Proceedings of Agent 2007, Chicago, IL, pp. 221-230

MODELING COLLECTIVE COGNITIVE CONVERGENCE

H.V. PARUNAK,* T.C. BELDING, R. HILSCHER, S. BRUECKNER
NewVectors Division of TechTeam Government Solutions, Inc.

ABSTRACT

When the same set of people interact frequently with one another, they tend to think more and more along the same lines, a phenomenon we call "collective cognitive convergence" (C^3). In this paper, we discuss instances of this phenomenon and why it is advantageous or disadvantageous; review previous work in computational social science and evolutionary biology that sheds light on C^3; define a computational model for the convergence process and a quantitative metric that can be used to study it; report on experiments with this model and metric; and suggest how the insights from this model can inspire techniques for managing C^3.

Keywords: Groupthink, cognitive convergence, social simulation

INTRODUCTION

When the same set of people interact frequently with one another, they tend to think more and more along the same lines. We call this phenomenon "collective cognitive convergence" (C^3), since the dynamics of the collective lead to a convergence in cognitive orientation.

C^3 is seen in many different contexts, including research subdisciplines, political and religious associations, and even persistent adversarial configurations such as the cold war. Tools that support collaboration, such as blogging, wikis, and communal tagging, make it easier for people to find and interact with others who share their views, and thus may accelerate C^3. This efficiency is sometimes desirable, since it enables a group to reach consensus more quickly. For instance, in the academy, it enables coordinated research efforts that accelerate the growth of knowledge.

But convergence can go too far, and lead to collapse. It reduces the diversity of concepts to which the group is exposed and thus leaves the group vulnerable to unexpected changes in the environment. Here are two examples.

In academia, specialized tracks at conferences sometimes become unintelligible to those who are not specialists in the subject of a particular track, and papers that do not fit neatly into one or another subdiscipline face difficulty being accepted. The subdiscipline is increasingly sustained more by its own interests than by the contributions it can make to the broader community or to society at large.

* Corresponding author address: H. Van Dyke Parunak, NewVectors, 3520 Green Court, Suite 250, Ann Arbor, MI 48105; e-mail: van.parunak@newvectors.net.

In military operations, the force-on-force orientation developed during the Cold War left both the former Soviet Union and the United States ill-prepared to deal with insurgencies and asymmetric warfare.

Groups that have undergone cognitive collapse will only produce output conforming to their converged set of ideas, and will be unable to conceive or explore new ideas. In the worst case, collapse may lead a group to focus its attention on a cognitive construct with little or no relation to the real world. For example, highly specialized academic disciplines become increasingly irrelevant to people outside of their own circle.

We became interested in this phenomenon by observing the increasing balkanization of the research field of multi-agent systems.[1] Since we work in the area of multi-agent simulation, it occurred to us that some light might be shed on the phenomenon, and on how it can be managed, with a multi-agent model. This paper presents some preliminary results.

Section 2 discusses previous work related to our effort. Section 3 describes our model, and a metric that we use to quantify C^3. Section 4 outlines a series of experiments that exhibit the phenomenon and explore possible techniques for managing it. Section 5 suggests directions for further research, and Section 6 concludes.

PREVIOUS WORK

Our research on C3 builds on and extends two previous bodies of work, in computational social science and evolutionary biology.

In computational social science, our work merits comparison with Axelrod's adaptive culture model (Axelrod 1997) and its antecedents. What he calls "culture" corresponds to our notion of an agent's cognitive interests. Axelrod studied the transmission of cultural traits, represented as the elements of a numerical vector, between neighboring agents distributed on a 2-D lattice. Agents had a chance, proportional to their cultural similarity, of copying a trait from a neighbor. He found that a small number of large and stable homogeneous regions, or "cultures", would form. His work exhibits the emergence of disjoint regions of cultural (cognitive) homogeneity as agents interact with those who are adjacent to them spatially. Our model differs from his in several ways.

- His agents always interact with the same neighbors. Our agents can change their interaction partners as a result of the system's dynamics.

- His agents interact on the basis of spatial contiguity. Our model offers a much wider range of drivers for interaction.

- The nature of the interaction in his model is the same at every round. Our model modulates the strength of the interaction by the size and convergence of the emerging group.

[1] We are grateful to Simon Thompson for initial discussions that led to this project.

