Submitted to SASO 2011

Socially-Constrained Exogenously-Driven Opinion Dynamics

Explorations with a Multi-Agent Model

H. Van Dyke Parunak, Elizabeth Downs, Andrew Yinger
Vector Research Center
Jacobs Engineering Group
Ann Arbor, MI 48105 USA
van.parunak@jacobs.com

***Abstract*—A number of studies have explored the dynamics of opinion change among interacting knowledge workers, using different modeling techniques. We are particularly interested in the transition from cognitive convergence (a positive group phenomenon) to collapse (which can lead to overlooking critical information). This paper extends previous agent-based studies of this subject in two directions. First, we allow agents to belong to distinct social groups and explore the effect of varying degrees of within-group affinity. Second, we provide exogenous drivers of agent opinion in the form of a dynamic set of documents that they may query. We exhibit a metastable configuration of this system with three distinct phases, and develop an operational metric for distinguishing convergence from collapse in the final phase. Then we use this metric to explore the system's dynamics, over the space defined by social affinity and precision of queries against documents, and under a range of different functions for the influence that an interaction partner has on an agent.**

Keywords-opinion dynamics; cognitive convergence; cognitive collapse; agent-based modeling; emergent behavior

I. INTRODUCTION

Humans form opinions and interests by interactions with their peers (collaboration) and with information sources (search). The pattern of these interactions emerges from the structure of a person's environment (social network, access to information sources) and the person's current opinions. In turn, the evolution of a person's interest may shape the environment, resulting in a complex feedback process. As a result of such dynamics, collective cognitive effects may emerge at the system level (across groups of people) that can dominate the individuals' opinion evolution without the person's being aware of them. One common phenomenon is alignment of opinions, a process that is sometimes called "consensus formation" [1] or "collective cognitive convergence" [2, 3]. This phenomenon contributes to the power of collaborative groups, but it poses a threat if a group's convergence turns into collapse, blinding it to new ideas or data contrary to its current opinion.

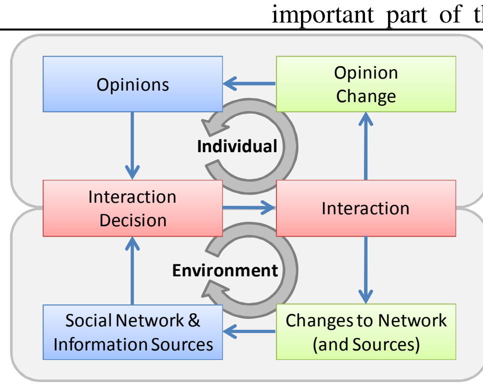


Figure 1. Stigmergic feedback between actors and their shared environment leads to the emergence of system-level features.

Such emergence of global features through direct and indirect feedback loops among autonomous actors in a shared environment is a common feature of stigmergic systems [4]. While stigmergy as a "design pattern" is best known in systems such as social insect colonies, there are many examples of humans coordinating through simple interactions in a shared environment to accomplish a common goal [5]. Figure 1 illustrates how a shared social and information environment couples individuals' opinion formation processes. One individual may select an interaction partner (person or information source) based on her current opinions. The interaction may change not only her opinions, but also the social network relations, affecting subsequent interaction decisions by others.

The objective of our research is to learn how to measure cognitive convergence and modulate it by making appropriate changes to a group of knowledge workers. We and others have studied the dynamics of shared opinions using agent-based models. The models developed to date are driven solely by the initial opinions of the agents, and social connections, if represented at all, form a connected graph among all agents. The specific setting that motivates our model requires extending such a model in two ways. In many government and business settings, a population of *analysts* is responsible for formulating recommendations for policy makers. While internal discussions among analysts are an important part of their work, they also consult exogenous information, in the form of a dynamic collection of *documents*. Furthermore, the analyst population is divided into separate *communities*, within which analysts interact preferentially. Each community starts with a *tasking*, a document that describes the subject that they are to explore. Exploring the dynamics of such a system requires two extensions that go beyond previous work by ourselves and others: interaction of disjoint social groups, and the influence of exogenous information.

Section II surveys previous work on opinion dynamics, and highlights the new contributions of our research. Section III outlines the structure of our model and formal measures we use to observe its behavior. Section IV reports experiments with the model over the space defined by varying levels of group affinity and varying precision in retrieval of exogenous information sources, leading to two suggestions for modulating convergence among knowledge workers. Section V concludes.

II. Previous Research on Opinion Dynamics

One recent review of computational studies of consensus formation [6] traces relevant studies back more than 50 years [7], including both analysis and simulation. These studies differ in the belief model and the topology, arity, and preference of agent interactions.

An agent's **belief** can be either a single variable or a vector, with real, binary, or nominal values. Vector models can be either *independent*, in which an agent can hold any combination of beliefs concurrently, or *correlated*, in which there is pressure for consistency among an agent's beliefs.

Different **topologies** can constrain interactions. Some models constrain interactions by agent location in an incomplete graph, usually a lattice (though one study [14] considers scale-free networks). In others any agents can interact (the "choice" model).

Interaction **arity** can allow agents to interact only two at a time, or as larger groups.

The likelihood of agent interaction may be modulated by their **preference** for similar agents.

Table 1 characterizes several studies in this area in terms of these dimensions. Our work extends this field in two ways. First, it supports multiple disjoint social networks. Second, it provides exogenous influences, in the form of a collection of documents that agents can query. These extension allows us to model a situation in which groups of agents are collectively analyzing information from a changing collection of information sources.

TABLE I. Representative Studies in Opinion Dynamics

Study	Belief	Topology	Arity	Preference?
Krause [1]	Real variable	Choice	Many	Yes
Sznajd-Weron [8]	Binary variable	Lattice	Two	No
Malinchik [9]	Real variable	Lattice, Random, or Hierarchy	Two	No
Deffuant [10]	Real variable	Choice	Two	Yes
	Binary vector, independent	Choice	Two	Yes
Axelrod [11]	Nominal vector, independent	Lattice	Two	Yes
Bednar [12]	Nominal vector, correlated	Choice	Many	No
Lakkaraju [13]	Real vector, correlated	Complete, Lattice, Regular, Small-world	Two	No
Parunak [2]	Binary vector	Choice	Many	Yes
This paper	Real vector, independent	Arbitrary, Unconnected	Two	Yes

III. An Agent-Based Model

This section describes our model and the metrics we use to monitor its dynamics. A wide range of configuration parameters are available to configure the initial set-up of a scenario (discussed under "model components") and govern the execution cycle (discussed under "model execution"). For each model component and execution step, we identify the main parameters that our model exposes.

