ANL/DIS-03-1

Proceedings of the Agent 2002 Conference on

Social Agents:
Ecology, Exchange, and Evolution

Co-sponsored by
Social Science Research Computing
The University of Chicago

Decision and Information Sciences Division
Argonne National Laboratory

Edited by
Charles Macal and David Sallach

October 11-12, 2002
The University of Chicago

February 2003

ISBN 0-9679168-3-6

Available by request from

Decision and Information Sciences
Division Office
Argonne National Laboratory
9700 South Cass Avenue
Argonne, Illinois 60439-4832
Telephone: (630) 252-5464
Home page: http://www.dis.anl.gov/

CONTENTS

Invited Speaker: Kathleen M. Carley

Ecological Simulation

SATURDAY, OCTOBER 12, 2002

Invited Speaker: Lars-Erik Cederman

Electricity Applications

MacroEvolution

Invited Speaker: Scott E. Page

Closing Panel

FOREWORD

Welcome to the *Proceedings* of the third in a series of agent simulation conferences cosponsored by Argonne National Laboratory and The University of Chicago. The theme of this year's conference, *Social Agents: Ecology, Exchange and Evolution,* was selected to foster the exchange of ideas on some of the most important social processes addressed by agent simulation models, namely:

- The translation of ecology and ecological constraints into social dynamics;

- The role of exchange processes, including the peer dependencies they create; and

- The dynamics by which, and the attractor states toward which, social processes evolve.

As stated in the *Call for Papers*, throughout the social sciences, the simulation of social agents has emerged as an innovative and powerful research methodology. The promise of this approach, however, is accompanied by many challenges. First, modeling complexity in agents, environments, and interactions is non-trivial, and these representations must be explored and assessed systematically. Second, strategies used to represent complexities are differentially applicable to any particular problem space. Finally, to achieve sufficient generality, the design and experimentation inherent in agent simulation must be coupled with social and behavioral theory. Agent 2002 provides a forum for reviewing the current state of agent simulation scholarship, including research designed to address such outstanding issues.

This year's conference introduces an extensive range of domains, models, and issues — from pre-literacy to future projections, from ecology to oligopolistic markets, and from design to validation. Four invited speakers highlighted major themes emerging from social agent simulation.

In *Varieties of Emergence*, Nigel Gilbert introduces multiple ways in which agent models can address social emergence, which clearly is one of the strengths of the paradigm. When multiple forms of social emergence are chained together, models with multi-layer, micro-macro processes become possible. Lars-Erik Cederman reiterates this theme in his presentation, *Levels of Complexity: Endogenizing Agent-based Modeling*. The discussions linked these two sessions together.

In *Simulating Society: The Tension between Transparency and Veridicality*, Kathleen Carley frames one of the fundamental axes of tension within agent modeling, that is, the counterposition of simple transparent models with complex, empirically informed models. She posits a shared infrastructure for social and organizational models, including shared toolkits; shared data sets; and databases linking papers, models, algorithms, and data.

Finally, Scott Page addresses the role of diversity in model design and development. His presentation, *The Interplay of Differences*, provides insight, paradoxes, and cautionary tales with which to guide our efforts in the years to come.

We believe that Agent 2002 contributes to further progress in computational modeling of social processes, and we hope that you find these *Proceedings* to be stimulating and rewarding. As the horizons of this transdiscipline continue to emerge and converge, we hope to provide similar forums that will promote development of agent simulation modeling in the years to come.

Charles Macal, Director
Center for Complex Adaptive System Simulation
Decision and Information Sciences Division
Argonne National Laboratory

David Sallach, Director
Social Science Research Computing
The University of Chicago

ACKNOWLEDGMENTS

Victor Lofgreen coordinated the logistics, with help from Lisa Reyes, and managed the audiotaping of the workshop. Kathy Ruffatto handled registration and administration, and Bob Baker managed the conference Web site. Argonne's Information and Publishing Division prepared the document for publication. Margaret Clemmons served as project manager and editor, with assistance from Michele Nelson (graphic design), the Document Processing Center (word processing support), and Gary Weidner (print production).

ORGANIZING COMMITTEE

David Sallach, The University of Chicago (co-chair)
Charles Macal, Argonne National Laboratory (co-chair)
Thomas Wolsko, Argonne National Laboratory
Victor Lofgreen, The University of Chicago
Michael North, Argonne National Laboratory
John Padgett, The University of Chicago
Randall Picker, The University of Chicago

Thursday, October 10, 2002

Methods, Toolkits, and Techniques

AGENT-BASED METHODS, TOOLKITS, AND TECHNIQUES

M.J. NORTH, Argonne National Laboratory, Argonne, IL*
R.M. BURKHART, Deere & Company, Moline IL

ABSTRACT

Several leading agent-based modeling toolkit developers and users met on October 7 and 10, 2002, at the Agent 2002 Conference on Social Agents: Ecology, Exchange, and Evolution in Chicago, Illinois, to discuss the state of the art and future directions of this emerging field. The discussions covered the Repast, Swarm, and NetLogo toolkits/techniques, as well as several others. The primary objective was to consider the capabilities of the various toolkits and techniques and discuss how they can best be used to meet the general needs of the agent-based modeling community. The demonstrations and discussions during these two days covered many topics, including possible ways to coordinate efforts across the various agent-based toolkits. Further, the formation of a new national social simulation society was announced.

SOCIAL SIMULATION SOCIETY

A national initiative is underway to form a new professional society that will focus specifically on computational social science. The purpose of this organization will be to explore advances in computational and organizational science. Both the toolkit-oriented sessions and the main conference are intended to be early activities that will lead to the formation of the North American Association for Computational Social and Organizational Science (NAACSOS).

The objectives of NAACSOS will be as follows:

- To encourage the international advancement of theory and research based on social simulation;
- To promote cooperation among researchers in the field;
- To maintain and list conferences, meetings, and workshops that are related to social simulation, with the aim of reducing conflicts in scheduling;
- To coordinate the organization of a regular international conference; and
- To support the development and enhancement of educational programs in the field and to publicize their availability.

---

* *Corresponding author address*: Michael J. North, Decision and Information Sciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439; e-mail: north@anl.gov. Roger M. Burkhart can be reached at BurkhartRogerM@JohnDeere.com.

Membership in NAACSOS will be open to scholars, practitioners, and students who (1) agree with the objectives of the society and (2) pay annual dues appropriate for their membership category.

The new NAACSOS group expressed interest in engaging with the entire community of agent-based modelers-researchers concerned with tools, ideas, and issues relevant to their own community. Discourse within and among the agent-based modeling community will broaden the focus of the group, rather than limiting discussions to particular toolkits.

Kathleen Carley (Carnegie Mellon University) is arranging for *Computational Mathematical and Organization Science* (CMOT) to be the official journal of NAACSOS. CMOT is published by Kluwer Academic Publishers.[1]

NAACSOS will consist of three subject areas, or sections, that support intellectual interchange in specific areas of modeling and research, including the sponsorship of specialized conferences or conference tracks. The current section categories are listed below, along with the contact person for that focus area:

- Computational Social Theory, David Sallach, contact lead (sallach@uchicago.edu);

- Methods, Techniques, and Toolkits, Michael North, contact lead (north@anl.gov); and

- Simulation Applications, Charles Macal, contact lead (macal@anl.gov).

The NAACSOS web site is http://www.dis.anl.gov/naacsos/.

Note: The inaugural conference for the newly formed NAACSOS society will take place in Pittsburgh, Pennsylvania, on June 22–25, 2003. For information about this conference, visit the Web site at http://www.casos.ece.cmu.edu/conference2003/.

TOOLKIT DEVELOPER'S MEETING

The toolkit sessions were organized on the basis of their potential role in support of the NAACSOS section on Methods, Techniques, and Toolkits. Paper sessions on Friday and Saturday were also loosely organized on the basis of the other sections within NAACSOS.

The Toolkit Developer's Meeting held on October 7, 2002, included representatives from academia, industry, and government. These representatives are listed below:

- Michael North, Argonne National Laboratory, organizer

- Charles Macal, Argonne National Laboratory

[1] The journal's Web site is http://www.kluweronline.com/issn/1381-298X.

- David Sallach, University of Chicago Social Science Research Computing, Repast
- Nick Collier, University of Chicago Social Science Research Computing, Repast
- Tom Howe, University of Chicago Social Science Research Computing, Repast
- Roger Burkhart, Swarm Development Group, Swarm
- Laszlo Gulyas, Harvard University/Loránd Eötvös University, various toolkits
- Seth Tisue, Northwestern University, NetLogo[2]
- Uri Wilensky, Northwestern University, NetLogo

In addition to the toolkit representatives, an expanding group of approximately six people who were attending a Repast tutorial the next two days arrived early to listen in and participate in the panel discussions.

The toolkit meeting was organized as a series of four panels:

- Where Are We Today?
- Where Should We Be in One Year?
- Where Should We Be in Five Years?
- How Do We Get There?

Each was framed and introduced by one of the toolkit representatives, with three or four additional representatives who were responsible for expanding their own points. The panel lineup is listed in a table at the end of this paper.

The panel reached agreement in a number of areas, as listed below:

- There is a sense that we are getting much closer to where we need to be in a first generation of tools. This belief is supported by the stability being reached in the toolkits. The current stage of toolkit development, however, may reflect only a “local optimum” that indicates the possibility of moving to other places entirely in the toolkit space.
- The gap between the modeler and the model builder or programmer is a big issue. We need methodologies that can better capture and translate the intent or concepts of a model into its implementation, or even eliminate this

---

[2] NetLogo is a new version of the original MIT StarLogo that is being developed at the Center for Connected Learning at Northwestern University (http://ccl.northwestern.edu/netlogo/).

translation entirely. Although more declarative ways of capturing the model structures might be prepackaged, there continues to be a cost (i.e., less flexibility in constructing custom models, such as that Swarm and Repast). A major goal of NetLogo is to eliminate any difference between the modeler and programmer, possibly by restricting flexibility but greatly increasing accessibility. There is an enormous difference between projects that have one modeler or programmer compared with larger projects that have multiple people assigned to specialized roles.

- Agent-based modeling is definitely “coming into its own,” as it is being accepted as a modeling and research technique. Although it may still be viewed as “on the edge of the periphery,” entire communities are emerging, often across disciplines as aided by the common toolkits. The toolkits have served an initial purpose, that is, aiding people in doing what they need to do.

- Newcomers to the community are still very confused when they encounter agent-based modeling through the various toolkits they are pointed to. Opportunities abound for training and community discussions that focus on principles, methods, and techniques that are not linked to specific toolkits. This opportunity is supported by the number of people who cross-subscribe to the various toolkit lists.

- Individual disciplines have made little progress in building their own domain-oriented frameworks or libraries, with some exceptions such as social network models in Repast. Although organized efforts have been few, new software techniques such as aspects or patterns may be useful in bridging the abstraction gaps of frameworks that still have to drive executable code out of the high degrees of customization that may be generated from cross-cutting domain specifications.

- To continue its advance on many fronts, agent-based simulation needs to expand to integrate with traditional simulation techniques (both discrete-event and continuous simulations, for example) and to related roles of agent-based computing such as agent-based optimization. Agent-based modeling still seems largely disconnected from multi-agent systems researchers, though this may differ somewhat in Europe. Real-time and “people-in-the-loop” simulations (being developed by NetLogo researchers, for example) are also possibilities that could to take agent-based modeling and simulation beyond current boundaries.

- Methods for effective use of agent-based simulation in various roles of both research and application need to be treated more systematically, including applying such techniques as validation and verification, which are sometimes not even addressed or mentioned in published models.

These points summarize the state of agent-based modeling at this time (2002). The following section looks at the various time frames of where we need to go, including specific action proposals. These ideas were discussed not only during the panels, but also throughout the conference, especially when the developers reconvened to present overviews of their respective toolkits.

TUTORIALS

The Repast tutorial attracted a capacity turnout of nearly 30 people. Significantly more people arrived for the "Methods, Toolkits, and Techniques" session of the main conference, followed by continued increased attendance for the paper sessions.

METHODS, TOOLKITS, AND TECHNIQUES

Individual toolkit overviews were presented on October 10 for the larger general audience. These sessions provided a more detailed description of the status and directions of individual toolkits than discussed in the panels held earlier in the week. The four toolkit presentations and demos looked at Swarm, Repast, Ascape, and NetLogo. Brief summaries of these presentations are given, followed by a broader discussion of future toolkit directions.

The presentation, *Swarm: An Eight-Year Design Perspective*, given by Roger Burkhart, combined a general history and overview with a demo, a review and assessment of original design goals, and some options for future directions. Many of the basic design principles of Swarm, including dynamic schedules of actions on an object-oriented representation, have proven successful and also have served as a model for other toolkits. The role of a common toolkit in forming a nucleus for communities of agent-based modelers has also been well proven.

Regarding specific technical goals for Swarm, many of its more elaborate structures have not been extensively utilized by most models. These structures support complex mixing of schedules and activities under explicit concurrency semantics, including distribution across multilevel swarms. As a result, some of these structures are not as usable or fully implemented as originally intended. They were also intended to support execution on parallel and distributed hardware; this remains a future option. Further open challenges include the support of agents that build their own definitions of structure and behavior at runtime to create a capacity for true open-ended evolution, and for the general model of concurrent agent interaction to stand as a self-defining model of computation in its own right, rather than resting on some other programming layer. At this time, these elements of original design vision may be more appropriate for new research than for direct incorporation into a production toolkit that has its own user base.

For a more realistic set of options for future directions, Roger Burkhart also used some of the slides from *Next Generation Swarm*, which Marcus Daniels presented at the ALife VII conference in August 2000 (available at http://www.swarm.org/alife7/img0.htm). Marcus presents the option to run Swarm as a browser plug-in for Web or desktop delivery under the Mozilla framework, including representation of the model as an XML document tree with multilanguage scripting capability against a COM interface. These capabilities have already been developed and demonstrated in various forms, including in collaboration with the IMT project of Ferdinando Villa, and so might be included in upcoming releases of pending Swarm code.

Seth Tisue, along with several of the graduate students who are also part of Uri Wilensky's Center for Connected Learning group at Northwestern University, gave an overview and demo of NetLogo (home page at http://ccl.northwestern.edu/netlogo/). NetLogo runs in a pure Java environment like the most recent MIT StarLogo version and implements the same Logo language; however, it is a separately developed modeling environment with its own funding directed by students. A library of 80+ extensively documented model examples includes many classic agent-based simulations as well as others that help the goal of system-oriented thinking. Like the original StarLogo, NetLogo provides fully interactive model development. The developers also want to go beyond strict two-dimensional spaces and are adding features like a BehaviorSpace for parameter sweeping of more controlled experiments. HubNet is a new project for classroom participation as part of agent-based models using handheld wireless devices such as TI-83+ calculators. NetLogo has a large and growing user community because it is used in schools. It is available free for educational and research use, without the source of the underlying implementation at this time.

Miles Parker of BiosGroup discussed Ascape, including an updated status on its development. BiosGroup has licensed the rights to Ascape and invested in significant additional development. A significantly upgraded version (version 3.0) was released to the public recently. In addition to generalizing the abstract patterns of its organizing “scape” concept, capability is being added in specific areas such as GIS. BiosGroup envisions an entire suite of Ascape “Line of Business” modules around a common core. The common core will remain available and free of license fees for academic use, but other licensing has not been resolved. A new feature permits the adjustment of the observed running speed of a simulation. Miles views Ascape as representing an 80/20 or 95/5 solution for the agent modeling features that people typically use, with some trade-off in flexibility.

Nick Collier and Tom Howe presented a summary of Repast, including current developments. A new version 2.0 is almost ready to be released. [Note: Repast version 2.0 was released in late 2002.] Scheduled upgrades include floating point time values and support of asynchronous threads, such as North would like to use to target parallel execution. Its scheduling model now includes events that occur over a duration, which aids distribution. The addition of GIS capability, both raster and vector, has been a frequent request, and work is in progress as part of a goal to generalize the topology of models. Major new effort has gone into a new SimBuilder interactive model builder (successor to Evolver) that includes that ability to write object behavior in a scripting language called NQPython (for Not Quite Python) that Nick Collier translates to the underlying Java environment.

CONCLUSIONS

All of the discussions were very constructive in seeking ways to unify efforts and build on the strengths of the various toolkits and their user communities. Some specific possibilities were made for continuing action and coordination, including the following:

- Developers are already building on the “Swarm-like” model of schedules, actions, and objects (especially Repast and Swarm, but possibly others). They could also explore the possibility of documenting and standardizing the common concepts expressed in each of them. This comment was a direct continuation of discussions already underway, following recent suggestions by Glen Ropella, who posited that this common structure might go to the point of standard Java APIs that could be used by developers of either Java Swarm or Repast. The group also expressed interest in the possibly of taking this common structure to the point of common APIs, but broader interest in capturing the concepts in a more language-independent or abstract form that could serve additional needs besides just programming uniformity, such as targeting different execution back-ends from a common GUI simulation builder, and making more explicit the assumptions behind models in a more explicit or declarative form than just program code. There is some interest in reducing models to underlying mathematical formalisms and using formal specification languages such as Object Z, but also possibly expressing model concepts in UML or as vocabularies in XML (which could serve as a form in which to generate model “documents” for execution by an engine). Exploring these various paths is one of the most direct recommendations to come from the discussions. The Repast and Swarm groups have offered to try to organize a follow-up activity probably including face-to-face meetings, but further details are still undefined.

- New agent-based modelers could benefit from a less fragmented path of entry into the various toolkit communities, including places where more general principles and important techniques and issues that span the toolkits can be discussed. In the past year, some of the classes and workshops held by Argonne, University of Chicago, and Santa Fe Institute have tried to explain and position the various toolkits, but this work can be expanded to help fill the need for more complete and accessible training. Existing events such as SwarmFest have long tried to address the broader agent modeling community (not just Swarm). The Swarm Development Group, however, would have to modify its charter to shift fully to a broader cross-toolkit role than its primary mission of supporting Swarm. As part of the new NAACSOS section, Argonne may develop some mailing lists that focus on broader modeling issues than are tied only to toolkits. The toolkit events and discussions, however, offer an unusually broad umbrella across disciplines. This unique interaction is a real value that should not be lost, but the quality and relevance of individual application models are best evaluated within the disciplines themselves. While participants basically agreed on the general needs, these have not yet been translated into a more fully coordinated plan for reaching out across the different agent modeling communities. From the discussions, it is not clear how much is desired or realistic. It becomes increasingly difficult to identify the current communities and the relative usage of different methods and tools and current needs. Roger Burkhart suggested that the upcoming SwarmFest 2003 (April 13–15, 2003) might be well-timed to follow up on cross-toolkit strategies and discussions, as that conference has previously helped to focus these efforts. Concrete evidence of progress includes common course materials, published books, Web sites, discussion lists, repositories of best practices, patterns and architectural templates, and minutes or proceedings of various events.

- On the technical platform for models, Web delivery is an important need regardless of language. For both quick prototyping and research, scripting, not low-level programming, is generating increased interest. Encouraging access and delivery of models, from modelers to model builders to users, will help the larger community to grow. The engine behind the scenes could increasingly be hidden behind an integration and delivery framework. The need for hybrids of agent-based models with GIS and other nonagent models could prioritize interoperation across boundaries that divide current implementations. The toolkits should continue to explore the possibilities and share results in making different kinds of models work together, including methods for direct interoperation across toolkits.

- We should not assume that any of the toolkits is at an endpoint, or that a narrower and more consolidated range of toolkits is a good thing. The current toolkits could be local optima that are due to be replaced completely, especially the longer the horizon of possible futures being considered. The impact on how we even conceptualize the problems being modeled is too early to think we can draw boundaries around anything. We need to continue our exploration to determine when the toolkits could run out of added value or create actual barriers that make it better or easier to just program a model directly.

- The most concrete suggestion was to revisit the question of whether we are at a local optimum one year from now.

- Finally, the agent-based modeling group could look at ways to obtain funding for the wide range of enhancements suggested during these sessions.

TOOLKIT DEVELOPER'S PANEL LINEUP

Toolkit Developer's Meeting Welcome
Michael North

Where Are We Today?
Moderator: Nick Collier
Panel: Roger Burkhart, Tom Howe, Charles Macal, Seth Tisue/Uri Wilensky

Where Should We Be in One Year?
Moderator: Roger Burkhart
Panel: Laszlo Gulyas, Tom Howe, Charles Macal, Seth Tisue/Uri Wilensky

Where Should We Be in Five Years?
Moderator: Laszlo Gulyas
Panel: Nick Collier, Roger Burkhart, Michael North

How Do We Get There?
Moderator: Tom Howe
Panel: Michael North, Laszlo Gulyas, Seth Tisue/Uri Wilensky

Design and Methods

UNDERSTANDING THE DIFFERENCE THAT SPACE CAN MAKE: TOWARD A GEOGRAPHICAL AGENT MODELING ENVIRONMENT

D. O'SULLIVAN, The Pennsylvania State University, University Park*

ABSTRACT

The geographical environments in which agents interact in models are typically very simplified. Many models run in completely aspatial worlds, such as markets, or in simplified representative spaces. In particular, grid-based lattices are the dominant spatial form in agent models. It is argued that richer representations are required to reflect the range of spatial forms that social interactions can take. This argument is supported by reference to an earlier study by the author examining the effects of deforming the grid structure of two cellular automata (CA): a majority-rule segregation CA and the 'Game of Life.' The findings demonstrate that spatial configuration can affect spatial dynamics, so that it is important to develop ways of understanding the difference that spatial configuration makes to the dynamics of social systems. Adding geographical sophistication complicates agent model architecture. Such models are more complex, and they also risk sacrificing the potential for learning about general system dynamics by observing model behavior. Thus, the analytical tools required to study geographically sophisticated models are also complex. Challenges facing the development of a geographical agent modeling environment to address issues of spatial representation and subsequent model analysis are briefly discussed.

INTRODUCTION: SPACE IN AGENT-BASED MODELS

There is a tendency in social science to regard space as merely a container within which social processes play out. This tendency might be characterized as the "All the world's a stage" view of life; that is, space is simply a backdrop, or at most a frame of reference, within which locations can be assigned coordinate values. This Newtonian perspective has been powerfully reinforced by the activities of cartographers, national cadastres, and more recently, by the global positioning system and geographical information system (GIS). These technologies focus attention on where things are with respect to a fixed frame of reference.

Recent approaches in human geography and, increasingly, in social science have rejected the idea of space as a neutral container. In human geography, this tendency emerged in the related notions of cognitive geography, cognitive maps, and behavioral geography (Golledge and Stimson, 1997). In a groundbreaking study of the 'intelligibility' of urban environments, urban planner Kevin Lynch (1960) argues that people develop internal representations of the environment, which affects behavior over time. Work on cognitive maps and cognitive mapping in robotics (Kuipers, 1979; Gopal and Klatzky, 1995), and in planning and psychology (Gärling, 1995), builds on these ideas. Kitchin (1996) provides an overview of these studies. For now, the important point is that representations of spatial structure affect the behavior of individual agents.

---

* *Corresponding author address*: David O'Sullivan, The Pennsylvania State University, Department of Geography, 302 Walker Building, University Park, PA 16802; e-mail: dbo2@psu.edu.

More germane to contemporary geography is an encounter in the mid-1980s with Anthony Giddens's structuration theory. This theory was partly a result of Torsten Hägerstrand's *time geography* (Hägerstrand, 1970; 1982). In time geography, the interaction of time and space in the daily routines of individuals is recognized. A person cannot be in two places at one time but can have interrelated social rights, responsibilities, and obligations that require presence in a number of places over the course of a day (or week, month, year, or lifetime). The details of the interaction between a person's daily routines in time and space—his or her *life-path*—and the life-paths of others are important. Giddens (1985) explicitly relates his influential theory of structuration to time geography, arguing that the spatial locations (*locales*) where social activities occur are at once the outcome of social activities and also influence the activities that occur. In human geography, the locale as an emergent phenomenon has been applied at scales from the personal to the regional (Thrift, 1983).

Although the time-geography framework has much to offer multiagent simulation modeling, little or no attempt has been made to deal with these complexities in constructing agent models. This may be excusable where agent modeling is deliberately very abstract, in the hope of uncovering general 'laws of motion' of social systems. However, the question remains open regarding how much difference spatial configuration and socio-spatial structure makes. More pragmatically, as agent modeling is increasingly directed to policy applications, there is a pressing need to be able to represent complex spatial configurations. As a step in the right direction, this paper proposes development of agent models that can accommodate more complex representations of spatial environments. This work can be seen as running loosely in parallel with attempts to represent social structure and organizations in agent models (see Prietula, et al., 1998).

This paper is organized in two parts. First, to demonstrate the difference that space can make, even in simple cases, the results of experiments with cellular automata (CA) are reported. These establish that the spatial structure of a model can make a difference to outcomes. Second, suggestions for a geographical agent modeling environment (GAME), and some of the issues to be faced in developing it, are discussed.

EXPLORING THE DIFFERENCE THAT SPACE MAKES: EXPERIMENTS WITH GRAPH-BASED CELLULAR AUTOMATA

To demonstrate the importance of spatial configuration to the behavior of complex dynamic systems, experiments on varying the lattice structure of two familiar CA models are described. For the interested reader, more details are reported in O'Sullivan (2000, 2001a). (Flache and Hegselmann [2001] report on similar work.) Anticipating the difference that lattice structure makes, Duncan Watts experimented with CA running on small world network structures (Watts, 1999, Chapter 8).

These experiments use an irregular or *graph-based* cellular automaton. In a conventional cellular automaton, cells are located at points on a regular lattice, and, except for edge effects, every cell neighborhood is identical. Typically, in two-dimensional grid lattices, each cell has four orthogonal neighbors (the von Neumann neighborhood) or, optionally, an additional four diagonal neighbors (the Moore neighborhood). Frequently, lattices are 'toroidal,' wrapping around in the east-west and north-south directions, so that all cell neighborhoods are equivalent.

In a graph-based cellular automaton, cells are treated as vertices in a graph $G(V, E)$, with vertex set $V = \{v_i\}$ and edge set $E = \{e_{ij}\}$, where each edge e_{ij} represents a neighbor relation between two vertices v_i and v_j. The neighborhood $N(v_i)$ of vertex v_i is the set $\{v_j \mid e_{ij} \in E\}$. The regular lattice of a conventional CA is a special case of this more general structure. This formalism is presented in more detail in O'Sullivan (2001b).

Altering the lattice structure of a CA (or agent model) begs the question: how does lattice structure affect system dynamics? This question is not easily answered. The current approach is to take an existing well-known CA on a regular lattice and to 'deform' its lattice, observing resulting changes in behavior. However, neither lattice deformation nor changes in behavior are readily parameterized, so it is difficult to concisely summarize the results of such experimentation. These difficulties should be borne in mind when considering the procedures described below.

Taking a cue from Watts and Strogatz's (1998) small world lattice rewiring process, one way to deform a regular lattice is to randomly select edges in the graph representing the regular grid lattice. One end-vertex of a selected edge is retained, and a new vertex at the other end is randomly selected from the graph. One difficulty arises because a path-dependency effect occurs in the small world rewiring process; that is, as some vertices acquire more neighbors, these vertices become increasingly likely to acquire still more neighbors, and a strongly skewed neighborhood size distribution develops. This development is undesirable because it can lead to difficulties in defining automaton update rules such that the same set of rules is applicable to both regular lattices and to lattices with varying neighborhood sizes. Given the geographical origins of this research, a rewiring process that biases rewiring in favor of nearby vertices is also desirable.

A rewiring process consistent with these desiderata is *edge-pair swapping* (see Figure 1). Four vertices v_0, v_1, v_2, and v_3 are randomly selected such that v_0v_1 and v_2v_3 are graph edges and v_0v_2 and v_1v_3 are not. Edges v_0v_1 and v_2v_3 are then replaced by v_0v_2 and v_1v_3. Restrictions are placed on how remote from each other the four vertices can be. Thus, v_2 is chosen so that it is no more than two edges from v_0, and v_3 is a randomly selected neighbor of v_2 that is not adjacent to v_1. This system ensures that the 'local coherence' of the graph is reduced only slowly by the deformation process. That is, cells that start as neighbors are likely to remain close to one another as graph deformation progresses. This is in contrast with small world rewiring, which rapidly reduces average distances between vertices in a graph. The effect of repeated application of the deformation process can be seen in Figure 1. Figure 2 shows the effect of this deformation

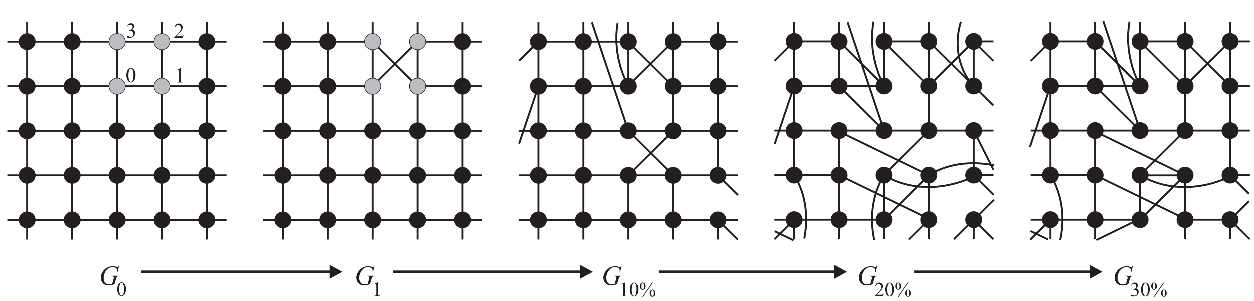

FIGURE 1 Deforming an Irregular CA Lattice (Subscripts indicate the number or percentage of graph edges that have been rewired in each case.)

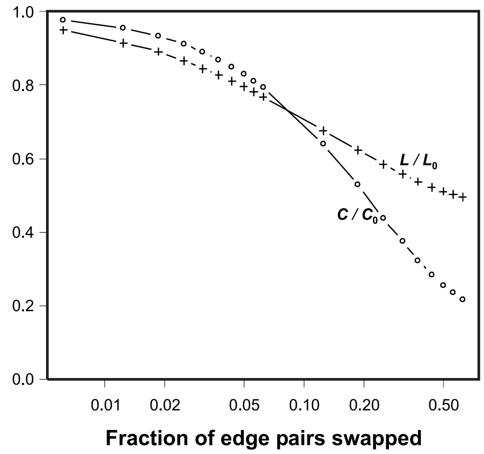

FIGURE 2 The Effect of Edge Swapping on Path Length, *L*, and Clustering Coefficient, *C*

process on the small world measures of graph structure, characteristic distance, and mean clustering coefficient. Because path length and clustering vary similarly, no small world structure arises (see Watts and Strogatz, 1998, for a discussion).

Describing the complex dynamics of a CA is also problematic. A single numeric parameter describing system behavior is desirable because it could be regarded as a function of the severity of deformation. In practice, the parameter most suitable for characterizing model dynamics strongly depends on the dynamic effects observed as a model runs, and no generalized measure is available. In these experiments, a *spatial information* measure (Wuensche, 1998) proved useful. This measure has the relative entropy form,

$$\sum p \frac{\ln p}{\ln q} \ ,$$

where *p* values refer to relative frequencies of occurrence of each possible cell neighborhood state, and *q* values refer to an expected frequency of occurrence calculated for each possible cell neighborhood state. In both cases described below, with 2 possible cell states and 9 cells in each neighborhood, there are 10 possible neighborhood configurations (from 0 through 9 cells in one state, with the balance in the other state). High information system configurations correspond to highly ordered arrangements of cells, where the CA rules lead to a distribution of cell neighborhood states different from a random arrangement. Low information configurations are indistinguishable from random on the basis of this measure. 'High' and 'low' values are relative, since the information measure is a dimensionless number whose value is theoretically unbounded and must be determined empirically. The specific information values attained are not important; rather, the evolution of this value as the CA state changes over time is of interest. The use of the measure becomes clearer in the discussion below.

Equipped with these two tools (a relative measure of lattice deformation and a means of summarizing system dynamics), we briefly examine the effect of deforming the lattice structure of two well-known CA rules in the following sections.

Majority-rule–based Segregation: A Spatially *Robust* Process

In a majority-rule CA, each cell adopts at the next time step whichever state is in the majority in its neighborhood. Starting from a random assignment of two cell states to lattice locations, a CA rapidly segregates into a stable arrangement with contiguous regions of cells in one or another of the possible states. Occasionally, all cells in the lattice end up in the same state, although starting from configurations where either cell state is equally likely, this occurrence is unusual. These dynamics are summarized in Figure 3, which shows the evolution of the spatial information measure for 50 random starting configurations of a two-state, 20×20 toroidal grid lattice.

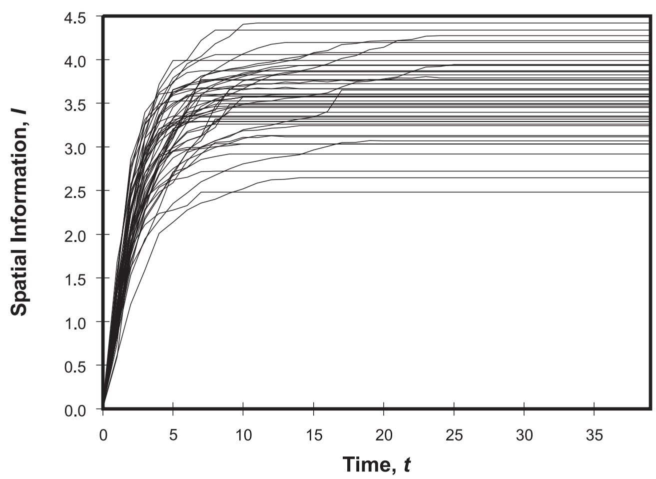

FIGURE 3 Evolution of Spatial Information for the Segregation CA on a Regular Lattice

For any set of starting configurations, the model behavior is summarized using the mean final spatial information attained by the CA. If the lattice is deformed by swapping pairs of edges, the effect on the final spatial information value attained is summarized in Figure 4. These plots show the effect on mean final spatial information for a set of 20 random starting configurations, for 11 different sequences of edge-pair swapping deformations, up to a total of 1,000 edge-pair swaps. In a lattice with only 1,800 edges, this extreme deformation effectively makes the lattice a random graph.

In the left plot, a large number of edge-pair swaps can be made before any appreciable change in behavior is observed. The right plot (note the exaggerated vertical axis) illustrates that little change is seen for small deformations. In fact, closer scrutiny of what is happening behind these summary data reveals that much of the fall in mean final spatial information is attributable to an increased tendency for the system to settle in a state where all cells are in one state or the other. The clearest outcome of this experiment is that the segregation CA is relatively *robust* under changes in its spatial or relational structure.

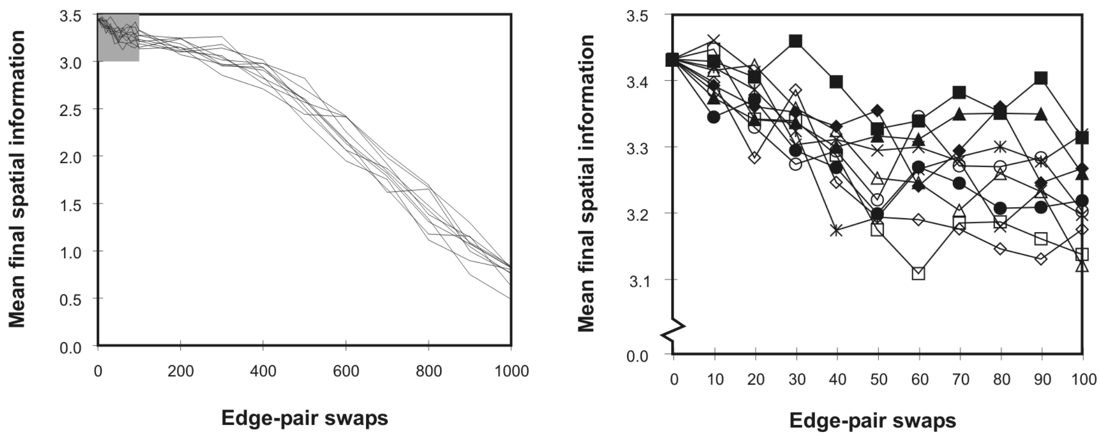

FIGURE 4 Changes in Mean Final Spatial Information as Segregation CA Lattice Is Deformed (The right plot is a magnified view of the shaded region in the left plot.)