C3 can be considered a cultural analog of biological speciation, and so we look for insight to research in this field as well (see (Futuyma 1998) for a review). The most commonly proposed speciation mechanisms are allopatric speciation, sympatric speciation, and parapatric speciation. In allopatric speciation, genetic barriers gradually evolve between two or more geographically isolated species. This might happen for instance between organisms living on separate islands. These barriers could evolve either through natural selection or through other means such as the founder effect (i.e., differences in genes between populations due to the small sample sizes of the founding populations). This is analogous to different specialized communities developing in isolation from each other in C3. One specific type of natural selection that can cause speciation is sexual selection (Fisher 1930; Andersson 1994), a social process by which female mate choice influences the evolution of male traits. In extreme cases, this can become a runaway process that leads to extravagant features that are detrimental to survival (thus leading to a shorter lifetime and fewer opportunities to mate). Similarly, in C3 a social process can lead to the development of academic specializations with little practical relevance.

In parapatric speciation, there is no discrete barrier between populations; individuals are distributed along a geographic continuum and are separated by distance. Finally, sympatric speciation refers to instances where a single population with no gene flow barriers divides into separate species. While the relative importance and frequency of these speciation mechanisms in nature are still heavily debated, the mathematical prerequisites for each mechanism have been extensively studied; this work could be adapted to predict when and how C3 will develop, and how it can be prevented.

There has also been much theoretical work done to study the amount of gene flow or migration that is necessary to prevent isolated populations of organisms from diverging or losing diversity due to genetic drift, or sampling error (Hartl and Clark 1989). Sewall Wright argued in his Shifting Balance Theory that a subdivided population with intermittent migration could exhibit more rapid evolutionary change than a single cohesive breeding population (Provine 1986). The mathematical frameworks for studying migration could be applied to modeling the exchange of ideas or individuals between groups in C3, and the amount of exchange that is necessary to prevent intellectual isolation.

A MODEL AND METRIC

We have constructed a simple multi-agent model of C^3 to study this phenomenon. Our model represents each participant's interests as a binary vector. Each position in the vector corresponds to an atomic interest. A '1' at a position means that the participant is interested in that topic, while a '0' indicates a lack of interest. At each step, each participant

- identifies a neighborhood of other participants based on some criteria (which may include proximity between their interest vectors, geographical proximity, or proximity in a social network),
- learns from this neighborhood (by changing an interest j currently at 0 to 1 with probability $p_{interest}$ = proportion of neighbors having interest j set to 1), and
- forgets (by turning off an interest j currently at 1 to 0 with probability $1 - p_{interest}$).

One boundary condition requires attention. If an agent has no neighbors, what should $p_{interest}$ be? We take the view that interests are fundamentally social constructs, persisting only when maintained. Thus an isolated agent will eventually lose interest in everything, and in our model, a null community leads to $p_{interest} = 0$ for all interests. Alternative assumptions are certainly possible, and would lead to a different model.

We need a quantitative measure of agent convergence to study C^3 systematically. To derive our measure, we cluster the population hierarchically based on cognitive distance between agents (in our case, the Jaccard distance between their interest vectors). Each node of the resulting cladogram forms at a specific distance (the "diameter" of the cluster represented by that node). The root has the highest diameter. In a random population of agents, the distances at which lower-level nodes join the tree is not much less than the diameter of the root (Figure 1), while in highly converged populations, the diameters of lower-level nodes are much less than the diameter at the root (Figure 2, where agents grouped at diameter 0 have identical interest vectors). Thus we compute the ratio of node diameter to root diameter (the "min-max ratio") for each node, and use the median of this ratio as a measure of overall system convergence. A ratio of 0 (as in Figure 1) means that more than half of the agents belong to groups within which all interest vectors are identical.