A. Model Components

Our model has five main components.

Topic Space.—Analysts and documents live in an abstract Euclidean space constructed from a set of topics. In our model, these topics have no semantics, but in the real world, a topic is a probability distribution over lexicographic terms (e.g., domain-relevant key words), constructed from a large collection of relevant documents using techniques such as Latent Semantic Analysis (LSA) [15] or Latent Dirichlet Allocation (LDA) [16]. The topic space is a hypercube of dimensionality equal to the number of topics, with a range of [0, 1] on each dimension. A given location in this space is a Topic Model Vector (TMV). A *theme* is a region in topic space. We generate analysts or documents associated with a theme by sampling a Gaussian with configurable mean and variance, resampling when the tails yield a location with a coordinate outside of [0, 1]. Relevant parameters are:

- Number of Topics: Dimensionality of topic space
- Theme Mean
- Theme Variance

Social Network.—We organize analysts into (static) groups where members of a group are likely to interact with other group members but less likely with members from other groups. This group structure models organizational and geographical constraints that externally influence the likelihood that two analysts interact. Additional internal interaction preferences within these constraints arise from the preferential selection by analyst interest. Parameters are:

- Number of Analysts
- Number of Groups
- Group Themes

Document.—A major innovation in our model relative to previous work on opinion dynamics is the explicit representation of exogenous influences on agent opinions in the form of documents. A document is a Topic Model Vector (TMV). In the real world, a document's TMV is discovered using topic modeling. Real-world document repositories typically contain documents from different sources and with different concerns. We model this clumping of documents with the notion of a theme, and generate a population of documents by sampling from several themes with specified

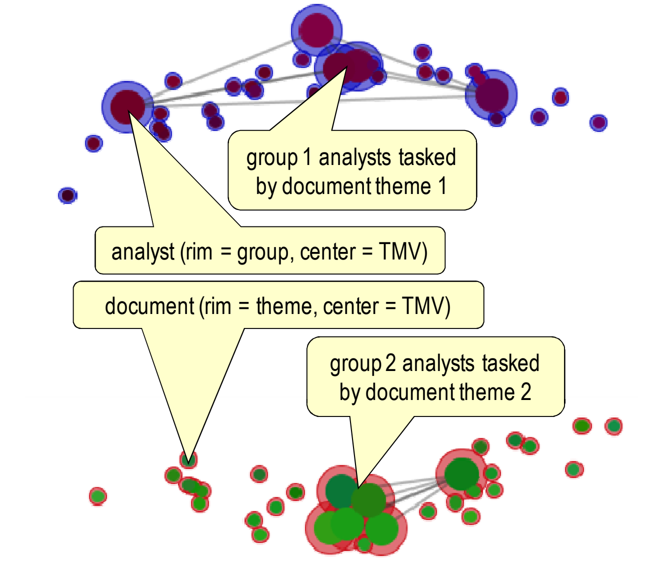


Figure 2. Visualization.

means (locations in topic space) and variances. Real-world document repositories are not static, but continually grow as new documents are discovered. We model the arrival of new information during the runtime of the analyst agents as the delayed introduction of documents sampled from a new theme. Parameters are:

- Number of Documents
- Number of Themes
- Document Themes (means and variances)

Analysts.—An analyst's current interest is also a Topic Model Vector (TMV). The tasking given to a community of analysts is defined as a theme, and we generate a community of analysts working on a given tasking by sampled that theme. The central object of our study is the movement of the analyst's TMV through topic space, relative to the TMVs representing documents and other analysts. Parameters are:

- Tasking Themes
- Number of Analysts

Document Search.—Real-world analysts use commercial or custom search engines to select documents for review. Depending on the search engine, queries may have different representations (e.g., key words, forms, relation graphs), but they always define topics of interest. Our model includes a representation of Document Search. An analyst poses a query as a subset of topics. The search weights documents by the strength of their entries on those topics, and probabilistically selects and returns a single document. Noise in this selection process models a real-world analyst's willingness to review documents that were not ranked first in their search results. Parameters are:

- Query Temperature
- Document Selection Temperature

B. Model Environment

Our model runs in the OPEN framework, a Java-based environment for model configuration, visualization, parameter exploration, and data collection. Figure 2 shows a visualization of the documents (small circles) and analysts (large circles with links indicating the social groupings). In this example, we instantiate two themes and two groups. The color of the rim of document nodes identifies the theme from which they were sampled, and the rim of the analyst nodes shows group membership. These colors do not change. The color of the inner document/analyst circle is defined by aggregating the elements of the entity's TMV. As analyst interest evolves, the color will change in the visualization. Document color does not change, since each document is static (though the set of documents can change).

We dynamically determine the drawing location of documents and analysts in the visualization through force-based graph layout. This process does not affect the opinion dynamics; it just determines the layout of the nodes in the visualization. Each drawing element (here document and analyst nodes) is associated with a simple agent that continuously updates the element's location. These updates are calculated as the vector sum (in 2D drawing space) of attractive and repulsive forces between the agents. We apply an exponentially growing repulsive force as two agents approach each other. The attractive force between two agents decreases with distance and thus acts on elements that are already close to each other. We modulate the strength of the attraction by the similarity of the agents between which the force is computed. That similarity is defined as the distance between the agent locations defined by theirTMVs in high-dimensional topic space. Thus relative node distance on the 2D screen reflects relative distance in topic space, a form of Multi-Dimensional Scaling (MDS).

C. Model Execution

First we configure a scenario. Then analysts repeatedly execute four steps: choose interaction type, assemble interaction options, select interaction target, and execute interaction. Analysts execute in random order with replacement.

Configuration.—We instantiate a topic space with a specified number of dimensions (10 in the experiments reported here), then a specified number of analysts in a specified number of groups, each with a tasking theme with specified mean and variance, and finally a specified number of documents from a specified number of themes, each with specified mean and variance.

Choose Interaction Type.—The analyst chooses probabilistically whether to interact with another analyst or with a document. On a given step, an analyst interact either with a document or with another analyst. The parameter is:

- Document Query Probability: The probability p_D that an analyst queries a document in this cycle. With probability $1 - p_D$, it interacts with another analyst.