The Game of Life: A Spatially Fragile Process

Similar experiments were performed on the Game of Life CA, also working with a 20×20 toroidal grid lattice. The Game of Life (Berlekamp, et al., 1982) is an additive CA rule, where cells in the '0' state switch to '1' if they have exactly three neighbors in the '1' state; otherwise, they remain at '0.' Cells in the '1' state remain in the '1' state if they have two or three neighbors also in the '1' state; otherwise, they 'die' and switch to the '0' state. Repeated application of these rules on a grid lattice results in very diverse dynamic behavior that is impossible to summarize using mean final spatial information. A more meaningful summary statistic is the 'transient time' before a starting configuration settles to a stable state. Figure 5 shows this statistic for three different starting configurations with transient times greater than 200, around 100, and less than 50 time steps.

To use transient times as a summary measure for CA behavior, we track how the observed distribution of transient times changes as the lattice is deformed. This is shown in Figure 6 for limited deformation of the Game of Life lattice. These box plots are based on only 20 random starting configurations; nevertheless, it is clear that even minor deformation of the lattice affects the behavior of the CA, resulting in reduced variability in transient times, with many more configurations settling to stable states in not more than 100 time steps. On the regular lattice, the observed median transient time is more than 100 time steps. More notably, very long lived configurations (more than 200 time steps) are only observed when 10 or fewer edge-pair swaps have been applied to the lattice.

The mechanism by which only modest lattice deformations alter behavior so dramatically can be explained with reference to the 'glider' configuration (Poundstone, 1985). On a regular toroidal lattice, a glider, once launched, can travel across the lattice indefinitely until it collides with other active cells. This mechanism often contributes to ongoing activity of a Game of Life configuration, extending transient times. On a lattice with even one irregularity, it is very likely that movement of gliders will be impeded because gliders often break up or halt and adopt a stable configuration when they encounter a lattice imperfection. The dependence of such

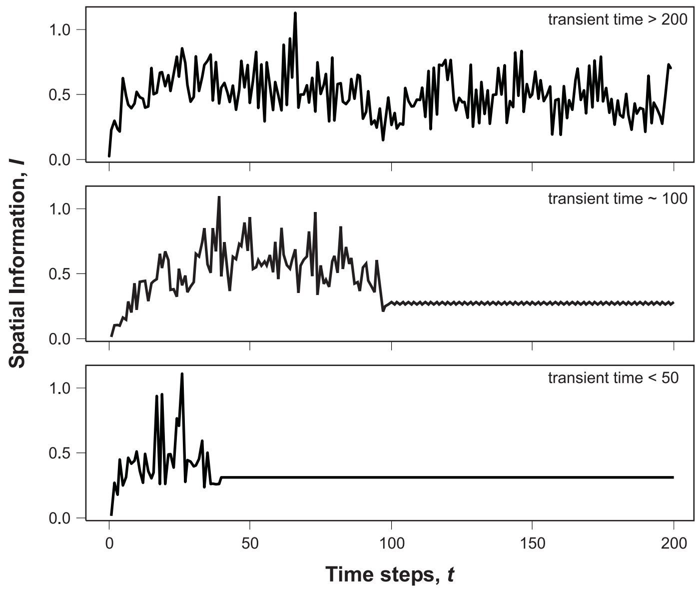

FIGURE 5 Three Spatial Information Time Series for the Game of Life CA

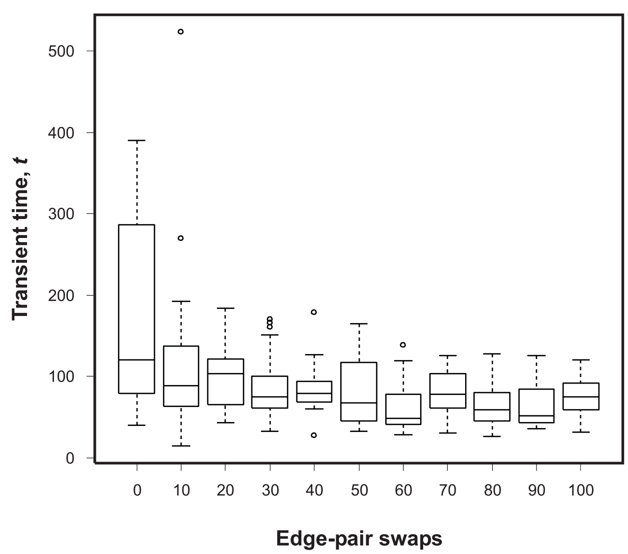

FIGURE 6 Summary Box Plots for Game of Life Transient Times under Lattice Deformation

patterns in the Game of Life CA on a regular grid results in a system that is *fragile* under spatial deformation.

Discussion

The inescapable conclusion of these experiments — regardless of the details — is that CA behave differently on the basis of their neighborhood structures. *This fact is hardly surprising*! However, while little work has been done to explore this issue, most CA and agent models routinely use one or the other of a limited number of spatial structures and fail to explore the implications of alternatives. *Spatial structure is a model parameter*, and, in the same way that variables used to determine agent behavior are systematically varied in experiments, tools are required that enable experimentation with the spatial structure of models. This is a pragmatic argument for the importance of enabling exploration of how the spatial structure of models affects observed behavior. Arguably, such development is also necessary on purely theoretical grounds as outlined in the introductory remarks.

Current multiagent simulations are not spatial in the sense implied here, with effects that can be detrimental to their use in understanding social systems, although there are exceptions to this generalization. In work over a number of years, Randy Gimblett and colleagues have developed simulations of human recreational behavior (Deadman and Gimblett, 1994; Gimblett, et al., 1996; Gimblett, et al. 2002). Westervelt and Hopkins (1999) integrate specialized environmental agent modeling software into the open-source *GRASS* GIS to assist in herd management, and Lake (2000) presents a custom-programmed agent model, also in *GRASS*, in a study of the foraging activity of Mesolithic societies. An example with no direct link to a GIS is the 'virtual Anasazi' work (Dean, et al., 2000). While these examples make a compelling case for geographical agent modeling, in all cases, spatially explicit data are used to represent the environment across a grid, and no spatial structural or relational representation is attempted. Some agent-based work on pedestrian behavior uses more detailed representations of geography (Dijkstra, et al., 2001; Helbing, et al., 2001; Kerridge, et al., 2001). For example, the STREETS model (Haklay, et al., 2001) operates on a high-resolution grid for obstacle avoidance and visual capabilities, but it uses a network of 'waypoints' for route planning and embeds the whole in a vector representation of urban space. Simultaneous use of various spatial representations is strongly reminiscent of the 'layers' typical of a GIS. However, again, a more complex, relational representation of spatial structure is not considered.

Some researchers are working with spatially structured agent models. In an investigation of residential segregation in Israel, Juval Portugali and colleagues introduce a Delaunay triangulation to represent neighborhoods in the built environment (Benenson, et al., 2002). Equally, it is clear that the use by agents of GIS functionality such as viewshed generation extends the capabilities of agents *and* the implied spatial structure (Itami, 2002). Indeed, it might be argued that it is inappropriate to include explicit spatial structure, but that spatial structures should emerge from the interactions of agents. (Batty [2001] presents an example where this happens.) This argument recalls debates in social science about the relationship between social structure and individual agency but has a similar 'which came first, the chicken or the egg?' quality to it. Whatever the outcome of that debate, pragmatically, it is important to enable the construction of models that explicitly represent spatial structure, *so that the implications of including or omitting such effects can be explored*.

TOWARD A GEOGRAPHICAL AGENT MODELING ENVIRONMENT

As a contribution to further research in this area, the development of a GAME is proposed. Such an environment would enable investigators to run multiagent simulation models (and CA models) on a variety of spatial structures. Figure 7 is a schematic block diagram for a GAME. The major elements required are an agent modeling toolkit, with its activity scheduling functionality, and a GIS, with its topological and spatial processing and visualization functionality. From a practical point of view, it is desirable to work with pre-existing toolkits and/or application programming interfaces that provide these capabilities; some open-source tools that could be used in this development are identified by name in Figure 7.

By using the spatial and topological processing capabilities of GIS software to operate on collections of spatial objects representing an environment, it should be possible to develop a 'lattice-builder' to provide readily reconfigured model structures for experimentation. Examples of possible lattice structures include:

- *Delaunay triangulations* (see Okabe, et al., 2000). This structure produces a planar lattice with no intersecting edges and has been proposed as a generalized cellular model for geographical work (see Semboloni, 2000, and Shi and Pang, 2000).

- *Distance-based graphs*. In this structure, two spatial locations are considered neighbors if they are within a fixed distance of each other. A variation on this approach is to weight the relationship between locations so that near locations have a strong relationship and distant locations have a weak (or no) relationship.

- *Nearest-neighbor graphs*. In this structure, each location has as neighbors the k locations nearest to it. This lattice construction rule can be used to construct the standard regular lattices. For example, the von Neumann grid is a k = 4 nearest-neighbors graph on a set of locations arranged on a grid. The Moore grid is a k = 8 nearest-neighbors graph on the same grid.

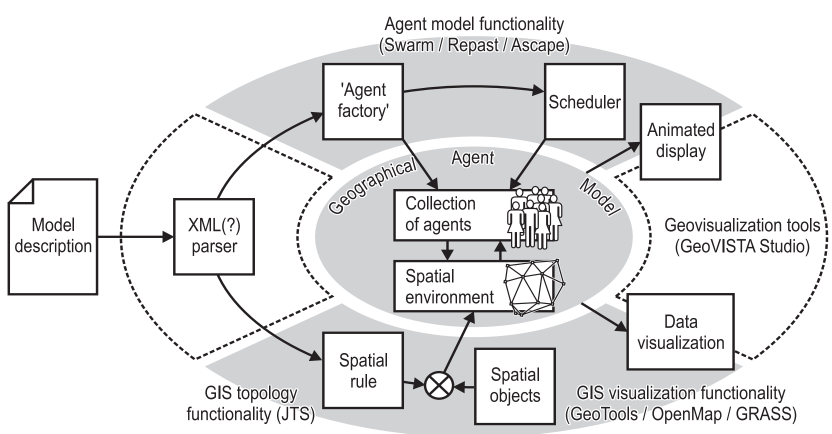

FIGURE 7 Schematic Showing Major Components of a Geographical Agent Modeling Environment (Core components are in gray; dashed lines denote additional components.)

With a lattice-builder module in place, the problem with working with awide variety of model structures representing the same or similar spatial data is effectively solved. The possibilities listed are only the most obvious ones, and others could be developed. For example, visibility graphs where any pair of mutually visible locations is joined by an edge might be a useful structure in models of human and other movement behavior (Turner and Penn, 2002). More complex structures might be developed where the strength of relations between locations (and also between agents) changes over time depending on agent-agent, place-place, and agent-place interactions.

Another major challenge is the problem of understanding and learning from agent models (not just the geographical kind). Even simple agent or cellular models that depart from regular lattice structures are formidably complex, given that, in principle at least, *any* model structure might be used. In addition to the extra complexity for presentation and analysis of results for an individual simulation run, there is the serious difficulty of inferring from the behavior of a model running on a particular spatial structure the likely behavior of the same model on other spatial structures. For example, how much can be inferred from a model of some social phenomenon based on, for example, the spatial organization of a particular urban neighborhood, about that phenomenon in other neighborhoods or other cities? Although there is no technical ‘fix’ for the problem of inference, the development of geographical agent models makes it more important than ever that sophisticated visualization tools are applied to the study of multiagent simulations. Only in a richly interactive simulation environment is it likely that investigators will be able to identify the recurrent spatio-temporal patterns that reveal what is happening in a model beyond just watching it unfold on the screen (see, for example, DiBiase, et al., 1992, who discuss user interactivity in animated mapping). With this in mind, further development of a GAME will depend on providing linkages to visualization displays such as scatter plot matrices, parallel coordinate plots, and other multivariate displays. A candidate tool for providing this functionality is GeoVISTA *Studio*, a ‘geovisualization workbench’ that enables dynamic visualization applications to be built from software components (see Gahegan, et al., 2002).

This paper has pointed out the importance of moving beyond the simplified abstract representations of space in most contemporary agent models and toward representations that better reflect the ways in which space itself structures and thereby alters social processes. It is imperative that these issues are explored and that tools are developed to support such research, if the dangers of working with wholly abstract ‘toy’ models are to be avoided.

REFERENCES

Batty, M., 2001, “Polynucleated Urban Landscapes,” *Urban Studies* **38**(4):635–655.

Benenson, I., I. Omer, and J. Portugali, 2002, “Entity-based Modeling of Urban Residential Dynamics: The Case of Yaffo, Tel Aviv,” *Environment and Planning B: Planning & Design* **29**(4):491–512.

Berlekamp, E.R., J.H. Conway, and R. Guy, 1982, *Winning Ways for Your Mathematical Plays*, New York: Academic Press.

Deadman, P., and H.R. Gimblett, 1994, “The Role of Goal-oriented Autonomous Agents in Modeling People-environment Interactions in Forest Recreation,” *Mathematical and Computational Modeling* **20**(8):21–33.

Dean, J.S., G.J. Gumerman, J.M. Epstein, R.L. Axtell, A.C. Swedlund, M.T. Parker, and S. McCarroll, 2000, “Understanding Anasazi Cultural Change through Agent-based Modeling,” in *Dynamics in Human and Primate Societies*, edited by T.A. Kohler and G.J. Gumerman, 179–205, Oxford: Oxford University Press.

DiBiase, D., A.M. MacEachren, J. Krygier, and C. Reeves, 1992, “Animation and the Role of Map Design in Scientific Visualization,” *Cartography and Geographic Information Systems* **19**(4):201–214.

Dijkstra, J., H.J.P. Timmermans, and A.J. Jessurun, 2001, “A Multi-agent Cellular Automaton System for Visualising Simulated Pedestrian Activity,” in *Theoretical and Practical Issues on Cellular Automata: Proceedings of the 4th International Conference on Cellular Automata for Research and Industry, Karlsruhe, 4–6 October 2000*, edited by S. Bandini and T. Worsch, 29–36, London: Springer.

Flache, A., and R. Hegselmann, 2001, “Do Irregular Grids Make a Difference? Relaxing the Spatial Regularity Assumption in Cellular Models of Social Dynamics,” *Journal of Artificial Societies and Social Simulation* **4**(4); available online at www.soc.surrey.ac.uk/ JASSS/4/4/6.html.

Gahegan, M., M. Takatsuka, M. Wheeler, and F. Hardisty, 2002, “Introducing GeoVISTA *Studio*: An Integrated Suite of Visualization and Computational Methods for Exploration and Knowledge Construction in Geography,” *Computers, Environment and Urban Systems* **26**(4):267–292.

Gärling, T., ed., 1995, *Urban Cognition*, Readings in Environmental Psychology, London: Academic Press.

Giddens, A., 1985, “Time, Space and Regionalisation,” in *Social Relations and Spatial Structures*, edited by D. Gregory and J. Urry, 265–295, Basingstoke, Hampshire, England: Macmillan.

Gimblett, H.R., B. Durnota, and R.M. Itami, 1996, “Spatially-explicit Autonomous Agents for Modelling Recreation Use in Complex Wilderness Landscapes,” *Complexity International* **3**; available online at http://www.csu.edu.au/ci/vol03/gimetal3/ gimetal3.html.

Gimblett, H.R., M.T. Richards, and R.M. Itami, 2002, “Simulating Wildland Recreation Use and Conflicting Spatial Interactions Using Rule-driven Intelligent Agents,” in *Integrating Geographic Information Systems and Agent-based Modeling Techniques for Simulating Social and Ecological Processes*, edited by H.R. Gimblett, 211–243, Oxford: Oxford University Press.

Golledge, R.G., and R.J. Stimson, 1997, *Spatial Behavior, A Geographic Perspective*, New York: The Guilford Press.

Gopal, S., and R. Klatzky, 1995, "NAVIGATOR: A Psychologically Based Model of Environmental through Navigation," in *Urban Cognition*, edited by T. Gärling, 79–102. London: Academic Press.

Hägerstrand, T., 1970, "What About People in Regional Science?," *Papers of the Regional Science Association* **24**, 7–21.

Hägerstrand, T., 1982, "Diorama, Path, and Project," *Tijdschrift Voor Economische En Sociale Geografie* **73**, 323–339.

Haklay, M., T. Schelhorn, D. O'Sullivan, and M. Thurstain-Goodwin, 2001, "So Go Down Town": Simulating Pedestrian Movement in Town Centres," *Environment and Planning B: Planning & Design* **28**(3):343–359.

Helbing, D., P. Molnár, I.J. Farkas, and K. Bolay, 2001, "Self-organizing Pedestrian Movement," *Environment and Planning B: Planning & Design* **28**(3):361–383.

Itami, R.M., 2002, "Mobile Agents with Spatial Intelligence," in *Integrating Geographic Information Systems and Agent-based Modeling Techniques for Simulating Social and Ecological Processes*, edited by H.R. Gimblett, 191–210. Oxford: Oxford University Press.

Kerridge, J., J. Hine, and M. Wigan, 2001, "Agent-based Modeling of Pedestrian Movements: The Questions That Need to Be Asked and Answered," *Environment and Planning B: Planning & Design* **28**(3):327–341.

Kitchin, R.M., 1996, "Increasing the Integrity of Cognitive Mapping Research: Appraising Conceptual Schemata of Environment Behaviour Interaction," *Progress in Human Geography* **20**(1):56–84.

Kuipers, B.J., 1979, "Modelling Spatial Knowledge," *Cognitive Science* **2**, 129–153.

Lake, M.W., 2000, "MAGICAL Computer Simulation of Mesolithic Foraging," in *Dynamics in Human and Primate Societies: Agent-based Modeling of Social and Spatial Processes*, edited by T.A. Kohler and G.J. Gumerman, 107–143, New York: Oxford University Press.

Lynch, K., 1960, *The Image of the City*, Cambridge, MA: MIT Press.

Okabe, A., B. Boots, K. Sugihara, and S.N. Chiu, 2000, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd Ed., Chichester, England: Wiley.

O'Sullivan, D., 2000, *Graph-based Cellular Automaton Models of Urban Spatial Processes*, Unpublished Ph.D. thesis, Faculty of the Built Environment, University College London.

O'Sullivan, D., 2001a, "Exploring Spatial Process Dynamics Using Irregular Graph-based Cellular Automaton Models," *Geographical Analysis* **33**(1):1–18.

O'Sullivan, D., 2001b, "Graph Cellular Automata: A Generalised Discrete Urban and Regional Model," *Environment and Planning B: Planning & Design* **28**(5):687–705.

Poundstone, W., 1985, *The Recursive Universe*, New York: Morrow.

Prietula, M.J., K.M. Carley, and L. Gasser, eds., 1998, *Simulating Organizations: Computational Models of Institutions and Groups*, Menlo Park, CA/Cambridge, MA: AAAI Press/MIT Press.

Semboloni, F., 2000, "The Growth of an Urban Cluster into a Dynamic Self-modifying Spatial Pattern," *Environment and Planning B: Planning & Design* **27**(4):549–564.

Shi, W., and M.Y.C. Pang, 2000, "Development of Voronoi-based Cellular Automata—An Integrated Dynamic Model for Geographical Information Systems," *International Journal of Geographical Information Science* **14**(5):455–474.

Thrift, N.J., 1983, "On the Determination of Social Action in Space and Time," *Environment and Planning D: Society and Space* **1**(1):23–57.

Turner, A., and A. Penn, 2002, "Encoding Natural Movement as an Agent-based System: An Investigation into Human Pedestrian Behaviour in the Built Environment," *Environment and Planning B: Planning & Design* **29**(4):473–490.

Watts, D.J., 1999, *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton Studies in Complexity, Princeton, NJ: Princeton University Press.

Watts, D.J., and S.H. Strogatz, 1998, "Collective Dynamics of 'Small-world' Networks," *Nature* **393**(4 June):440–442.

Westervelt, J.D., and L.D. Hopkins, 1999, "Modeling Mobile Individuals in Dynamic Landscapes," *International Journal of Geographical Information Science* **13**(3):191–208.

Wuensche, A., 1998, "Classifying Cellular Automata Automatically," Working Paper 98-02-018, Santa Fe Institute, Santa Fe, NM.

SITUATED SOCIAL ECOLOGY: AN INTEGRATED DESIGN HERMENEUTIC

D.L. SALLACH, The University of Chicago*

ABSTRACT

For computational modeling and social theory to co-evolve effectively, it will be beneficial to develop customized design methodologies. This paper integrates several design techniques into an overall hermeneutic. The intended purposes of the design hermeneutic are (1) to assure sufficient range in scope from the broadest ecological context to all relevant physiological assumptions; (2) to facilitate ontological experimentation, using situation theory where appropriate, and to advance social theory thereby; and (3) to use behavior design to govern the complexity balance for a given class of social simulations. Successful application of the proposed design hermeneutic may facilitate the modeling of meaning-production practices in social interaction.

INTRODUCTION

First-generation social agent simulation has demonstrated that simple rules are capable of generating diverse aggregate effects (Schelling, 1978; Epstein and Axtell, 1996; Axelrod, 1997). The insights generated thereby are promising and must be appreciated. At the same time, it seems unlikely that the epistemological potential of social simulation will be fully realized through models of discrete agents controlled by exogenous rules.

Scientific revolutions in other disciplines have been accompanied by the emergence of new conceptual entities such as quanta, genes, and tectonic plates (Thagard, 1992). In a similar way, social simulation will likely move beyond the premises of folk sociology. In addition, the ability to simulate endogenous agent production and management of meaning is likely to be necessary to more fully realize the potential of the methodology. Thus, arguably, to facilitate a prospective breakthrough, future research programs need to be in active dialog with the forefront of theoretical sociology.

One implication of such an evolution may well be the development of frameworks that transcend methodological individualism. That is, the extent to which social entities can be adequately modeled as discrete agents will need to be explored. While some of the richer sociological traditions regard the actor as a socially generated and defined entity (cf., Forgas, et al., 2001), agent simulation models have not addressed such a representation (Padgett, 2000). Similarly, although Parsonian functionalism posited a normative order that is effective and largely autonomous (Hilbert, 1992), the second half of the twentieth century is populated with efforts to comprehend the endogenous emergence of social norms (Garfinkel, 1967, 2002; Collins, 1981b, 2000). These insights have yet to be adequately incorporated in social simulation research programs (Collins, 1994; Sallach, 2003).

---

* *Corresponding author address*: David L. Sallach, Social Science Research Computing, The University of Chicago, Chicago, IL 60637; e-mail: sallach@uchicago.edu.

The present paper introduces a framework, a methodology, and a hermeneutic for modeling social agents situated in their settings. 'Framework' is intended to convey, not a theory, but a comparative structure in which a variety of theories can be explored. 'Methodology' is intended to convey an approach to software design that can span the distance between the hardware architecture and the social world being simulated. 'Hermeneutic' is intended to convey a multilevel interactive dialogue capable of realizing controlled models of social complexity.

A major purpose of this paper is to introduce a Situated Social Ecology (SSE) design framework. The SSE framework has three components: (1) an external-internal-external (EIE) hermeneutic that interleaves ecology, social interaction, and agent orientation; (2) Layered Formalism and Zooming Analysis; and (3) Behavior-oriented Design. Each of these components is well-conceived to be part of such a dialog and, ultimately, aligned with social theory, as described below.

A DESIGN HERMENEUTIC

The first issue to be addressed is why a design hermeneutic should be the goal of this discussion rather than, for example, the more prevalent design 'strategy.' It has long been recognized that the richness of social processes is infinite in its detail (Weber, 1949, pp. 72–73). Thus, an epistemological strategy based on social simulation inevitably imposes simplicity upon the domain being modeled. Indeed, it is frequently argued that the resulting simplicity is a strength that prevents resulting insights from being lost in a haze of complexities (Axelrod, 1997). However, there is also an abiding concern that some of the richness thus abstracted away is highly relevant to an adequate understanding of the process being modeled. Approaching design as a hermeneutic process emphasizes the importance of capturing the construction, communication, preservation, and transformation of meaning as a vital, inescapable aspect of modeling social processes.

The Nature of a Hermeneutic Process

The craft of hermeneutics originates in textual interpretation. Originally applied to sacred texts, in the modern era the techniques became generalized to literature, historical documents, and other secular texts (Palmer, 1969; Andersen, 2000). More recently, the method has been recognized as extending to technically mediated or enhanced communication (Ihde, 1998). In applying hermeneutics to software design, the present discussion proposes a further generalization of hermeneutic methods.

Hermeneutic method arises from a mutually defining tension between a communicative whole and its constitutive parts. The meaning of a sentence, for example, is determined by its component words, whereas the sense of the words is shaped by the enveloping sentence. Because the semantic effects flow both upward and downward, there is no privileged starting point for analysis, which is why the procedure is frequently referred to as a hermeneutic circle.

The hermeneutic process extends to multiple layers. Thus, sentences are within paragraphs that are within Chapters that are within books that are within oeuvres, etc. Both the larger and smaller layers establish a context that contributes to the meaning of the unit. As a result, a small change in one layer, or in the interpretation of a particular layer, can result in a significant alteration of meaning.

A Scale-based Design Hermeneutic

Extending hermeneutic techniques to software or simulation design produces several desirable results. First, the resulting design process is inherently multiscale in nature. Second, because hermeneutics is a meaning-capturing method, it lends itself to modeling human communication, and the associated diversity of understandings (and misunderstandings).[1]

Whether implicitly or explicitly, the part/whole tension is always scale-based in form. Both larger and smaller units provide constraints that shape meaning at its own level while, at the same time, requiring constraints from adjacent (and successive) levels to further disambiguate an interpretation. In this, the levels resemble the type of complex systems described by Juarrero (1999, esp. pp. 131–150), in which context-sensitive constraints from below and above together define the dynamics of the system. The bottom-up constraints define emergent capabilities and make them available. The reciprocal top-down constraints shape and select lower level components in ways that facilitate the efficacy of the emergent system. The synergy of the two processes warrants further exploration (cf., Sallach, 2000).

As a design process, hermeneutics allows the possibility of establishing environments in which agents, either singly or in coordination, can use constraint programming (Wallace, 2002), constraint-based coherence (Thagard, 1999), convergence (Agre, 1995), and other techniques to overcome ambiguities present, by design, in the representation of natural and social ecologies. Such capabilities for meaning extraction and resolution will potentially support richer semantic processes than agent models have heretofore produced.

The EIE Hermeneutic

The focus of the EIE design hermeneutic moves from the opportunities and constraints inherent in (1) natural ecologies, (2) social ecologies, and (3) agent mechanisms, and back again. Because social dynamics is the primary focus of the method, it begins with representations of an (artificial) ecology and moves toward agent physiology, passing twice through the intermediate (and, by design, more complex) realm of social interaction, once on the way in and once on the way back. EIE is selected[2] to allow agent assumptions to be relatively simple, while still recognizing the complexity of the natural and social environments.

The artificial external (ecological) and internal (physiological) levels can be seen as defining boundary structures for a given class of agent simulations. Within the constraints of these enveloping levels, a broad set of experiments in social processes can be explored. For example, semiosis between ecology and physiology (Hoffmeyer, 1996), as simulated with tropistic and hysteretic agents (Genesereth and Nilsson, 1987, pp. 307–313; Ferber, 1999,

---

[1] This capability potentially exists for both the designer and the agent. That is, the designer can capture and implement the types of meanings that are available in particular domains, while agents can capture the situated meanings that are available in their local setting.

[2] In principle, any of the three SSE foci could be the initial and terminating points of the hermeneutic, internal-external-internal, and social-out-social as well as EIE. Selection of any of the three directions constitutes a design strategy.

pp. 192–207), results in models of simple, adaptively constrained societies (Lake, 2000; Kohler, et al., 2000; Dean, et al., 2000).[3]

The EIE hermeneutic shapes three moments of the design process: (1) the overall concept, (2) the formulation of relevant mathematical models (Layered Formalism and Zooming [LFZ]), and (3) the design and development of behavior-based capabilities. The first moment defines the qualitative structure of the domain. The second allows experimentation on the aggregate effects of alternate ontologies. The third defines the nature and grain of agent capabilities. Successive passes can be used to align and integrate the overall design.

The focus of EIE is scale-oriented, so it passes from the largest phenomena considered (e.g., cosmologies, ecologies), to the intermediate level (social interaction), to the smallest scale (cognition, emotion, memory). However, as one first proceeds inward (toward the small), only the large is known. So the first pass faces social interaction design issues only relative, for example, to ecological concerns. After the needed cognitive, emotional, and neurophysiological structures have been designed, and the focus starts to move back up in scale, additional (individual level) detail can be incorporated at the social interaction level.[4] The process is not as sequential as it sounds, especially since it is in the nature of hermeneutics that one moves up and down the scale repeatedly until achieving an acceptable level of coherence.

EXPLORING ONTOLOGIES

In the natural sciences, scientific progress has been associated with the identification of an appropriate mathematical model. In the social sciences as well, game theory illustrates how a single formalism can result in a proliferation of productive research programs. However, questions of interest in most social science domains are too complex to be fully axiomatized. LFZ analysis was developed, in part, to explicitly address that fact (Devlin and Rosenberg, 1996). Accordingly, it is natural for LFZ to work in conjunction with a controlled strategy of agent simulation.

An analytical process based on LFZ is summarized as follows[5]:

1. At each stage of the analysis, define a formalism that is minimal.

2. At each stage of the analysis, utilize minimal precision within the formalism.

3. Refine the analysis iteratively, increasing the formalism and the precision until a promising model is obtained.

---

[3] There are also similarities with behaviorism, which posits a direct relationship between environmental stimuli organism responses.

[4] One might, for example, consider Goffman-type questions about how interaction is under constraints to establish, preserve, and/or repair the “social self.”

[5] The present summary draws on Devlin and Rosenberg (1996, esp. pp. 126–150).

4. When unresolved complications persist, zoom (shift levels of abstraction) until they can be satisfactorily modeled.

5. At each stage, align the analysis with guiding theories and relevant ontologies.

It can be seen that the iterative, multiscale nature of LFZ design is quite compatible with the EIE hermeneutic. The latter differs in more specifically clarifying the levels incorporated and the order of their incorporation. It also differs in explicitly acknowledging the mutual interacting constraints emanating from various levels. Finally, the EIE addresses the necessity to define behavioral capabilities. So, while LFZ is compatible with the encompassing hermeneutic, the latter supplements it with a broader design-oriented focus.

LFZ design is based on and applies the formalism of situation theory (Barwise, 1989; Devlin, 1991). This formalism has a number of strengths that make it highly appropriate for social modeling and simulation (Sallach, 2003). First, it controls social complexity by restricting its models to particular situations. Second, the situations considered are continually being transformed by the introduction of new actors, definitions, and resources. This allows situation theory to capture the indexicality of social dynamics, in which the meaning of the same social act is actively shaped and defined by the context in which it occurs, including prior communications and acts (Devlin and Rosenberg, 1993; Devlin, 1994).

Defining Situations

The context-sensitivity of social action has long been addressed by social scientists. One early contributor, W.I. Thomas (1967, p. 42), famously noted the unique efficacy of situational definitions:

> Preliminary to any self-determined act of behavior, there is always a stage of examination and deliberation which we may call *the definition of the situation.* And actually not only concrete acts are dependent on the definition of the situation, but gradually a whole life-policy and the personality of the individual himself follow from a series of such self-definitions.

However, as the quote indicates, Thomas tends to focus on stable, cumulative aspects of situational definitions, including the effects of cultural definitions.

In the second half of the twentieth century, a number of scholars began to emphasize contingency and emergence in situational definitions (Garfinkel, 1967, 2002; Collins, 1981b, 2000). In the words of Rawls (2002, p. 30), “Every situation has different patterns of order that are required for the coherence of action within that situation.” It has been suggested (Collins 1981a, 1994; Sallach, 2003) that understanding the micro/macro relationship between emergent situations and large-scale historical structures has the potential to contribute to a breakthrough in the social sciences.

Consideration of the fluid nature of situations and, in particular, how shared definitions are socially achieved, disturbed, and restored raises issues concerning the methods and competencies used in the process by social actors. Such questions suggest a promising line of simulative research and are considered in the next section.

Generally, there appear to be two design approaches for exploring the explanatory possibilities of simulating dynamic situation definition. First, the designer could make available a variety of prespecified, but possibly parameterizible, situations that could be invoked by situated agents. These situations might be differentially relevant to agents based on emotional commitments, cognitive models, self-interest, and their position within a social structure, as well as more ephemeral contingencies. This approach would be a natural extension of exogenously defined agent simulation, while allowing for greater complexities and more ambiguities in the dynamics of agent interaction.

Second, the design ontology could be specified at a higher level of abstraction, as supported by situation theory (Devlin, 1991), and allow agent construction of situation definitions in response to emergent circumstances. This generative strategy would be more difficult to implement but would have the advantage of representing the process of situation definition endogenously and, thus, achieve a closer dialog with microsociological theory.

Exploring the possible role of situations within social simulation illustrates one of the natural strengths of situation theory, but the formalism has broader ontology modeling capabilities. Diverse entities (individuals) and relationships, and their types, can be distributed at a variety of spatial and temporal levels of abstraction, providing a formalism that can potentially be aligned with any social domain of interest. It is this extensive capability that carries the potential for theory-driven design.

Complex System Dynamics

As a second example, consider the application of models of complex system dynamics to social system emergence. Drawing on the emergence of cognitive systems from a neurological substrate, Juarrero (1999) generalizes a basic framework. The core of this generalized process is comprised of a first-order process giving rise to a second-order process, which, in turn, constrains the first. In Juarrero's terms, constraints at the first level create a repertoire of capabilities that are available at the emergent level, while the latter selects elements and, thus, reciprocally constrains the first-order process.

As Sawyer (2001) argues, the emergence of mind from brain can be seen as analogous to the emergence of social processes from individual action. Applying the complex systems framework to a social example, the organization of an army coordinates the action of individuals and, thereby, creates coercive resources that did not previously exist. At the same time, the (socially emergent) army shapes and selects the individual (first-order) elements that compose it.

The stability of such a process depends on mutually reinforcing interactions between the two levels. Anything that disturbs the pattern of multilevel interactions has the potential to disrupt or even disintegrate the emergent army of the example. Among the sources of perturbation are parallel military emergents that compete, whether for recruits or in conflict that directly attacks its components.

In natural settings, considered from a microsocial perspective, the skein of interdependencies for an army, or for the many other social emergents that might be investigated, is dauntingly complex. One advantage of using agent simulation to apply the model of complex system dynamics is the ability to control the complexity of the process under consideration.

The Interaction Order

A third example of experimentation in ontologies that might prospectively undergird a more compelling analytical framework is provided by the idea of an interaction order (Goffman, 1983). As formulated by Rawls, the interaction order is produced locally under the constraints imposed by mutually constituted intelligibility (1989) and the presentational self (1987).

Since Rawls views the interaction order as *sui generis*, there are issues concerning its relation to other social entities and processes. The interaction order and social structure, although operating at distinct social levels, place reciprocal demands on, and are never found in isolation from, the other. Specific forms of interaction between the two levels might well be clarified by computational experimentation.

Nor need the three examples be viewed as mutually exclusive. The interaction order might provide a layer that mediates between emergent social structures and the social selves that compose it. Defining and managing situations might be regarded as essential capabilities of social agents who are engaged in creating and evolving social emergents. Such questions are not resolved here, but they illustrate the potential importance of ontology experimentation.

BEHAVIORAL COMPETENCES

The ethnomethodological critique of standard social science methodologies contends that reducing the empirical to the conceptual results in the complete loss of the emergent social phenomenon (Rawls, 2002, p. 50). Despite the fact that social simulation generates rather than captures data, it faces a similar methodological issue: how best to design an architecture capable of emulating the meaning production process. Syntactic or algorithmic models can structure probability-driven choice points to which meaning can be imputed by human analysts but, in such models, meaning-conveying, meaning-preserving, and meaning-recovering practices that are inherent in natural social processes are not endogenous to the simulated agent interaction. The means by which to achieve such a demanding goal remain largely undefined.

The methodology outlined here is designed to capture the multiple considerations that frame social communication and decision processes. Generation of such contextual contingencies is a prerequisite to the development of effective models of meaning-production. While it also remains necessary to control the balance between simplicity and verisimilitude, this can largely be achieved through the specification of the grain of agent perception and action. In the design of situated agents, what can be sensed, and the capabilities available in developing a response, largely determines the level of sophistication of the simulation. This specification of agent grain can be regarded as the boundary structure for a class of simulations, the analog of experimental boundary conditions in the natural sciences.