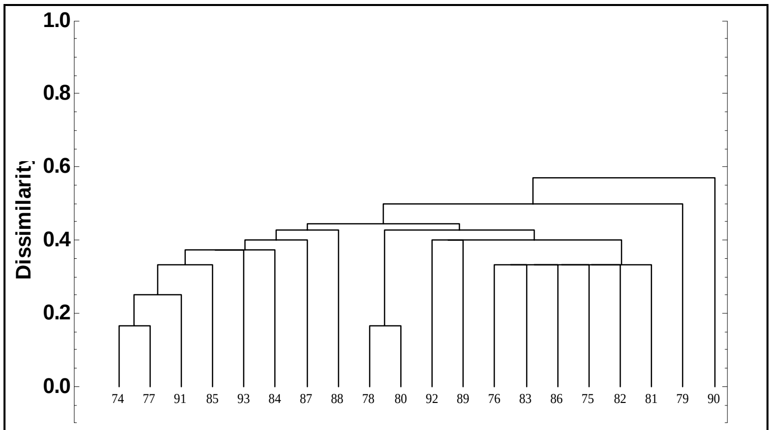

Figure 1 Cladogram of random interest vectors. The median ratio of the dissimilarity at which a node joins the tree to the dissimilarity of the root is 0.583.

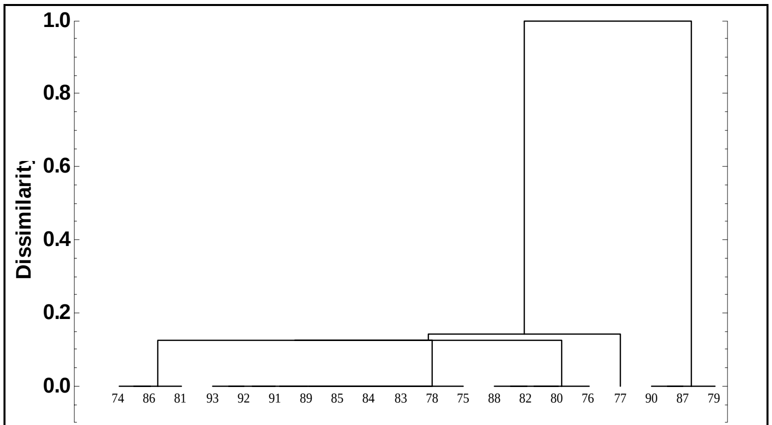

Figure 2 A highly converged population, whose median min-max ratio is 0

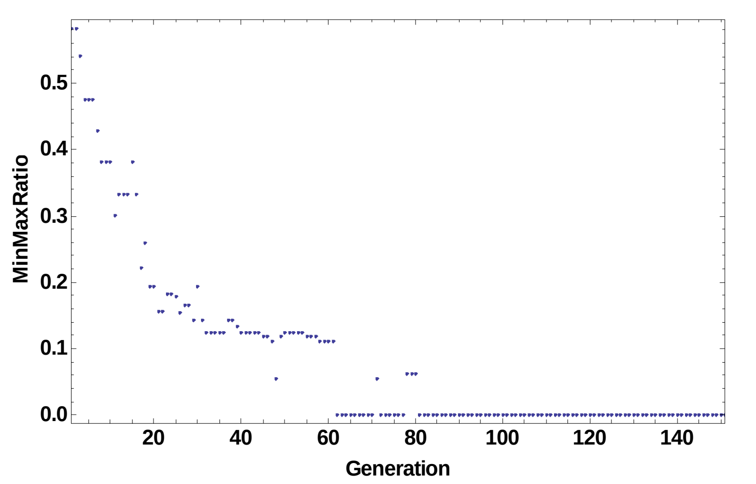

Figure 3 Evolution of 20 agents with length-10 interest vectors, neighborhoods defined by similarity > 0.5

Figure 3 shows the behavior of this measure over a sample run of the system with 20 agents and interest vectors of length 10, where the probability of learning and forgetting is equal, and where agents are considered to be in the same group if the similarity between their interest vectors (the similarity threshold) is greater than 0.5. It takes only about 80 generations for the median min-max ratio to reach 0. (A generation consists of selecting one agent, choosing its neighbors, choosing with equal probability whether it shall attempt to learn or forget, selecting a bit in its interest string at random, then if it is learning and the bit is 0, flipping the bit with probability p_{learn} * $p_{interest}$, or if it is forgetting and the bit is on, flipping the bit with probability p_{forget} * (1 $p_{interest}$.)) Figure 2 shows the state of this system at generation 300. By generation 370 it has collapsed into two groups of completely homogeneous agents of sizes 3 and 17 respectively.

SOME EXPERIMENTS

Armed with this model and metric, we can explore the dynamics of C[3] under a variety of circumstances. As we might expect, forming neighborhoods based on similarity of interest leads to rapid cognitive convergence. But surprisingly, other sorts of neighborhoods also lead to convergence.