Assemble Interaction Options.—If an analyst is interacting with a document, all its interaction options (possible targets for interaction in this cycle) are documents currently in the document space. If the agent is interacting with another analyst, all its interaction options are analysts from one of the disjoint groups in the social network. The agent picks from its own group with probability defined by its Affinity parameter, and otherwise picks from another randomly selected group. The parameter is:

- Affinity: the probability that an analyst will choose to interact with a member of its own group rather than an analyst in another group.

Select Interaction Target.—The analyst agent selects one interaction target from the options assembled in the previous step. As both document content and analyst interest are represented as TMVs, this step is identical for documents and analysts. This step by the agent models both the analyst's decision what query to construct based on its current interest and what search result to select. In terms of social interaction decision, it is the analyst's choice what issues to explore with other analysts and then what person to interact with.

An analyst agent constructs a query by probabilistically selecting a subset of topics from its TMV. Guided by a model parameter, the agent decides how many topics should make up this query. The more topics are in a query, the more specific it is. Then, the agent selects as many topics as it needs to populate the query. This choice favors topics that are currently of high interest to the analyst, but we add a temperature noise to that selection. For zero temperature, the top-N topics (N is the size of the query in this cycle) of interest are chosen. For high temperatures, this choice is practically random. Parameters are:

- Search Topics: Number of topics on which to query
- Topic Selection Temperature: Amount of randomness in topic selection (via Boltzman normalization)

Then an interaction target is selected based on the query. In this "Search Execution", the TMVs of the interaction options are sorted by their values in the topics of the query – TMVs with higher values rank higher. Thus, we create a relevance ranking of the TMVs relative to the given query. Based on that ranking, but again with a temperature noise parameter, we select one TMV from the top. Here the temperature noise models an analyst's result-selection behavior. The parameter is:

- Search Temperature: Amount of randomness in document selection (via Boltzman normalization)

Execute Interaction.—With the interaction target (document or analyst) selected in the previous step, the analyst now updates its interest model. If the analyst interacts with another analyst, then it samples the Learning Style parameter to determine the personality it should assume in this interaction, standard or curmudgeon. The selected personality sets the update rule for updating each topics interest level as a function of the difference in interest on that topic between the agent and the selected interaction target. Section IV.D explores the form of this rule. In the standard personality, the agent shifts its interest level in updated topics to be closer to the interest level in the interaction target. In the curmudgeon personality, it shifts away from the other interest level. Most of our experiments are performed with the standard personality. If the interaction target is a document, then the agent always uses the standard personality. Thus, document content always draws the analyst closer to the document. Parameters are:

- Learning style: probability that agent acts as a curmudgeon
- {Analyst or Document} {Learning or Forgetting} Rate: the amount an analyst increases a topic value (learning) or reduces it (forgetting) on interaction with an analyst or document, respectively.

D. *Performance Metrics*

We define two kinds of metrics: a set of component metrics, and a single aggregate metric.

Component Measures.—The set of topics in a given model span a high-dimensional metric space with valid locations (TMVs) limited to the [0, 1] interval for each topic. As analyst agents update their TMV through interactions with other analysts or documents, they move through this topic space. We developed metrics that measure aspects of the analyst movement to detect dynamic characteristics that indicate cognitive collapse. In the following, we specify the initial set of measures that apply to a single step (TMV update) by an analyst agent.

The most fundamental measure on the analyst movement through topic space is the magnitude of a single TMV update, which is the length of the vector between the agent's prior and new location in each cycle. "Encoded" in a step are the agent's choice of the interaction mode (document or analyst), the agent's limitation of the set of possible interaction partners (In-Group/Out-of-Group for analyst interactions), the emulation of query construction (based on current interest = current location) and relevance selection, and the application of a "personality" in the actual calculation of the TMV update as a function of the agent's location and the location of the interaction target.

The length of a step in topic space conveys the **absolute** magnitude of the impact a particular interaction had on the analyst's interest. It does not show the nature of the step **relative** to the other analysts. A second measure is the distance of the analyst's location (after the step) to the Center of Gravity of all analysts in the model, that is, the mean over the TMVs of all analysts regardless of group affiliation. The mean TMV may not be near any analyst. Movement of the analysts shifts the location of the mean TMV, thus successive "distance to mean TMV" measurements, unlike "step-length," are not statistically independent.

We explored other measures on the step-by-step movement of analyst agents, such as the length of the step vector projected onto the vector from the agent's prior location to the mean TMV, and the distance between the agent's initial and current locations. We found that the model dynamics of interest are sufficiently observable in the first two metrics defined above.

An Aggregate Measure.—All the measures in the previous section concern a single step on the part of an analyst agent. Initial explorations based on these measures show that we also need to discover a directed walk, in which an agent's successive steps are correlated with one another. In previous work [17], we applied information-theoretic (entropy) measures to detect a directed walk, but encountered idiosyncrasies from the specific definition of the system states whose probabilities are measured in the entropy calculation. For the current research, we developed an aggregate metric that measures the "directedness" in an agent's movement through topic space without the complications of the entropy calculations discussed in [17]. The delayed step length metric adds the step vectors (delta TMV) for a single agent over the most recent n cycles (configurable, 50 in the results reported here). The vector sum of steps of a **random walk** is on the order of $\sqrt{n}$, while the vector sum of steps that generally point in the same direction (**directed walk**) tends to be on the order of n.

IV. Experiments with the Model

We have conducted numerous experiments with this model, exploring the space. This section walks through an example scenario, exhibits the system's metastability, derives an objective way to measure the cognitive collapse of a knowledge community, and uses this measure to explore the space defined by community affiliation and precision of interaction with exogenous information. Both of these dimensions are new to the simulation study of opinion dynamics, and our experiments explore only a small portion of the space that they define. Nevertheless, our results suggest two practical principles for managing convergence and preventing collapse among knowledge workers.

