirical testing, this is one of the first studies to directly test these models within a dynamic framework. Although agent-based models such as this will need to be coupled in the future with
empirical data on policy networks, the results illustrate conceptually the importance of considering both network dynamics along with cognitive biases in understanding why segregated networks emerge and persist over time.

Findings support the core ACF hypothesis, that biased assimilation—or a polarization in beliefs when individuals with competing “core” values or beliefs exchange information—is an important driver of polarization in networks. Surprisingly, the degree to which policy actors seek to avoid communication with dissimilar actors is much less important than the learning processes that govern how beliefs are diffused within a policy network.

The dynamics of the model outlined here depend on only two parameters. This boosts tractability and generalizability, but on the other hand also leaves room for building in greater realism. One possible elaboration is to provide agents with feedback that signals whether their beliefs are correct. As it stands, this model is perhaps appropriate for modeling belief change in policy realms characterized by high levels of uncertainty and/or prominent systems of normative beliefs. But it will be important to also understand how groups of agents may also come to solve complex problems and arrive at the “correct” or “optimal” policy solutions.

Ultimately, this is all about decision-making. Within the ACF, as well as in other frameworks, decisions are made where there must be some degree of consensus. This could be the adoption of a policy or agreement to stay with a quota for commons harvesting. The question then becomes, what are the dynamics by which subsystem actors come to a consensus? When do they remain divided, or even become more polarized? This paper helps us move towards explicit models of these processes, emphasizing in particular the importance of co-evolutionary models of network change and social learning. These formal models will then enable us to explore the types of institutional arrangements that may help to promote consensus in the face of complex policy problems, either by intervening in the structure of policy networks or through arrangements that attenuate the negative effects of biased assimilation.

**Literature cited**

Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton University Press.

Baldassarri, D., & Gelman, A. (2008). Partisans without constraint: Political polarization and trends in American public opinion. *American Journal of Sociology, 114*(2), 349-399.

Butts, C. T. (2006). sna: Tools for social network analysis. Retrieved from http://erzuli.ss.uci.edu/R.stuff.

Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., Jäger, J., et al. (2003). Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences, 100*(14), 8086-8091.

Dietz, T., & Henry, A. D. (2008). Context and the commons. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 7821-7826.

Henry, A. D. (2007). Simulating the evolution of policy-relevant beliefs: Can rational learning lead to advocacy coalitions? *The challenge of self-governance in complex, globalizing economies*, IFE Working Paper Series 47 (pp. 135-160). Freiburg, Germany: Institute of Forestry Economics, University of Freiburg.
Henry, A. D. (2009). The challenge of learning for sustainability: A prolegomenon to theory. *Human Ecology Review, 16*(2), 131-140.

Henry, A. D., Lubell, M., & McCoy, M. (2010). Belief systems and social capital as drivers of policy network structure: The case of California regional planning. *Journal of Public Administration Research and Theory*, in press. doi:10.1093/jopart/muq042.

Henry, A. D., Pralat, P., & C.-Q. Zhang. (2011). Emergence of segregation in evolving social networks. *Proceedings of the National Academy of Sciences, 108*(21), 8605-8610.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America, 101*(46), 16385-16389.

Innes, J. M. (1978). Selective exposure as a function of dogmatism and incentive. *The Journal of Social Psychology, 106*(2), 261-265.

Johnson, P. E., & Huckfeldt, R. (2005). Agent-based explanations for the survival of disagreement in social networks. *The Social Logic of Politics* (pp. 251-268). Philadelphia: Temple University Press.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*(11), 2098-2109.

McNie, E. C. (2007). Reconciling the supply of scientific information with user demands: An analysis of the problem and review of the literature. *Environmental Science & Policy, 10*(1), 17-38.

Munro, G. D., & Ditto, P. H. (1997). Biased assimilation, attitude polarization, and affect in reactions to stereotype-relevant scientific information. *Personality and Social Psychology Bulletin, 23*(6), 636-653.

Munro, G. D., Ditto, P. H., Lockhart, L. K., Fagerlin, A., Gready, M., & Peterson, E. (2002). Biased assimilation of sociopolitical arguments: Evaluating the 1996 U.S. presidential debate. *Basic and Applied Social Psychology, 24*(1), 15.

Ohtsuki, H., Hauert, C., Lieberman, E., & Nowak, M. A. (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature, 441*(7092), 502-505.

R Development Core Team. (2006). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org.

Sabatier, P. A., & Jenkins-Smith, H. C. (1999). The advocacy coalition framework: An assessment. *Theories of the policy process* (1st ed., pp. 117-166). Westview Press.

Sabatier, P. A., & Jenkins-Smith, H. C. (Eds.). (1993). *Policy change and learning: An advocacy coalition approach*. Westview Press.

Sabatier, P., Hunter, S., & McLaughlin, S. (1987). The devil shift: Perceptions and misperceptions of opponents. *The Western Political Quarterly, 40*(3), 449-476.

Schneider, M., Scholz, J., Lubell, M., Mindruta, D., & Edwardsen, M. (2003). Building consensual institutions: Networks and the national estuary program. *American Journal of Political Science, 47*(1), 143-158.

Weible, C. M., & Sabatier, P. A. (2005). Comparing policy networks: Marine protected areas in California. *Policy Studies Journal, 33*(2), 181-201.