Things that Don't Work

We might think that highly tolerant agents, those that consider all agents their neighbors, might be more robust to convergence. Figure 4 shows the evolution of the same population of agents when two agents consider one another neighbors if their similarity is greater than 0 (that is, they have at least one bit position in common). This configuration might be a model for a conference that has only plenary sessions. The population still collapses.

Perhaps the problem is that as agents converge, their neighborhoods increase in size. Figure 5 shows the effect of

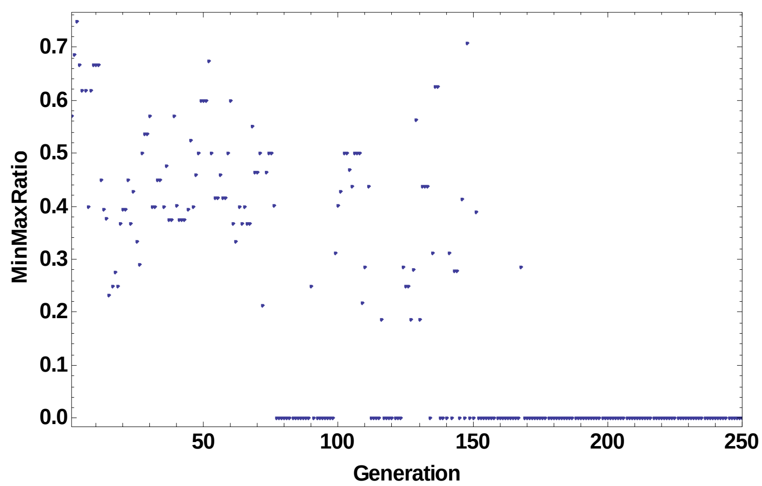

Figure 4 Evolution with similarity threshold = 0

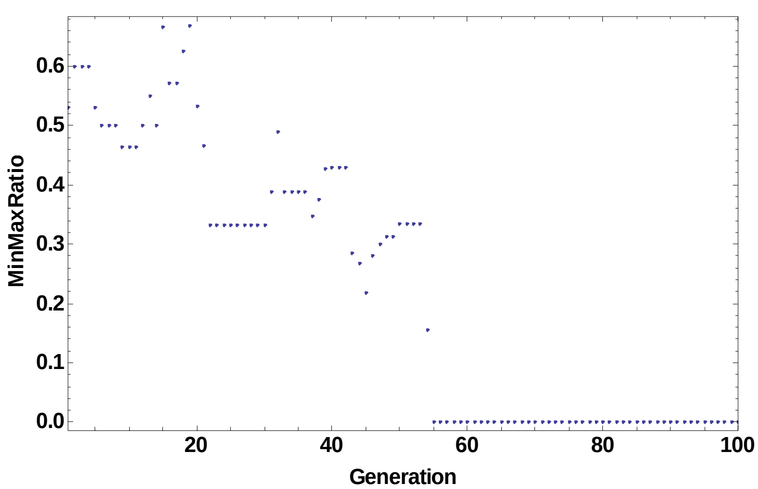

Figure 5 Fixed-size agent neighborhoods (four closest agents)

defining an agent's neighborhood at each turn as the group of four other agents that are closest to it. This configuration models a conference with separate tracks. Though agents base their adaptation at each turn on only 20% of the other agents, the population still collapses.

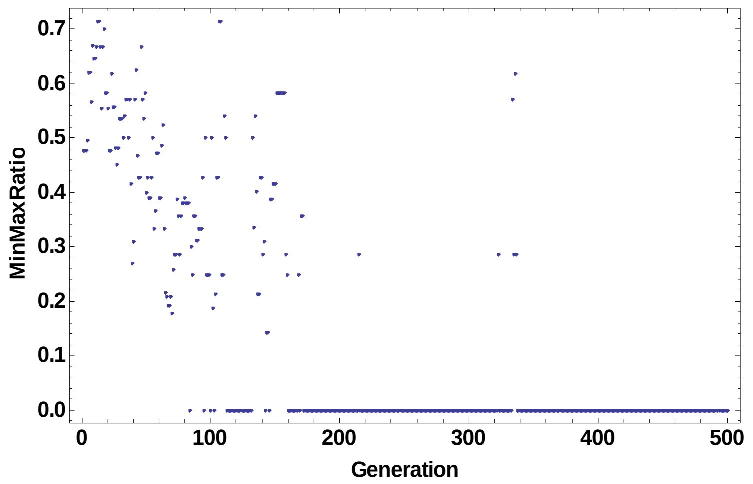

Figure 6 Neighborhoods of four randomly-chosen agents

Figure 6 shows an even more radical approach. Here an agent's neighbors at each step are four randomly chosen agents. Imagine a conference at which papers are assigned to tracks, not by topic, but randomly. In spite of the mixing that this random selection provides, the population again collapses.