A. A Representative Scenario

We consider a small scenario with two distinct document themes and two groups of analysts. We sample 25 documents for each theme. One theme is the tasking (sample initial interest vector) for all 6 analysts of the first group, and the other theme initializes the 5 analysts of the second group. The documents and analysts are embedded in a 10-dimensional topic space. Figure 3 shows the Topic Model Vectors (TMVs) for each document and each analyst agent in the model. Each row in this visualization is either a document (upper half of Figure 3) or an analyst (lower half of Figure 3). Each column is one of the topics (here 10) that make up the topic space of the model. Thus, a cell in this matrix shows the level of interest in (analyst) or relevance for (document) a particular topic. We shade the cells dark for high topic levels and light for low levels. The two document themes (and thus the two document subsets and the two analyst groups) are distinctly different, but somewhat overlapping in two topics. Figure 4(1) shows the initial layout graphically. We clearly see the association of each group with a subset of the documents available to all analysts. The analysts in the larger group (at the bottom of

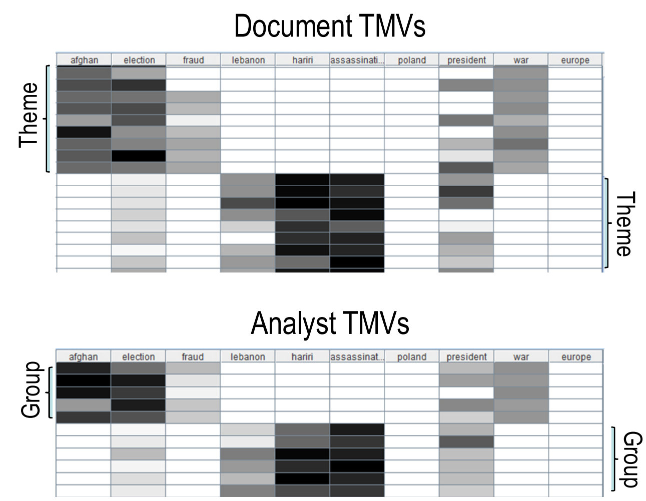


Figure 3. Document and analyst TMVs.

the figure) have higher settings for their affinity parameter (more likely to interact within their own group) than those in the smaller group.

The screenshots in Figure 4 illustrate the evolution of the model. As the simulation runs, we visualize the recent interactions of agents with other agents (red lines) and documents (blue lines) fading away into history (line transparency). The screen shots in Figure 4 show the interactions in the four most recent cycles.

In #2, the agents of each group that were initially spread out in their respective tasking theme converge on their common interest and thus form tighter clusters in their respective group. While there are also cross-theme/group interactions (blue/red lines crossing the gap in the center of the view), most interactions occur within the tightly clustered groups and their surrounding theme.

In #3, interactions of low-affinity analysts with the other group eventually lead to the defection of two analysts from the interest pattern of the smaller group and their transition towards the larger group.

In #4, once the first analyst defects from the smaller, low-affinity group, others follow rapidly. Eventually, all analysts abandon their interest in the upper document theme.

In #5, both groups of analysts have converged on the same set of interests exemplified by the document set in the lower part of the screen. Interactions among analysts within a group are no different from Out-of-Group interactions.

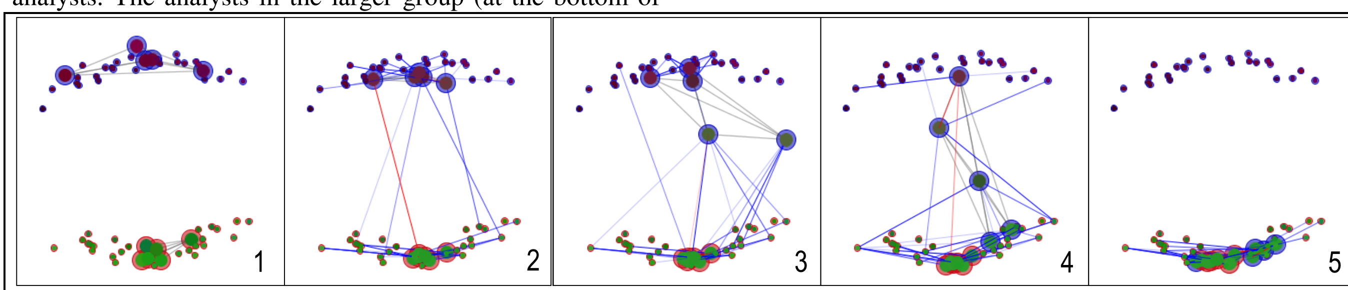


Figure 4. Stages in the evolution of the model in the example scenario.