Behavior-oriented design is a form of simulation design that focuses on the management of multiple simultaneous agent priorities in potentially complex domains (Bryson and Stein, 2001a,b; Bryson, 2001). Accordingly, action sequences and cognitive competencies that may serve as responses to the opportunities and constraints of natural and social ecologies can be identified and implemented. The natural ecology, when its elements are relevant, defines opportunities and constraints relative to which social interaction can focus. Together, natural and social ecologies define situations discerned and processed by adaptive agents, situations that form the context of ongoing social processes.

Behavior-oriented design can be summarized in a series of discrete steps:

1. Specify at a high level what the agent is intended to do.
2. Describe likely activities in terms of sequences of actions.
3. Identify sensory, action, and communicative primitives from these sequences.
4. Identify the state necessary for these primitives, clustering them by the shared state.
5. Identify and prioritize goals or drives that the agent might need to attend to (prototype drive roots).
6. Select the next behavior to implement.

In summary, step recursively down in the level of detail, then step back up.

In the EIE design hermeneutic, agent capabilities are developed relative to the experimental ontology developed using LFZ modeling. LFZ yields the mathematical models; through behavior-oriented design, the models become computational.

Ultimately, the goal of the SSE framework is to tighten the connection between social theory and software design. In addition to the application of endogenous social theories to a variety of domains, the SSE framework allows their potential integration with broader theories such as situation (Barwise, 1989; Devlin, 1991) and/or information flow (Barwise and Seligman 1997). Such theoretical cross-fertilization has the potential to contribute to breakthroughs in the social sciences.

CONCLUSION

Agent simulation requires software and architectures that are a product of the design sciences (Simon, 1996). As a social research methodology, the resulting designs must be closely aligned with the theories and empirical insights that guide them. As substantive insights evolve, so will social simulation architecture.

The SSE framework developed in the preceding discussion is guided by four priorities: (1) an integrated multilevel scope, (2) a design process that focuses on the endogenous production of meaning, (3) mathematically grounded ontological models that support dynamics arising from contingency and interaction, and (4) the establishment of boundary structures that can define simulation classes. While design methodologies for social simulation will evolve, these SSE priorities are likely to have continuing relevance.

ACKNOWLEDGMENTS

I am grateful to Nick Collier for comments on an earlier draft of this paper.

REFERENCES

Agre, P.E., 1995, "Computational Research on Interaction and Agency," pp. 1–52 in P.E. Agre and S.J. Rosenschein, eds., *Computational Theories of Interaction and Agency,* Cambridge, MA: MIT Press.

Andersen, P.B., 2000, "Genres as Self-organizing Systems," pp. 214–260 in P.B. Andersen, C. Emmeche, N.O. Finnemann, and P.V. Christiansen, eds., *Downward Causation: Minds, Bodies and Matter,* Aarhus, Denmark: Aarhus University Press.

Axelrod, R., 1997, *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration,* Princeton, NJ: Princeton University Press.

Barwise, J., 1989, *The Situation in Logic,* Stanford, CA: CSLI Publications.

Barwise, J., and J. Seligman, 1997, *Information Flow: The Logic of Distributed Systems*, New York, NY: Cambridge University Press.

Bryson, J.J., 2001, *Intelligence by Design: Principles of Modularity and Coordination for Engineering Complex Adaptive Agents*, unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Bryson, J.J., and L.A. Stein, 2001a, "Architectures and Idioms: Making Progress in Agent Design," pp. 73–88 in C. Castelfranchi and Y. Lespérance, eds., *Seventh International Workshop on Agent Theories, Architectures, and Languages,* New York, NY: Springer.

Bryson, J.J., and L.A. Stein, 2001b, "Modularity and Design in Reactive Intelligence," pp. 1,115–1,120 in *Proceedings of the 17th International Joint Conference on Artificial Intelligence,* San Mateo, CA: Morgan Kaufmann.

Collins, R., 1981a, "Micro-translation as a Theory Building Strategy," pp. 81–108 in K. Knorr-Cetina and A.V. Cicourel, eds., *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-sociologies*, Boston, MA: Routledge & Kegan Paul.

Collins, R., 1981b, "On the Microfoundations of Macrosociology," *American Journal of Sociology* **86**(5):984–1,014.

Collins, R., 1994, "Why the Social Sciences Won't Become High-Consensus, Rapid-Discovery Science," *Sociological Forum* **9**(2):155–177.

Collins, R., 2000, "Situational Stratification: A Micro-Macro Theory of Inequality," *Sociological Theory* **18**(1):17–43.

Dean, J.S., G.J. Gumerman, J.M Epstein, R.L. Axtell, A.C. Swedlund, M.T. Parker, and S. McCarroll, 2000, "Understanding Anasazi Culture Change through Agent-based Modeling," pp. 179–207 in T.A. Kohler and G.J. Gumerman, eds., *Dynamics in Human and Primate Societies: Agent-based Modeling of Social and Spatial Processes*, New York, NY: Oxford University Press.

Devlin, K., 1991, *Logic and Information,* New York, NY: Cambridge University Press.

Devlin, K., 1994, "Situation Theory and Social Structure," pp. 197–237 in M. Masuch and L. Polos, eds., *Knowledge Representation and Reasoning under Uncertainty*, Berlin, Germany: Springer-Verlag.

Devlin K., and D. Rosenberg, 1993, "Situation Theory and Cooperative Action," pp. 213–264 in P. Aczel, D. Israel, Y. Katagiri and S. Peters, eds., *Situation Theory and Its Applications,* Vol. 3, Stanford, CA: CSLI Publications.

Devlin K., and D. Rosenberg, 1996, *Language at Work: Analyzing Communication Breakdown in the Workplace to Inform Systems Design*, Stanford, CA: CSLI Publications.

Epstein, J.M., and R. Axtell, 1996, *Growing Artificial Societies: Social Science from the Bottom Up,* Cambridge, MA: MIT Press.

Ferber, J., 1999, *Multi-agent Systems: An Introduction to Distributed Artificial Intelligence,* Harlow, England: Addison-Wesley.

Garfinkel, H., 1967, *Studies in Ethnomethodology,* Englewood Cliffs, NJ: Prentice-Hall.

Garfinkel, H., 2002, *Ethnomethodology's Program: Working Out Durkheim's Aphorism.* Lanham, MD: Rowan & Littlefield.

Genesereth, M.R., and N.J. Nilsson, 1987, *Logical Foundations of Artificial Intelligence,* San Mateo, CA: Morgan Kaufmann.

Goffman, E., 1983, "The Interaction Order," *American Sociological Review* **48**(1):1–17.

Hilbert, R.A., 1992, *The Classical Roots of Ethnomethodology: Durkheim, Weber and Garfinkel,* Chapel Hill, NC: University of North Carolina Press.

Hoffmeyer, J., 1996, *Signs of Meaning in the Universe,* Bloomington, IN: Indiana University Press.

Ihde, D., 1998, *Expanding Hermeneutics: Visualism in Science,* Evanston, IL: Northwestern University Press.

Juarrero, A., 1999, *Dynamics in Action: Intentional Behavior as a Complex System,* Cambridge, MA: MIT Press.

Kohler, T.A., J. Kresl, C. Van West, E. Carr, and R.H. Wilshusen, 2000, "Be There Then: A Modeling Approach to Settlement Determinants and Spatial Efficiency among Late Ancestral Pueblo Populations of the Mesa Verde Region," pp. 145–178 in T.A. Kohler and G.J. Gumerman, eds., *Dynamics in Human and Primate Societies: Agent-based Modeling of Social and Spatial Processes*, New York, NY: Oxford University Press.

Lake, M.W., 2000, "MAGICAL Computer Simulation of Mesolithic Foraging," pp. 107–143 in T.A. Kohler and G.J. Gumerman, eds., *Dynamics in Human and Primate Societies: Agent-based Modeling of Social and Spatial Processes*, New York, NY: Oxford University Press.

Padgett, J.F., 2000, "Closing Panel: Results and Prospects," pp. 259–262 in C.M. Macal and D.L. Sallach, eds., *Proceedings of the Workshop on Agent Simulation: Applications, Models, and Tools*, Argonne National Laboratory, Argonne, IL.

Palmer, R.E., 1969, *Hermeneutics: Interpretation Theory in Schleiermacher, Dilthey, Heidegger, and Gadamer*, Evanston, IL: Northwestern University Press.

Rawls, A.W., 1987, "The Interaction Order Sui Generis: Goffman's Contribution to Social Theory," *Sociological Theory* **5**(2):136–149.

Rawls, A.W., 1989, "Language, Self and Social Order: A Reformulation of Goffman and Sacks," *Human Studies* **12**:147–172.

Rawls, A.W., 2002, "Editor's Introduction," pp. 1–64 in *Ethnomethodology's Program: Working Out Durkheim's Aphorism*, Lanham, MD: Rowan & Littlefield.

Sallach, D.L., 2000, "Classical Social Processes: Attractor and Computational Models," *Journal of Mathematical Sociology* **24**(4):245–272.

Sallach, D.L., 2003, "Social Theory and Agent Architecture: Prospective Issues in Rapid-Discovery Social Science," *Social Science and Computer Review*, Summer.

Schelling, T.C., 1978, *Micromotives and Macrobehavior*, New York, NY: W.W. Norton.

Simon, H.A., 1996, *The Sciences of the Artificial*, Cambridge, MA: MIT Press.

Thagard, P., 1992, *Conceptual Revolutions*, Princeton, NJ: Princeton University Press.

Thagard, P., 1999, *Coherence in Thought and Action*, Cambridge, MA: MIT Press.

Thomas, W.I., 1967, *The Unadjusted Girl*, New York, NY: Harper & Row.

Wallace, M., 2002, "Constraint Logic Programming," pp. 512–532 in A.C. Kakas and F. Sadri, eds., *Computational Logic: Logic Programming and Beyond*, Berlin, Germany: Springer.

Weber, M., 1949, "Objectivity in Social Science and Social Policy," pp. 49–112 in E.A. Shils and H.A. Finch, eds., *The Methodology of the Social Sciences*, New York, NY: Free Press.

Friday, October 11, 2002

Invited Speaker:

Nigel Gilbert

VARIETIES OF EMERGENCE

N. GILBERT, University of Surrey, UK*

ABSTRACT**

The simulation of social agents has grown to be an innovative and powerful research methodology. The challenge is to develop models that are computationally precise, yet are linked closely to and are illuminating about social and behavioral theory.

The social element of social simulation models derives partly from their ability to exhibit emergent features. In this paper, we illustrate the varieties of emergence by developing Schelling's model of residential segregation (using it as a case study), considering what might be needed to take account of the effects of residential segregation on residents and others; the social recognition of spatially segregated zones; and the construction of categories of ethnicity. We conclude that while the existence of emergent phenomena is a necessary condition for models of social agents, this poses a methodological problem for those using simulation to investigate social phenomena.

INTRODUCTION

Emergence is an essential characteristic of social simulation. Indeed, without emergence, it might be argued that a simulation is not a *social* simulation. However, the notion of emergence is still not well understood (but see Sawyer 2002). In this paper, we consider the idea of emergence in a very simple way. We start with a simple model that can be applied to a wide variety of different phenomena, not just societies, but even atomic particles. We discuss how this model seems to show emergence and then suggest that to be useful as a simulation of social phenomena, the model needs to be made somewhat more complicated; and so we explore the consequences of adding several refinements. This will enable us to consider a number of different varieties of emergence. Finally, we draw some conclusions about the notion of emergence and make a methodological point.

THE SCHELLING MODEL OF RESIDENTIAL SEGREGATION

The example used here is already rather well known. Schelling (1971) published a paper in the *Journal of Mathematical Sociology* proposing a theory about the persistence of racial or ethnic segregation despite an environment of growing tolerance. He suggested that even if

---

* *Corresponding author address*: Nigel Gilbert, Centre for Research on Simulation in the Social Sciences (CRESS), School of Human Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK; e-mail: n.gilbert@soc.surrey.ac.uk.

** This is an edited transcript of the introductory talk given at the Workshop on Agent 2002 Social Agents: Ecology, Exchange, and Evolution Conference on October 11 and 12, 2002.

individuals tolerate[1] racial diversity, if they also remain uneasy about being a minority in the locality, segregation will *still* be the equilibrium situation.

The Schelling model consists of a grid of square patches. In the examples in this paper, the grid consists of 500 × 500 patches. There are 1,500 agents located on this landscape, initially at random, with no more than one on any patch. The majority of the agents, 70%, are green, and a minority are red. The remaining patches, shown in black in Figure 1, are vacant.

Each agent has a tolerance parameter. Green agents are "happy" when the ratio of greens to reds in its Moore neighborhood — the eight immediately adjacent cells or patches — is more than its tolerance. The reverse applies to the reds. So we can calculate in a straightforward way what percentage of agents are happy, given any particular configuration.

EMERGENCE OF CLUSTERS

If agents are randomly assigned to patches, an average agent has about 58%, or roughly 5 out of the 8, of its surrounding neighbors that are of its own color. In this situation, about 18%

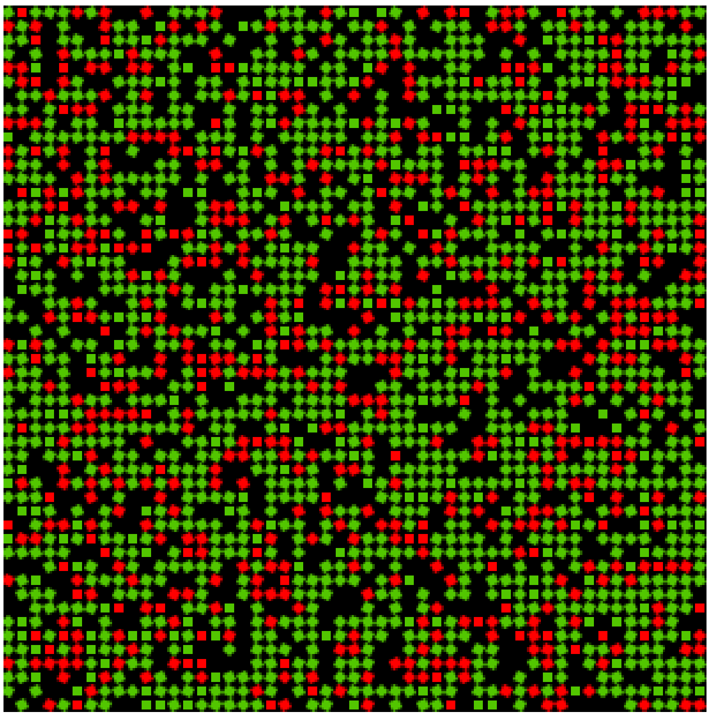

FIGURE 1 Initial Random Distribution of the 1,500 Agents: 70% Green and 30% Red

[1] The choice of the word *toleration* here is strange. We continue to use it because the literature talks about toleration. Nevertheless, we find the idea that minorities can only be 'tolerated' (rather than, for example, welcomed or celebrated) slightly repugnant.

of the agents are "unhappy." The exact percentage of unhappy agents for a particular configuration depends on the random distribution of the agents.

In this initial arrangement, there are no dynamics, no emergence, and no patterns of segregation. We just have an aggregation of cells where the number of unhappy agents can be explained analytically without much difficulty. Things get slightly more interesting when the unhappy agents are allowed to move. There are a variety of ways in which this can be implemented, the simplest being for the agent to select vacant patches at random until a congenial one is found. This can result in a phenomenon known as tipping, because when agents move to a position where they are happy, they may make other agents unhappy. These in turn will need to move, and so on.

The result is that, with moderate to low values of tolerance, the agents relocate so that they form clusters of agents all of the same color (Figure 2). The clustering, a feature of the grid as a whole, has *emerged* as a consequence of the rules obeyed by the individual agents. The extent of clustering can be measured by using statistics developed by geographers, such as the join count or Moran's contiguity ratio (Cliff and Ord, 1981; Cressie, 1991). However, we are only interested here in the fact that clustering has occurred, and this is clear from inspection of Figure 2.

Schelling showed that clustering occurs when we give the agents any value of tolerance much above 30%. As noted above, randomly allocating the agents to patches results in an average of about 58% of an agent's neighbors being of the same color. As a result of allowing the unhappy agents to move and the emergence of clusters, the percentage of same color neighbors rises to between 75% and 80%.

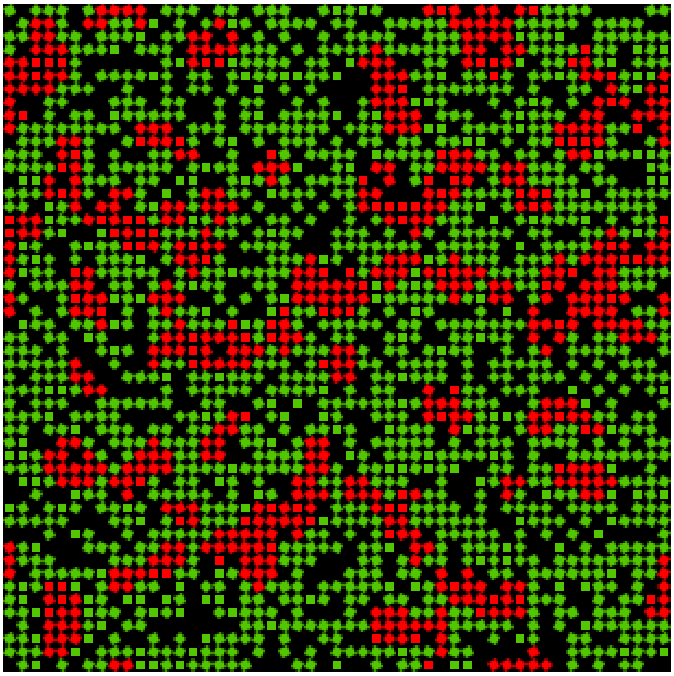

FIGURE 2 Emergence of Clustering after Unhappy Agents Have Been Allowed to Relocate by Random Walk

However, once the agents have located themselves in places where they are happy, all motion stops, giving a static, ‘frozen’ equilibrium. But that is an odd kind of model for a dynamic social world where agents are constantly on the move in some way or other. A more acceptable notion of emergence as far as social simulation is concerned is one in which emergence occurs *despite* the fact that the agents themselves are moving.

To illustrate this idea, John Holland (1975) suggests the physical analogy of the bow wave in front of a boat moving across water. Water particles constantly flow past the boat, but the bow wave itself is relatively stationary. However, few conventional definitions or descriptions of emergence insist on the need for emergent features to be maintained despite changes in the identities of the underlying elements.

What happens in the Schelling model if the agents are constantly being replaced? Let’s repeat the simulation exactly as before, except that a random 5% of the agents are substituted by agents of random color at every time step. The clusters remain, despite the fact that after about 20 steps, most of the agents have been replaced by other individuals. Emergent social phenomena persist, even though the agents themselves may come and go.

VALIDATION

In the United States, the level of residential segregation has remained high, despite the fact that the income inequality between blacks and whites is decreasing. There are antidiscrimination laws, affirmative action policies, and generally less discriminatory attitudes by whites. The Schelling model has been used as an explanation for the persistence of residential segregation despite all these positive, progressive social policies. Although the model is usually related to racial discrimination in the United States, there are other examples of residential segregation where it could be relevant. For example, in many cities in Europe, there are districts where Chinese or Turkish restaurants are found exclusively; in Majorca, there are segregated communities of English and German immigrants, and there is religious segregation as in Northern Ireland.

There is an increasing body of scholarship that relates the Schelling model to empirical data (e.g., Clark, 1991; Portugali, et al., 1994; Portugali, 1999; Sander, et al., 2000). The recurring theme of this work is to elaborate the basic model to take more account of the implications of the fact that the agents being modeled are human and members of society. For example, the effect of what has been called ‘downward causation’, in which the emergent clusters cause changes to the behavior of the individual agents, may need to be considered.

The clusters themselves can often act as though they were agents; for example, neighborhoods can lobby city governments. Moreover, because the agents represent not particles, but people, they often recognize and name the clusters/neighborhoods, and this might have some effect on their behavior in ways that affect the development of segregation. The agents in the basic Schelling model are all exactly the same. What happens if we introduce some degree of heterogeneity? People have the ability to talk and to interact symbolically. What difference could that make? In the remainder of this paper, we explore how one might add these complications to the basic Schelling model.

DOWNWARD CAUSATION

As we have seen in the basic model, individual actions can lead to emergent features, such as clusters and neighborhoods, visible at the societal or macro level. But we should also consider the ways in which such features can influence or constrain individual action. As an example of downward causation (Campbell, 1974), let us take a typical macro-level effect: the crime rate. A crime *rate* is necessarily a macro-level attribute because it is defined as the number of crimes committed by a population per unit time. A crime rate is not a meaningful measure for individuals. Let us assume that that cost of a home in each neighborhood depends in part on the crime rate (housing is cheap in areas with high crime rates) and that the crime rate depends on the ratio of reds and greens in the locality (the more reds, the higher the crime rate). Let us also propose that, instead of choosing new locations at random, agents can only move to spots where they can afford to buy or to rent, so that they are restricted by the property value of the new location relative to the value of their old location.

Figure 3 illustrates the typical result of running such a model, and its most noticeable characteristic is that it still has clusters. The poorer reds are forced to stay in their poor red districts. The richer greens have the ability to move where they want, but they like to be around other greens in green areas. There are a very few poor greens who are surrounded by reds and who cannot move to more desirable green areas.

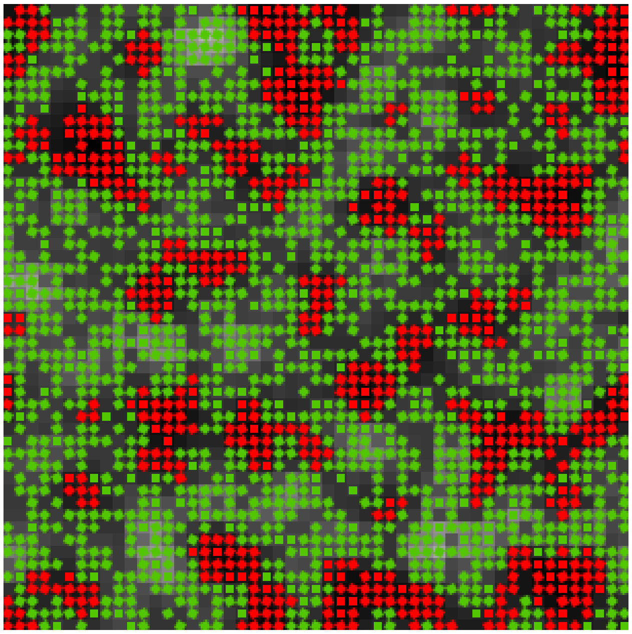

FIGURE 3 Model with Downward Causation [Background gray shade marks crime rate (black: high crime rate, low property values; white: low crime rate, high property values).]

SECOND ORDER EMERGENCE

People may recognize the neighborhoods in which they are living as having discernible boundaries, a name, and perhaps even a special history or culture. They may find the neighborhood particularly desirable for this reason (for example, fashionable neighborhoods in cities) or particularly undesirable. In other words, not only the researcher, but also agents themselves, can detect the presence of emergent features and act accordingly. And this, in turn, can affect what they do. This idea is known as second order emergence (Gilbert, 1995) or the double hermeneutic (Giddens, 1986). More precisely, second order emergence occurs when the agents recognize emergent phenomena, such as societies, clubs, formal organizations, institutions, localities, and so on, where the fact that you are a member, or not a member, changes the rules of interaction between you and other agents.

We can elaborate Schelling's model in a way that illustrates what one might mean by second order emergence by allowing patches to be labeled as red or green according to their past history. The agents recognize what is a good patch for them in terms of the labels that have been applied. The analogy is with a city district that may be generally recognized to be a good or bad place to live depending partly on its current characteristics, but also partly on its history. The result is shown in Figure 4. The picture looks familiar because once again, we have clear clustering.

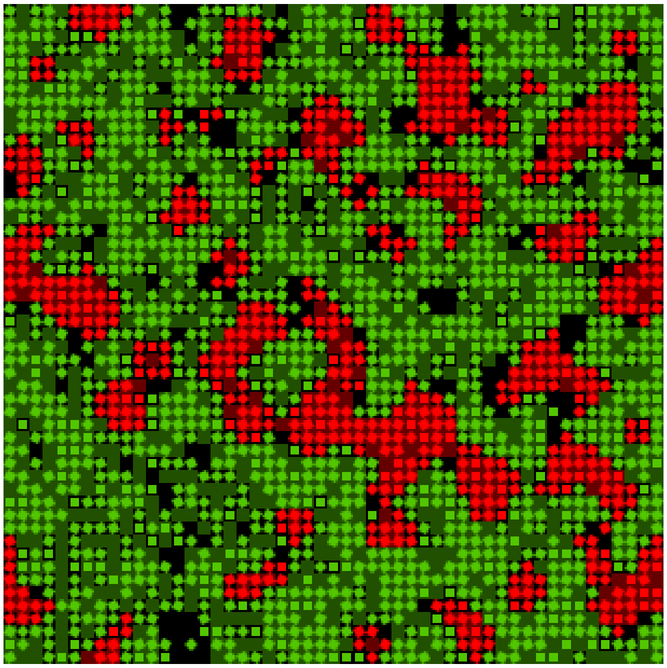

FIGURE 4 Model with Second Order Emergence [The colors of the patches (dark red or green) show the labels applied to the districts as a result of the color of the agents that were there previously or are there now.]

HETEROGENEITY

In all the models so far, the agents are identical, except for their location and color (red or green). They all have exactly the same tolerance. One can experiment with either random or systematic variations in tolerance, to correspond with environmental differences and inherited class differences.

If the tolerance for individual agents is randomly varied between agents, we get an even stronger clustering than before. If the tolerance value is arranged to correlate with the color of the agent, so that reds have a higher tolerance than greens, the reds become much more clustered than the greens (Figure 5).

How might correlations between tolerance and color arise in real populations? We might build into the model ideas of socialization, inheritance and class, and evolution or learning. However, these possibilities are not pursued here.

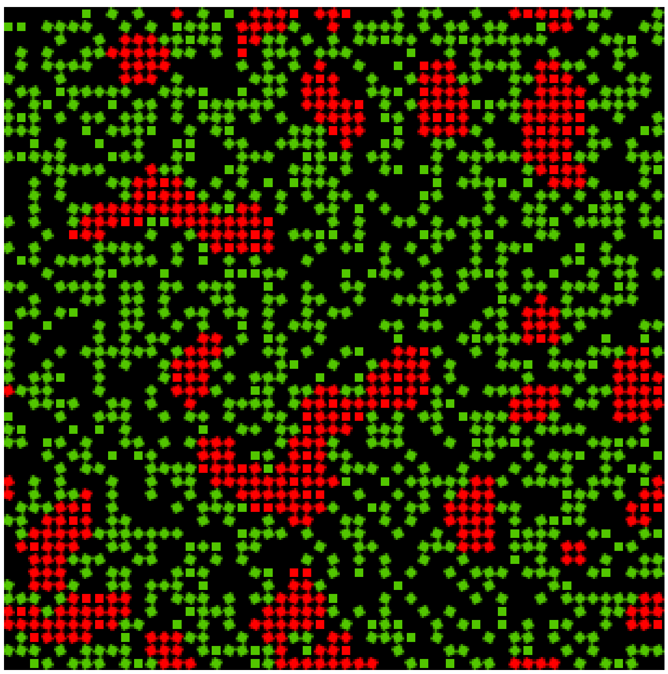

FIGURE 5 Model with Tolerance Related to Color (With tolerance at 55% for reds and 25% for greens, the reds become much more clustered than the greens.)

INTERACTION

Some of these models have depended on the idea that individual agents can conceptualize notions of neighborhood, recognize them, and communicate. But that in turn implies that we are dealing with agents that have some capacity for symbolic interaction. How might we represent this? There is a developing body of work on 'tag models' (e.g., Hales, 2001; Riolo, et al., 2001) in which agents have binary valued tags that can be interpreted in terms of color, ethnicity, class, education, gender, and so on. The agents act according to their tags and can also perceive the tags of other individuals.

This is rather like the Schelling model, except that instead of the modeler having chosen *a priori* that it is going to be color that marks the difference between the agents, the agents themselves decide, as it were, which of all their tags will become their significant characteristic. It could be 'color' or 'gender' or something else.

Here is a simple version. Each agent is given three binary tags. Agents are happy only if their neighbors are sufficiently similar to themselves, where similarity is measured by the Hamming distance between the agents' tags. The outcome is again a familiar one: the agents are clustered (Figure 6). However, in this simulation, the feature shared by the agents within each cluster varies from one cluster to another. This could represent a city in which, for example, one district is ethnically black, another is united because everybody speaks Japanese, and a third is dominated by stock traders.

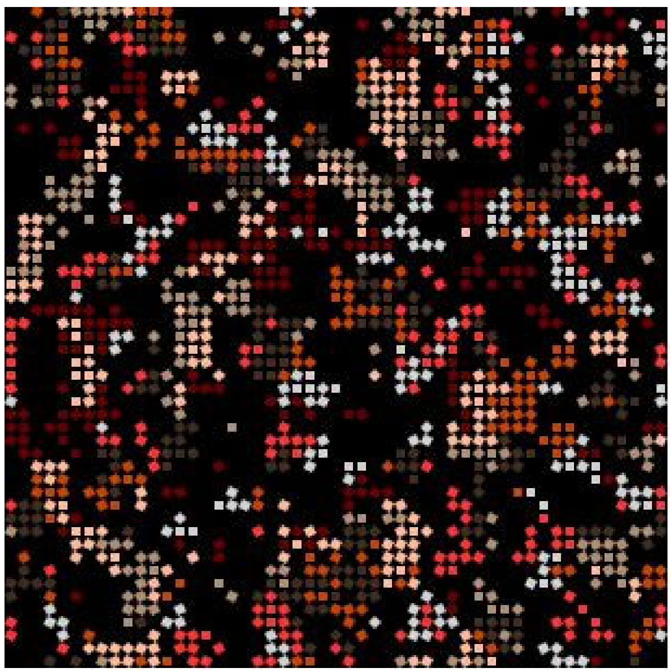

FIGURE 6 Model with Agents That Have Tags (Agents are colored according to the value of their tags, treated as a binary number.)

CONCLUSION

In this paper, we have illustrated some of the philosophical discussions about varieties of emergence, using a very simple computational model. We have tried to be straightforward about this, because there have already been some very illuminating although rather complex philosophical discussions about emergence in societies (Alexander, et al., 1987; Coleman, 1990; Archer, 1995; Sawyer, 2002). We have shown that verbal descriptions of types of emergence can be instantiated as rather simple computational models.

There is also a methodological conclusion from this exercise. All the models mentioned here seem to be adequate at some level of abstraction. Although the basic Schelling model is very simple, it did illustrate a surprising phenomenon: that 'tolerant' households could generate residential segregation through their locational decisions. We then showed that other features could be added to the model that seem to be fundamental to human societies, such as the ability to recognize emergent features. However, *all* the models yielded the same type of clusters of similar agents. The results of the simulations vary slightly in the form of the clusters and the degree of clustering, but not so much that it is plausible to conclude that one must be a better model of residential segregation than another.

The fact that we have observed emergence in all of these models cannot therefore be the sole criterion for choosing among them. The *Journal of Artificial Societies and Social Simulation*,[2] of which I am the editor, has published many papers that include an argument along the following lines: "I have developed and run a model, which shows some emergent features. The emergent features correspond to features in the real world, and since I have shown the correspondence of these features with empirical data, my model is therefore correct." A similar argument can be found in much of our social simulation literature.

We hope to have demonstrated that this kind of argument is not adequate. One has to validate a model at *both* the individual level *and* at the macro level before one can suggest that the simulation is a good representation of the social processes it is aiming to model.

ACKNOWLEDGMENTS

You may have recognized the technique of starting with a simple model and refining it, as I have done here. It was pioneered by Epstein and Axtell (1996) in *Growing Artificial Societies*, and I thank them for the idea. The models from which the figures in this paper were obtained were written in NetLogo (http://ccl.northwestern.edu/netlogo/), and I thank the developers for an excellent system.

REFERENCES

Alexander, J.C., B. Giesen, R. Münch, and N.J. Smelser, 1987, *The Micro-Macro Link*, Berkeley: University of California Press.

Archer, M.S., 1995, *Realist Social Theory: The Morphogenetic Approach*, Cambridge: Cambridge University Press.

---

2 http://www.soc.surrey.ac.uk/JASSS/

Campbell, D.T., 1974, "Downward Causation' in Hierarchically Organized Biological Systems," in *Studies in the Philosophy of Biology*, edited by F.J. Ayala and T. Dobzhansky, London: Macmillan.

Clark, W.A.V., 1991, "Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model," *Demography* **28**:1–19.

Cliff, A.D., and J.K. Ord, 1981, *Spatial Processes: Models and Applications*, London: Pion.

Coleman, J.S., 1990, *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.

Cressie, N.A.C., 1991, *Statistics for Spatial Data*, New York: Wiley.

Epstein, J.M., and R. Axtell, 1996, *Growing Artificial Societies: Social Science from the Bottom Up*, Cambridge, MA: MIT Press.

Giddens, A., 1986, *The Constitution of Society: Outline of the Theory of Structuration*, University of California Press.

Gilbert, N., 1995, "Emergence in Social Simulation," pp. 144–156 in *Artificial Societies: The Computer Simulation of Social Life*, edited by N. Gilbert and R. Conte, London: UCL Press.

Hales, D., 2001, *Tag Based Cooperation in Artificial Societies*, University of Essex, UK, Colchester.

Holland, J.H., 1975, *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI: University of Michigan Press.

Portugali, J., 1999, *Self-organization and the City*, Heidelberg: Springer.

Portugali, J., I. Benenson, and I. Omer, 1994, "Socio-spatial Residential Dynamics: Stability and Instability within a Self-organized City," *Geographical Analysis* **26**:321–340.

Riolo, R., M.D. Cohen, and R. Axelrod, 2001, "Cooperation without Reciprocity," *Nature* **414**:441–443.

Sander, R., D. Schreiber, and J. Doherty, 2000, "Empirically Testing a Computational Model: The Example of Housing Segregation," pp. 108–115 in *Proceedings of the Workshop on Simulation of Social Agents: Architectures and Institutions*, edited by D. Sallach and T. Wolsko, Chicago, IL: University of Chicago; ANL/DIS/TM-60, Argonne National Laboratory, Argonne, IL.

Sawyer, K.R., 2002, "Emergence in Sociology: Contemporary Philosophy of Mind and Some Implications for Sociological Theory," *American Journal of Sociology* **107**:551–585.

Schelling, T.C., 1971, "Dynamic Models of Segregation," *Journal of Mathematical Sociology* **1**:143–186.

DISCUSSION: OPENING SESSION*

R.K. SAWYER, Washington University, Moderator

David Sallach: This opening session might be conceived of as being in the realm of computational social theory. I'm sure that we are all pleased to get underway, so let me introduce the moderator of the opening session, Keith Sawyer, from Washington University.

Keith Sawyer: No one is better suited to start the conference than Nigel Gilbert. Agent-based social simulation is a new paradigm, and most new paradigms, I think, are classic convergent phenomena. They're not the responsibility of any one individual; rather, they're collective phenomena. Sometimes it seems as if certain individuals almost single handedly created the paradigm. If any one person had that claim, Nigel would qualify. He had one of the first edited volumes on the topic in 1994, and he was the author of a textbook in 1999. A classic sign that a paradigm is coming of age is the introduction of the first textbook. He's a sociologist as well as a computer modeling person. It's my honor to introduce Nigel Gilbert, all the way from the United Kingdom.

[Presentation by Gilbert]

Sawyer: Emergence has been an outstanding issue in all of the social sciences going back to the 19th Century and the founding theorist of economics, Karl Menger. The founding theorists of sociology, like Emile Durkheim all struggled with this issue more than 100 years ago. Now it seems that with this new methodology, we have a way to rediscover some of these same concerns and perhaps also a new way to help resolve and address some of the theoretical problems that are so long-standing in the social sciences. We have a few minutes now for questions.

Michael North: Mike North from Argonne National Laboratory. Of the models you've looked at, you said there's no "best" model. It would probably depend on the question you're trying to ask; that is, what are you trying to achieve with the model? It would also depend on including the important features of what you're trying to achieve. In that regard, would you say that everything else being equal, that simpler is better?