These figures differ in how long it takes the system to converge to a min-max ratio of 0. The time to convergence is highly variable, even within a single configuration. Repeated runs show that we should not assume that because (say) Figure 5 converges faster than Figure 4, small groups will always lead to faster convergence than highly tolerant agents. The one constant across all runs is that the system does converge, in fewer than 500 generations (often far fewer).

Introducing Variation

The collapse of agent interests is due to the lack of any mechanism for introducing variation. Once the population loses the variation among agents, it cannot regain it. We have explored three mechanisms for adding variation to the population: random mutation, curmudgeons, and interacting subpopulations.

The simplest approach is mutation. At each generation, with some small probability p_{mutate}, after learning or forgetting, the active agent selects a bit at random and flips it. This mechanism models spontaneous curiosity on the part of agents. Figure 7 shows an extended run with parameters the same as in Figure 3 (neighborhoods defined by a similarity threshold of 0.5), but with $p_{mutate} = 0.03$. Mutation is certainly able to reintroduce variation, but the level is critical. If mutation is too low (say, 1%), it

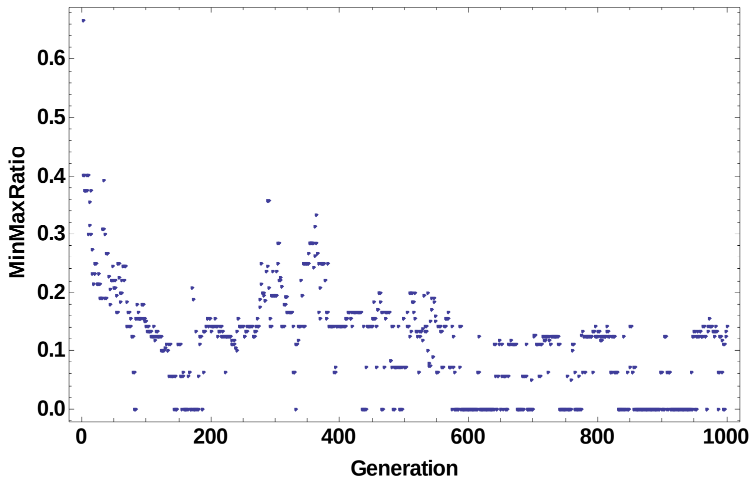

Figure 7 Adding 3% mutation

is unable to keep up with the pressure to convergence, while if it is too high (10%), the community does not exhibit any convergence at all (and in effect ceases to be a community). The nature of its contribution follows a clear pattern. When it is in the critical range, the system occasionally collapses to a min-max ratio of 0, but then discovers new ideas that reinvigorate it.