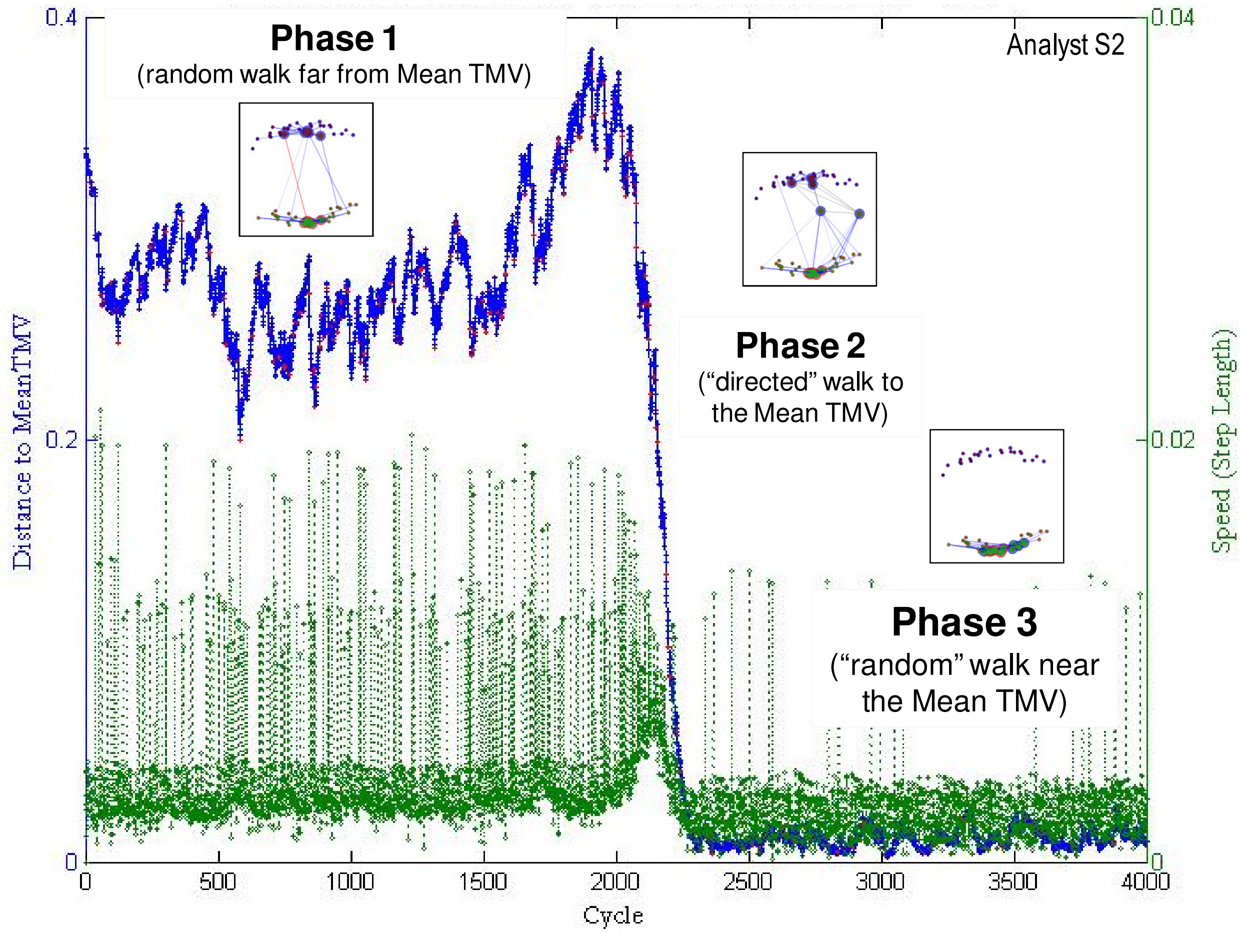


Figure 5. The dynamics of a single analyst from the top group. Both, "step length" (green) and "distance to mean TMV" (blue) display three distinct phases that align with the observed convergence of interests among the two groups.

Interactions with documents are (mostly) confined to the theme all analysts converged on.

B. *A Metastable Transition*

Our metrics allow us to detect three distinct phases of interest evolution dynamics (Figure 5, metrics applied to a single analyst from the upper, smaller group). The agents from the smaller (upper) group with lower affinity first remain in their separate interest area (Phase 1), eventually defect one-by-one to the interest of the other group (Phase 2), and then explore the other interest area jointly with the agents from the larger group (Phase 3). The three distinct phases in the model dynamics are reflected in the "step length" metric and in the "distance to mean TMV" metric.

Phase 1 is characterized by a large distance to the center of gravity of all analysts and a relatively high frequency of long steps. The large distance to the mean TMV reflects the separation of analysts' interests in this initial phase into two groups, so that the center of gravity of all analysts is far from any individual analyst. Individual steps are of three types. Short steps (most frequent) are interactions with other analysts from the same group (and thus similar interest) or documents near the analyst's initial tasking. Medium length steps are interactions with documents from the other theme. We set the {Analyst or Document} {Learning or Forgetting} Rates so that document interactions have less impact on analyst interest than analyst-to-analyst interactions. The longest steps, and the least frequent, are interactions with analysts from the other group. In Phase 1, the roughly constant distance to the mean TMV shows that the agent's successive steps with respect to the center of gravity are random (not correlated).

Phase 2 corresponds to the agent's defection from the region of its original tasking to the region occupied by the larger group. The distance to the mean TMV rapidly shrinks. The step-length metric shows an increase in the magnitude of high-frequency interactions as the agent moves away from the documents and analysts in its own group, due to the TMV update rule, which computes larger changes for larger differences between the analyst's TMV and the TMV of its interaction partner.[1] At the same time, the magnitude of the lower-frequency steps that correspond to interactions with the other document set and out-of-group analysts decreases, as the analyst moves closer to those entities. In this phase, the agent's successive steps are correlated, as the rapidly falling distance to the mean TMV shows.

[1] Section IV.D explores alternative update rules.

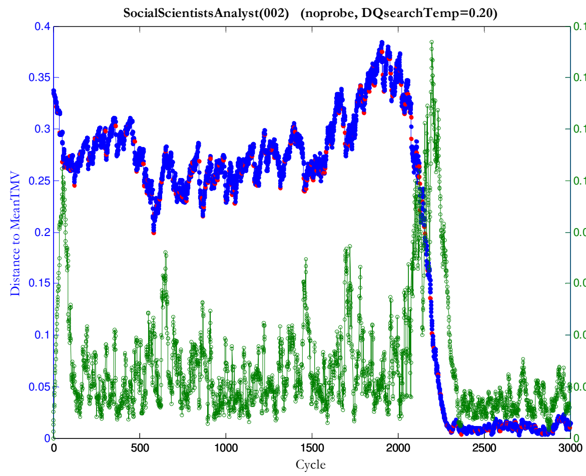


Figure 6. Re-plotting Figure 5 with Delayed Step Length reveals a strong peak for phase 2.

Phase 3 is qualitatively similar to Phase 1 in that the agent's successive steps are again uncorrelated. In this phase, all analysts of both groups occupy the same region in topic space and share a common interest in documents of the second document theme. The analysts' distance to the mean TMV is small as they are now all tightly clustered and most of the interactions result in only minor changes to an analyst's TMV (small step-length) as they select either nearby documents or analysts that are close by regardless whether they are inside or outside of the analyst's group. The analysts still interact (infrequently) with documents from the other theme (larger steps), but those interactions have no lasting effect on the analysts' relative location to each other.

Figure 6 shows again the "distance to the mean TMV" (blue) for a single analyst as in Figure 5, but compares its time series with that of the "delayed step length" metric. As expected, a strong peak in the delayed step length corresponds to the rapid fall-off in distance to the mean TMV (around cycle 2200). Phase 2 in the agent dynamics is indeed characterized by a directed walk while Phases 1 and 3 are (generally) less directed. The delayed step length metric is an indicator for Phase-2 dynamics.

We also note a significant fine-structure in the metric leading up to the merger of the analyst with the other group's interest. While this structure invites further investigation, we hypothesize the following. The first peak in the metric (around cycle 50) corresponds to the initial convergence of interests within the analyst's group, dispatching of the noise in the group-member's probabilistic initialization. Subsequent peaks (e.g., near cycles 600, 1500, and 1700), corresponding to significant drops in the agent's distance to the mean TMV, are failed "attempts" of the analyst to free itself from its group that are thwarted in subsequent interactions with its group members and its current document theme. Eventually, the agent succeeds in defecting to the other group. No such fine-structure was observed in the simple gradient climber in [17].