Nigel Gilbert: I've gone on record, saying that, yes, in these models simplicity is indeed a virtue. I always go for the simple models rather than the complicated ones. I think, though, that there's actually more to it than that. Bob Axelrod is presenting a paper in which he's advocating the KISS principle — Keep It Simple Stupid. I agree with that, but it all depends on what you're trying to do. Quite clearly, there are policy-related models where simplicity is a virtue, but achieving that virtue is extraordinarily hard, while still making it relevant for the policy concerns — if you talk about, let's say, segregation as a model, as an issue.

* Editor's note: The discussion sessions were recorded with the speakers' knowledge and then transcribed. The transcripts were edited for continuity and ease of reading; every effort was made to identify speakers and interpret comments accurately.

It would be nice to be able to say that the model we should choose depends on the question. Unfortunately, in real science, we don't often have a question formulated before we start. The question comes out of the research. It's an emergent property, if you like. So using the question to determine how complicated the model can be may not be a terribly helpful answer, although it's one that's easy to say.

Robert Reynolds: Bob Reynolds from Wayne State University, University of Michigan. Have you looked at extending the model to include goals for your agents and intentionality? And do you think that adding that would keep these clusterings, maintain them, or modify them in some way?

Gilbert: The direct answer to your question is that I have thought about extending. The reason that I didn't extend the model was that I wasn't sure what I was going to learn from it because there's a sense in which I've already done enough to demonstrate the kind of methodological point that I wanted to make. But we're still going to get clusters. I would be amazed if we didn't, in fact, get certain clusters, unless the goals that we gave the agents were really very strange goals. So one could say that we can, of course, take this basic model and elaborate it forever!

Reynolds: In your talk you mentioned "downward causation." It seems that, if these clusters are in some sense going to modify the actions of those agents, that giving the agents some intentionality would be effectively a way to do that.

Gilbert: So the fact that an agent is in a cluster affects the kinds of goals that they seek, for example.

Reynolds: Exactly.

Gilbert: I think that it would be an interesting thing to think about. As a sociologist, I would then say that's probably not the end of the matter, because the way in which people describe what they're doing and why they're doing it is itself a socially contextual matter. And so it isn't that you can get a set of objective intentions or goals from people, but those goals and intentions are themselves socially created. There may be end factors or a double layer there as well.

I suppose that what I'm saying in all of this — and what you're hinting at — is that this kind of exercise can be useful simply as a way of thinking through these kinds of sociological issues, even if at the end you get yet another boring model with another set of partly random clusters. It might actually be a useful exercise. Thank you for helping me make that point.

Charles Macal: Charles Macal also from Argonne National Laboratory. In your talk, it appeared in your concluding statements that you're shifting the burden onto the process or the notion of validation at both the micro and macro levels. I'm curious to see if you could describe the notion of validation in regard to a social model. Is it possible to validate or prove that a model is correct?

Gilbert: Yes, I think that's a very good question, which means I probably can't adequately answer it! Can I put the question back to you? What do *you* think I ought to mean by *validation?* Do you, in raising that question, want to be skeptical about the real possibility of validation of these models?

Macal: I would say that yes, I am skeptical about whether it's possible to validate social agent models and social system models, but I think that there's still a useful or constructive activity in terms of providing what I would call tests of *in*validation to models, because we can prove that a model is *not* correct, but we can't prove that a model *is* correct. Perhaps some generally agreed-upon series of invalidation tests that have been applied to a model would be acceptable to the larger community.

Gilbert: I have a lot of sympathy with that position, so perhaps I ought to say that I would argue that we would need to examine the *in*validity of our models at both the macro and the micro level.

Doug Lauen: Doug Lauen from University of Chicago. I'd like to follow up on the last question. As a person who's just getting into this field, I'm trying to figure out the relationship between agent-based modeling and equation-based solutions. I've been thinking that perhaps it's a way to develop good theoretical expectations; it's a way to build rigorous theory through deductive thought experiments. The last question was related to how we might work at it the other way. That is, we run an experiment through agent-based modeling and then look to actual data, empirical data, to validate the question. So my question is, what is the relationship between agent-based modeling and basically building theory and making equation-based validations of this type of procedure?

Gilbert: I'd like to say, and perhaps it is a good question to sum up, that the advantages of agent-based modeling, in my experience, are that it is *much* easier to observe emergent phenomena. It's only when you start doing agent-based modeling that the kinds of issues behind my talk become pressing. It is possible to say things about emergence if one is simply, or not simply, writing theoretically about emergence, or indeed if one is doing equation-based modeling. Keith is a good example of the former, but I would say that it really hits you between the eyes if you're doing agent-based modeling.

Adaptation and Networks

SOME METHODOLOGICAL ISSUES IN MODELS OF REINFORCEMENT LEARNING

J. BENDOR, Graduate School of Business, Stanford University
D. DIERMEIER, Kellogg School of Management, Northwestern University
M. TING, Department of Political Science and SIPA, Columbia University*

ABSTRACT

Behavioral game theory has become increasingly popular in social science applications. We discuss some of the methodological challenges in using this approach to study interactive decision making. We discuss its relationship to classical game theory, show that many existing applications of behavioral models lack empirical content, and provide a solution to this problem. Finally, we discuss the promise of behavioral models in solving long-existing puzzles in strategic interactions.

1 INTRODUCTION

The recent decade has witnessed a revolution in economic methodology. Increasingly, rational actor assumptions are being replaced by behaviorally oriented approaches ranging from learning models (both Bayesian and non-Bayesian) to myopic adaptation and stimulus-response models and evolutionary approaches. These ideas have even been introduced into game-theoretic models (e.g., Fudenberg and Levine, 1998; Young, 1998). This new field, *behavioral game theory*, promises to integrate the formal analysis of strategic interaction with psychologically plausible decision mechanisms. As any new approach, however, it also raises some important methodological problems.

In this paper, we focus on one particular class of behavioral models: reinforcement learning. Intuitively, reinforcement learning models are designed to capture the fundamental "Law of Effect" (Thorndike, 1911) whereby positive reinforcement increases the tendency to play an action, whereas negative reinforcement decreases it.

This class of learning models can be specified as follows. At each (discrete) time period t, an agent i can be described by the pair $[p_{i,t}(\alpha^i), a_{i,t}]$, where $p_{i,t}(\alpha^i)$ is i's propensity (i.e., probability) to play some action α^i, and $a_{i,t}$ is i's aspiration level. At $t = 1$, each player begins with initial (exogenously specified) aspirations and propensities. Players receive their stage-game payoffs $\pi_{i,t}$ and then can adjust both propensities and aspirations in response to their experiences. An outcome is coded as a "success" if $\pi_{i,t} \geq a_{i,t}$, and as a "failure" otherwise. Successes tend to increase the propensity of the chosen action; failures tend to decrease it. Specific models then differ as to how the adjustment process is defined. For example, propensity adjustment may be random or deterministic, and aspirations may be exogenously fixed or adjust to experience as well.

* *Corresponding author address*: Michael Ting, Department of Political Science and SIPA, Columbia University, New York, NY 10027-6902; e-mail: mmt2033@columbia.edu.

Perhaps the most well-known reinforcement learning model is the Bush-Mosteller model (1955). It is widely used in applications in sociology.[1] It is defined as follows. If an actor who takes action α^i codes the outcome as successful, then

$$p_{i,t+1}(\alpha^i) = p_{i,t}(\alpha^i) + \delta[1 - p_{i,t}(\alpha^i)],$$

where $\delta \in [0, 1]$ represents the speed of reinforcement learning or adaptation, given a successful outcome. Similarly, if the outcome was coded as a failure, then

$$p_{i,t+1}(\alpha^i) = p_{i,t}(\alpha^i) - \beta p_{i,t}(\alpha^i),$$

where $\beta \in [0, 1]$ represents the speed of inhibition. Finally, aspiration adjustment can be implemented by stipulating that tomorrow's aspirations are a weighted average of today's aspiration level and today's payoff (Cyert and March, 1963):

$$a_{i,t+1} = \lambda a_{i,t+1} + (1 - \lambda)\pi_{i,t},$$

where $\lambda \in [0, 1]$. Note that in the case $\lambda = 1$, the aspiration level is constant and thus determined exogenously.

The Bush-Mosteller model, however, not only relies on a particular functional form that specifies propensity and aspiration adjustment, but also requires adjustment to be deterministic. In contrast, reinforcement models in general may include probabilistic elements: random adjustment, random errors, trembles, and so forth. Rather than assuming a specific adjustment process, we may want to specify axioms that capture various models of reinforcement learning consistent with the Law of Effect. In Bendor, et al. (2002, 2003), we suggest the axioms presented in Sections 1.1 and 1.2.

1.1 Assumptions about Propensity Adjustment

For each agent i, let $\{\alpha^i_j\}_{j=1,\ldots,m(i)}$ denote her set of actions. In addition, let $\underline{p}_i$ and $\overline{p}_i$ denote agent i's lowest and highest feasible propensity levels, respectively.

A1 (positive feedback): If i used action α^i in t and if $\pi_{i,t} \geq a_{i,t}$, then $\Pr[p_{i,t+1}(\alpha^i) \geq p_{i,t}(\alpha^i)] = 1$; if $p_{i,t}(\alpha^i) < \overline{p}_i$ and $\pi_{i,t} > a_{i,t}$, then $\Pr[p_{i,t+1}(\alpha^i) > p_{i,t}(\alpha^i)] = 1$.

A2 (negative feedback — direct effect): If i used action α^i in t and if $\pi_{i,t} < a_{i,t}$, then $\Pr[p_{i,t+1}(\alpha^i) \leq p_{i,t}(\alpha^i)] = 1$; if $p_{i,t}(\alpha^i) > \underline{p}_i$, then $\Pr[p_{i,t+1}(\alpha^i) < p_{i,t}(\alpha^i)] = 1$.

A3 (negative feedback — indirect effect): If i used action α^i_r in t and if $\pi_{i,t} < a_{i,t}$, then for every other action a^i_s (where $s \neq$ r), $\Pr[p_{t+1}(a^i_s) > 0] > 0$.

[1] For a recent example and further references, see, for example, Macy and Flache (2002).

Axioms A1 and A2 formalize the core idea of reinforcement learning: the propensity to take an action responds to positive and negative feedback. Note that both A1 and A2 allow for probabilistic adjustment rules. In addition, A3 requires that no action is *a priori* excluded; rather, each other action must be reachable with some (arbitrarily small) probability. Note that A3 does *not* require that there be any new propensity vector that i moves to in $t + 1$ in which *all* actions (other than the one used in t) receive positive weight. Instead, there could be a set of propensity vectors, one in which α_1^i gets positive weight, another in which α_2^i does not, and so on. Note also that if an agent has only two actions, then A3 is implied by A2.

1.2 Assumptions about Aspiration Adjustment

Many applications of reinforcement learning assume a fixed aspiration level (e.g., Macy, 1991). However, positing fixed aspirations precludes an important kind of learning: aspirations should reflect one's payoff experience. Indeed, to assume otherwise — to keep aspirations constant in the face of discrepant evidence — seems inconsistent with the spirit of the underlying research program: agents learn from experience. Thus, reinforcement models should be flexible enough to capture endogenous aspirations, as expressed in the following three axioms:

A4: If $\pi_{i,t} = a_{i,t}$, then $\Pr(a_{i,t+1} = a_{i,t}) = 1$.

A5: If $\pi_{i,t} > a_{i,t}$, then $\Pr(a_{i,t} < a_{i,t+1} \leq \pi_{i,t}) = 1$.

A6: If $\pi_{i,t} < a_{i,t}$, then $\Pr(\pi_{i,t} \leq a_{i,t+1} < a_{i,t}) = 1$.

In addition to rules on propensity and aspiration adjustment, reinforcement models also may want to allow for *inertia*. That is, while agents learn by experience, these codings might not invariably lead to adjustments in propensities or aspirations (the agent may be engaged with other matters). With some (perhaps very small) probability, agents may not change their propensities or aspirations. Including randomness and inertia has two benefits. First, it allows the model to capture more general behavioral assumptions. Second, these two features also dramatically enhance the model's predictive power. However, to see this second consequence, we need to define the explanatory concept of reinforcement learning models.

2 EQUILIBRIA

When formally analyzing a social system, we first need to specify the possible states of the system S. In a normal form game, these are, for instance, the strategy combinations. Deterministic theories (such as classical game theory[2]) then identify a set of predicted outcomes from the set of possible states. In its simplest form, an explanatory concept E is a correspondence that selects from each set of possible outcomes S of a collective decision process a (possibly empty) subset $E(S)$. In empirical terms, an observed outcome in $E(S)$ is consistent with the theory, whereas an outcome outside of $E(S)$ is not. The set $E(S)$ thus constitutes the empirical content of the theory. Note that $E(S)$ need not be a singleton; that is, the theory may not predict a unique outcome.

---

[2] For the moment, we consider only pure strategies. The issue of mixed strategies is discussed below.

In identifying the predicted set, different theories rely on different features of the social system. Classical game theory, for example, assumes that each social system can be represented as a game consisting of a game form and preferences for each actor.[3] In the case of normal form games, the game form consists of a finite set of players N, and for each agent, $i \in N$ is a nonempty set Σ_i of actions available to i. Hence, $\Sigma := \times_{j \in N} \Sigma_j$ serves as the state space S. Preferences are usually given by a von Neumann-Morgenstern utility function for each agent i: $u_i : \Sigma \rightarrow \Re$. Thus, a normal form game is given by a tuple $[N, (\Sigma_i)_{i \in N}, (u_i)_{i \in N}]$. Any explanatory concept in noncooperative game theory is therefore of the form $E[\Sigma; N, (\Sigma_i)_{i \in N}, (u_i)_{i \in N}]$.

Classical game theory uses various explanatory concepts of this form, such as minmax, rationalizability, etc. By far the most widely used solution concept, however, is Nash equilibrium. Nash equilibrium identifies its explanatory set as follows. For any action profile $\sigma := (\sigma_i)_{i \in N}$ with $\sigma_i \in \Sigma_i$, let σ_{-i} denote the partial profile of σ for all players except for i. Then we can rewrite $\sigma = (\sigma_i, \sigma_{-i})$. We then arrive at Definition 1.

Definition 1: A tuple σ is a Nash equilibrium (i.e., $\sigma \in E_{\text{Nash}} [\Sigma; N, (\Sigma_i)_{i \in N}, (u_i)_{i \in N}]$ if for all i and $\sigma'_i \in \Sigma_i$:

$$u_i(\sigma_i, \sigma_{-i}) \geq u_i(\sigma'_i, \sigma_{-i}).$$

As an example, consider the two-person Prisoners' Dilemma where C denotes "cooperation" and D denotes "defection." Then $S = \Sigma = \{(C, C), (C, D), (D, C), (D, D)\}$ and $E_{\text{Nash}}[\Sigma; N, (\Sigma_i)_{i \in N}, (u_i)_{i \in N}] = \{(D, D)\}$.

Theories can fail in two ways. If $E(S) = S$, the theory has no explanatory power. If $E(S)$ is empty, the theory does not have any implications in situation S. For an empirical theory, this means it does not make any predictions. Its domain is restricted to decision contexts other than S.

One of the advantages of using Nash equilibrium to model strategic interaction is that in many applications $E_{\text{Nash}}(\Sigma; .)$ is nonempty, particularly if the set of actions is enlarged to allow for mixed strategies.[4] On the other hand, in some important applications Nash equilibrium lacks empirical content. Perhaps the most well-known example is repeated interaction. The problem is formally expressed in the following "folk theorem": any feasible payoff vector that guarantees each player her individually rational payoff (i.e., what she can achieve irrespective of the other players' actions) can be sustained as a (subgame-perfect) Nash equilibrium provided players are sufficiently patient with respect to future payoffs (Fudenberg and Maskin, 1986).[5]

The issue of which explanatory concept to use is less settled in models of reinforcement learning. Macy and Flache (2002), for example, have recently suggested an analogue to Nash equilibrium for the Bush-Mosteller model, called Self-Reinforcing Equilibrium (SRE). In an SRE, players' propensities to try certain actions generate payoffs, and hence feedback, which are

---

[3] For details on the definition of noncooperative games and utility functions, see, for example, Osborne and Rubinstein (1994). For simplicity, we focus on normal form games.

[4] For the most important existence results, see, for example, Osborne and Rubinstein (1994). They also provide a detailed discussion on how to interpret mixed strategy equilibria.

[5] Note that problem applies not only to Nash equilibria, but also to subgame-perfect Nash equilibria.

consistent with those original propensities. Aspirations, in turn, must be consistent with payoffs generated by a given propensity vector. Thus, the combination of propensities, aspirations, actions, and payoffs form an equilibrium in which all these different elements reinforce each other.

Definition 2: A tuple $(p_{i,t}; a_{i,t})_i$ is an SRE if for all i, t, and α^i:

(*i*) $p_{i,t+1}(\alpha^i) = p_{i,t}(\alpha^i)$

and

(*ii*) $a_{i,t+1} = a_{i,t}$.

We say an outcome is *stable* if it is generated by an SRE. In an SRE, the state space now consists of tuples $(p_{i,t}; a_{i,t})_i$. An SRE thus induces a (possibly degenerate) distribution over outcomes determined by $(p_{i,t})_i$. Only in the case of *pure* SREs (i.e., for each i an α^i exists such that for all (t : $p_{i,t}$ $(\alpha^i) = 1$), do we have a well-defined (induced) explanatory concept over Σ. This limits general existence properties for SREs.[6] In addition to the parameters N, $(\Sigma_i)_{i\in N}$, and $(u_i)_{i\in N}$, such an explanatory concept would also depend on the rules for changing propensities and aspirations.

As discussed above, explanatory concepts can be deficient in two ways. In addition to the lack of general existence properties, they may lack empirical content. As shown in Bendor, et al. (2002), this is the case for SRE. Consider the following very general axiom on positive reinforcement.[7]

Axiom 1 (A1*): For all i, t, and action α^i chosen by player i in period t, if $\pi_{i,t} \geq a_{i,t}$, then $p_{i,t+1}(\alpha^i) \geq p_{i,t}(\alpha^i)$.

We can then show the following two theorems:

Theorem 1 (Bendor, et al., 2002): Consider any normal form game in which players adjust their action propensities by any arbitrary mix of adaptive rules that satisfy Axiom (A1*) and where aspirations are exogenously fixed. Any outcome of the stage game can then be sustained as a stable outcome by some pure SRE.

Theorem 2 (Bendor, et al., 2002): Consider any normal form game in which players adjust their action propensities by any arbitrary mix of adaptive rules that satisfy Axiom (A1*) and adjust their aspirations by any arbitrary mix of rules that satisfy Axiom (A4). Any outcome of the stage game can then be sustained as a stable outcome by some pure SRE.

Thus, any reinforcement learning model satisfying axioms (A1*) and (A4) has no empirical content.

---

6 See Bendor, et al. (2003) for examples of 2×2 games that only possess equilibria with pure propensities. Examples include Chicken or the Prisoners' Dilemma for certain parameter values.

7 Note that (A1*) is weaker than (A1).

It is worth noting that both theorems have very large domains in several important respects. First, they hold for any number of players, including one-person decision problems. Second, each player can have any number of actions, finite or infinite. Third, the game can be symmetric or asymmetric. Hence, the players can, for example, have completely different sets of actions. Fourth, the results do not even require that players continue to use the same adaptive rule over time: a person could switch to different methods of adjusting his/her action propensities or aspirations, provided only that new rules continue to satisfy Axiom (A1) and Axiom (A2), respectively.[8] In particular, the Bush-Mosteller model satisfies both axioms.[9]

3 LIMITING DISTRIBUTIONS

A possible solution to the methodological problems of using SREs as explanatory concepts is to give up equilibria as explanatory concepts altogether and use a stochastic process approach. That is, rather than identifying a set of *states* as the empirical content of the model, we can use a set of probability distributions. That is, we model reinforcement learning as a (stationary) Markov chain and then use the chain's stationary distributions as our explanatory concept.

Specifically, for each i and t, we assume that there is a finite, time-invariant number of propensity levels (not necessarily the same for each individual i). To represent random propensity adjustment, we define for each i a family of random variables $\{P_{i,t}\}_{t\in N}$ with values drawn from the list of feasible propensity levels. Propensity adjustment then corresponds to a (stochastic) dynamic process. By putting point mass on one of the possible realizations, we can also capture deterministic adjustments (e.g., the Bush-Mosteller rule).

As in the case of propensities, we assume that for each i and t, there is a finite, time-invariant number of aspiration levels (not necessarily the same for each individual i). Again, we allow for random adjustment (with point mass in the case of deterministic adjustments). So, formally for each i : $\{A_{i,t}\}_{t\in N}$ is a family of (possibly degenerate) random variables. We assume that $\{P_{i,t}\}_{t\in N}$ and $\{A_{i,t}\}_{t\in N}$ are mutually independent stationary processes.

A third potential source of randomness can originate from stochastic payoffs (i.e., the assumption that the payoff to a player is not completely determined by the choices of all players, but also has a random component). That is, payoffs are modeled as a nondegenerate (conditional) probability distribution with finite support for each action profile. For each action profile outcome o, we denote realized payoffs by $\pi_{i,t}(o)$ with corresponding stationary random variables Π_i. Let $\underline{\pi}_i(o)$, denote agent i's minimal possible payoff given outcome o and $\overline{\pi}_i(o)$ her maximal payoff. For example, in the two-person Prisoners' Dilemma (PD) $\pi_{i,t}(C, C)$ denotes agent i's payoff at time t, given that both agents have cooperated. We assume that payoff realizations are mutually independent across agents and time.

---

[8] Indeed, Theorem 1 is even more general. In contrast to Theorem 2, it also allows for changing payoffs. See Bendor, et al. (2002) for details.

[9] Bendor, et al. (2002) prove an analogous result for theories of satisficing.

Different payoff assumptions then correspond to different restrictions on the respective distribution, such as assumptions on the ordering of expectations or the supports of the random variables. These restrictions can be applied to different aspects of the distribution. For example, we can require that each agent's expected payoff from mutual cooperation in the two-person PD is strictly higher than the expected payoff from mutual defection. Formally,

$$E[\Pi_i\,(C, C)] > E[\Pi_i\,(D, D)].$$

Alternatively, we can assume that distributions are ordered in terms of their best or worst possible realizations. For example, we can assume that each agent's maximal payoff from mutual cooperation in the two-person PD is strictly higher than the *maximal* payoff from mutual defection:

$$\bar{\pi}_i\,(C, C) > \bar{\pi}_i\,(D, D).$$

Of course, which one of these assumptions makes sense depends on the phenomenon being modeled.

We can now describe a full cycle of learning. In each period t, an agent is endowed with a vector of propensity levels $p_{i,t}$ and an aspiration level $a_{i,t}$. Initially (i.e., for $t = 1$), these levels are assigned arbitrarily. Given the realized action of each agent, each agent receives a randomly drawn payoff conditional on the outcome of the election and her own action. We assume that propensity adjustment is inertial with probability $\varepsilon_p > 0$ and aspiration adjustment with probability $\varepsilon_a > 0$, and that these probabilities are i.i.d. across agents and periods. This leads to a propensity adjustment with probability $1 - \varepsilon_p$ and to an adjustment of the agent's aspiration level with probability $1 - \varepsilon_a$. So, with probability $\varepsilon_a \varepsilon_p$ the agent is completely inertial.[10] We assume that agents' behavior can be described by some (not necessarily the same) rules for propensity and aspiration adjustments that satisfy axioms (A1) through (A6).

Our model then defines a discrete-time, finite-state Markov process. That is, we have a family of random variables $\{X^t : t \in N\}$, where X^t assumes values on the state space $S = \times_{i=1,\dots,N} S_i$, and S_i consists of elements of the form $(p_i, a_i) =: s_i$. Generic states are thus of the form $(p_i, a_i)_{i=1,\dots,N}$, denoted s. Note that given the independence assumptions on $\{P_{i,t}\}_{t\in N}$ and $\{A_{i,t}\}_{t\in N}$, $\Pr(X^t = s'|s) = \Pi_{i=1,\dots,N} \Pr(X_i^t = s'_i|s)$, where $\{X_i^t\}$ is the (decomposed) family of random variables assuming values on S_i. Since transitions $\{P_{i,t}\}_{t\in N}$ and $\{A_{i,t}\}_{t\in N}$ are stationary, we have a stationary Markov process.

Whereas (classical) game theory uses an equilibrium approach as its explanatory concept, our behavioral approach uses probability distributions over states of the form $(p_i, a_i) =: s_i$. Note that any such probability distribution induces a probability distribution over Σ. We now need to investigate the properties of this process.

Bendor, et al. (2002) show that, under certain mild conditions, the stochastic process is ergodic: it has a unique limiting distribution. Specifically, we prove the following result.

10 Note that an agent may be inertial with respect to only propensities or aspirations or both.

Theorem 3 (Bendor, et al., 2002): An aspiration-based adaptive process has a unique limiting distribution if any of the following conditions hold:

1. **Action trembles:** With a positive probability (which is i.i.d. across periods and independent across players), player *i*, instead of doing what he intended to do, "experiments" by randomly playing some action given by a totally mixed vector of probabilities over feasible actions. (This vector is i.i.d. across periods and independent across players.) Further, in the stage game, there is an outcome *o* in which nobody gets his minimal payoff (i.e., $\pi_i(o) > \underline{\pi}_i$ for all i).

2. **Extreme propensities excluded:** Neither 0 nor 1 is a feasible propensity value for any action for any player. Further, in the stage game, there is an outcome in which nobody gets his minimal payoff (i.e., $\pi_i(o) > \underline{\pi}_i$ for all i).

3. **Stochastic payoffs:** Every vector of actions produces a (nondegenerate) distribution of payoffs for every agent, where each distribution is finitely valued. Payoffs are i.i.d. across periods and independently distributed across players.

4. **State trembles:** With positive probability (again i.i.d. over periods and independently across players), *i*'s state can randomly tremble to any neighboring state on his grid.

Thus, we not only can ensure existence, but also uniqueness. On the other hand, the explanatory concept is inherently probabilistic. While this may actually be an advantage for empirical work, it does make a comparison with the results of classical game theory more difficult.

4 COMPARISON WITH CLASSICAL GAME THEORY

To facilitate a comparison with classical game theory (e.g., Nash equilibrium), it is useful to use an equilibrium concept defined on Σ. This can easily be derived using the following approach. Suppose we use the Bush-Mosteller rule with state trembles (i.e., an instance of case 4 in Theorem 3).[11] Now consider an initial tremble probability and then gradually reduce it toward zero, holding all other parameters constant.[12] This yields a sequence of (unique) limiting distributions and their associated statistics (e.g., the population's average propensity to cooperate). As the tremble probabilities get sufficiently small, by continuity, further diminutions in these probabilities can have only negligible effects on the associated limiting distributions. In short, as the trembles go to zero, the limiting distributions converge. In the limit, we are left with a distribution that assigns non-zero probability only to finitely many states (often a unique state). However, when the tremble probability is exactly zero (not arbitrarily small, in the limit), the corresponding learning rule would be subject to Theorem 2. That is, it would lack empirical content.

[11] Similar approaches can be defined for all versions of stochastic reinforcement learning.

[12] This approach corresponds to Foster and Young's (1990) concept of a stochastically stable state. However, in contrast to our reinforcement learning model, they consider perturbed best response dynamics.

This approach again leads to an explanatory concept defined on sets of states, which allows us to directly compare its predictions with the predictions of classical game theory. As we show in a series of related papers (Bendor, et al., 2002, 2003, n.d.), the predictions are fundamentally different. That is, in contrast to other forms of learning (e.g., Bayesian learning, perturbed fictitious play),[13] reinforcement learning cannot serve as a behavioral microfoundation for Nash equilibrium. This is true both for point predictions and for comparative analysis. For example, as shown in Bendor, et al. (2003), reinforcement learning in general does not select Nash equilibrium in 2×2 games like Chicken or PD. In PD, for instance, reinforcement learning may select the strictly dominated strategy combination (C, C) as the most likely state.

As in classical game theory, we can conduct comparative statics analysis, that is, change the parameters (e.g., the payoffs) of the game form, and then investigate how the model's predictions (Nash equilibria, in classical analysis; limiting distributions, in our case) change in response. Consider the case of mixed Nash equilibria versus mixed propensities. One of the most notorious predictions of classical game theory is that in a mixed strategy equilibrium an agent's mixing probabilities do not respond to changes in his payoffs, but to the changes in his *opponent's* payoffs. This does not hold in reinforcement models. In those models, each agent's payoff changes alter his or her *own* propensities.[14]

5 CONCLUSION

This paper discusses some methodological issues in models of reinforcement learning. We show that a plausible variation of equilibrium analysis (Self-Reinforcing Equilibrium) due to Macy and Flache (2002) lacks empirical content. We then show that a stochastic version of reinforcement learning allows us to make a unique, but probabilistic, prediction. This insight permits us to define an induced equilibrium concept that can be directly compared to the predictions of Nash equilibrium. We then discuss differences between reinforcement learning and the classical analysis of games. That is, reinforcement learning offers a conceptually and empirically distinct alternative to the classical, rational choice-based approach. This suggests that we can use reinforcement learning modes to solve persistent puzzles in the application of game theoretic models to empirical phenomena. Examples include cooperation in the Prisoners' Dilemma (Bendor, et al., 2003) or the solutions to the turnout paradox in models of electoral participation (Bendor, et al., n.d.).

6 ACKNOWLEDGMENTS

We would like to thank Adam Meirowitz for his very helpful comments on an earlier draft and Sunil Kumar, Dilip Mookherjee, and Debraj Ray for many illuminating conversations over the years.

[13] See, for example, Fudenberg and Levine (1998) or Young (1998) for a detailed discussion of this literature.

[14] See Bendor, et al. (2003) for details and other examples.

7 REFERENCES

Bendor, J., D. Diermeier, and M. Ting, n.d., "A Behavioral Model of Turnout," *American Political Science Review* (forthcoming).

Bendor, J., D. Diermeier, and M. Ting, 2002, "The Empirical Content of Adaptive Models," Mimeo, Stanford University (unpublished).

Bendor, J., D. Diermeier, and M. Ting, 2003, "Recovering Behavioralism: Adaptively Rational Strategic Behavior with Endogenous Aspirations," in K. Kollman and S. Page (eds.), *Computational Political Economy*, Cambridge: MIT Press, pp. 213–274.

Bush, R., and F. Mosteller, 1955, *Stochastic Models of Learning*, New York: John Wiley and Sons.

Cyert, R., and J.G. March, 1963, *A Behavioral Theory of the Firm*, Englewood Cliffs, NJ: Prentice-Hall.

Foster, D., and H.P. Young, 1990, "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology* **38**:219–232.

Fudenberg, D., and E. Maskin, 1986, "The Folk Theorem for Repeated Games with Discounting and Incomplete Information," *Econometrica* **54**:533–554.

Fudenberg, D., and D. Levine, 1998, *The Theory of Learning in Games*, Cambridge: MIT Press.

Macy, M., 1991, "Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange," *American Journal of Sociology* **97**:808–843.

Macy, M., and A. Flache, 2002, "Learning Dynamics in Social Dilemmas," *Proceedings of the National Academy of Sciences*, **99**:7229–7236.

Osborne, M., and A. Rubinstein, 1994, *A Course in Game Theory*, Cambridge: MIT Press.

Thorndike, E.L., 1911, *Animal Intelligence: Experimental Studies*, New York: MacMillan.

Young, H.P., 1998, *Individual Strategy and Social Structure*, Princeton: Princeton University Press.

VIABILITY OF COOPERATION IN EVOLVING INTERACTION STRUCTURES

N. HANAKI, Columbia University, New York, NY
A. PETERHANSL, Columbia University, New York, NY*

ABSTRACT

The emergence and sustainability of cooperation are examined in multiperson local public good provision games in which partner selection is endogenized. The games take place on a social network, that is, a collection of players, each of whom is acquainted with a subset of the other players (i.e., "partners"). The network evolves over time as players sever ties and create new ties among themselves based on who their current partners recommend. The share of contributors of the public goods is analyzed, with a specific focus on the explicit dependence between old and new ties. The interaction structure necessary for cooperation and cooperation is shown to co-evolve in the system.

1 INTRODUCTION

> A modern economy is characterized by interaction, both direct and indirect, between individuals. Three aspects of this interaction are important. First, individuals interact in different ways. Second, agents learn over time. Third, interaction takes place through networks. — Kirman (1997, p. 340)

This paper attempts to capture Kirman's view of the economy by focusing on the three aspects addressed above. We explicitly model (1) players interacting on multiple levels, (2) players learning over time, and (3) players interacting through a network. These three aspects, taken together, are intimately related to the four important features in our framework:

- Agent interactions are local.
- The decision process of an agent for the interaction (that is, the agent's strategy) is based on local information.
- An agent's decision process is shaped by his neighborhood (i.e., the local environment of other players) in which he resides.
- The local environment of an agent is shaped by the agent's decision process.

The relational structure of an agent (vis-à-vis the other players) and an agent's preferences have simultaneous feedback to each other. Thus, we incorporate both an agent's relational structure into the decision-making process and the agent's decision-making process into the evolution of the relational structure.

* *Corresponding author address*: Alexander Peterhansl, Department of Economics, Columbia University, 1022 International Affairs Building, 420 West 118th Street, New York, NY 10027; e-mail: ap11@columbia.edu.

This interplay between strategy and structure[1] is really a restatement of the interplay between micro- and macrostructure that has long been recognized in economics and the social sciences in general (see Tesfatsion [2002] for references). We are interested in investigating this interplay between micro- and macrostructure from an evolutionary perspective. That is, given a few simple adaptive rules, how do strategies and structure co-evolve?

In the following sections, we approach this problem by investigating it in the context of providing local public goods. In particular, we present a new model for studying the emergence and sustainability of cooperation in the game. The literature on the sustainability of cooperation has looked at a number of mechanisms that yield sustained cooperation (see Eshel, et al. [1998] for references). What remains uninvestigated, however, is the sustainability of cooperation in a population evolving in its strategy and its structure.

This paper takes Kirman's idea of the importance of the network to the extreme. It does so by incorporating social structure into the agent decision problem. Thus, we offer one formalization of Mark Granovetter's embededness argument (Granovetter, 1985). Raub and Weesie (1990) and Montgomery (1998) offer two additional formalizations of embededness. Our model has elements of both of these approaches, though we do not discuss this here. In a sense, we are 'socializing' the agent decision problem and show that 'socialization' can serve as an important coordination mechanism.

2 THE EVOLVING NETWORK AND PREFERENTIAL TRIAD FORMATION

This section briefly and informally describes the network of players and how the network evolves, as ties are severed and created among players. The central issue in this 'rewiring' process is how to evolve the network contingent on agent preferences. Because our algorithm is quite particular, this section serves as a narrative prelude for the formal presentation in Section 3.

Interaction among players in the population takes place between players, each of whom interacts with a number of partners, a subset of the other players. The connection between a player and a player's partners does not need to be spatially interpreted. It is meant to serve as a more general framework of a player's relational structure relative to the other players.

Players have the ability to evaluate and *change* their partners. In other words, players are not passively subject to their initial partners but can take active measures to alter their environment so as to surround themselves with more suitable players.

Players have opinions of their partners. When players update their partners, they take two pieces of information into account: (1) the opinions they have of their partners and (2) their strategy for the game. Together, these determine which partner is least desirable.[2] The tie to this partner will be severed and a new partner will be sought out.

---

[1] Here, we use strategy in a general manner, that is, strategy that determines a player's actions from one period to the next. Structure is also used generally and refers to the relational structure among agents. It is, in short, the topology on which the agents reside.

[2] The candidate link to be served is actually least desirable for *both* players giving up the link, as discussed in Section 3.

The process of new partner selection is based on "preferential triad formation." The triad is a fundamental sociological concept that looks at a subset of three actors (players in our case) and the possible connections among them. Our algorithm is inspired by balance theory and the idea of transitivity: if player i likes player j and player j likes player k, then player *i will most likely meet and like* player k.[3] The economics literature (e.g., Jackson and Watts, 2002) tends to have endogenously formed new connections to random agents. We believe that our process is a meaningful way to break out of this assumption of randomness. Evolving social connections are, after all, characterized by their distinct non-randomness (see Figure 1 for an example). Players i and j interact with one another. Players j and k also interact with one another. If all of these players have high opinions of each other, then our algorithm stipulates that it will be very likely that players i and k will also forge a mutually productive interaction.