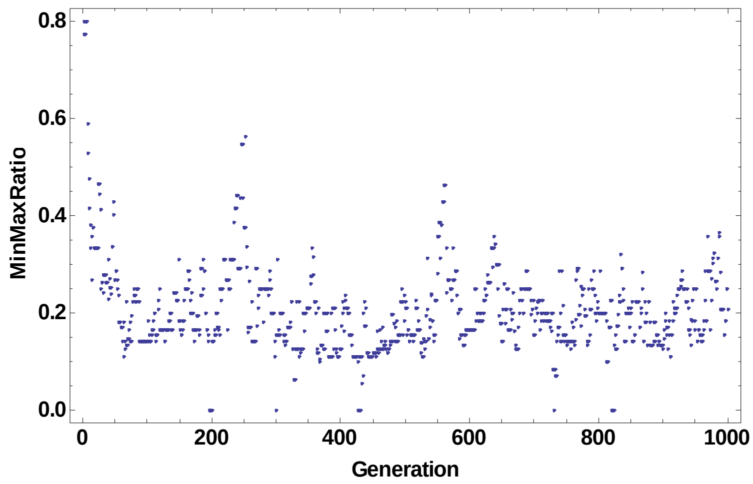

Figure 8 10% curmudgeons

A curmudgeon is a non-conformist, someone who regularly questions the group's norms and assumptions. Recall that ordinarily agents learn by flipping a 0 bit to 1 with probability $p_{interest}$, the proportion of neighbors that have the bit on, and forget by flipping a 1 bit with probability equal to 1 – $p_{interest}$. To model curmudgeons, when an agent decides to learn or forget, with probability p_{cur},, it reverses these probabilities. That is, its probability of forgetting when it is curmudgeonly is $p_{interest}$ (instead of 1 - $p_{interest}$ in the non-curmudgeonly state), and its probability of learning is 1 – $p_{interest}$.

Figure 8 shows the effect of 10% curmudgeons, again with the baseline configuration of Figure 3. The system clearly converges, but seldom reaches a min-max ratio of 0. Furthermore, p_{cur} can achieve this balancing effect over a much wider range than p_{mutate}. As much as researchers may resent reviewers and discussants who "just don't get it," curmudgeons are an effective and robust way of keeping a community from collapsing.

The third source of variation is even more robust. So far, our agents have chosen a new set of neighbors at every step, based on their current set of interests. What would happen if we assign each agent to a fixed group at the outset, using a fixed similarity threshold that allows groups of various sizes to form?

If the threshold is very high, each agent will initially be a group unto itself. With no neighbors to reinforce its interests, it will begin to forget them, and the agents will independently approach the fixed point of an all-zero interest string.

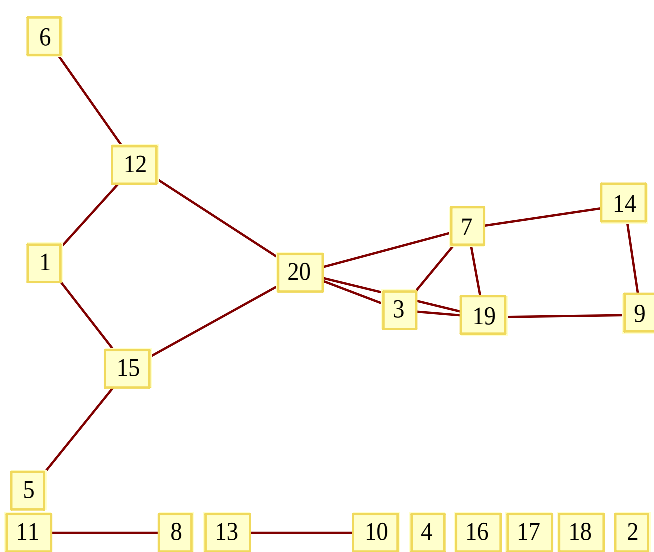

Figure 9 Neighborhood relations for threshold of 0.5

If the threshold is very low, all agents will form one large group, and converge as in Figure 4.