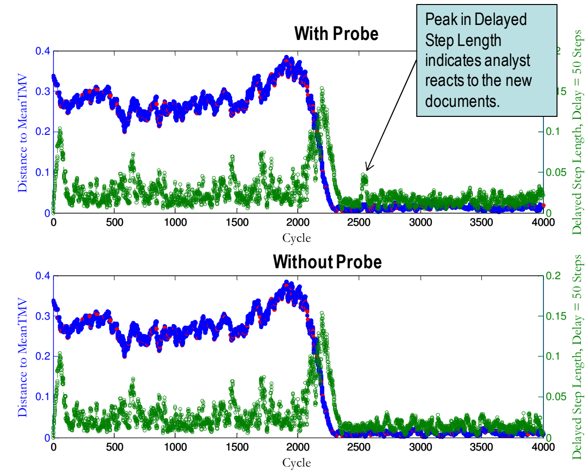


Figure 7. Visual inspection of response to new information in the delayed step length metric.

C. *Defining and Measuring Collapse*

Thus far, we have analyzed cognitive convergence in the interest evolution of the analyst agents in the two groups in our reference model. The delayed step length metric peaks as an agent distinctly reacts to information exposed by the other group and the other document theme. We have yet to define and detect cognitive collapse.

We have informally defined "cognitive collapse" as the inability of an agent or a group of agents to respond to new information. We can now operationalize this definition:

An agent is in cognitive collapse if it is not in phase-2 dynamics and if it does not return to phase-2 dynamics when qualitatively new information is introduced.

We introduce *qualitatively new information* by adding new documents to the model at runtime. We add documents from a third theme at a time when all agents have fallen into Phase-3 dynamics and are cognitively converged. This new document theme (located between the two initial themes in topic space) probes the analyst dynamics. If analysts in Phase 3 are collapsed and not just converged, they should not respond to new information by returning to Phase 2. In other words, the self-reinforcement of interests among the analysts should outweigh the "pull" from the new information.

We detect Phase-2 dynamics using the delayed step length metric. An initial visual inspection of this metric with (Figure 7, top) and without (Figure 7, bottom) a probe at cycle 2500 indicates that new information in the form of new documents may result in a return to Phase 2. Since the information space is otherwise static in our model, that return is short-lived and the agents quickly converge again.

To facilitate automated exploration of a system's tendency to collapse, we use the nonparametric Mann-Whitney-U test [18] to compare step lengths before and after probe insertion. The baseline configuration used in the results reported so far has a document query temperature of 0.12 and an analyst affinity of 0.4 (location *a* in Figure 8). We explore the region of parameter space that increases these parameters up to a query temperature of 0.32 and an

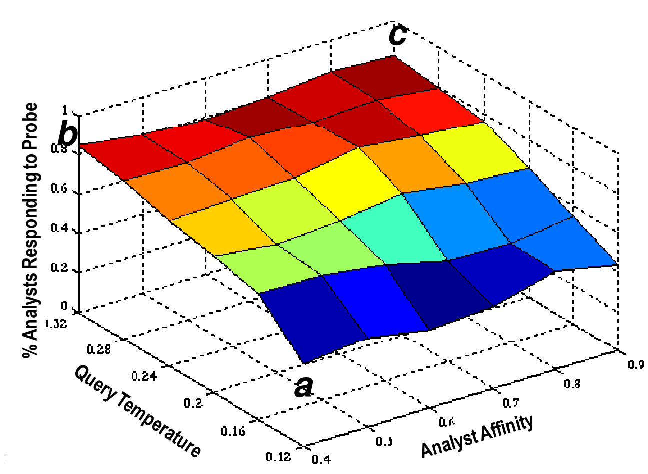


Figure 8. Collapse as function of search temperature and analyst affinity

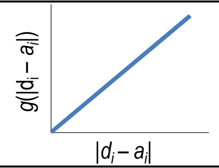
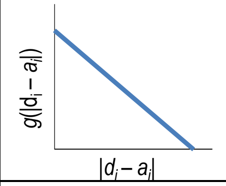
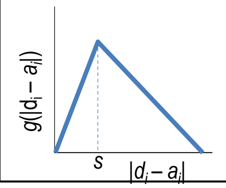
Cognitive Rationale	$g(x = \lvert d_j - a_j \rvert)$	Sketch
Curiosity	x	
Homophily	$1 - (1 - x)$	
Mixture	$s \in [0, 1]$ = "sweet spot" parameter $x \le s$: x/s $x > s$: $(1 - x)/(1 - s)$	

Figure 9. Alternative models for the TMV update rule g.

affinity of 0.9, and run 20 replications at each point. We allow each configuration to reach Phase 3, then insert a probe, and compute the percentage of analysts who respond to the probe, indicating that they are not collapsed. Figure 8 shows this metric over the parameter space. Increasing the query temperature, and thus exposing analysts to unexpected documents, dramatically reduces collapse. Surprisingly, the likelihood of collapse does not vary systematically as we change the affinity of analysts for their own group. Practically, if one wishes to modulate the rate of convergence among analysts, adding a variable quantity of noise to their queries against exogenous information appears to be more effective than motivating more or less interaction with other teams of analysts.

D. Exploring the Update Rule

In our model, the TMV update rule translates the difference in TMV elements between an analyst and an interaction partner into the length of the analyst's step (toward the partner for an ordinary analyst, and away for a curmudgeon). In all experiments reported thus far, the magnitude of the change in the interest in a particular topic in the TMV is proportional to the difference in that topic between the agent and the interaction target. This modeling assumption reflects **curiosity**: an interest very different from mine stimulates my interest and has a larger effect on me than an interest that is very similar to my own. (Parents of college freshmen often observe this principle when their children return home at their first college vacation.) An alternative model is **homophily**: I am more likely to move toward ideas that are close to my own than toward those that are different. Here the magnitude of interest change in the agent model would be inversely proportional to the difference in interest between the agent and the interaction target. In real-world analysts, the correct model is likely to be a **mixture** of these two effects: interests too far from mine are threatening, and interests too close to mine are boring, so my response will be greatest somewhere in the middle.

Figure 9 summarizes these options. We parameterize the mixture model with s, which indicates the hypothesized difference that will lead to maximum movement. When s = 0, we recover homophily, while s = 1 yields curiosity.