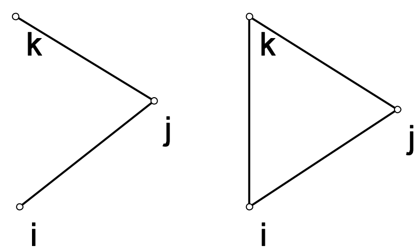

FIGURE 1 Preferential Triad Formation

3 SETUP

Games are played on a network, as represented by an undirected weighted graph, where players are represented by vertices, and connections among them are represented by edges. The graph at time t, $G_t = \{V, E_t\}$, is described by a set of players (vertices) $V = \{1, 2, 3, ..., n\}$ and a set of weighted connections among them (edges), $E_t \subseteq \{(i, j, \hat{\rho}_{ij,t}) \mid i, j \in V \text{ is } \hat{\rho}_{ij,t} \in (0, 1)\}$. The set of players connected to player i is that player's partners and is denoted by $\hat{N}_{i,t}$. The cardinality of $\hat{N}_{i,t}$ is referred to as the "degree" of vertex i, $k_{i,t} = |\hat{N}_{i,t}|$. At any time t, a number of games are being played on the graph. In a game, players play only against their partners.

Players have opinions of their partners. More specifically, all players rate the desirability of their connection to their partners. This is modeled with $\rho_{ij,t}$, which represents player i's opinion of partner j. Note that the opinion of two players is usually asymmetric; i.e., $\rho_{ij,t} \neq \rho_{ji,t}$ (as discussed in Section 3.3).

The weight of a connection $\hat{\rho}_{i,jt}$ is given by the geometric mean of each player's opinion, $\hat{\rho}_{ij,t} = (\rho_{ij,t}\, \rho_{ji,t})^{½}$. Thus, $\hat{\rho}_{ij,t}$ is the mutual desirability of a connection between players i and j. Unlike the simple arithmetic mean, the geometric mean is a plausible way to model mutual desirability for two reasons. First, it penalizes the discrepancy between opinions. Second, it penalizes the discrepancy at an increasing rate. This means that a discrepancy between two

[3] See Wasserman and Faust (1994) for a detailed explanation of these ideas.

players' opinions of each other has a negative impact on mutual desirability. In addition, a higher discrepancy exacerbates this negative impact.

Games on the network are driven by two adaptive rules. One rule determines how players' partners evolve. We refer to this rule as the "partner updating" rule. Another rule determines how players' strategies are defined given their partners. We refer to this rule as the "strategy updating" rule. Players' strategies and partners come to bear on one another in a locally played game. We refer to the game as "local interaction." Each of these is discussed in the following subsections.

3.1 Local Interaction

Players play a "local public good provision" game with their partners. In each period, agents receive an endowment $c = 1$, which they can choose to contribute to the public good in their neighborhood. If a player contributes, the net cost of the contribution is c. This contribution generates a benefit of $b > 1$ that is shared equally by all of the player's partners.[4] Since the net benefit of b strictly outweighs the net cost of 1, efficiency requires everyone to contribute. From a player's point of view, however, not contributing always yields a higher payoff (that is, 1) regardless of the actions taken by its partners. Contributing is strictly dominated by not contributing.

A player's choice of actions at time t is $\tilde{A}_{i,t} \in \{0,1\}$, where $\tilde{A}_{i,t} = 0$ is "do not contribute," and $\tilde{A}_{i,t} = 1$ is "contribute" to the public good based on the player's type (or strategy), $\tau_{i,t} \in (0, 1)$, and realized action of partners $A_{j,t-1}$ in the previous period:

$$\tilde{A}_{i,t} = \begin{cases} 0 & if \quad \dfrac{\sum_{j \in \hat{N}_{i,t-1}} A_{j,t-1}}{k_{i,t-1}} < \tau_{i,t} \ . \\ 1 & otherwise \end{cases}$$

For local interaction, we define a player's realized action $A_{i,t} \in \{0,1\}$, which is the action determined above — subject to a "trembling hand" probability of making a mistake. If a player's strategy leads that player to contribute, there is a probability $\lambda_A \ll (1)$ that it will not contribute, and vice versa.[5] The dynamics of type (strategy) is discussed in Section 3.2.

Thus, the total payoff to player i at time t is:

$$\pi_{i,t} = c(1 - A_{i,t}) + \sum_{j \in \hat{N}_{i,t}} \frac{b}{k_{j,t}} A_{j,t} \ . \tag{1}$$

It is simply the sum of the player's period-to-period endowment c (if it chooses not to contribute) and the benefit received from the partners, should they choose to contribute. The

---

[4] The fact that c is the *net* cost of contributing is merely a simplification of the game for our purposes.

[5] The value of $A_{i,0}$ is randomly determined.

game is a variant of the widely studied multiperson prisoner's dilemma game with local interaction.

3.2 Updating Strategy

Players update their choice of strategy by observing how they and their partners did by contributing or not contributing in the rounds of play.

Each player tracks two variables for strategy updating. These variables, $\overline{\pi}_{i,t}^{C}$ *and* $\overline{\pi}_{i,t}^{NC}$, define the average payoffs from contributors and noncontributors, respectively, for players and all of their *partners.* These variables represent one way to capture how one action is doing compared with the other in terms of payoffs to players. The difference in average payoffs from noncontributors and contributors is defined as $\Delta_{i,t} = \overline{\pi}_{i,t}^{NC} - \overline{\pi}_{i,t}^{C}$. On the basis of this difference, player i's strategy at time t is determined as follows:

$$\tau_{i,t+1} = \frac{1}{1+e^{-\sum_{s=0}^{t} \delta^{t-s} \Delta_{i,s}}},$$

where $\delta \in [0,1]$ is a discounting coefficient, and the initial strategy $\tau_{i,1}$ lies in the interval [0.05, 0.95].[6] If player i observes that, on average, contribution fares better, then the player becomes more likely to contribute by lowering τ, and vice versa. In case either strategy's average payoff is equal, the strategy becomes closer to the neutral value 0.5. This updating rule is in line with the one used by Eshel, et al. (1998).

3.3 Updating Partners

The updating of players is a two-step process. Step 1 involves losing a relationship, which consists of removing a connection between two players. Step 2 is the creation of a relationship, which consists of adding a connection between two players. Each of these steps is carried out for all players with a particular frequency. Section 3.4 describes these updating rates and their importance in the results in more detail.

3.3.1 Opinions and First- and Second-degree Mutual Desirability

For player i, the opinion of partner j is defined as follows:

$$\rho_{ij,t} = \frac{1}{1+e^{-\sum_{s=0}^{t} \partial^{t-s} b\left(\frac{A_{j,s}}{k_{j,s}} - \frac{1}{\overline{k}}\right)}}. \tag{2}$$

[6] Based on the randomly determined $\tau_{i,1}$, the corresponding $\Delta_{i,0}$ is assigned.

The $\rho_{ij,t}$ depends on the discounted sum of the surplus that player i has received from partner j above what an average contributor in the system can provide,

$$\sum_{s=0}^{t} \partial^{t-s} b\left(\frac{A_{j,s}}{k_{j,s}} - \frac{1}{\bar{k}}\right),$$

where $\delta \in [0,1]$. This is to capture the idea that one would like to interact with contributors with less partners. How does one define the maximum number of partners a contributor can have without lowering one's partners' opinions? One way is to let each player use local information in determining this threshold.[7] Another way is to derive the threshold from some global parameter of the model. We take the latter approach[8] and set the threshold to the overall average number of partners each player has in the system, $\bar{k}$.

On the basis of two players' opinions, we define *first-degree mutual desirability* among them, $\hat{\rho}_{ij,t} \in (0,1)$. It characterizes how strong the players feel about each other:

$$\hat{\rho}_{ij,t} = (\rho_{ij,t}\rho_{ji,t})^{1/2} . \tag{3}$$

First-degree mutual desirability is only defined for players who are directly connected to each other and thus are partners.

Next we define *second-degree mutual desirability* between two players i and k who are connected via a set of mutual partners $M_{ik,t} = \hat{N}_{i,t} \cap \hat{N}_{k,t}$. We call a player k for whom $M_{ik,t} \neq \phi$ and $k \notin \hat{N}_{i,t}$ a second-degree partner of i. The set of such players is denoted by $\hat{N}_{i,t}$ (see Figure 2 for an example):

$$\hat{\hat{\rho}}_{ik,t} = \frac{1}{\left|M_{ik,t}\right|} \sum_{m \in M_{ik,t}} \left(\hat{\rho}_{im,t}\hat{\rho}_{mk,t}\right)^{\frac{1}{2}} . \tag{4}$$

Second-degree mutual desirability is defined for players that are separated by one player. It can be thought of as the mutual opinion of two players who have been referred by a common friend.

3.3.2 Losing a Relationship

Losing a relationship is simply based on first-degree mutual desirability. The tie is severed[9] to a partner for whom $\mathrm{Min}[\hat{\rho}_{ij,t} - \tau_{i,t}] < 0$. When the relation is severed, the agent creates a new relationship with another agent in the population. This process is governed by the chain of personal references and is described below.

---

[7] For example, the historical average of the benefit the player has received from its contributing partners.

[8] As the total number of edges stays constant over time, this approach is robust.

[9] If there is more than one such partner for a player, a random tie among those will be severed.

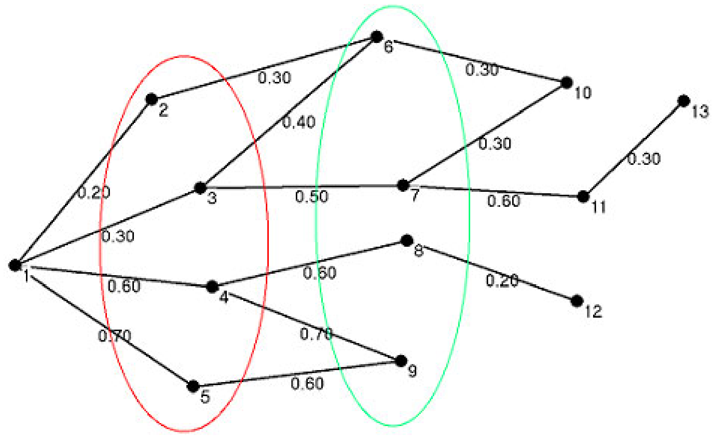

FIGURE 2 Sample Graph Nodes (These nodes represent players, and edges connecting two nodes represent an active relationship between two players. The left oval encloses player 1's partners, $\hat{N}_{1,t} = \{2,3,4,5\}$. The right oval encloses player 1's second-degree partners $\hat{N}_{1,t} = \{6,7,8,9\}$. The weights of the edges are mutual desirabilities, as discussed above.)

The strategy of a player can be interpreted to reflect the player's restlessness and dissatisfaction with his partners. The higher a player's $\tau_{i,t}$, the higher the willingness to change partners. In addition, a high value for $\tau_{i,t}$ can reflect a player's opportunistic and egoistic character, which makes that player undesirable for other players, as seen below.

3.3.3 Creating a New Relationship

The creation of a new relationship is governed by second-degree mutual desirability. Players know how desirable partners of their partners are. Using this local information, a relation between two currently unconnected players i and $k \in \hat{N}_{i,t}$ such that $Max\left[\hat{\hat{\rho}}_{ik,t} - \tau_{i,t}\right] > 0$, is established.[10] If no such players are found among second-degree partners, a new relationship is formed between player i and player h, who is randomly chosen from the rest of population. This mechanism allows the formation of new ties to depend on the existing sets of ties. In previous studies, new ties have been created independent of existing ties.[11] The initial $\rho_{ik,t} = \rho_{ki,t} = \hat{\rho}_{ik,t}$

---

[10] The substraction of τ is to incorporate the tolerance level, which determines the loss of relationship in the creation of a relationship.

[11] For example, Jackson and Watts (2002) consider randomly chosen ties to be severed or created by revealing information of the two involved parties to each other. The exception is Watts (1999) by whom we were inspired. Watts does not, however, consider the simultaneous evolution of strategies and network structure.

will be set to $\hat{\rho}_{ik,t}$ if a new partner is from the second-degree partners; otherwise, $\rho_{ih,t}$, $\rho_{hi,t}$, and $\hat{\rho}_{ih,t}$ are set to 0.5.

When $\tau_{i,t}$ is high, a player *i* is more likely to find a new partner outside of the second-degree neighborhood. Given that $\tau_{i,t}$ represents how opportunistic or egoistic player *i* is, it is more difficult to create a mutually agreeable relationship among second-degree neighbors who have some information about player *i*. As a result, player *i* must resort to other people.

3.4 Updating Rates and Asynchronicity

In each time step of the game, all players are selected in a random order, and their strategies and partners are updated probabilistically. The results of the earlier updates are utilized for players who are updating later. This asynchronous updating is especially relevant in the partner updating process in which, in case the relationship is lost, the partner involved should know about it when it is his turn to update partners. In each time step, the players' partners are updated with probability υ_P, and the players' strategies are updated with probability υ_S. The ability to control the partner and strategy updating speeds of the players — an important aspect of this model — is discussed below.

4 RESULTS

Our setup focused on providing a general definition of a topology: an undirected weighted graph, or what we call a network.[12] All previous topologies on which games have been analyzed are subsumed as special cases of this setup (i.e., rings, lattices, etc.). The motivation behind our broad definition was to allow us to investigate the sensitivity and dependence of cooperation on the structural parameters of the environment in which it occurs. How does cooperation fare in structureless environments (i.e., random graphs)? How does cooperation fare in more structured environments? What are the structural conditions for cooperation? Can they be endogenously created to bring about cooperation? The initial results for a number of these questions are provided in the following subsections.

4.1 Structural Measures

We utilize two simple measures of graphs as indicators of "structure" in our network: the clustering coefficient of graph and the variance of the degree distribution of the players.

The clustering coefficient of a player measures how densely the partners of that player are connected. It is defined as follows. The clustering coefficient of vertex *i*, γ_i, characterizes the extent to which vertices adjacent to vertex *i* are adjacent to each other. It is defined as

$$\gamma_i = \frac{\left|E(\hat{N}_i)\right|}{\binom{k_i}{2}},$$

[12] In this paper, the words, network and graph, are used interchangeably.

where $|E(\hat{N}_i)|$ is the number of edges within the neighborhood of vertex i and $\binom{k_i}{2}$ is the total number of possible edges among them. The neighborhood of vertex i in our setup is partners of player i. This is the probability that two vertices in $\hat{N}_i$ will be connected. The clustering coefficient of the *graph* γ is γ_i averaged over all the vertices in the graph.[13] The clustering coefficient for the graph takes a value between 0 and 1; $\gamma = 0$ implies that no neighbor of any vertex i is adjacent to any other neighbor of i, and $\gamma = 1$ implies that the graph consists of several disconnected but individually complete subgraphs.[14] Also, for a random graph with n vertices and average degree $\bar{k}$, the clustering coefficient is approximately $\bar{k}/n$.[15]

The second measure, the variance of the degree distribution of the players, is derived from the distribution of the number of partners that each player has. Since the total number of edges in the graph is constant (it is a parameter of the setup), the first moment of this distribution is fixed. The second moment, however, gives us a measure of how the number of partners varies from one player to another.

4.2 Structure and Cooperation in a Static Environment

This section investigates the relationship between graph structure and cooperation in a graph in which players' partners are fixed from the outset. Thus, players do not change their partners during play.

Graphs in this environment are generated by using the β model of Watts and Strogatz (1998). The β model generates a graph by starting from a ring with degree k and randomly rewires each edge with probability β. For small values of β, the model generates a graph with a high clustering coefficient and low-degree variance. For larger values of β, the resulting graph approaches a random graph (low clustering coefficient and high-degree variance). For each value of β, we repeat the following process 100 times: generate a graph and let the game be played on the graph starting with 50% of players contributing and a uniformly distributed strategy lying in the interval of 0.05 and 0.95. We measure the clustering coefficient of the graph and variance of degree distribution as well as the steady-state share of contributors for each realization and average them over the 100 realizations.

Figure 3 plots the steady-state share of contributors (the percentage of the number of players contributing) for an increasing clustering coefficient generated by lower values of β. Each line represents a different benefit-cost ratio used *(b* = 6 and 12). We observe that the amount of contribution is sensitive to the clustering coefficient. Observation 1 summarizes this statement.

Observation 1. *There is a positive relationship between the clustering coefficient of a fixed network and the percentage of steady-state contributors.*

---

[13] This definition is taken from Watts (1999), p. 33.

[14] A graph is complete when every vertex in the graph is connected to every other vertex in the graph.

[15] This is the probability of the number of random edges that will be in a neighborhood of size $\bar{k}$.

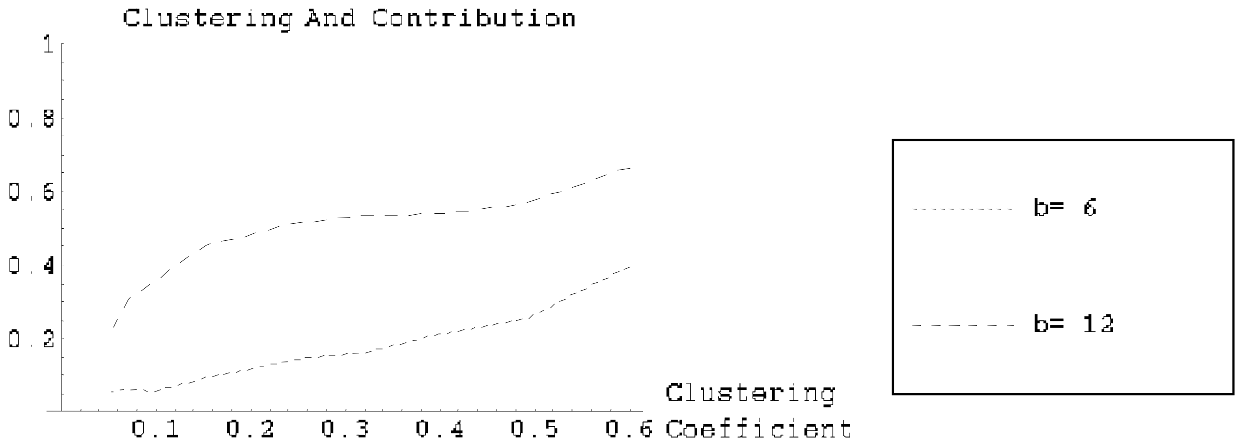

FIGURE 3 Clustering and Contribution — Steady-state Share of Contributors on the *y* Axis and Clustering Coefficient on the *x* Axis

In addition, the different benefit-cost ratios fall into a rough order with b = 6 on the bottom and b = 12 on the top. Contributions from generous players (who have a low value for τ) bring about large benefits for their partners. These partners then evolve into generous players, and clusters of contributors are formed. With larger benefits coming from an increase in b, these clusters are easier to form.

For clustering coefficients below 0.5, it is not possible to sustain cooperation at the high initial level of 50%. Except for the regime of high benefit-cost ratio (b = 12) and high clustering coefficient, the number of contributors in the steady state falls to levels considerably below 50%. The intuition for this follows the reasoning above in reverse. The benefits players receive from their partners are not sufficient to turn them into generous players. As opportunistic players, they generally do not contribute.

From Figure 3, we conclude that structural parameters have an important influence on the steady-state share of contributors in local public goods games. The highly structured topologies used in the literature (associated with high cooperation, in line with our results) should be regarded with caution. They represent a special case among an array of structural possibilities.

4.3 Emergence of Structure I

Another important feature of our setup (beside the generality of the topology) is the ability for players to update their partners by losing ties and creating new ties to players. In this case, the system has the potential to evolve its topology in a directed manner (via preferential triad formation), according to the rules established above.

The results presented in this section and in Section 4.4 are 50 realizations of graphs that are structureless at the outset. That is, we start with random graphs and iterate them for 5,000 periods. In contrast to Section 4.2, the results here (and in Section 4.4) are a time series where we can observe the share of contributors, the strategy of the players, and structural variables changing over time.

In this section, we investigate a scenario where players update partners without updating their strategies. Thus, strategies are fixed from the outset. In Section 4.4, players update both partners and strategies.

Figure 4 shows that under a fixed strategy, partner updating (set at $\upsilon_P = 0.05$) is effective. The share of contributors and the average strategy stay constant. The mean clustering coefficient rises and then drops off to quickly reach a steady state below its original value. The variance of the degree distribution has a monotonic initial rise and then flattens quickly to a steady state.

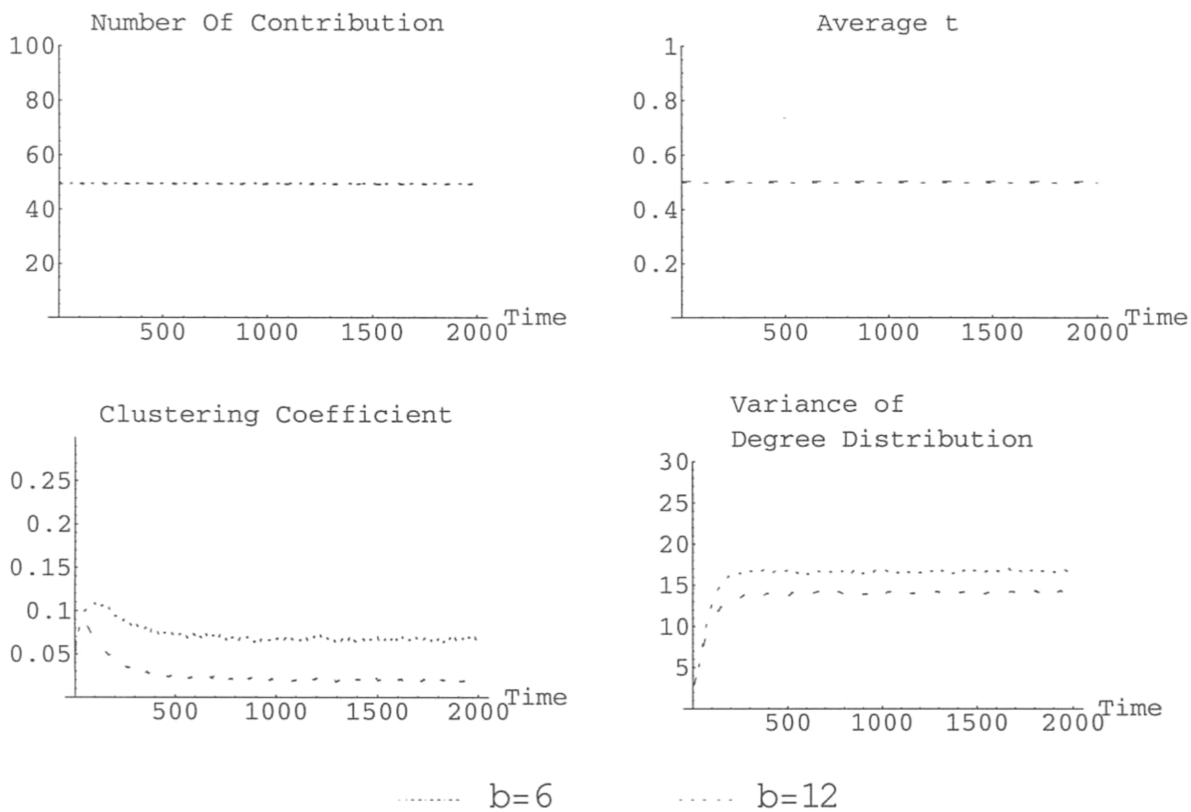

FIGURE 4 Number of Cooperators (players who contribute), Average Strategy, Clustering Coefficient, and Variance of Degree Distribution over Time (averaged over 50 different realizations for each benefit-cost ratio) $N = 100$, $\upsilon_S = 0$, $\upsilon_P = 0.05$, $\delta = 0.8$, and $\lambda_A = 0.01$

The share of contributors in the network is independent of the benefit-cost ratio. The average strategy (τ) is constant by definition. The clustering coefficient and the variance of the degree distributions are clearly functions of the benefit-cost ratio.

The simple dynamics of the structural variables suggests that a kind of "sorting" is taking place. Generous players (with a low τ value) lose ties to opportunistic players (with a high τ value) and surround themselves with other generous players, thereby creating clusters. The

opportunistic players are driven out of these newly formed clusters (see Figure 5). This raises the clustering coefficient. As the average degree of the generous players rises (at the expense of the degree of less generous players), the variance of the degree distribution also rises. The generous players are then inevitably surrounded by too many partners, which leads to a loss of some ties to the generous players. We thus have a drop in the clustering coefficient and a flattening of the variance of the degree distribution. The result is summarized in Observation 2.

Observation 2. *Under fixed player strategies, partner updating forms a cluster of generous players who sustain cooperation.*

Under fixed player strategies, partner updating changes the structural properties of the graph so as to "lock in" the initial levels of contributions. Generous players are clustered with other generous players to initiate contribution. Opportunistic players stop contributing.[16]

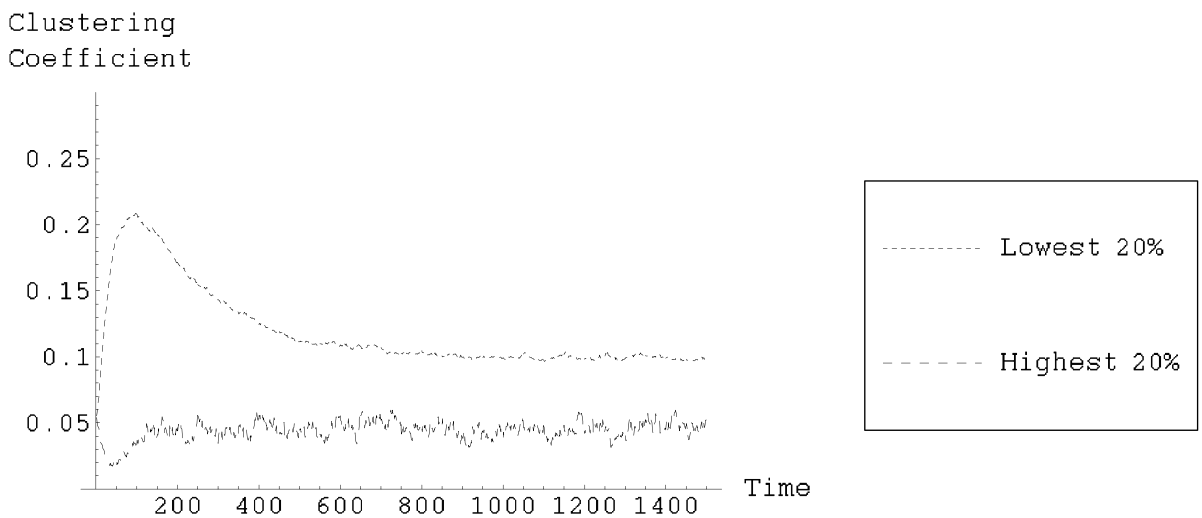

FIGURE 5 Average Clustering Coefficient for Players with the Lowest 20% and the Highest 20% Value of τ in the Population (averaged over 50 different realizations) b = 6, N = 100, $\upsilon_S = 0$, $\upsilon_P = 0.05$, $\delta = 0.8$, and $\lambda_A = 0.01$

4.4 Emergence of Structure II

This section shows the full extent of the features of the model. The graph is random at the outset and endogenously produces structure over time as players update their strategies and their partners. Here, the share of contributors, the average strategy, the clustering coefficient, and the variance of the degree distribution are all functions of the benefit-cost ratio.

[16] The results for higher partner updating rates, $\upsilon_P = 0.1$, $\upsilon_P = 0.3$, and $\upsilon_P = 0.5$, are qualitatively the same as above.

Figure 6 displays the results for strategy updating set to $\upsilon_S = 0.5$ and partner updating set to $\upsilon_P = 0.05$. Here, we see the potential that partner updating has when paired with strategy updating. Under a regime of a high benefit-cost ratio (*b* = 12), the share of contributors can be substantially raised above its initial level. Here, the initial share was 50%. Under lower benefit-cost ratio regimes *(b* = 6), the share of sustained contributors is below the initial value, although it is significantly higher than the share that could have been sustained under a random graph, namely, 0%.

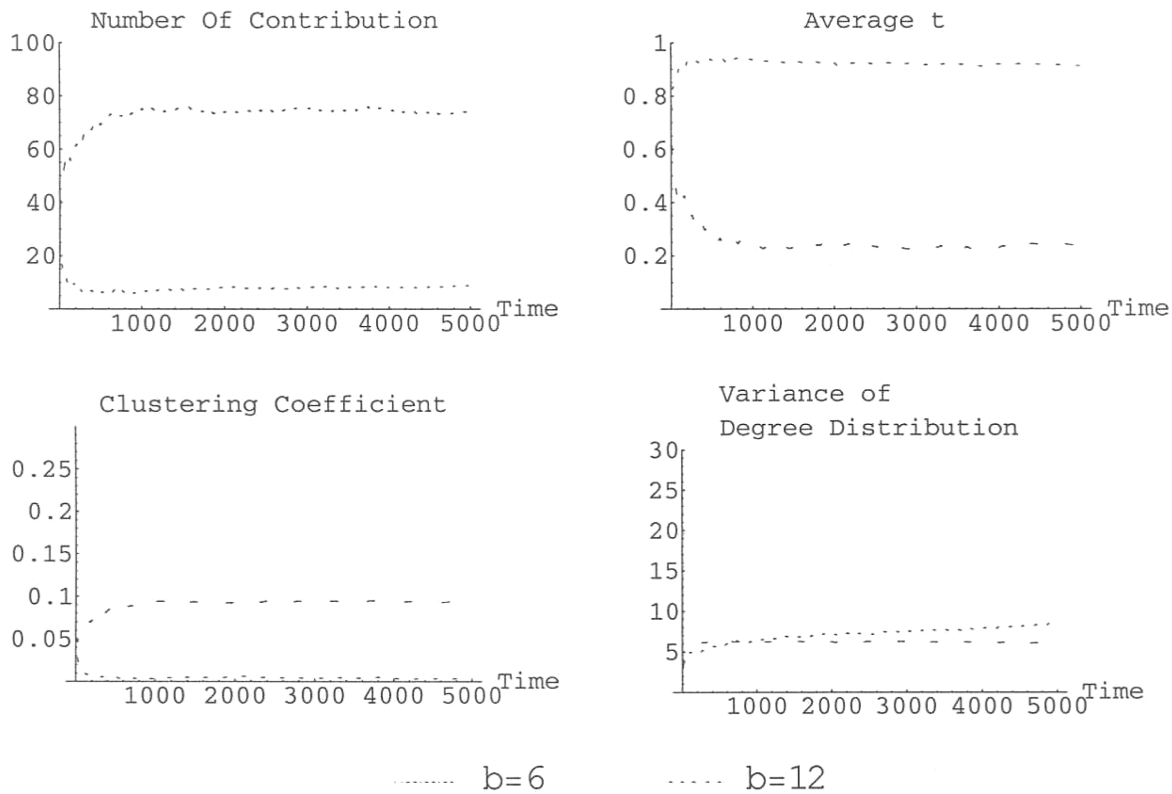

FIGURE 6 Number of Cooperators (players who contribute), Average Strategy, Clustering Coefficient, and Variance of Degree Distribution over Time (averaged over 100 different realizations for each cost-benefit ratio) $N = 100$, $\upsilon_S = 0.5$, $\upsilon_P = 0.05$, $\delta = 0.8$, and $\lambda_A = 0.01$

The share of contributors, the average strategy, and the clustering coefficient rise monotonically and then quickly settle into steady-state values. For the lowest benefit-cost ratio (*b* = 6), however, the variance of the degree distribution does not reach a steady state even after 5,000 periods. A higher rate of updating partners exacerbates the instability of the variance as shown in Figure 7.

We conjecture that the increasing variance of the degree distribution is due to a polarization of the strategy of players. As discussed in Section 4.3, the cooperations are sustained through cooperators forming a cluster. When strategies are updated, players in the

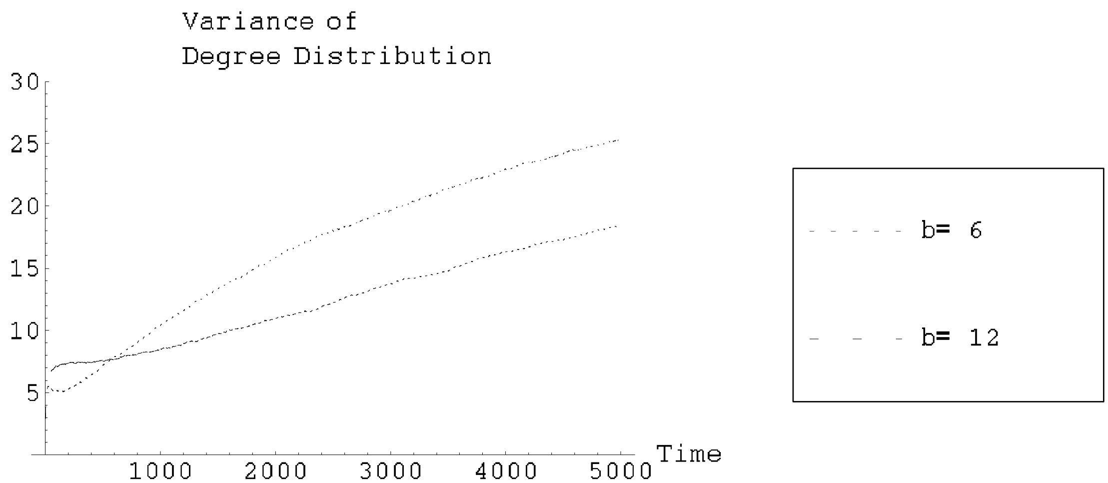

FIGURE 7 Variance of Degree Distribution over Time (averaged over 100 different realizations for each benefit-cost ratio) N = 100, υ_S = 0.5, υ_P = 0.3, δ = 0.8, and λ_A = 0.01

cluster of the cooperators are likely to become more cooperative, and those on the outside are likely to become less cooperative. This response is accelerated by cooperative players attracting even more partners and less cooperative players further losing partners (which increases the variance in the degree distribution). However, a more detailed exploration of the dynamics of the model with strategy and partner updating is required to sharpen this reasoning.

5 CONCLUSION

Two conclusions can be drawn from the simulations. First, this study shows that the interaction topology has an important effect on the outcome of simple games, such as the provision of a local public good. By using a broad definition of an interaction structure, we have shown that adaptive rules are sensitive to the topology in which they operate. As it pertains to the local public good game, we linked contribution directly to structural parameters. An important consequence of this conclusion is a call for a closer investigation of the interaction structures. Though the literature has made some inroads into investigating topological effects, the results turn out to be special cases with limited significance. A study in which the topology itself does not evolve, has, in effect, fixed an important variable controlling the share of contributions. Second, this study shows that considerations of locality and the ability to change locales can provide important sources of coordination. In the local public good game, the ability to change partners leads to significant levels of contribution, starting with topologies where no contribution was expected.

Our results are based on two very simple structural measures. There is much room here to expand these measures and to investigate their implications for the strategy dynamics in simple games, even in the fixed interaction structure. Also, utilizing other structural measures might help us to understand the dynamics in cases where the interaction topology and individual strategies co-evolve.

6 REFERENCES

Eshel, I., L. Samuelson, and A. Shaked, 1998, "Altruists, Egoists, and Hooligans in a Local Interaction Model," *American Economic Review* **88**:157–179.

Granovetter, M., 1985, "Economic Action and Social Structure: The Problem of Embededness," *American Journal of Sociology* **91**:481–510.

Jackson, M.O., and A. Watts, 2002, "The Evolution of Social and Economic Networks," *Journal of Economic Theory* **106**(2):219–241.

Kirman, A., 1997, "The Economy as an Evolving Network," *Journal of Evolutionary Economics* **7**:339–353.

Montgomery, J.D., 1998, "Toward a Role-Theoretic Conception of Embededness," *American Journal of Sociology* **104**:92–125.

Raub, W., and J. Weesie, 1990, "Reputation and Efficiency in Social Interactions: An Example of Network Effects," *American Journal of Sociology* **96**:626–654.

Tesfatsion, L., 2002, "Economic Agents and Markets as Emerging Phenomena," *Proceedings of the National Academy of Sciences* **99**:7191–7192.

Wasserman, S., and K. Faust, 1994, *Social Network Analysis*, Cambridge, United Kingdom: Cambridge University Press.