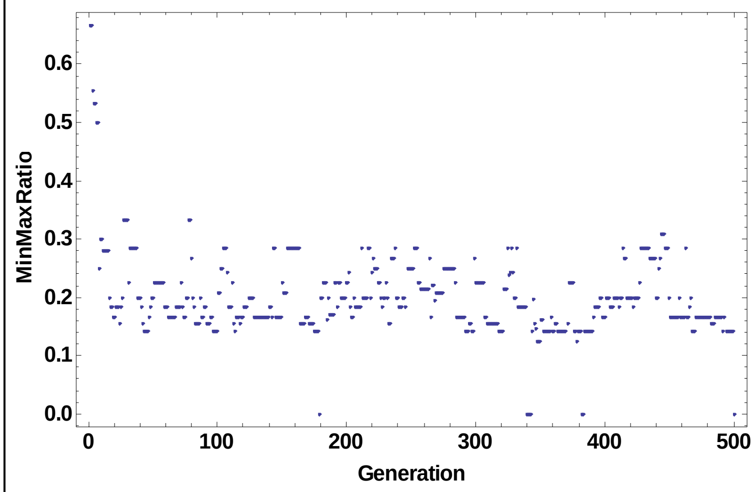

Figure 10 Fixed neighborhoods induced by threshold 0.5

For intermediate thresholds, the agents form a number of neighborhoods. Importantly, some agents ("bridging agents") belong to more than one neighborhood. Figure 9 is a graph of the agents, with an edge between two agents if those agents are neighbors of one another. Because neighborhoods are fixed over the run, each neighborhood can converge relatively independently of the others, but the bridging agents (in this case, notably agent 20) repeatedly displace each neighborhood's equilibrium with the emerging equilibrium of another group, a phenomenon noted by Page (Page 2007). As a result, the system shows convergence without collapse (Figure 10). This mechanism, like curmudgeons and unlike mutation, provides robustness against intermittent collapse. This system reflects a community with subdisciplines, but subdisciplines that recognize the value of members who bridge with other subdisciplines and exchange ideas between them. Such members are likely to be tolerated better by subgroups than would curmudgeons, because the source of the variation introduced by the bridging individuals is perceived as resulting from their multidisciplinary orientation rather than their orneriness

NEXT STEPS

Our simple model has shown a surprisingly rich space of behaviors. A number of directions for further work suggest themselves. For example:

- How can convergence be monitored in practice? Our metric, while effective for simulation, is impractical for monitoring actual groups of people. One might monitor the amount of jargon that a group uses, or lack of innovation, as indicators of convergence.

- We have suggested that convergence is a two-edged sword. What is the ideal degree of convergence, to allow the production of specialist knowledge without compromising the ability to escape collapse?

- How does convergence vary with group size? Recent work (Palla, Barabási et al. 2006) suggests that convergence in small groups requires specialized knowledge, while convergence in large groups requires a general knowledge base.

- We have assumed homogeneous tendencies to learn, forget, mutate, or behave curmudgeonly over all agents. How does the system respond if agents vary on these parameters? In particular, what is the impact of these parameters for bridging individuals in comparison with non-bridging individuals?

CONCLUSION

It is natural for groups of people to converge cognitively. This convergence facilitates mutual understanding and coordination, but if left unchecked can lead the group to collapse cognitively, becoming blind to viewpoints other than their own. Experiments with a simple agent-based model of this phenomenon show that seemingly obvious mechanisms do not check this tendency. In the domain of academic conferences, these well-intended mechanisms include plenary sessions, special tracks, or even random mixing. A source of variation must be introduced to counteract the natural tendency to converge. Mutation is effective if just the right amount is applied, but tends to let the system intermittently collapse. Curmudgeons are more robust, but socially distasteful. Perhaps the most desirable mechanism consists of bridge individuals who provide interaction between individually converging subpopulations.

Further understanding of C^3 could give important guidance in monitoring and managing collaboration. For example, consider a team of analysts searching for information.

- If a group's searches are sparsely distributed in search space, guide more analysts to join this group to cover more areas in this search space.
- If a group's searches are not specific enough, artificially promote the splitting of groups to create smaller, specialist groups (for example, by introducing specialists).
- If a certain convergence threshold is reached (perhaps because the search space has been exhausted), artificially introduce a curmudgeon to guide the group into a new area of the search space.
- If in a group only a few individuals drive convergence, artificially encourage less active individuals to participate more.
- If in a group the majority of people prevent the exploration of novel areas in search space, artificially encourage these people to be more adventurous.

REFERENCES

Andersson, M., 1994, *Sexual Selection*. Princeton, NJ, Princeton University Press.

Axelrod, R., 1997, "The Dissemination of Culture: A Model with Local Convergence and Global Polarization." *Journal of Conflict Resolution* **41**(2 (April)): 203-226. http://www-personal.umich.edu/~axe/research/Dissemination.pdf.

Fisher, R. A., 1930, *The Genetical Theory of Natural Selection*. Oxford, UK, Clarendon Press.

Futuyma, D. J., 1998, *Evolutionary Biology*. Sundarlund, MA, Sinauer.

Hartl, D. L. and A. G. Clark, 1989, *Principles of Population Genetics*. Sunderland, MA, Sinauer.

Page, S. E., 2007, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, NJ, Princeton University Press.

Palla, G., A.-L. Barabási, et al., 2006, "Quantifying social group evolution." *Nature* **446**: 664-667. http://arxiv.org/abs/0704.0744.

Provine, W. B., 1986, *Sewall Wright and Evolutionary Biology*. Chicago, University of Chicago Press.