To explore the effect of the update rule, we focus on two observable events in our model:

a. The convergence of the interests of the two groups
b. The response to a new-information probe

We instantiate a large number of systems for each value of s and measure the probabilities $p(a)$ of convergence and $p(b)$ that the system responds to a probe (or in other words, is not collapsed). We hypothesize that as we increase s from 0 (homophily) to 1 (curiosity), these probabilities will vary qualitatively as sketched in Figure 10. At s = 0, homophily dominates and strongly limits interest changes of the analysts, so we expect convergence of the two groups to be unlikely ($p(a)$ ~ 0) and the agents are likely to be collapsed within their own groups ($p(b)$ ~ 0). At the other extreme (s = 1, high curiosity), the agents are highly mobile in topic space and thus are likely to converge on a common interest ($p(a)$ ~ 100%), but we still expect a significant risk for cognitive collapse ($p(b) < 1$) as new information may be "drowned out" by the wealth of stimuli already accessible to the agent. Intermediate values of s lead to intermediate levels of convergence, but with low risk of collapse ($p(b)$ ~ 1).

Figure 11 shows the results for low affinity and high query temperature (b in Figure 8), qualitatively confirming our hypothesized curve. We again plot the average over 20 random seeds for each data point. Low s (homophily) yields no interest convergence between the groups but decreasing likelihood of collapse (increasing likelihood of non-collapse). Our hypothesis did not take into account that group convergence is a phenomenon with a critical threshold. Thus, instead of a gradual rise of the convergence probability, the plot remains at 0% until a critical value of s is reached. At that point, the probability of all analysts converging on the same interest region rapidly increases, but

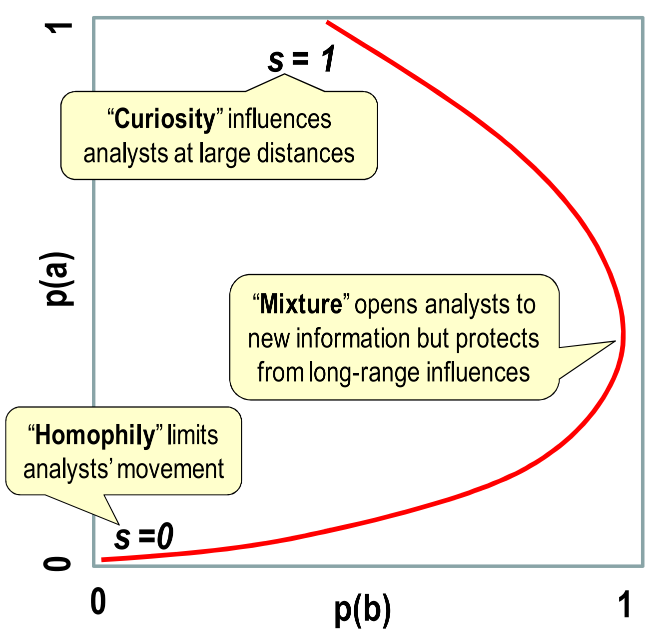


Figure 10. Hypothetical Effect of "Sweet Spot" Parameter.

the analysts remain equally open to new information (probability of collapse remains low). Finally, as curiosity dominates (large *s*, the TMV update behavior used in earlier examples), the likelihood of collapse begins to rise to the 20% level observed at Figure 8*b*. The increase in risk of cognitive collapse towards the end of the "sweet spot" sweep suggests another practical lesson for real-world knowledge workers: a mixed learning strategy that is most sensitive to information that is neither completely novel nor entirely familiar is less vulnerable to collapse than either extreme.

Figure 12 shows the corresponding plot for our baseline location of low affinity and low query temperature (Figure 8*a*). That configuration was tuned to exhibit a high

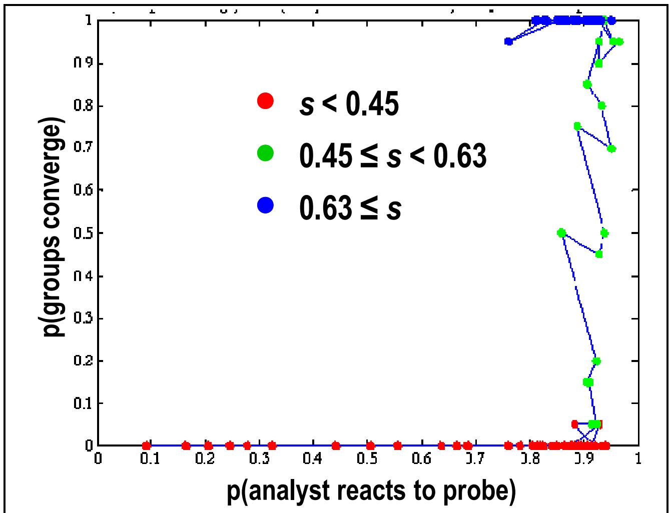


Figure 11. Sweep over the "sweet spot" parameter s for a configuration that shows low probability of collapse at the "Curiosity" extreme (high search noise but low affinity, Figure 8*b*).

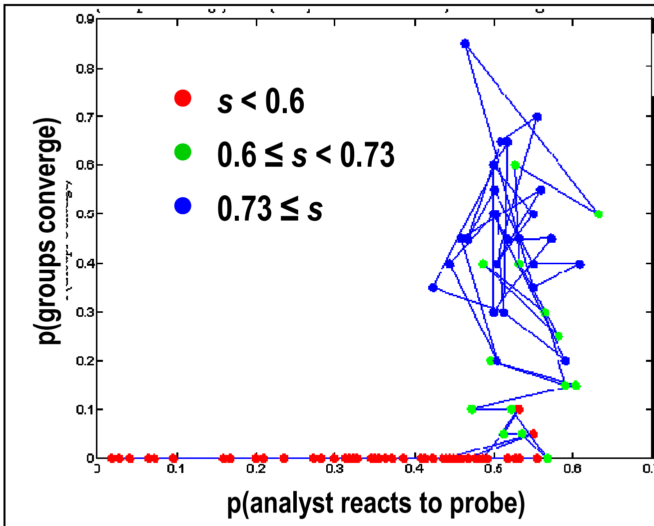


Figure 12. "Sweet spot" sweep for low search noise and low affinity (Figure 8*a*).

probability of cognitive collapse ($p(b)$ is low), so we know from the previous sweep that the end-point of our *s*-curve should be more to the left. Indeed we find that the probability of event a (x-axis) for large s is only around 50% (note the different axes scaling compared to Figure 11).

In this sweep, volatility in the likelihood of group convergence is very high (Figure 13). For large *s*, we see strong fluctuations in the probability of group convergence (blue line), even though each data point is the average over 20 individual runs. As group convergence is a threshold phenomenon, where only the defection of the first analyst triggers a stampede of the rest of the group, dominating curiosity (large *s*) combined with the other settings of the model parameters seems to be taking the model to a phase boundary. The specific nature of this system sensitivity to a cognitive characteristic of the analysts remains open for further exploration.