Watts, D.J., 1999, *Small Worlds: The Dynamics of Networks between Order and* Randomness, Princeton, NJ: Princeton University Press.

Watts, D.J., and S.H. Strogatz, 1998, "Collective Dynamics of 'Small-World' Networks," *Nature* **393**:440–42.

ENDOGENOUS NETWORK FORMATION AND THE EVOLUTION OF PREFERENCES

T.K. AHN, Workshop in Political Theory and Policy Analysis, Indiana University*
A. LAZAR, Computer Science and Information Systems, Youngstown State University
D. CHAVALARIAS, Center for Research on Applied Epistemology, Ecole Polytechnique

ABSTRACT

Analytical and computational models were developed to study the conditions for the stability of a population consisting of agents with heterogeneous preferences. The analytical models that utilize an indirect evolutionary approach show that the ability to detect others' types is critical for the evolution of reciprocal preferences. The computational models incorporate agents' memories and endogenously built social networks into the evolutionary dynamics. The simulations based on the computational models show that the strength of the social network is a critical factor for the success of non-selfish preferences. A fully heterogeneous population consisting of egoists, reciprocators, and altruists can be stable for a range of parameter conditions.

INTRODUCTION

Many social situations that require cooperation among multiple individuals to achieve a common goal, benefit those who free-ride on others' efforts. If there were any biological or social selection mechanisms that favor those who gain by cheating, societies would most likely be inhabited by selfish individuals. In both economics and political science, the modern mode of thinking is to assume that everyone is selfish and to devise rules and institutions that still deliver tolerable social outcomes. However, self-reflection, careful observation of other human beings, and experimental evidence from the social sciences, indicate that our societies are not composed entirely of selfish individuals, but rather of three diverse types: selfish, fair, and altruistic. Where does this heterogeneity come from? How do those with non-selfish motivations survive?

While this question has been widely addressed by evolutionary game theorists (Axelrod and Hamilton, 1981; Axelrod, 1981; Bendor and Swistak, 1997), their models often underestimate the cognitive capability of human agents and the flexibility of human behavior. In this paper, the indirect evolutionary approach (Güth and Yaari, 1992; Güth and Kliemt, 1998; Güth, et al., 2000; Ahn, 2001) is utilized, which combines the features of standard non-cooperative game theory and standard evolutionary game theory. The agents in the indirect evolutionary models are rational in the sense that they have utility functions rather than fixed behavioral rules, and they make choices based on the utility maximization principle. In terms of motivations, agents are heterogeneous; some agents have utility functions that do not map the material payoffs into utilities in a linear manner. In other words, they care about the social consequences of their actions.

* *Corresponding author address*: T.K. Ahn, Workshop in Political Theory and Policy Analysis, 513 North Park Ave, Bloomington, IN 47408; e-mail: tahn@indiana.edu.

In an indirect evolutionary process, selection operates on material payoffs. The types of individuals that are more successful materially increase over time. A mathematical formula of evolution is used that is consistent with both biological and cultural interpretations of the evolutionary process.

A variety of social interactions have the material payoff structure of the Prisoner's Dilemma in which individuals face the temptation to defect, cheat, or free-ride. If all the individuals behave selfishly, however, everyone is worse off than they would be in at least one other outcome in which some of the individuals cooperated. Figure 1 shows a public good provision problem involving two individuals. Each of the two individuals has an initial endowment of p ($0 < p < 0.5$) and makes a binary choice of whether to contribute (cooperation) or not (defection) for the provision of a public good. Contribution costs 1 to the contributor but returns $1-p$ to each of the two individuals. No matter what the other does, an individual is always better off when he or she does not contribute. Therefore, if both individuals are selfish neither will contribute. Then each would receive a material payoff of p, which is smaller than the $1-p$ that each of them would obtain if they both contributed.

		Individual *j*	
		Cooperation	Defection
Individual *i*	Cooperation	$1-p$, $1-p$	0, 1
	Defection	1, 0	p, p

FIGURE 1 Two-person Public Good Provision Problem

MODELING MOTIVATIONAL HETEROGENEITY AMONG RATIONAL AGENTS

Experimental evidence strongly supports the hypothesis that there is a significant proportion of individuals whose preference ordering over the four possible outcomes of the action situation is not linear to the amount of material payoff he or she obtains in each of the four outcomes shown in Figure 1 (Ahn et al., forthcoming; Ahn et al., 2001; Cho and Choi, 2000; Clark and Sefton, 1999; Hayashi et al., 1999). In particular, most of the non-selfish individuals seem to have an assurance preference with the following ordering over the four outcomes: u(C,C) > u(D,C) > u(D,D) > u(C,D). Those who have an assurance type preference are *reciprocators* in the sense that they cooperate if their partners cooperate but defect if their partners defect.

A relatively small proportion of individuals show a preference ordering of u(C,C) > u(D,C) > u(C,D) > u(D,D), which implies unconditional cooperation. These individuals are *altruists*. In most of the experiments, about a half of individuals reveal a self-interested preference ordering of u(D,C) > u(C,C) > u(D,D) > u(C,D). They are *egoists*. Other possible types are empirically and analytically insignificant. Figure 2 is the utility payoff matrix that models the three preference types.

		Individual j Cooperation	Individual j Defection
Individual i	Cooperation	$1-p$	$0+\beta_i$
	Defection	$1-\alpha_i$	p

$$0 \le \beta_i \le \alpha_i \le 1$$

FIGURE 2 Utility Payoff Matrix for Individual

In Figure 2, if α_i is greater than p, individual i prefers to cooperate when j also cooperates. If β_i is larger than p, individual i prefers to cooperate even when j defects. The restriction $\beta_i \le \alpha_i$ implies that no individual has a preference ordering by which he or she prefers to cooperate when the other defects, but prefers to defect when the other cooperates. Substantively, p can be interpreted as the relative magnitude of the material temptation to defect.

One's preference type (egoist, reciprocator, or altruist) is a joint function of one's *generic* type (α_i, β_i) and the material payoff parameter (p). For a given generic type, one is more likely to be an egoist when p is large. A population can be characterized by a probability distribution function F (α_i, β_i). For a given F, the proportion of behavioral reciprocators (δ) and that of altruists (γ) are again functions of p.

INDIRECT EVOLUTION

In an indirect evolutionary process, agents interact in the action situation shown in Figure 1 on the basis of their preferences shown in Figure 2. Evolution selects those who are more successful materially. The question is whether or not any non-selfish types can survive and, if so, which type would. In this section, the indirect evolutionary process is analyzed under four different conditions.[1] In the next section, the simulation model is extended to incorporate repeated interactions, memory, and social networks.

It is assumed that, at each evolutionary stage, each player plays the game only once with another player who is randomly drawn from a population of infinite size. There are four possible ways under which such a game can be played. The key factors are (1) whether the game is played under complete or incomplete information regarding players' types and (2) whether the game is played simultaneously or sequentially. From these two dichotomies result four different evolutionary conditions: simultaneous, complete information (SC); simultaneous, incomplete information (SI); sequential, complete information (QC); and sequential, incomplete information (QI).

The expected material payoff for an egoist (reciprocator, altruist) at time t will be denoted as $\pi_{e,t}$ ($\pi_{r,t}$ $\pi_{a,t}$). At each evolutionary stage, a reasonable solution concept of non-cooperative game theory is used to derive players' behavior.[2] When multiple equilibria exist, it is assumed that a cooperative equilibrium (i.e., one in which at least some players cooperate) is played.

---

1 For complete analyses of all the four conditions, see Ahn (2001).

2 A Nash equilibirum for SC, a Bayesian equilibirum for SI, a subgame perfect equilibirum for QC, and a sequential equilibirum for QI.

To simplify mathematical analysis, for a given behavioral type, the values of α and β are assumed to be the same across players. This facilitates studying the population dynamic of $F^t(\alpha,\beta) \rightarrow F^{t+1}(\alpha,\beta)$ in a simpler dynamic of $(\delta,\gamma)^t \rightarrow (\delta,\gamma)^{t+1}$ in which the proportion of reciprocators (δ) and that of altruists (γ) at time t+1 are calculated by following time-independent replicator functions:

$$\delta_{t+1} = \frac{\delta_t \pi_{r,t}}{(1-\delta_t - \gamma_t)\pi_{e,t} + \delta_t \pi_{r,t} + \gamma_t \pi_{a,t}} \tag{2}$$

$$\gamma_{t+1} = \frac{\gamma_t \pi_{a,t}}{(1-\delta_t - \gamma_t)\pi_{e,t} + \delta_t \pi_{r,t} + \gamma_t \pi_{a,t}}. \tag{3}$$

A type's relative proportion in an evolutionary stage is exactly proportional to its relative proportion in the immediately preceding stage times its relative success measured in terms of the obtained material payoffs. This evolutionary dynamic may occur either genetically or culturally. The entire evolutionary process, regardless of the original population condition $(\delta,\gamma)^0$, can be approximated by a continuous-time dynamic of which the vector derivatives are

$$[\dot{\delta} = \delta_{t+\Delta t} - \delta_t, \quad \dot{\gamma} = \gamma_{t+\Delta t} - \gamma_t]. \tag{4}$$

Figure 3 illustrates the evolutionary dynamics of all of the four possible single-play environments. Only the evolutionary dynamics under the QI condition are discussed in more detail. The sequential, incomplete condition is more common than other conditions in the real world. That is, agents in the real world can hardly be sure of the exact motivational types of others.

Under the QI condition, a player plays the game as a first mover with probability 0.5 and as a second mover with the same probability. Since agents are rational, their behavior is not deterministic. The utility maximizing behavior is a function of the material incentive, p, and the composition of types within a population $(\delta,\gamma)^t$. The lower-right panel of Figure 3 shows three different equilibrium zones under the QI condition as functions of p, δ (Rec), and γ (Alt).

In all the three zones, the behavior of second movers is a direct function of their types: egoists always defect, reciprocators copy the choice of the first mover, and altruists always cooperate. The difference across types is in their behavior as first movers. In Zone I, all three types of first mover cooperate. Since the proportion of reciprocators is relatively large compared with that of altruists and egoists combined, it pays for the first-mover egoists to cooperate. In Zone II, egoist first movers defect, but reciprocator first movers still cooperate. In Zone III, there are too few reciprocators and altruists, thus no equilibrium exists in which reciprocator first movers cooperate. In all three zones, egoists obtain the highest average payoff. Altruists decrease in all three zones. The relative proportion of reciprocators decreases in Zones I and II, but increases in Zone III. Therefore, stable states exist along the horizontal axis with the proportion

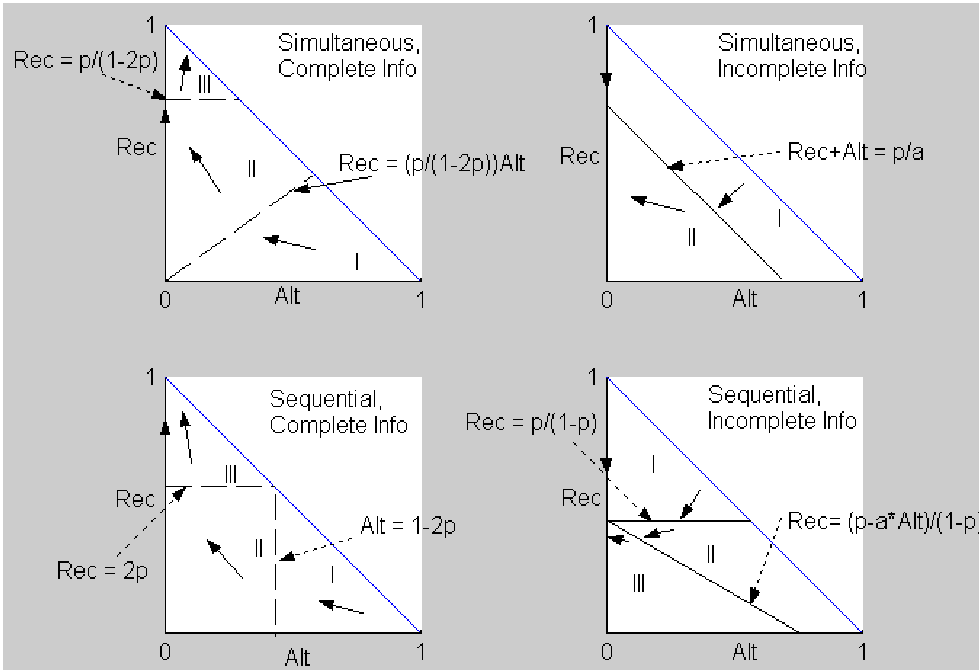

FIGURE 3 Information, Play Sequence, and Evolution of Preferences (Under the SC and QC conditions, stable states exist along the line $\delta + \gamma = 1$. However, $\delta = 1$ is the only attractor. In both incomplete information conditions, $\delta = \gamma = 0$ is the only attractor. Rec = δ; Alt = γ; a = α [see Figure 2].)

of reciprocators being smaller than $p/(1-p)$. However, if there is a constant stream of invasions by altruistic mutants, the stable states are absorbed into the attractor in which only egoists exist.[3]

RATIONAL AGENTS IN SMALL WORLD NETWORKS

The indirect evolutionary models of the previous section use one-shot social dilemma games as the conditions of interaction. This section examines the effects of endogenous network formation on the evolution of preferences using computer simulations. The sequential incomplete information is used as the baseline interaction. Sequential interactions are by far more common than simultaneous interactions. However, when agents are assumed to be completely ignorant of another agent's type in a given interaction, the evolutionary result is one in which egoists prevail. In real social settings, incomplete information is not necessarily the most common information condition. Any increase in an agent's informational capacity favors the evolution of non-selfish preference types. For example, a model in Ahn (2001) assumes that, in an indirect evolutionary

[3] Random mutations at these stable states imply that the relative proportion of reciprocators to egoists, $\delta/(1-\gamma-\delta)$, remains the same after an arbitrarily small invasion by altruists. The disturbance caused by this kind of mutuation reaches back to another stable state that is only slightly removed from the original stable state. However, since the existence of altruists favors egoists, the recovered stable state inhabits a larger proportion of egoists than that of the original stable state. After a long sequence of disturbances and recoveries, the population converges to the attactor in which only egoists remain.

dynamics, agents know the type of another agent with probability q. This mixed-information assumption results in a stable mixed population for a wide range of parameter values.[4]

The incomplete information condition that favors egoists was used as the baseline condition of interaction to highlight the effect of endogenous networks. For the network building to be possible, it is also necessary to allow agents to live more than one evolutionary stage and to have memory of past interactions. A strict single-play game situation would imply either a perfect anonymity or a perfect certainty regarding the future — players are perfectly sure that there would be no more interaction among currently interacting players. One-shot games are a useful approximation to account for specific cases in which the probability of a future encounter between a pair of players is very small. However, there are also many social interactions characterized by ongoing relationships that are built on past interactions.

Specifically, it was assumed that agents have memories of past interactions that partly condition their interactions with others. Thus, when an agent is in the position of playing the game as the first mover, it recalls past interactions and searches for those who behaved in a trustworthy manner in the past and offers to play the game by taking the first move. It does not necessarily mean that agents always cooperate as first movers. An agent may not have any past incidents of cooperative interactions; in that case, the agent has to play the game with another agent chosen randomly. In addition, if the agent is egoistic and he or she knows an altruistic agent from past experience, the egoist first mover will choose to defect to the altruistic second mover.

Second, it was assumed that each agent dies with a probability $1 - \theta$ after each evolutionary stage. When an agent dies, it is replaced by its offspring whose type is probabilistically determined by the distribution of types at the moment of its birth.

The two added assumptions, that agents live for an uncertain length of time and that they have memory, expands the definition of cooperation. Cooperation is not merely an outcome in the $\{0,1\}^2$ strategy space of a single-shot game. Cooperation cannot be dissociated from all the possible future worlds it brings about. For this potentiality to be effective, agents must have memory while the future must be uncertain. The sequential incomplete information game was then extended by incorporating the two additional assumptions.

MEMORY, ENDOGENOUS NETWORKS, AND THE EVOLUTION OF PREFERENCES

Consider a population of N agents that play the basic social dilemma game under the QI condition. An agent plays the game several times with different partners during his or her lifetime. After an agent plays a game — or multiple games — in an evolutionary stage, it dies with a probability of $1 - \theta$, giving each agent a mean life expectancy of $1/(1 - \theta)$ evolutionary stages. When an agent dies, it is immediately replaced by a new agent. The type of the new agent is determined probabilistically in the manner specified in Equations 2 and 3.

---

[4] Güth, et al. (2000) provide elegant analyses of the models in which agents develop informational capabilities.

Each agent has a perfect memory of past interactions. An "address book" is maintained in which the names of other agents who cooperated with it in the past are written. The agents whose names appear in an agent's address book are called the "relationships" of the agent. When an agent's name is in the address book of another agent, the latter is a "friend" of the former. At each evolutionary stage, an agent randomly chooses a name from his or her address book and plays a sequential social dilemma game as a first mover. If an agent's address book is empty, it interacts as a first mover with another agent randomly chosen from the population. At each evolutionary stage, an agent plays only once as a first mover. However, he or she can play multiple times as a second mover depending on the number of "requests" it receives. This reflects the fact that being a first mover of an interaction usually takes much more time than just reacting to another's initiative as a second mover.

This gives rise to the possibility of a lock-in by which a pair of players play the game for the entire duration of their lifetimes, which is not very realistic. Rather, it is assumed that even when his or her address book is not empty, an agent interacts with someone outside his or her address book with a probability $e \in [0,1]$. In real life, when a pair of agents interacts too often and only between themselves, the returns from the interaction decrease. Therefore, to diversify their information and optimize their payoffs, sometimes agents have to go beyond their established relationships. Individuals can also be forced to interact with someone they do not know; a large e in this case reflects instability due to political and economics reasons. The parameter e reflects how often this voluntary or involuntary *exploration* occurs. For now, it is assumed that for a given configuration of other parameters, there is a value of e, which optimizes the expected payoff for an agent.

The model outlined above defines a directed adaptive network endogenously built by the agents. It will evolve in time as agents live and die. Setting $e = 1$ and $\theta = 0$ results in the baseline single-play condition. In other words, the baseline single-play condition is a special case of the more general model outlined here. If the address book of agent i is not empty, it is assumed that i has some relations. Then, as a first mover, i interacts, with the probability $1-e$ with one of the agents in the address book. With a probability e, it interacts with a randomly chosen agent from outside of its relationships. Table 1 summarizes the behavior of the agents. The agents make their decisions at a stage to maximize their expected utility. This is why the decisions by egoists and reciprocators, when they play the game as first movers with someone chosen outside of their address books, are functions of the distribution of types in the current population.

TABLE 1 Behavior in the Presence of Relations and Friends

	As a First Mover		As a Second Mover
	From Address Book	Outside Address Book	
Egoists	Pick an altruist and defect. If there are no altruists, cooperate.	Cooperate iff $\delta > \frac{p}{1-p}$.	Always defect.
Reciprocators	Pick at random and cooperate.	Cooperate iff $\delta > \frac{p-\alpha\gamma}{1-p}$.	Copy first mover's behavior.
Altruists	Pick at random and cooperate.	Always cooperate.	Always cooperate.

SIMULATION RESULTS

This model has been simulated in Java. For this series of simulations, the focus was on examining the influence of the exploration parameter e and the initial distribution of types while keeping $p = 0.3$, $\alpha = 0.4$, $\theta = 0.99$, and N (the number of agents at any given evolutionary stage) = 1,000 for all simulations. Table 2 shows the nine parameter conditions with which the simulations are run and reports the mean proportions of altruists and reciprocators, and their standard errors in parentheses. Each simulation is run for 5,000 evolutionary stages, which corresponds to about 50 generations, given that the life expectancy of an agent is approximately 100 evolutionary stages. The results reported in Table 2 are best viewed by comparing three columns for each of the three rows. Alternatively, one can also compare three rows of a given column to see whether or not, for a given network strength, the population dynamics differ depending on the initial condition.

The three columns of the first row of Table 2 address the question of whether or not a small combined proportion of altruists and reciprocators can invade a population mostly inhabited by egoists. When $e = 0.1$ and, thus, an agent, as a first mover, interacts 9 out of 10 times with someone in its current address book whenever the address book is not empty, the non-selfish preference types successfully invade the egoistic population. In fact, the egoists are completely driven out of the population by the 5,000th stage. Figure 4 illustrates the average evolutionary trajectory of the five simulations with initial conditions of the first row and first column in Table 2. The rectangular space in each panel of Figure 4 is only a relevant subspace of the entire state space shown in the panels of Figure 3. The evolutionary age of the population is marked on the trajectory for each 1,000th stage. Starting from the initial population state, which consists of 10% altruists, 10% reciprocators, and 80% egoists, the population steadily evolves toward northeast, signifying that both the altruists and reciprocators increase. The graph shows both the direction and speed of the evolution. Once not many egoists remain in the population, the evolution is slow; that is why the distance between the 4,000th and the 5,000th stage in the graph is very short. The two smaller panels of Figure 4 show the standard errors of the proportions of reciprocators and altruists at a regular interval.

TABLE 2 Simulation Parameters and Results: Mean Proportions and Their Standard Errors at the 5,000th Evolutionary Stage

		Network Strength		
		Strong Social Network: $e = 0.1$	Modest Social Network: $e = 0.5$	Weak Social Network: $e = 0.9$
Initial Population	Altruists:10 % Reciprocators: 10%	$\gamma = 0.70(0.11)$ $\delta = 0.29(0.12)$	$\gamma = 0.21(0.11)$ $\delta = 0.13(0.07)$	$\gamma = 0.0(0.00)$ $\delta = 0.11(0.06)$
	Altruists: 33.3% Reciprocators: 33.3%	$\gamma = 0.45(0.16)$ $\delta = 0.55(0.16)$	$\gamma = 0.39(0.10)$ $\delta = 0.57(0.09)$	$\gamma = 0.03(0.03)$ $\delta = 0.32(0.11)$
	Altruists: 45% Reciprocators: 45%	$\gamma = 0.49(0.05)$ $\delta = 0.51(0.05)$	$\gamma = 0.60(0.08)$ $\delta = 0.40(0.08)$	$\gamma = 0.13(0.07)$ $\delta = 0.40(0.11)$

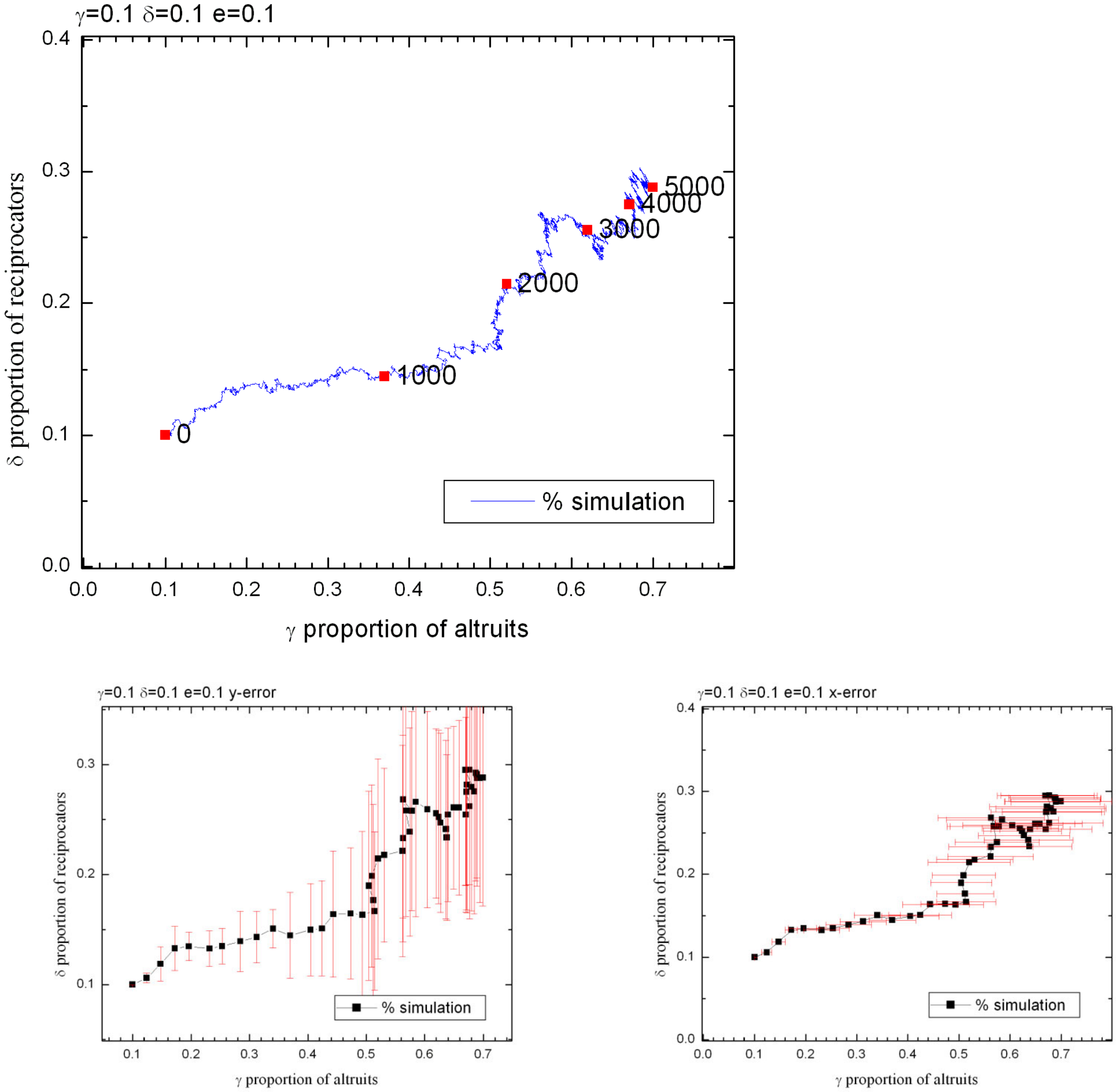

FIGURE 4 Strong Social Networks (Can nonselfish preferences invade an egoistic population?)

The eight panels of Figure 5 show the average evolutionary trajectories of the remaining eight simulation conditions. Each panel has the initial population composition and the network strength at the upper-left corner. When networks are weak, the invasion by altruists and reciprocators into an egoistic population is either slow and incomplete (e = 0.5), or impossible (e = 0.9). With networks of modest strength, both the proportions of altruists and reciprocators increase slightly; the egoists, however, maintain the majority at the 5,000th evolutionary stage. Whether or not the non-selfish preference types eventually take over the entire population cannot be answered within the simulation data. Given the slow and tortuous evolutionary trajectory (see the left panel of the first row in Figure 5), however, it can be conjectured that this is a range of parameter configuration in which a fully mixed population composed of all three types could be stable.

In contrast, when the network is weak (e = 0.9), meaning that agents interact with someone outside of their address book 9 out of 10 times, the invasion is not successful. Altruists are completely driven out of the population and reciprocators survive only because they remain as neutral mutants. This can be attributed to the large proportion of egoists, which makes reciprocators not cooperate as first movers and, thus, behave exactly the same as egoists.

When the three types are evenly distributed at the initial stage of an evolution, the evolutionary trajectories show two patterns depending on the network strength. This can be observed by comparing the three columns of the second row in Table 2, and corresponding panels in Figure 5. When networks are either strong or modest, egoists are driven out of the population. However, when the force of endogenous network formation is weak because of the high probability of exploration, altruists are driven out of the population and the reciprocators remain as neutral mutants.

The final question is whether or not egoists can invade a population consisting of altruists and reciprocators. The third row of Table 2 provides the answer to this question. When networks are strong or modest, egoists' invasion is unsuccessful. In fact, they are completely driven out of the population. However, when networks are weak, egoists can successfully invade the population and drive out altruists and neutralize reciprocators.

CONCLUSION

Indirect evolutionary models that explore the conditions for the evolution of different preference types have been developed. Within one-shot game settings, either reciprocators or egoists are favored evolutionarily depending on the information conditions. In extended models with memory and endogenous social networks, the conditions for altruists and reciprocators to survive, and even invade, an egoistic population were also examined. While many assume that every type of preference except an egoistic one is not evolutionarily viable, other possibilities were explored. The computational models show that the presence of social networks endogenously built by agents who have memories can change the evolutionary dynamics. When agents have the cognitive capacity to classify their environment, social networks play an important role; social cooperation emerges at a substantial scale, and non-selfish preferences can flourish.

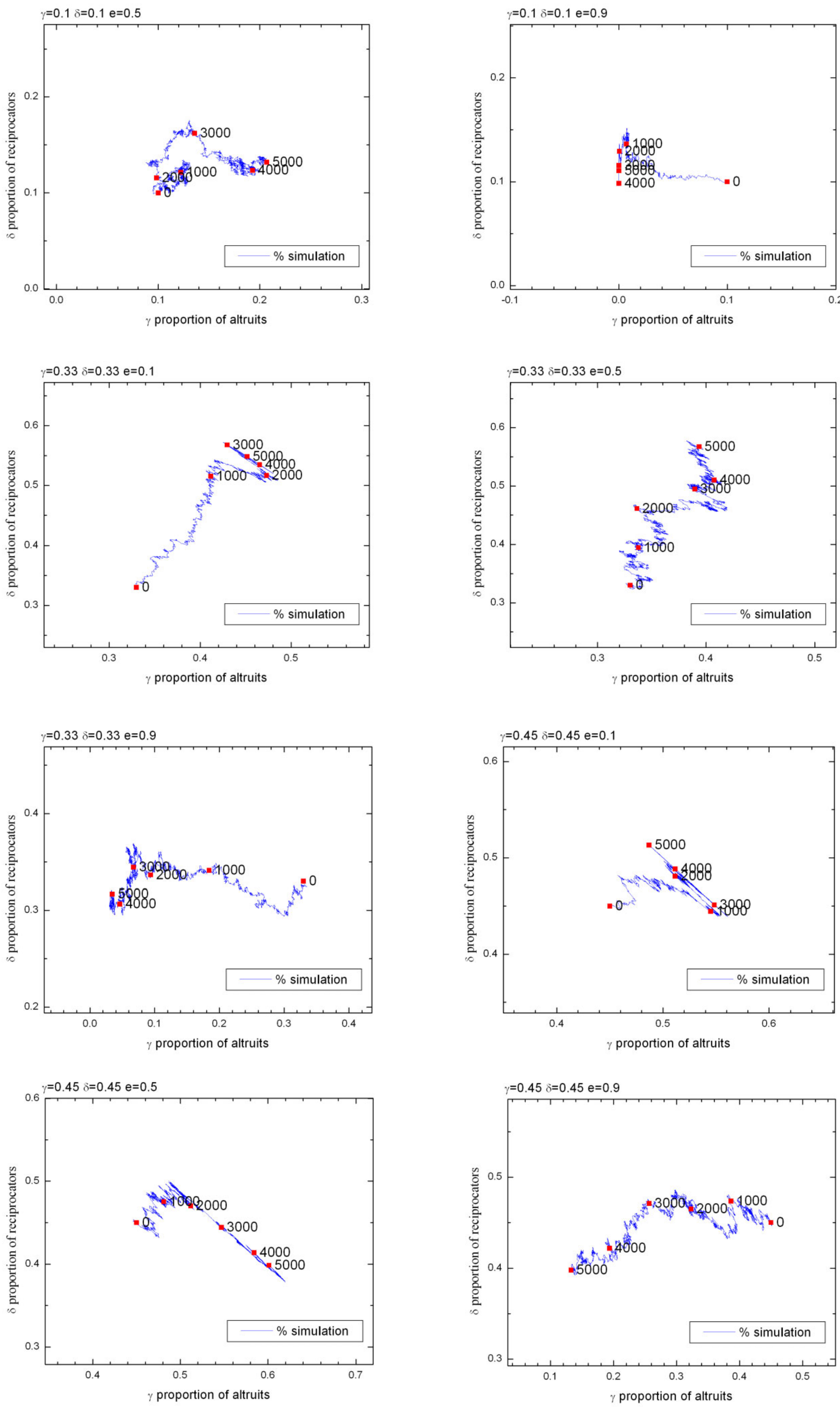

FIGURE 5 Average Evolutionary Trajectories from Five Simulations

REFERENCES

Ahn, T.K., 2001, “Conspiracy of Doves: An Indirect Evolutionary Approach to Institutions and Culture,” Working Paper W01-12, in *Workshop in Political Theory and Policy Analysis*, Bloomington, IN: Indiana University.

Ahn, T.K., E. Ostrom, and J. Walker, forthcoming, “Incorporating Motivational Heterogeneity into Game Theoretic Models of Collective Action,” *Public Choice*.

Ahn, T.K., E. Ostrom, D. Schmidt, R. Shupp, and J. Walker, 2001, “Cooperation in PD Games: Fear, Greed, and History of Play,” *Public Choice* **106**(1/2):137–155.

Axelrod, R., 1981, “The Emergence of Cooperation among Egoists,” *American Political Science Review* **75**:306–318.

Axelrod, R., and W.D. Hamilton, 1981, “The Evolution of Cooperation,” *Science* **211**:1390–1396.

Bendor, J., and P. Swistak, 1997, “The Evolutionary Stability of Cooperation,” *American Political Science Review* **91**:290–307.

Cho, K., and B. Choi, 2000, “A Cross-Society Study of Trust and Reciprocity: Korea, Japan, and the U.S.,” *International Studies Review* **3**(2):31–43.

Clark, K., and M. Sefton, 2001, “The Sequential Prisoner’s Dilemma: Evidence on Reciprocation,” *Economic Journal* **111**:51–68.

Güth, W., and M. Yaari, 1992, “An Evolutionary Approach to Explaining Reciprocal Behaviour in a Simple Strategic Game,” in *Explaining Process and Change*, edited by H. Kliemt, 23–34, Ann Arbor: University of Michigan Press.

Güth, W., and H. Kliemt, 1998, “The Indirect Evolutionary Approach: Bridging the Gap between Rationality and Adaptation,” *Rationality and Society* **10**(3):377–399.

Güth, W., H. Kliemt, and B. Peleg, 2000, “Co-evolution of Preferences and Information in Simple Games of Trust,” *German Economic Review* **1**(1):83–110.

DISCUSSION: ADAPTATION AND NETWORKS

M. HEANEY, The University of Chicago, Moderator

Michael Heaney: I'd like to open this discussion by saying that, after attending the panel, one thing about the presentations deeply disappointed me, and that is that Jonathan Bendor is not here in person to serve as the lightning rod of criticism and argumentation, as he has done so effectively in years past. Perhaps there will be some contention for the new lightning rod of criticism, but please, Professor Diermeier, if you would, tell Professor Bendor that he was certainly missed at today's panel. I'd like to make a few comments, allow time for the panelists to respond, and finally open the discussion to everyone for comments and questions.

First, this is an interesting panel to start with, because in a sense this set of scholars is not a traditional agent-based modeling group. They didn't show us any continuous run-time movies. In a sense, they haven't fully bought into the agent-based modeling framework, and that's part of what makes their work so interesting. We're seeing the beginning of the incorporation of some insights from agent-based modeling into ongoing fields of study. So throughout the conference, we should be asking ourselves, how can some of the partial insights that we gain from our models be incorporated or used in certain ongoing areas of theoretical investigation or empirical study?

Each of these authors' goals is to take something from agent-based modeling, or from related fields, and incorporate it into micro-economic theory, game theory, and political economics. The Diermeier paper looks at variations in the process of rationality — different ways of thinking about adaptation. The Hanaki and Peterhansl paper looks at evolving structures, and the Ahn paper looks at heterogeneity of types and builds in those ideas.

Looking at the Diermeier paper — and this is a meta-theoretical paper — I'd first like to ask a meta-theoretical question. Could you defend the way in which your paper approaches the question of empirical validity? As I understand it, you are seeking the kind of empirical validity that is a matching between the final result or the equilibrium result of the model and in a sense a final result or an equilibrium result in the empirical world. Is this necessarily a fruitful question to ask, given the difficulty that we have in doing this matching? Should we focus instead on whether there's empirical content or empirical validity to the kinds of social mechanisms that are at work or simply to the local interaction structures? It wouldn't necessarily have to just be a question of macro- or micro-level matching empirical validity, but also an issue of mezo-level empirical validity.

My second comment is that this paper takes the idea of assumptions very seriously, and there was one assumption in particular that I would be very interested in seeing you work with, and that is the assumption of social structure. Part of the methodology of this paper, and of other papers by some of these scholars, has been to ask what assumptions do we need in order to get a certain kind of result? With regard to the part of the paper that deals with social structure, in order to eliminate the possibility that, as you say, "anything can happen," my question would be, what kinds of social structure do you need? How much social structure do you need to get a certain kind of result?

For example, perhaps you don't need a completely connected social network. Perhaps all you need is the existence of certain opinion leaders in the network or the existence of certain brokers in the network. So you may have more social structure in your model than you already need in order to get the result. That kind of investigation would seem to be consistent with the spirit of some of the other aspects of the paper.

I have a few different comments, a couple of minor ones, for the Hanaki and Peterhansl paper. First of all, with regard to the partner-changing aspect to the paper, I think that you should make partner changing costly, and that will help to give you some interesting path dependencies in the model and also make the dynamics a bit more realistic. Second of all, your paper stimulated some interesting questions that I think you could investigate with just some slight variations in the model. For example, I'm interested in knowing how the model might deal with gossip; that is, third, not just individuals exchanging information within dyads, but also exchanging information about what may be going on in other dyads in the network.

Your paper also seems to stimulate the question of whether information can become a local public good. That is, if you start to see local regions within the network that are developing rich information — pools of information — and whether that begins to take on public-good characteristics and whether it leads to certain regions of the network to start performing in a superior fashion.

My major comment on this paper, though, is that I'd like to see you explore the notion of "over-embeddedness;" that is, the extent to which social structure can have these positive effects on cooperation and interaction, but also negative effects, which may lead to inferior results. Your paper begins to suggest that. I'm also interested seeing if you could take a little time to discuss the tension between Figure 3 and Figure 5 in your paper. Figure 3 shows a positive, monotonic effect of increasing clustering, of increasing social structure, but Figure 5 shows that there is some limit to the benefits of social structure, because clustering begins to go down over time. So could you think about, or could you incorporate within your model, not only the beneficial effects of social structure, but also the cost of social structure?

As to the Ahn paper, I would be interested in your thoughts on the value of the following distinction: is there a conceptual value in distinguishing between the types of actors who react differently under different contexts as you take various types of actors and you give them a certain context and see how they react differently? Is there a value in making the distinction between that and saying that actors can change their type in different contexts?

For example, is it that I am an egoist, and in contexts A, B, and C, I behave differently, or is it that I think about my world differently as it is framed to me in different ways? Do I think and act like an altruist in my family, but with regard to publication of my research, do I think and act like an egoist? Perhaps with respect to, say, participation in professional associations, I act like a reciprocator. Is it a matter of me changing my type in these different contexts under these different frames, or is it a matter of playing different strategies in different contexts? Is there a value in making a distinction between these two ideas?

Also, in your model, are there changes in the population of the agents? If I understand the model correctly, changes in the composition of the population occur only through death, and I wonder if it would be possible to build a model through which agents change their type within their own lifetimes. Also, what do you think the value or lack of value of that would be? It would be interesting to see you incorporate some second-order effects; that is, that agents start learning locally from what's going on among others in their local network.

If I could have just one effect on the research of the scholars in this panel, I'd like to see each project investigate social structure with a little more nuance. As I understand these models, the authors are conceptualizing social structure as being the kind of thing that is either more or less. So we have either a more structured network or a less structured network. However, there are different kinds of structures, and we might not be interested in more or less. Rather, we might be interested in certain characteristics of the structure and how those characteristics of the structure feed into the behavior of the actors.

For example, it might not just be a question of more or less connection within the network, but also more or less hierarchy, more or less cohesiveness, and also the existence of brokers and where those brokers are in the network. Investigating these types of questions would give the discussion of social structure more nuance, and it would avoid conclusions that are somewhat — or what I find to be somewhat — unhelpful, such as structure; well, this shows that structure matters. Personally, I believe that structure matters. And so how and under what conditions does structure matter? In order to begin to answer that question, we need to have more differentiated ideas about structure.

These papers seem to interact with one another and speak to one another in interesting ways. I'd like to propose that Diermeier and Hanaki and Peterhansl take the Ahn issue seriously and think about — or at least muse upon — how your models would be different if you incorporated Ahn's idea of multiple types of actors. Dr. Ahn, I propose that you take up the Diermeier idea of thinking about variations in adaptive rationality and how that might affect the results of your model. Finally, not for the discussion here, but perhaps for the drive or the flight home, it seems that each of these research programs could write at least one paper that takes these ideas and applies them within the context of prospect theory. How appropriate would that be, given the awarding of the Nobel Prize in economics this week? With that, I'll give each of the authors a chance to comment. Professor Diermeier.

Daniel Diermeier: [inaudible on tape]…of working in some very real social features into the particular games that have been studied very well, but it seemed like there was a lot of room for introducing embeddedness and a number of other features that we've tried to introduce in looking at these games and studying their outcomes. But who knows where we can take it from here? Maybe looking at empirical data eventually would be something very exciting to do, although, given the paper as it is now, it would be very difficult to make that link. I think you're correct.

David Sallach: David Sallach, University of Chicago. Professor Diermeier, regarding this tendency to try to begin to focus some of the models so that they move away from agents that are thoroughly abstract and manifest socially interesting phenomena, I suggest that, in addition to classical and mixed preferences, it would very interesting to see a similar kind of analysis of agents that have sociation effects, which are not only present in various physiological kinds of needs, but also in marginal utility.

Diermeier: Yes, absolutely. The way one could model that in this context would be, in a certain sense, to have balance on the aspiration levels — an upper bound, for example. From that point on, I'm satisfied; I'm not going to update anymore. Actually, it's fairly straightforward then to incorporate it into the theorems. Right now, for every possible payoff, we have an aspiration level, but you can restrict that. You would only need to reformulate the theorems and then basically … as well if the aspiration level is above or below in that wide range, and then you get the result again, or corresponding result, I should say.

Michael North: Michael North from Argonne National Laboratory. This comment is for Ahn's paper. It was very interesting that you said that you're obviously developing a simple agent-based model, but then you have another mathematician who is trying to find an analytic solution for the same problem. This comment may apply to the rest of the panel as well. I think it is important to do that type of docking, either between agent-based models or between, say, a computational technique and an analytical technique. In particular, perhaps each of you would comment on the rapprochement in the sense of how you would take what you have and try to compare it to some other type of technique, particularly if you're analytical or computational, if you take computational for analytical.

T.K. Ahn: If any computation model can, for example, be properly modified without sacrificing the substantive assets such that it can solve analytically, there's no reason why you shouldn't do that, and maybe for every agent-based model, especially in relation to an evolutionary one, it's better advice to check if that's addressed as it is or with proper modification analytically.

I've seen some papers or presentations, which, for me, it's definitely do-able analytically, and it can be solved. Probably, it's not just to show off math muscles, but you have a more general, complete understanding of the situation.

Alexander Peterhansl: I have a quick comment on this subject. Even though ultimately your ambition might be to model something analytically, one of the things that we've learned in putting our model together and writing our paper is that the computational method can serve as an incredibly powerful test bed to test out features, to play and get a feel for what's driving the model. In that way, it serves as an incredible filter for things that you might eventually want to put into a small analytical toy model that you can go around and present. I think ultimately, though, your lessons are learned on the computer, exploring parameter spaces, putting in new features, taking them out again, and so on. In terms of a formal coupling, though, I think it is an exciting area. I don't know how much has been done in this area, but I think it is very promising indeed.

Diermeir: You're running into open doors. I'm very much in favor of that. I think it's important in two respects. We've seen the general trade-off between analytical versus computational approaches. I think that it's pretty clear what that is, just as T.K. and then Alexander have pointed out.

There's something else, I think, which is unfortunately not done nearly enough, and people may not even be quite aware of it, but a couple of questions could be asked. What am I really doing when I'm doing a simulation or when I'm doing agent-based modeling? What properties am I implicitly assuming, or what type of solution concept am I really working with? That motivated our paper mainly because there's a large, very well respected literature in sociology that starts out, writes down the models, and then has 5 or 10 different starting values.

But what you want — and I think increasingly it will be important, particularly for this field — is to be taken seriously by people that work in economic theory or formal modeling and others, in a game theory. … Whenever somebody develops, say, a new solution concept, they show existence, and that's an important part of it. For example, what is the solution concept? Does it exist? What are its uniqueness properties; what are its robustness properties? Then they connect it to some mathematical theory that is appropriate in that case, whether it's a mark … theory or whether it's some kind of computer science approach that shows us that what's going on is not in any sense empty or too dependent on the details of the particular simulation. I think that is not done enough. And so, in a sense, that's what we're trying to do in our paper, at least for one class of models.

Invited Speaker:

Kathleen M. Carley

SIMULATING SOCIETY: THE TENSION BETWEEN TRANSPARENCY AND VERIDICALITY

K.M. CARLEY, Institute for Software Research International,
Carnegie Mellon University*

ABSTRACT

Computational organization science is a new scientific field whose researchers share a common methodological orientation to formal modeling, which, because of the complex and nonlinear nature of organizations, often results in the use of computational and mathematical models. However, the research community remains divided on the relationships that should exist among models, theory, and reality. There is also little agreement on the fundamental bases for judging the importance or quality of models. As the use of simulation grows in the social sciences, so does this debate. Underlying the diverse ways in which data can be linked to models is a fundamental tension between accuracy and simplicity. On the one hand, simplicity is valued, with the justifying argument being that, if models are to be explanatory, they should be reductions of reality. On the other hand, there is a belief that, if models are to be accurate, they should provide a match to the real world in sufficient detail for the problem at hand. This tension is often expressed in terms of arguments over transparency and veridicality. The paper discusses this tension and how it plays out in the computational social and organizational sciences. Findings from behavioral and cognitive psychology are used to explain the basic ways people respond to computational models.

INTRODUCTION

Computational analysis is dramatically reshaping the way we think about society and social processes. Everything from the impact of information technology to the fundamentals of cooperation and altruism is being addressed by using computational models. Computational models, often in the form of virtual worlds, are being used in social, technological, and engineering policy domains to address — via “what if” analysis — how technologies, decisions, and organizational and governmental policies influence the performance, effectiveness, flexibility, adaptiveness, and survivability of complex social and organizational systems. Computational models are being used increasingly in the classroom to demonstrate social processes and the impact of change to undergraduate and graduate students. New programs are rapidly springing up in which computational modeling and analysis play a role. Essentially, the nascent field of computational social and organizational science has been born.

The focus of this field is the study of societies and organizations as computational entities. Organizations and societies are viewed as inherently computational, as they are complex adaptive information processing systems incorporating search engines. As noted by Carley and Gasser (1999), computational organizations are seen as taking two complementary forms — natural and artificial. The natural, or human, organization or society is universally “informatted,”

---

* *Corresponding author address*: Kathleen M. Carley, Institute for Software Research International, Carnegie Mellon University, Pittsburgh, PA 15213; e-mail: kathleen.carley@cmu.edu.

that is, filled with information and devoted to continually acquiring, manipulating, producing, and disseminating information. It is a multiagent system in which information acquisition, dissemination, processing, and searches are carried out by the joint and interlocked activities of people and automated information technologies embedded in a specific organizational design. By contrast, an artificial organization or society is composed of multiple distributed, heterogeneous, socially intelligent adaptive agents. Each of these agents has organizational properties such as collective action; a task assignment; a set of data, skills, or abilities; and constraints on the agents each can interact with, when the interactions can take place, and about what. These agents can mutually influence, constrain, and support each other as they try to manage and manipulate the knowledge, communication, and interaction networks in which they are embedded.

Computational analysis is used to develop a better understanding of the fundamental principles of "sociality" (i.e., organizing, coordinating, adapting, and managing multiple information processing agents, whether they are humans, corporations, WebBots, or robots) and the fundamental dynamic nature of groups, organizations, institutions, and societies. Indeed, computational analysis plays a ubiquitous role in theory building, data collection, data analysis, education, and policy analysis. For example, a combination of model development, simulation, and virtual experiments is used to develop a better understanding of the fundamental principles of organizing multiple information processing agents and the nature of organizations as computational entities. Overall, the aims of research in this area are to (1) build new concepts, theories, and knowledge about organizing and organization, coordination and linkage, and communication and technology; (2) develop tools and procedures for the validation and analysis of computational organizational models; and (3) develop computational organizational tools that can be used as educational and management aids. It is important to note that computational analysis does not simply serve organizational and social theorizing; rather, computational theorizing about human phenomena is actually pushing the research envelope in terms of computational tools and techniques. Research in this area has resulted in a large number of models; an empirically grounded theory of organizational design and adaptation; better management tools; and a more complete understanding of the ways in which social, organizational, and knowledge networks interlink to help produce effective, robust, and adaptive organizational designs. A number of edited volumes (e.g., Carley and Prietula, 1994; Prietula, et al., 1998) and the journal, *Computational and Mathematical Organization*, discuss research in this area.

Computational organization science is a new scientific field having interdisciplinary roots. Despite differences in training, researchers in this area are methodologically committed to formal modeling, which, because of the complex and nonlinear nature of organizations, often results in the use of computational models. The formal models in this field are computational (e.g., simulation, emulation, expert systems, computer-assisted, numerical analysis) and mathematical (e.g., formal logic, matrix algebra, network analysis, discrete and continuous equations), with many researchers using whichever is appropriate to the research question being addressed. However, the community is not in agreement about the relationships among models, theory, and reality; furthermore, the community disagrees about the fundamental bases for judging the value, importance, or quality of computational models. As the use of simulation grows in the social sciences, so does the debate over the evaluation of models.

Formal models are used to develop and test theory. Some members of this community take the strong computational stance that the theory (i.e., the simulation model) should perform the task it seeks to explain. In this case, the models can actually take the place of agents (human, group, or organization) in an experimental setting. Because of the use of computational modeling, computational organization science is an important component of the curriculum in distributed artificial intelligence (Carley and Gasser, 1999). In this case, high veridicality is called for. Other members of the community opine that the model is the theory, while still others opine that many models can comprise a theory.

The relation of computational models to reality is complex. Underlying all of the diverse ways in which data can be linked to models is a fundamental tension — accuracy versus simplicity. This paper discusses this tension and how it plays out in the computational social and organizational sciences. Findings from behavioral and cognitive psychology are used to explain the basic way that people respond to computational models. On the one hand, there is a belief in simplicity. The basic argument is that, if they are to be explanatory, models should be reductions of reality, so apply Occam's razor in finding the simplest explanation. On the other hand, there is a belief in accuracy. The basic argument is that, if they are to be accurate, models should provide a match to the real world in sufficient detail for the problem at hand. Validation tests should be applied and a satisfactory explanation found that enables you to make decisions, set policies, etc., with minimal risk. Immediately, it should be obvious that the problem is a sociopsychological one; that is, "simple" and "satisfactory" are in the eye of the beholder. This tension is often played out in terms of arguments over transparency and veridicality.

Transparency means that it is "obvious" to the viewer how things work. The basic analogy is a glass clock with a transparent face lets you see the mechanism. In other words, transparency implies, "I understand it." Veridicality means that the model mirrors the workings of the real world (i.e., it portrays truth). In other words, veridicality implies, "I observe a match between the model and the real world."

Within the computational social and organizational sciences, models run the gamut from very simple to complex and detailed. For simple models, the authors often argue for the value of transparency, whereas, for more complex models, the authors often argue for the value of veridicality (see Figure 1).

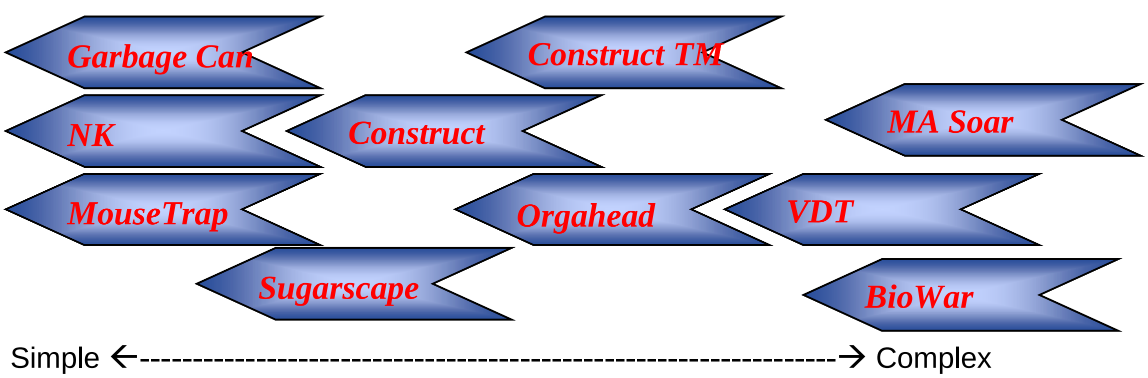

FIGURE 1 Models Run the Gamut from Simple to Complex

WHY THE TENSION IS HIGH IN THE ORGANIZATIONAL SCIENCES

The research in computational organization science spans all aspects of social and organizational science. In each domain, examples of simple and complex models exist. To anchor the discussion, consider two such models. The first is the Garbage Can Model (GCM) of Organizational Choice by Cohen, et al. (1972). This model is classic and very simple. It is readily reprogrammed in a couple of weeks by undergraduates in computational modeling courses. The second is BioWar, a very detailed, complex model that (to date) has taken five people-years to develop.

The purpose of the GCM is to illustrate that choice and energy lead to an organizational situation in which not all decisions are made. This was, in fact, an argument against optimization and rational behavior and for satisfying and boundedly rational behavior. The purpose of BioWar is to enable policy makers to evaluate privacy restrictions and containment policies, facilitate detection, etc., for weaponized biological attacks in cities. BioWar is a city-scale multiagent network model of weaponized biological attacks linked to census data, school district information, etc., and is capable of generating insurance claim reports, absenteeism, etc.

Basic tension transparency and veridicality are not unique to the social and organizational sciences. However, the state of the computational field here, the level of mathematical training, and the relative paucity of computation leads to a different balancing act than in engineering, physics, and chemistry. The basic difference is:

> *In engineered systems, people do not assume they know how things work but trust the "math" of "the physical world." In social systems, people assume they know how things work and do not trust the "math" of "the social world."*

The physical and engineering sciences extensively utilize simulation. In part, their greater acceptance of computation is due to their being older sciences and so have a greater understanding of the phenomena being studied. They are also "wealthier" sciences with greater budgets from foundations, funding agencies, and industry, which means that more work has been done. And, very importantly, they are simpler sciences mathematically. That is, phenomena are studied at a less complex level (fewer interacting parts), and the fundamental entities do not "learn." As a result, less data are needed for validation of a model than in the social and organizational sciences. When this is coupled with the fact that there is relatively less mathematical and computer sciences training in the social and organizational sciences, it should be obvious that there is a problem. Moreover, the fundamental nature of human cognition exacerbates this problem, leading to extended debates and possibly poor choices regarding transparency and veridicality.

The upshot is that in the social and organizational sciences, there is both "physics envy" and a distrust of mathematics. Consequently, social and organizational scientists tend to equate transparency with simplicity. Moreover, many people suggest that if a model is transparent, it has achieved sufficient accuracy with respect to the real world and is a meaningful model. In other words, basic human nature really means that transparency is not "I understand it," but "I think I understand it." Thus, the field is filled with people who assume that they understand a simple model. Transparency is perceived transparency, not necessarily actual transparency.

With veridicality, additional forces come into play. The lack of personnel and finances, as well as the relative youth of the field, means that there is relatively little data on the phenomenon of study. Consequently, trust often replaces proof. Highly veridical models, which are of necessity reasonably complex, are generally regarded with distrust. Essentially, the general distrust of math engenders a lack of trust in computational models. This situation is then exacerbated by the paucity of data, which leads to both divergent expectations, due to an inability to completely map the landscape of possibilities, and minimal levels of validation. This lack of trust often leads to arguments of the form: “This model (the highly veridical one) does not provide insight” or “If you had a good theory, you wouldn’t need this level of complexity.” Both arguments are specious. However, they are made due to a lack of education and a lack of agreement with the model. The latter often is due to people thinking, “Well, my data do not agree with your model.” Thus, what veridicality really means is, “I *believe* there is a match between the model and the real world.”

Rapid advances have been possible because of a unified approach to information processing, explicit attention to the findings of contingency theory, the use of canonical tasks, and the use of social network representation schemes and measures. This unified approach is beginning to pay off in that researchers’ models are now building on each other, and the models can be docked one to the other. Nevertheless, the problem just described is likely to affect many individuals’ careers, given the rapid increase in graduate students interested in computational modeling. To understand the ramifications of this problem, we take a sociocognitive perspective and explore how the basic tenets of human behavior affect the modeling community.

THE PSYCHOLOGY OF PERCEPTION

A number of findings have emerged within cognitive science and behavioral psychology in the last few decades about the nature of the human mind. Let us consider a few of these.

- *People automatically create interpretations of visual images.* Whether we are talking about blobs or networks, when faced with a picture, people are uniformly able to create a story, an interpretation of what they see. Simpler pictures lead to simpler stories. However, there is often little commonality among the stories generated by a given picture. In other words, pictures create meaning, but not necessarily shared meaning.

- *Chunking facilitates learning.* Basically, if you break down a complex story into self-contained segments, it will be easier to learn. Average attention span, age, gender, and countless other factors contribute to the size of the chunk that can be learned at once. Many educators suggest 15 minutes as the temporal size for a chunk. If learning takes longer than 15 minutes, divide it up.

- *People learn many things by experience.* With feedback, the more experience, the better the performance. This is typical Bush-Mosteller learning. The result is essentially an S-shaped learning curve.

- *When we have nothing else to go on, we assume others are like us.* Essentially, people generalize a lot and use analogical reasoning. These traits, coupled with experiential learning, result in people knowing themselves best. If people decide that they are “alike,” then they will assume that they will behave in the same way, know the same things, and share the same values.

- *People are overconfident about decisions even when they have little data.* Basically, people base the likelihood of things on their own experience without taking actual data into account. If there are more red than blue cars in my neighborhood, then I assume that there are more red than blue cars everywhere.

- *People's beliefs are a function of their social information processing (social influence).* The basic idea is that you are more likely to believe something I tell you if we are friends. Similarly, you are more likely to share the same beliefs as your friends. In effecting a change of opinion, therefore, social influence can be as or more important than the facts.

Now let us consider the implications of these findings for the modeling community. First, consider the implications of the fact that people create interpretations of visual images automatically. This means that visual images are being interpreted. Further, it means that each person will have his or her own interpretation. In fact, since you cannot understand something unless it relates to what you already know, interpretations will vary widely when individuals in a group vary widely in experience. So how do we know which interpretations are accurate? When does it even occur to us that our interpretation is not shared?

Accuracy of interpretation, at least in science in the United States, is typically judged by consensus. However, we appeal to the will of the majority only when in doubt. The higher the complexity of the visual image, the less likely it is to be completely processed by the viewer. People tend to become aware of their processing with time. Thus, the higher the complexity of the model and/or the higher the complexity of the visual aid, the more likely it is that people will be aware of not having processed everything. As such, it is more likely that people will think they don't understand a complex model. This means that viewers interpreting both simple and veridical models are more likely to be aware that they are interpreting highly veridical models and less likely of being confident in their interpretations.

Thus, simplicity facilitates visualization. Visualization increases perceived transparency. Thus, simple models are viewed as transparent. People think they understand them and that there is no room for interpretation. There is, therefore, no call for consensus. This, in turn, engenders extensive claims of applicability, as each viewer interprets and so applies the model in his or her own substantive context. In contrast, veridicality leads to either more complex visual images or to simple images containing proportionally less information (than an image of the same complexity for a simple model). When a complex image is used, people are more aware that they do not understand things. As a result, they are less likely to trust the model. When a high-level but simple image is used, people are more likely to think they understand the model and that there is no room for interpretation. However, they may be wrong.

Chunking exacerbates this process. The idea, again, is that people learn in short, contained chunks. Simplicity facilitates short presentations. There just is not that much to say. Consequently, simpler models, which are perceived as transparent, should be easier to learn. Simple models can be "learned" in fewer lessons than highly veridical models. From this, a common inference is likely to be that transparency promotes learning. This would, however, be a somewhat fallacious inference, as it is only perceived transparency and it is the simplicity that is the core cause. For highly veridical models, chunking implies that the model must be modularized before being presented. Since veridical models have more to them, they require more and/or longer presentations. Now, if we add the fact that most people are busy, this means that the chance of being present to learn all of a model is higher for a simple model. In addition, the chance of learning the model is higher for simple models because there would be fewer or smaller chunks. Additionally, given the limited number of contact hours we have with students, educators would be unlikely to teach veridical models in total, but may try to teach multiple simple models. As a result, over time there should be a broader community of scholars who think they understand simple models, have their own interpretation of them, and do not question them. Moreover, there should be a smaller community of scholars who think they understand or have even been exposed to more veridical models. This can lead to a widespread view as to the lack of utility of veridical models, while a smaller group of insiders emerges who fully believe in, and have found validation for, these same models. It also suggests that the more veridical models would be taught at very few institutions; most likely, only at those institutions where their developers teach.

Next, consider the role of social learning. There are three findings that need to be considered at the same time:

1. *Experiential learning*. "I live in the real world; therefore, I have learned how it works."

2. *Others are like me*. "My interpretation of how things work is shared."

3. *Overconfidence*. "I am right about how things work even though I am reasoning from personal experience."

These factors come together to suggest that the accuracy of a model is judged not by "objective" shared data, but by subjective experience.

For simple models, social learning means that simple models are perceived as transparent. Basically, people look at the simple model and go through an exercise like the following:

> I think I understand the model and that my understanding is shared by everyone. I don't expect the model to match the real world. I think my perception of what the model has to say about the real world is also shared by others. I am right, so I do not need to check my facts. Moreover, because there is a common understanding concerning the limitations of the model and what it has to say, we can use this model to set policy, make decisions, and educate students.

Since simple models are easily learned, taught, and communicated, many people will act this way. As a result, such models will be used to set policy without the users ever confirming that they really match the real world or that their interpretation is shared by everyone. It also means that policy setting will be based on storytelling, with the simulations used as a device for creating scenarios from which to reason. This is not meant to imply that this is a bad way to set policy or make managerial decisions. However, it is meant to suggest that human sociocognitive behavior will lead to the use of simple models even when they do not match the real world (are not veridical). Moreover, since, as you increase the simplicity of the model, you often increase the number of interpretations, this use of simple models under the guise of unspoken agreement means that decision makers may be acting on a presumed consensus that, in fact, does not exist.

Veridical models are difficult to learn, teach, and communicate, so there will be little consensus and little social error checking. The implications of social learning for veridical models depend on whether the model is presented at a high level, and so with perceived transparency, or in all its detail. When veridical models are perceived as transparent, people will look at the apparently simple model and go through an exercise like the following:

> I think I understand the model (I'm not sure, as I know stuff is being left out). I think my understanding is shared by everyone. I expect the model to match the real world (after all, the developers claim the model is veridical). I think my perception of what the model has to say about the real world is shared by others. I am probably right, so I don't really need to check with others, but if there is any disconfirming evidence, I will be ready to change my mind.

This line of reasoning means that acceptance of the model will hinge on the interpretation that people make of it. If the interpretation of the model does not match the user's view of the real world, the model would be judged wrong whether or not it matched any actual data on the real world. The reason that data would not outweigh opinion is that the model is sufficiently detailed that there would not be sufficient data to validate all aspects of the model. This means that accurate models — at least, models that are more accurate than opinion — may not be used to set policy. In contrast, if the user's interpretation of the model does match the user's view of the real world, then the model will be viewed as accurate, regardless of the force of evidence. In this case, decision makers may be overconfident about the model's predictions.

When veridical models are not perceived as transparent, the story will change as follows:

> I know I don't understand the model. My lack of understanding is shared by everyone. I expect the model to match the real world (after all, the developers claim the model is veridical). I think my perception of what the model has to say about the real world is shared by others. Again, I am not sure, as the model is complex, so I may or may not be right.

Again, acceptance of the model will hinge on the interpretation that people make of the model. If the user thinks his or her interpretation of the model does not match his or her view of the real world, then the user thinks that the model is "probably wrong," regardless of the evidence, although evidence could be amassed to change the user's opinion. As a result, accurate models may be distrusted. On the other hand, if the interpretation of the model does match the user's view of the real world, then the model would be viewed as "probably right." However, the tentativeness of this conclusion may lead to a lack of confidence in the model's predictions. Another influence may be that veridical models often take many "lessons" to be learned. Imagine that what you first learn of the model is the high-level, simple, and therefore transparent version. Then the social learning process may lead to the problem that as people learn more about a veridical model, their belief and confidence in the model decrease.

Finally, the research on beliefs demonstrates that beliefs are a function of the individual's previous beliefs, facts/information, and the beliefs of others that he or she interacts with. In addition, the impact of new information is a function of who sent the information, whether the information agrees with my current belief, and the weight/frequency of the information.

Social influence leads to simple and veridical models being believed, used, and thought of in very different ways. When a simple model is presented, it is likely to be perceived as transparent, which is not to imply that it is transparent. People then decide if they believe it. Because the model is simple, there are not multiple presentations; thus, there is a low flow of information. Consequently, there are few opportunities to change one's opinion. Due to the factors discussed earlier, people tend to assume that others share their understanding and interpretation. Thus, they do not seek information from others. A consequence is that people not only have beliefs about simple models, they have very strong beliefs. Due to the lack of presentations and the lack of information seeking, people rarely get contradictory information. Since people have strong beliefs, they require a huge amount of contradictory information to change their beliefs. Thus, simple models are likely to win or lose the day purely on the basis of whether they are presented to a sympathetic audience.

In contrast, when a veridical model is presented (if presented in a high-level fashion), it is likely to be perceived as transparent. People then decide if they believe it. Since it is a complex model, people will recognize that they might not completely understand it, so this will be a weak belief. Further, since it is a complex model, there are likely to be multiple presentations, meaning that there is a high flow of information and many opportunities to change one's belief. The result, at least initially, is an increase in people's uncertainty about the model. This leads to a general assumption that others do not understand the model or interpret it in the same way. Consequently, people are likely to seek information from each other, and this likelihood should increase, at least initially. Since beliefs are initially weak, it takes little information to change them. Now, if people have access to more people than information about the model, they will quickly take on the opinion of others. Thus, veridical models are likely to win or lose the day on the basis of the number of people in your social group or your access to information about the model.

THE VALUE OF REAL TRANSPARENCY AND ACTUAL VERIDICALITY

The application of the findings from cognitive science and behaviorial psychology thus suggest that the use of, belief in, and acceptance of models have more to do with social and cognitive processes than with the scientific process and the weight of evidence. This brings to the fore the question of whether or not there is any value to real transparency and actual veridicality.

There are, in fact, a number of benefits of real transparency. If models were truly transparent, they would be easier to teach, learn, and recode. Moreover, it should take less time and space to explain them, as no discussion of interpretations would be needed. However, just because a model is transparent does not guarantee comparability of the original and recoded results, due in part to both compiler issues and the fact that typical results presented for a model are the result of post-processing the model's results, and such post-processing is rarely presented. Finally, transparency does enable theory building.

There are also many benefits of veridicality. Veridicality is valuable in explaining the model to decision makers or policy makers, as you can appeal to the match with the real world. The closer the match, the more the decision makers are able to reason within the model. However, actual veridicality leads to an increase in time for learning the model and an increase in the amount of time and space needed to explain the model. The more veridical the model, the more specific the predictions it generates. As such, veridicality enables both policy analysis and managerial decision making. Further, the more veridical, the "easier" it is to validate, in the sense that fewer assumptions need to be made about which real-world data should be used to match the model. However, it is more difficult to validate, in the sense that more data are needed.

Interestingly, veridicality also creates transparent claims of applicability. Finally, veridicality enables theory construction.

The difference between the use and belief in models (as a function of social and psychological processes) and the true value of veridicality and transparency leads to a great irony. Simple models are perceived as transparent and to require little data to validate. However, they generate only generic predictions with a plethora of interpretations and so are difficult to falsify. However, people do not recognize this morass of interpretations so it is more likely that there will be greater belief in the truth of simple models. Moreover, they are likely to be viewed as having great utility and as improving theory.

In contrast, veridical models are perceived as being difficult to validate. However, they actually generate very specific predictions and are consequently more falsifiable. Yet, though they fit better into the scientific process, they are typically perceived as having less utility by basic researchers and as being further removed from theory and theory construction.

CONCLUSION

Human psychology, coupled with the state of the social sciences, has led to a misplaced trust in simple models. This lack of trust is retarding the development of social and organizational engineering, and could have serious social and political consequences, particularly if such simple models are used to set policy. This lack of trust in more veridical models is not shared by non-social scientists. Consequently, they are more likely to develop complex social and organizational models. Since they are subject to the same “naïve sociologist bias” and since they are unaware of findings, they are likely to generate intuitive but inaccurate models. However, if policy makers and managers suffer the same “physics envy” as social scientists, these complex models built by non-social scientists are likely to be believed simply by virtue of the discipline of the author.

So how do we solve the problem? Basically, we need a shared infrastructure for social and organizational models. We need shared toolkits; shared data sets; and databases linking papers, models, algorithms and data. We need, in addition, increased mathematical and computational training — not just in statistics — in the social sciences. We need courses and textbooks on validations and analysis. Increased training on how to read and present models and model results is also called for. We need tools for visualizing highly veridical models. Moreover, we need more on-line journals with links to models.

Model simplicity and complexity form an axis of tension. This tension plays out in complex ways since simple models are often perceived to be transparent, whereas complex models are often argued to be more veridical. Science needs both transparency and veridicality. However, fundamental social and cognitive processes lead to model development and use being based more on perceived transparency and believed veridicality rather than on the actual transparency and veridicality of the model. Fundamental advances need to be made before the community can climb out of this quagmire.

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation (NSF) IGERT in CASOS, the Office of Naval Research, United States Navy Grant No. N00014-97-1-0037, Army Research Laboratory, NSF ITR/IM IIS-0081219, NSF KDI IIS-9980109, and DARPA. Additional support was provided by the NSF IGERT for research and training in CASOS and by the Center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University (http://www.casos.ece.cmu.edu). The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Department of Defense, the NSF, or the U.S. government.

REFERENCES

Carley, K.M., and M. Prietula, 1994, "ACTS Theory: Extending the Model of Bounded Rationality," in K.M. Carley and M. Prietula (eds.), *Computational Organization Theory*, Hillsdale, NJ: Lawrence Earlbaum Associates.

Carley, K.M., and L. Gasser, 1999, "Computational Organization Theory," in G. Weiss (ed.), *Distributed Artificial Intelligence*, Cambridge, MA: MIT Press.

Cohen, M.D., J.G. March, and J.P. Olsen, 1972, "A Garbage Can Model of Organizational Choice," *Administrative Science Quarterly* **17**(1):1–25.

Prietula, M.J., K.M. Carley, and L. Gasser (eds.), 1998, *Simulating Organizations: Computational Models of Institutions and Groups*, Menlo Park, CA: AAAI Press/The MIT Press.

DISCUSSION:
SIMULATING SOCIETY

N. CONTRACTOR, University of Illinois Urbana-Champaign, Moderator

Noshir Contractor: It's my pleasure to introduce Kathleen Carley and a talk that she's giving titled "Simulating Society: The Tension between Transparency and ..." — I'm going to say this slowly — "Veridicality." One of Kathleen's challenges today is to say that word at her normal frenetic pace many times as she goes through the slides.

I discovered — of course, it's sort of cliché — that Kathleen needs no introduction. That being said, I've also discovered, having spoken with many people who know Kathleen with her many hats, that various people have different impressions of what Kathleen's strengths are and don't always appreciate the diverse portfolio that she brings to the academic and scholarly community.

Many of you may not know that Kathleen has a new position. She's still at Carnegie-Mellon but now is in the Department of Computer Science at ISRI, which is the Institute for Software Research International, and also has been associated with CASOS, which stands for Computational Analysis of Social and Organizational Systems. As we all know, Kathleen could sit in computer science just as comfortably as she has done for many years in social and decision sciences or sociology and so many other areas.

In addition to being a substantive leader in the area of computational organization theory and, more generally, in computational, social, and organizational systems, she has focused her own work on many computational models, some of which she'll be talking about today, including Construct and Orgahead and Lemans and Netwatcher, and so on. Obviously, she has also been in the field of organization science and has been somewhat of the pioneer for computational organization theory. In addition to her own substantive areas, she's taken on all kinds of administrative and leadership roles within this area, editing *CMAT*, which has been or will be shortly renamed to be …

Kathleen Carley: The *Computational Social, and Organization Science Journal*.