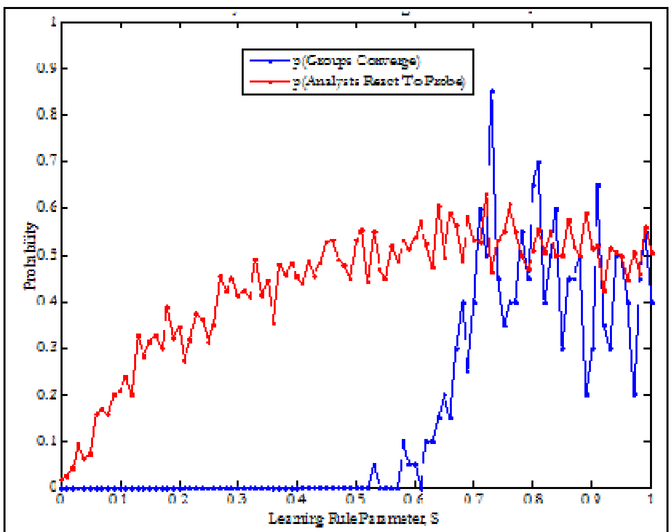


Figure 13. Independent plot of the event probabilities from Figure 12.

V. CONCLUSION

The opinion dynamics of multiple interacting knowledge workers are complex, often counter-intuitive, and yet critical for much collaborative work in the modern world, and enjoy the attention of a significant research community. Previous simulation studies focus on the evolution of an initial distribution of opinions across agents. While suggestive, such studies do not account for two critical features of knowledge workers in the real world.

- Their *social environment* is highly clustered, and they are more likely to interact with another agent in their cluster than with an agent in another cluster.
- Their *information environment* includes exogenous knowledge sources ("documents") in addition to other agents, and they seek out these documents with a query process.

Our new model implements both of these features. The resulting system exhibits an interesting metastability that allows us to formulate an operational measure of cognitive collapse. A preliminary exploration of the parameter space of social affinity and query precision with this measure yields two (very provisional) practical lessons.

First, query precision has much more influence on collapse than does social affinity. If one wishes to modulate the convergence of a community of analysts, managing the amount of noise added to their queries is a more promising method than changing their group membership.

Second, motivating analysts to prefer opinions that are neither completely new nor completely familiar will lead to more robust convergence without collapse than the extremes.

ACKNOWLEDGMENT

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via US Army Research Office contract number W911NF-10-C-0108. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, ARO, or the U.S. Government.

REFERENCES

[1] U. Krause, "A Discrete Nonlinear and Non-autonomous Model of Consensus Formation," in *Communications in Difference Equations*, S. Elaydi, *et al.*, Eds., ed: Gordon and Breach Science Publ., 2000.

[2] H. V. D. Parunak, "A Mathematical Analysis of Collective Cognitive Convergence," in *the Eighth International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS09)*, Budapest, Hungary, 2009, pp. 473-480.

[3] H. V. D. Parunak, *et al.*, "Understanding Collective Cognitive Convergence," in *Multi-Agent-Based Simulation IX: International Workshop, MABS 2008, Estoril, Portugal, May 2008, Selected Papers*. vol. LNAI 5269, N. David and J. S. Sichman, Eds., ed Berlin: Springer, 2009, pp. 46-59.

[4] H. V. D. Parunak and S. A. Brueckner, "Engineering Swarming Systems," in *Methodologies and Software Engineering for Agent Systems*, F. Bergenti, *et al.*, Eds., ed: Kluwer, 2004, pp. 341-376.

[5] H. V. D. Parunak, "A Survey of Environments and Mechanisms for Human-Human Stigmergy," in *Proceedings of E4MAS 2005*. vol. LNAI 3830, D. Weyns, *et al.*, Eds., ed: Springer, 2006, pp. 163-186.

[6] R. Hegselmann and U. Krause, "Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation," *Journal of Artifical Societies and Social Simulation,* vol. 5, p. none, 2002.

[7] J. R. P. French, "A formal theory of social power," *Psychological Review,* vol. 63, pp. 181–194, 1956.

[8] K. Sznajd-Weron and J. O. Sznajd, "Opinion Evolution in Closed Community," *International Journal of Modern Physics C,* vol. 11, pp. 1157-1165, 2000.

[9] S. Malinchik and D. Rosenbluth, "Paradoxical Dynamics of Population Opinion in Response to Influence of Moderate Leader," presented at the IEEE Symposium Series on Computational Intelligence (SSCI 2011), Paris, France, 2011.

[10] G. Deffuant, *et al.*, "Mixing beliefs among interacting agents," *Advances in Complex Systems,* vol. 3, pp. 87-98, 2000.

[11] R. Axelrod, "The Dissemination of Culture: A Model with Local Convergence and Global Polarization," *Journal of Conflict Resolution,* vol. 41, pp. 203-226, 1997.

[12] J. Bednar, *et al.*, "Conformity, Consistency, and Cultural Heterogeneity," in *Annual Meeting of the American Political Science Association*, Marriott, Loews Philadelphia, and the Pennsylvania Convention Center, Philadelphia, PA, 2006.

[13] K. Lakkaraju and A. Speed, "A Cognitive-Consistency Based Model of Population Wide Attitude Change," presented at the 2010 AAAI Fall Symposium Series, 2010.

[14] G. Long, *et al.*, "The evolution of opinions on scale-free networks," *Frontiers of Physics in China,* vol. 1, pp. 506-509, 2006.

[15] S. Deerwester, *et al.*, "Indexing by latent semantic analysis," *Journal of the American Society for Information Science,* vol. 41, pp. 391-407, 1990.

[16] D. M. Blei, *et al.*, "Latent Dirichlet Allocation," *Journal of Machine Learning Research,* vol. 3, pp. 993-1022, 2003.

[17] H. V. D. Parunak and S. Brueckner, "Entropy and Self-Organization in Multi-Agent Systems," in *The Fifth International Conference on Autonomous Agents (Agents 2001)*, Montreal, Canada, 2001, pp. 124-130.

[18] F. Mosteller and R. E. K. Rourke, *Sturdy Statistics: Nonparametrics and Order Statistics*. Reading, MA: Addison-Wesley, 1973.