Contractor: That's right. This morning David [Sallach] mentioned that Kathleen and others have been instrumental in creating this international association with the horrible-sounding acronym, ICSAS, or something like that. Eventually, it seems, all of us are going to be proud to be members of this association, even if we aren't comfortable with the acronym.

In addition, Kathleen has pioneered a lot of graduate education in this area. She has been involved with the National Science Foundation IGOT initiative, which trains people in this area. Many of the next generation of scholars in computational modeling are being trained at CMU through efforts of Kathleen and her colleagues. She has also been involved with a lot of research; some of it would almost be considered service, in the best sense of the term, with NSF, Office of Labor Research, National Security Agency, and other federal as well as private companies.

I've had the opportunity to get to know Kathleen very well, especially over the last three years, because we've been co-PI's on a National Science Foundation grant that was funded through the knowledge and distributed intelligence, titled "Co-Evolution of Knowledge Networks in 21st Century Organizational Forms: Computational Modeling and Empirical Testing." Much of what we've been doing involves issues of interest to all of us.

When I previewed her slides, I noticed that Kathleen's presentation is a very different Kathleen presentation than many of us have seen. It is much less technical, and it has fewer numbers. It still has lots of pretty pictures, but also lots of philosophical ideas. Kathleen may disagree, but as I read this, I can see that a lot of the issues that Kathleen is inviting us to discuss and debate are issues that have come out of the collaborative effort that we have on this project with Stanford, Illinois, Carnegie Mellon, and USC. We have different computational models that we are trying to explain to one another, to connect with one another, to relate to one another, and the word "transparency" has come up many times in that context. I can say with some degree of humility, and perhaps shame, that even after working for the first two years with Construct and Orgahead — two of Kathleen's models — in our own project, it was only in the last six months that we finally said, "Oh, I think we now understand exactly what Construct does and what Orgahead does." Maybe we are mere mortals, but I suspect that may be true for many individuals, not just in terms of Kathleen's models, but also in terms of one another's computational models. Please join me in welcoming Kathleen to this conference.

Carley: Thank you. As Nosh said, this talk is very different for me, and the reason it's very different is because I'm not going to show you a lot of simulation results. I want to talk about a very fundamental issue that pervades all of our research, and I think that's important for us as a field, and that is the relationship between transparency and veridicality.

[Presentation by Carley]

Contractor: Thank you, Kathleen. We'll see how far we can go in our discussion, given the time.

Unidentified Speaker: … You mentioned that evidence for a match is an important issue. In mathematics, there's a simulation science. You can think of a homomorphism as a sort of mapping between the model's components and some real system. Homomorphisms could be considered in terms of input/output behavior, in terms of the internal structure or processes. There are different ways in which you can create evidence for a match, and some ways might be more useful or easier to do than others. Have you had experience along those lines, or could you approach it in an incremental way rather than trying to get…?

Carley: For simple models, the most common type of match that people think of is face validity, that is, basically showing that the inputs feel right relative to the real world. For more complex models, especially where you're trying to set manager or policy decisions, you have to try to map outputs as well as inputs. The hardest thing to map is map processes, and that's where there's considerable debate between the agent-based community and the system dynamics community. In the agent-based community, we value such things as the ethnographic study because that lets us match processes, whereas in system dynamics that would be less important.

Nigel Gilbert: Nigel Gilbert from University of Surrey. Kathleen, I'm not going to be able to respond to this properly, because there's a huge amount of material that needs careful thought behind what you're saying. There's one area I wasn't clear about because you used the word "prediction" a number of times, in the last but one slide, for example. What kind of prediction did you have in mind? I doubt if you meant prediction in the sense that I could, if I had a sufficiently — I'm not sure I can say this word — a very ...

Carley: Veridical. The model is really true.

Gilbert: …veridical model, could I predict the next president? Obviously, I could not. What sense of prediction are you using when you talk about prediction?

Carley: With some of these models, there really is that sense of prediction — trying to predict the future. For example, in the SimVision model, they try to get information on the entire organization, basically like a design team. Then they do things like predict how long it will take the organization to design that product, predict where there will be cost overruns, and so on. So it is prediction in the sense of forward forecasting.

In the sense of bio-war, you're never going to be able to tell; you're probably never going to be able to test predictions in that sense. You're trying to see if your model can emulate events in the past, such as Merlov or influenza attacks or other contagious plagues and so on, and you also are trying to see if, in a forward sense, it can predict and match onto influenza attacks in the future, which of course is not "weaponized." There are both of those senses in the highly veridical models.

Gilbert: Isn't a word of caution called for as well? You've described the ambitions of some of these models, but are they realistic ambitions? I mean realistic not in the practical sense, but in a theoretical sense. Is it reasonable, given the nature of society — of social systems — that one could expect to build models that do that kind of prediction?

Carley: I usually think about it in this way. If we're trying to predict known events, like cost overruns, then it's very reasonable to be able to do that. If, on the other hand, we're trying to explain when something new will be created or an unintended consequence of something will happen, or whether we can determine the type of the next terrorist attack — the type, not when it will occur — those would involve creativity. I don't know if we'll ever be able to do that. I know a lot of very smart people are working very hard on it, such as projects at Argonne and other places. I don't know if we'd ever be able to do prediction of this type. I have doubts about it, but who knows?

Other kinds of prediction, where it's about *known* things, can probably be done, but only over a limited window of opportunity, because social change will require …. Well, it's like using these models to say, I can predict two or three years out, but I'm not going to predict 100 years out.

Gilbert: Just a thought. I don't want to dominate this discussion, but to come back to the issue, isn't this, or at least one's belief in the possibility of this prediction or this kind of prediction, actually another dimension of the dichotomy that you had between the simple and the complex models? It seems that the supporters of simple models are advocating them as models for understanding, *not* for prediction.

Carley: You're right.

Keith Sawyer: Keith Sawyer from Washington University. I'm also an advocate for more complex models, particularly with the ancient communication languages. I think that models are much too simple for actual human communication, but most of the models that exist seem to fall into your simple category. My impression of your talk is that it's a fairly damning indictment of the state of the field because you conclude that simple models are useful only for teaching purposes, but not for much else. Is that a fair interpretation?

Carley: I'm saying that human psychology — social psychology — certainly supports that complex models are extremely useful for teaching. I would say more that my conclusions, not that they're bad or anything, but if we're going to build these veridical models, we need to change the way we …. Well, we have a lot more work to do. It's more on that end.

Sawyer: So even though I'm an advocate of more complex models, do you think simple models could bootstrap us or lead in a linear fashion to gradually more complex models? Or should we just jump ahead and start with the more complex ones?

Carley: I think that the simple models are fine. I'd like to see them taught in grade schools and high schools, but in college I'd like to see only more complex models taught. That's my own personal thought, but that's a preference.

Unidentified Speaker: As I listen to you talk about the psychology of social, cognitive aspects, one of the first things that occurs to me is that instead of looking at simple versus complex models, we should be looking at an entirely new paradigm of how we teach modeling or how we're going to use models. As we look at these models, we do not believe that models can predict things right now, but what we would like to do is use those models for critical thinking.

In other words, we'd like to understand the very processes that drive some of the different nation-state or terrorist acts that are going on in this world to help us spot-check our faulty assumptions. We'd like to take something that we know is very detailed, abstract it out, and see where it is, and maybe that's something. I liked the fact that you included the different socio-computing in your presentation. Perhaps we need to take another look at something like that.

Carley: Part of me completely agrees with you. I don't think we have a good paradigm for the way we teach modeling. In fact, in our program, we've been exploring how to do it better. Part of it is that we need to borrow more of the techniques from the way a simulation is taught in engineering, but part of it is that we need better tools, like improved simulation tools, in the classroom than we have right now.

Claudio Cioffi-Revilla: Claudio Cioffi from George Mason University. You said something about the simple models and visualization, but I missed any reference you made to veridicality and realism in connection with visualization. I think it's a very important area, and I'd like you to say something about it, because one of the powerful contributions of effective visualization is to render understandable what are otherwise highly complex structures and outcomes and processes and so on. I think that this is an area that has been neglected, perhaps because the tools weren't around until recently or for any number of reasons.

It seems that we need to implement not necessarily more windows to monitor a complex process, but windows that have more efficient renditions of what's actually going on in complicated, complex processes. After all, we have only two eyes, and we're not likely to get any more, even with glasses. So this is an area where graphics that are specialized, technically specific to the complexity of social systems and processes, should be encouraged and developed as they have been in other domains. Could you say something about this?

Carley: First, I totally agree, and I would say that within the United States, the SEIS division of the National Science Foundation is actually funding, and is looking to fund more, proposals dealing with the visualization of complex social systems. The other thing is that the new work that's been done in data mining in the past few years has led to — not only linked to — new tools for visualizing the results of data mining. Some of those tools will be very useful within the social sciences.

David Sallach: David Sallach from the University of Chicago. You didn't say a lot about the level of abstraction of the model…[inaudible on tape] … data, complex data that can then be statistically analyzed and so on. I wonder if it isn't possible that theoretical progress in the social sciences will involve identifying the kinds of abstract models that generate the appropriate kinds of complexity.

Carley: Probably one of the ways some of that came out was in fluid flow and in — I'm trying to think of the exact words — these tanks for ships that were developed a long time ago. Basically, they started off modeling things at a very complex level, and then they did hundreds of thousands of simulations. Then they did a response surface analysis to get a more abstract version of what that model looked like. Once the model met a certain criterion for similarity, from then on they only used the response service approach. In the social sciences, we've been "messed up" because we've almost gone the other way. We said that our data will give us the response service; now let's build a complicated model to match it. That's misleading because you're only riding the very narrow region of the data you've got. I think that we need to go more to processing. I think we need to spend a lot more time with ethnographers.

Michael North: Mike North from Argonne National Laboratory. You mentioned at the start of your talk, and it's something you had talked about …[inaudible on tape]… and that's true. I'd like to add one thing to clarify things a bit. One of the things that we're very interested in looking at is combinations of things. I'm not saying that any one of those combinations will strictly happen, because that, as Nigel and others have mentioned, would be amplifying your point. So it's not directly predictive in the sense that we're not saying that that combination's going to occur; rather, we're looking at expanding people's range of things that they consider, especially combinations that could be particularly bad or particularly good, as well as being aware of those things.

Unidentified Speaker: Going along with that, and also your point of prediction, would you include that type of thing as a form of, perhaps a weak form of, prediction? That's not saying that any of these things *will* happen, but just to look at all, using the computer's ability to enumerate combinations?

Carley: Absolutely.

North: I would include that as one of the important areas of prediction. It's not traditional prediction in the sense of physics, such as here's your exact number. But it's based on the rules you've told us, or even on the complex rules you've told us. These are the things — the range of things — that could happen.

Carley: Yes, and moreover, you're putting a likelihood on the different areas.

North: Yes, exactly.

Daniel Diermeier: Daniel Diermeier at Kellogg Managerial, Economic and Decision Sciences. I want to connect to one specific aspect, which has to do with what I'm going to call "heuristics and biases" that come from the cognitive and social psychology dimension. I wasn't quite sure what the domain was for this. You were talking about editors, but most of the time you were talking about policymakers and managers.

I can see that this is a very important potential problem if you think about, for example, policy makers. Let's say that I teach at a business school. If you talk to managers, they understand supply and demand curves, but they don't understand complicated auction mechanisms, for example, and how this interacts. I'm less clear that these same heuristics and biases would occur in the context of the scientific community. If I think about price theory, for example, it's not really taught from the picture; it's people understanding the details of that. It doesn't seem that it is perceived transparency, but it's an understanding of what the limits are.

It seems that what you're saying is very relevant in terms of policy makers, but it's not clear to me the extent that it has to do with the adoption of these types of models within the scientific community.

Carley: I disagree. Actually, I've seen some of these same debates in different working groups around the country and also in the way students talk at various institutes. These same arguments come up whether or not they adopt or use new models and what kind of models they build. When they teach courses, these models affect what and how they teach, so I see it as important throughout social sciences. I think some of my examples were very tuned toward policymakers because the ramifications for us as a field are severe, but I think that this is important at all levels.

Contractor: I have a closing comment to end this discussion. We are in the Windy City, and those from out of town may not know that it's not called the Windy City because of the wind. It's because of the wind-backed politicians that have been in charge of this city for several decades, hence the name Windy City.

I bring that up today because Kathleen's talk echoes an earlier discussion that was triggered by a comment by Carl Wyke, who's a very famous organizational scholar. He argued that all theories, like Kathleen's computational models, fall into what is called the "GAS model," hence, the connection with windy. In the GAS model, if you think of a clock, you can put G at 12:00, A at 4:00, and S at 8:00, but you can never be at all of these places at the same time. G stands for general, A stands for accurate, and S stands for simple.

If you think of any theory, if you take any theory, and locate it somewhere within this circle, it can't be at the same time general, accurate, *and* simple. Much of the frustration that I see coming out of Kathleen's talk, as well as this discussion, actually centers on trying to have it all and trying to be at GAS at the same time, rather than just G, A, or some combination thereof.

Ecological Simulation

TOWARD A FAIR DISTRIBUTION OF LOSSES: SIMULATION OF A FLOOD SCENARIO

L. BROUWERS, Stockholm University and KTH, Sweden
H. VERHAGEN, Stockholm University and KTH, Sweden*

ABSTRACT

Natural disasters are increasing, possibly due to climate changes and changes in land use. Furthermore, as a result of the concentration of assets and population in high-risk areas, economic losses are escalating. This research focuses on how to handle these economic losses in a fair way at the individual property-owner level. This case study involves the Tisza River, the second largest river in Hungary. It flows through one of the poorest agricultural regions in Europe, and large areas have been repeatedly struck by floods. The Hungarian government is incurring huge costs for implementing flood mitigation measures and for economically compensating the victims. Because it is impossible to predict the time, location, and magnitude of a flood, a simulation model was used to evaluate new policies. The simulation model consists of the various stakeholders (i.e., individual property owners, insurance companies, central government). The behavior of the river and the financial consequences are simulated on a year-by-year basis. We extended the model by using the Consumat approach to model the property owners. The results were then compared with respect to wealth distribution in the case of Consumat agents and simple agents. In the Consumat case, the system is more dynamic and seems more realistic. Further investigation of these effects is planned with the objective of obtaining real-world data to verify the outcomes.

1 INTRODUCTION

There are strong indications that humans are gradually, but definitely, changing the earth's climate. Emissions from fossil fuels and greenhouse gases are altering the atmosphere, leading to an uncertain future of global warming (Jepma and Munasinghe, 1998). A possible correlation between climate change and the frequency and severity of natural disasters can be seen. As the number of catastrophes increases, the financial losses also escalate. From 1988 to 1997, major natural catastrophes cost the world's economies US$700 billion (Munich Reinsurance Company, 1998). These increasing costs cannot be explained solely by the higher frequency of catastrophes. Rather, the increased concentration of population and vulnerable assets in high-risk zones is said to be the main reason for the rise in the costs associated with economic damages (Loster, 1999). A key problem for policy makers is to identify ways to improve resilience and effectively protect society against the increasing risk. Questions of accountability and liability for preventing and absorbing the financial losses are on the political agenda in most countries.

In this paper, we focus on the distribution of wealth to determine whether floods that only affect part of the Palad-Csecsei Basin in Hungary have disproportional effects on the income and wealth of just a few agents. For this formulation, we use the Gini coefficient (Gini, 1912). As

* *Corresponding author address*: H. Verhagen, Forum 100, SE-16440 Kista, Sweden; phone: +46 8 16 1694; Fax: +46 8 703 9025; e-mail: verhagen@dsv.su.se; http://www.dsv.su.se/~verhagen/welcome.html.

part of our continuous development efforts, we implemented an agent-based model based on the Consumat approach (Janssen and Jager, 1999; Jager, 2000). Sections 2 and 3 describe the case and the Consumat approach in greater detail, respectively. Section 4 describes the simulation model. Section 5 presents the simulation results, and Section 6 discusses the conclusions and future research.

2 THE UPPER TISZA CASE

In Hungary, as much as 20% of its 93,000 square meters of territory is at risk for flooding. The Upper Tisza region is one of the largest natural riverside systems in Central Europe. Both international and Hungarian studies have indicated that floodwaters are becoming higher and more frequent, probably as a result of global warming and land-use changes.

The Tisza is the second largest river in Hungary, and the Upper Tisza River stretches to the county of Szabolcs-Szatmár-Bereg. There is no extensive lake system in the Carpathian Mountains, resulting in a large contrast between the maximum and minimum levels of water. The lack of lakes is the reason for the three annual floods in the Tisza. The first flood occurs in early spring, the second in early summer, and the third in the autumn. Except for minor or moderate annual floods, extreme floods occur every 10 to 12 years. During recent years, however, large floods seem to have become more frequent, occurring in 1970 (levee breaches), 1993, 1995, 1998, 1999, 2000, and 2001 (dike burst).

Within the framework of an international research project,[1] this case study was performed to identify flood management strategies that were acceptable to stakeholders. The stakeholders involved in the project consisted of water management bureaus, insurance companies, municipalities (represented through the mayors), catastrophe management organizations, and environmentalist nongovernmental organizations. To test different flood management policies, a small basin was modeled. During the final stakeholder workshop, which took place in September 2002, the stakeholders used the computer model as a tool for discussing and evaluating different policy alternatives.

The basin of study is located in a poor area where the residents depend on agriculture; however, the income from agriculture is not sufficient to support the local population. Shifting part of the economical responsibility from the government to the individual property owners is a challenging task because most people are too poor to afford insurance. A flood can actually be very rewarding for those with insurance, however, because of the current practice of double-compensating the victims; that is, some property owners receive compensation from both the government and the insurance company.

In the flood model used in the Tisza project, the property owner agents were not modeled as decision-making agents. It was assumed that all property owners who could afford insurance would buy it. The extended model presented here is a first step in the direction of making the model more realistic.

---

[1] International Institute for Applied Systems Analysis (IIASA), Austria; Stockholm University/KTH, Department of Computer and Systems Sciences, Sweden; and the Hungarian Academy of Sciences, Hungary.

3 THE CONSUMAT APPROACH

The Consumat approach, developed by Wander Jager and Marco Janssen (Janssen and Jager, 1999; Jager, 2000), is a model of human behavior with a focus on consumer behavior. It combines in an elegant way many of the leading psychological theories, such as theories about human needs, motivational processes, social comparison theory, social learning theory, theory of reasoned action, etc. These theories explain parts of human behavior but lack the generality to take all circumstances into account, thus rendering them less useful for an overall view. To rectify this problem, Janssen and Jager set out to develop a meta-theory, which in turn became the Consumat model.

The driving forces at the macro and the micro level determine the environmental setting for the Consumat behavior. The micro level is formed by the individual Consumats, who have needs that may be more or less satisfied; have opportunities to consume; and have various abilities to consume the opportunities. Furthermore, Consumats have a certain degree of uncertainty. Depending on the combinations, ‘satisfied/not satisfied’ and ‘certain/uncertain,’ the Consumats are engaged in four different cognitive processes: repetition, deliberation, imitation, and social comparison. When a Consumat is both certain and satisfied, there is, of course, no reason to change its behavior; thus, repetition is the strategy chosen. An uncertain but satisfied Consumat has a reason to change its behavior. In this case, the cognitive process chosen is imitation of its neighbors. A certain but unsatisfied Consumat, on the other hand, will deliberate. The final strategy is to consult the social network, the strategy chosen by uncertain and unsatisfied Consumats.

4 SIMULATION STUDIES

The simulation experiments performed on the flood simulation model described above were used to investigate the effects of various flood risk management strategies. The flood model has been used in a study about flood mitigation and loss sharing in northeastern Hungary in the Upper Tisza region (see Brouwers [2002] for a detailed description). Most of the data used in these agent-based social simulations are real data from the Palad-Csecsei Basin; in some cases, real data were not available (e.g., a geographically explicit income distribution) in which case we used fictive but realistic data.

The flood model simulates flood failures in the Palad-Csecsei Basin. A flood failure occurs when a levee breaks, floodwater overtops the levee, or water finds its way under the levee. The reason for restricting the simulations to flood failures is that insurance companies compensate damages caused by failures, but not damages caused by groundwater-related floods.

Nine different flood failure scenarios are implemented in the model. The number of scenarios is based on the assumption that the flood can be of three different magnitudes and that a failure can occur at three different locations. Financial damages are estimated for all flooded private properties for the nine failure scenarios. Even with a hydrological model, it is impossible to model when and where a levee failure will occur. This uncertainty is made explicit in the stochastic variables, Magnitude and Failure. For each simulation year, the stochastic variables are assigned new random values. Magnitude tells if there will be a 100-year flood, a 150-year flood, a 1,000-year flood, or no flood at all. The probabilities for these events are 1/100, 1/150, 1/1,000, and 1-(1/100 + 1/150 + 1/1,000). The second variable, Failure, tells if the flood will cause a levee failure at one of the three locations.

For each simulated year, the financial consequences for the property owner agents are computed. For a flood failure in the simulated year, the Catastrophe module calculates the depths of the floodwaters and the land areas (cells in the grid) inundated. The Palad-Csecsei Basin is geographically represented in the form of a grid, in which every cell represents an area of 10 square meters. There are $1{,}551 \times 1{,}551$ cells in the grid. Only private properties are considered in these experiments, so all other cells are filtered out. If a flood failure occurred in the simulated year, the Catastrophe module is consulted. The financial damages are calculated for each inundated cell. The losses for an individual property owner depend on the prevailing loss-sharing policies. In some countries, the government compensates the victims to 100%, whereas other countries are more restrictive. In addition, the property owner can buy flood insurance. The wealth of all property owner agents is updated in the agent module every year after consulting the policy module to find the current loss-sharing strategies.

4.1 Description of Agent Decision-making Model

As described above, two different types of agents are available for comparison. The first has a simple decision-making model, which means that if an agent has enough financial means to buy insurance, it does. The second is based on the Consumat approach. Thus, agents have four alternatives:

1. Agent is satisfied and certain: Repetition.

2. Agent is satisfied but uncertain: Imitate neighbors (if more than two neighbors are insured, the agent also buys insurance).

3. Agent is not satisfied but is certain on flood risk: Deliberate (change strategy if the agent can afford to buy insurance).

4. Agent is not satisfied and is uncertain on flood risk: Imitates Social Network (goes with the majority in its network).

Agent satisfaction is coupled with the agent's financial means. The agent is satisfied if its wealth is larger than its satisfaction threshold and if its wealth is larger than the previous year. The uncertainty of the agent is coupled with its risk profile and the number of years since the last flood failure. Section 4.2 specifies all functions.

4.2 Simulation Setups

Simulation setups fall into three types of assumptions: general, social, and simulation. These assumptions are discussed in the following subsections.

4.2.1 General Assumptions

- Income = random distribution with a mean of $36{,}900 \times 12$; that is, $12 \times$ average monthly income (which is 36,900 Hungarian Forints, statistics from 1998) using a normal distribution.

- Flood frequency = 4. Because statistical records do not reflect last decades' increased flooding, the return period for floods has been decreased. A flood frequency of 4 means that the probability that a 100-year flood will occur is $1/100 \times 4$.

- Premium size for insurance = 0.02% of the property value. The size of the insurance premium does not reflect the underlying flood risk; it is based on the property value alone. This corresponds with existing premium pricing in Hungary.

- Penetration rate = 0.6. The fraction of property owners who carry flood insurance (bundled with property insurance). The average penetration rate for property insurance in Hungary is 60%.

- Expenses = 0.9. The figure 0.9 is an estimate; however, the area simulated is a very poor area. Thus, 90% of an agent's yearly income is spent on direct expenses.

- Content threshold = 10.000 HUF. This figure corresponds roughly to one-third of a monthly income. An agent who has less money to spend (for an entire year) is not content.

- Flood compensation from the government = 0.5. This figure shows the trend to reduce compensation from the government. Flood compensation used to be much higher (90 to 100% of damages).

- Flood compensation from insurer = 0.8. Property owners with insurance contracts are compensated for a fraction of the damages. The size of the fraction is determined by using different coverage or deductibles. For simplicity, we assume that the companies deduct 20% of the damages and only compensate to 80%.

4.2.2 Social Assumptions

- Minimum number of contacts in social network = 2.

- Maximum number of contacts in social network = 50.

- Number of social nodes = 10.

- Probability that a property owner knows a social node = 0.9.

- Number of neighbors = 5.

4.2.3 Simulation Assumptions

- Time period simulated = 30 years.
- Number of property owner agents = 2,580.
- Series of simulations = 2.
- One series of 9 × 30 years with Consumat model for decision making on insurance.
- One series of 5 × 30 years with simple model for decision making on insurance.
- Wealth transformation function for property agents (an agent cannot have a negative wealth in these experiments).
- Flood Failure
    - No flood failure this year: Wealth year n = max (0, Wealth year n – 1 + Income × (1– expenses) – Insurance premiums)
    - Flood failure occurred this year: Wealth year n = max (0, Wealth year n – 1 + Income × (1– expenses) – Insurance premiums – Flood Damages + Gov Compensation + Insurance Compensation).
- Risk function.
- Risk for flooding = RiskValue – log2 (number of years since the last flood). If the risk is higher than 0, a flood is expected. The risk values are randomly distributed between 0 and 5. A risk value of 0 means that the agent will never expect a flood because the risk function is always below 0. A risk value of 5 means the agent will always expect a flood even if it has not occurred within the last 30 years (which is the maximum number of years in the simulation).

4.3 The Gini Coefficient

The Gini coefficient is the most frequently used measure for inequality. The Gini coefficient was used in this study to analyze the results of the different simulation settings with respect to the distribution of wealth within the agent population. Since we do not have the corresponding data for the real population, we are only interested in trends.

5 SIMULATION RESULTS

The simulations were run only a couple of times to obtain an indication of the possible results. The base model, where agents buy insurance when they can afford it, produced a rather static society. Fewer and fewer agents bought insurance because most of the uncertain agents were those who suffered from floods, whereas their neighbors did not and most of time did not buy insurance. The results for the five runs of this model are depicted in Figure 1.

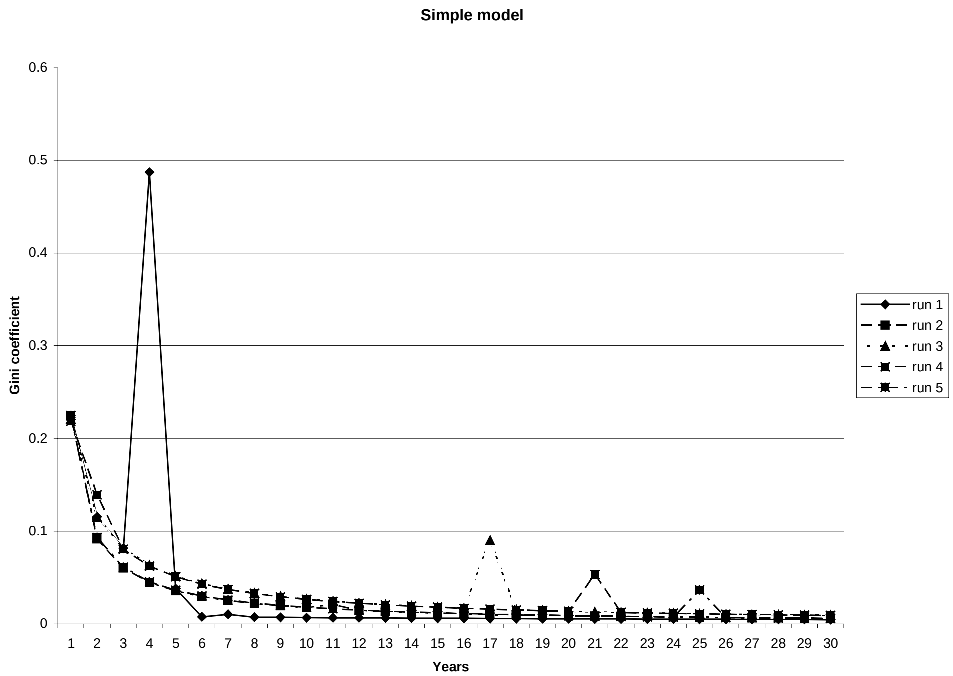

FIGURE 1 Gini Coefficient for the Base Case Simulations

The Consumat-based simulations, on the other hand, show a more dynamic, or even a chaotic, society. Most floods resulted in changes in insurance-buying behavior and in a skewered wealth distribution. The results are depicted in Figure 2.

6 DISCUSSION AND FUTURE RESEARCH

The extension of the model has been successful, since the results are more in line with the real world. Even if the Gini coefficient values are not in the range usually found in an entire society, in our case most inhabitants are very poor and have about the same amount of money to spend. We plan to further investigate these by (1) approaching insurance companies to try to access their statistics and (2) interviewing a representative selection of the inhabitants in the Palad-Csecsei Basin to investigate their social network and decision-making procedures.

7 ACKNOWLEDGMENTS

The authors wish to thank Wander Jager and Marco Janssen for inspiring discussions and clarification of the Consumat approach. One author (L. Brouwers) would like to thank Karin Hansson and Love Ekenberg for cooperating with the Tisza project.

This work was in part sponsored by VINNOVA through the project Software agent technology for Ambient Intelligent Systems, part of the “Nätverksbaserad programvaruteknik” program.

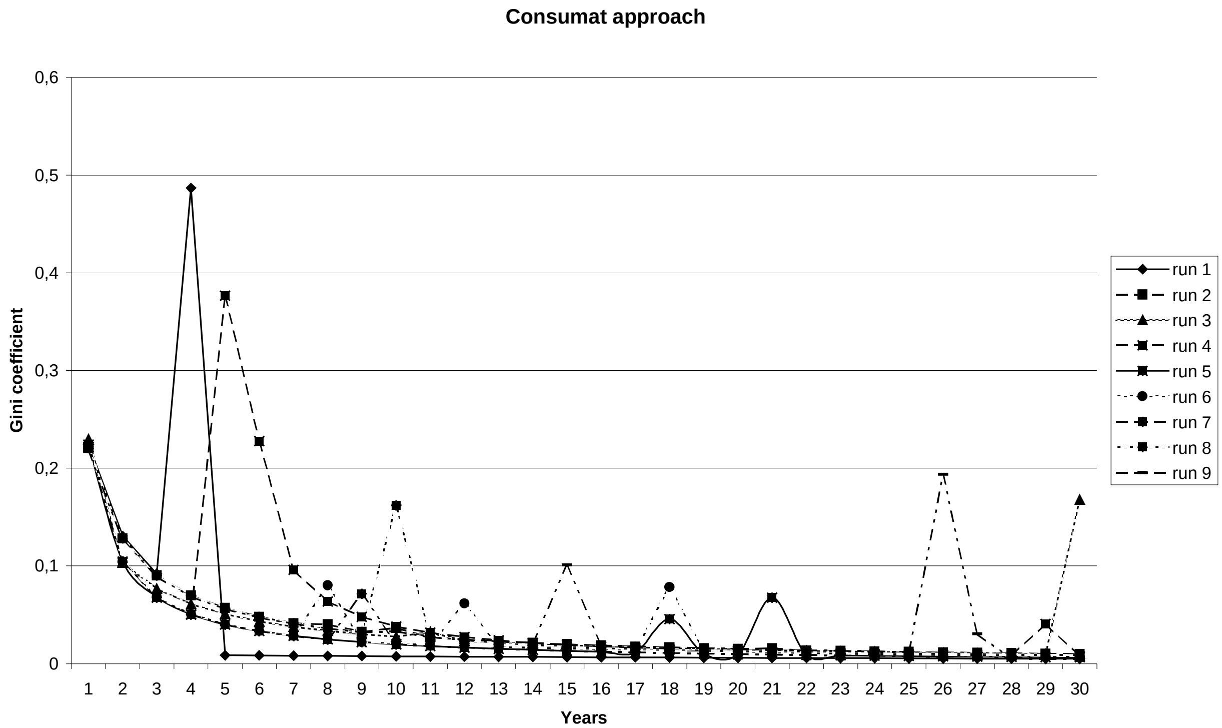

FIGURE 2 Gini Coefficient for the Consumat-based Simulations

8 REFERENCES

Brouwers, L., 2002, *Flood Risk Management Policy from an Insurance Perspective*, Ph.L. thesis, Department of System and Computer Sciences, The Royal Institute of Technology and Stockholm University, Sweden.

Brouwers, L., 2002, *Spatial and Dynamic Modelling of Flood Management Strategies in the Upper Tisza*, Young Science Summer Program Report, International Institute of Applied Systems Analysis, Austria.

Gini, C., 1912, “Variabilitá e mutabilita,” reprinted in *Memorie di metodologica statistica*, E. Pizetti and T. Salvemini, eds., Rome: Libreria Eredi Virgilio Veschi.

Jager, W., 2000, *Modelling Consumer Behaviour*, Ph.D. thesis, Veenendaal, The Netherlands: Universal Press.

Janssen, M.A., and W. Jager, 1999, “An Integrated Approach to Simulating Behavioural Processes: A Case Study of the Lock-in of Consumption Patterns,” *Journal of Artificial Societies and Social Simulation* **2**(2). Available at http://www.soc.surrey.ac.uk/JASSS/2/2/2.html.

Jepma, C.J., and M. Munasinghe, 1998, *Climate Change Poli*cy, Cambridge, United Kingdom: Cambridge University Press.

Loster, T., 1999, “Flood Trends and Global Change,” *Proceedings of the IIASA Conference on Global Change and Catastrophe Management: Flood Risks in Europe*, International Institute of Applied Systems Analysis, Austria.

Munich Reinsurance Company, 1998, *Annual Review of Natural Catastrophes 1997*, Technical Report, Munich, Germany.

EMPIRICAL FOUNDATIONS FOR AGENT-BASED MODELING: HOW DO INSTITUTIONS AFFECT AGENTS' LAND-USE DECISION PROCESSES IN INDIANA?

L. CARLSON, M. JANSSEN, T. MYINT, E. OSTROM, and A. YORK*
Center for the Study of Institutions, Populations, and Environmental Change
Indiana University, Bloomington, IN

ABSTRACT

The use of agent-based modeling (ABM) has recently been extended to the study of natural resource management and land-use and land-cover change. Many ABM applications have been at a conceptual and abstract level, which helps scholars to recognize how macro patterns can emerge from simple rules followed by agents at a micro level. ABM has a greater potential than many other approaches to capture the dynamic relationships between social and ecological systems. This paper contributes to a larger effort to explore how individual decision making by a heterogeneous set of landowners, given local biophysical conditions, led to the particular aggregate pattern of land-cover change in Indiana, with an emphasis on forest-cover change. In our preliminary effort, we created a model structure that allowed examination of the institutional impact of government programs on individual land-use decisions. Our model is based on the concept that an initial condition endows an agent with a particular set of beliefs and desires that could lead to any number of intentions, actions, and outcomes. Institutions have the potential to intervene in an agent's decision-making process and alter its beliefs and desires by providing information and incentives. The next crucial step in our effort will be to extend this model to study the impact of other political institutions, such as taxation and zoning, as well as utilize the conceptual model to facilitate implementation of institutions in the agent-based model.

BACKGROUND

The use of agent-based modeling (ABM) has recently been extended to the study of natural resource management and land-use and land-cover change (Parker, et al., 2003; Janssen, 2003). Many ABM applications have been at a conceptual and abstract level, which helps scholars to recognize how macro patterns can emerge from simple rules followed by agents at a micro level. ABM has a greater potential than many other approaches to capture the dynamic relationships between social and ecological systems. This tool should be useful in helping to develop a theory that relates how institutions affect land-cover change because of ABM's power to model individual agent decision making over time. A crucial next step in our effort will be using ABM to understand the linkage between social and biophysical systems at multiple levels, thereby establishing a methodology that links empirical findings to model construction.

This paper contributes to a larger effort of the Biocomplexity Project of the Center for the Study of Institutions, Populations, and Environmental Change (CIPEC). As part of this project,

* *Corresponding author address*: Abigail York, Center for the Study of Institutions, Populations, and Environmental Change, 408 N. Indiana Avenue, Bloomington, IN 47408; e-mail: amyork@indiana.edu.

we are developing an agent-based model of the decisions of rural landowners in Monroe County, Indiana, USA. Our objective is to use the model to explore how individual decision making by a heterogeneous set of landowners, given local biophysical conditions, led to the particular aggregate pattern of land-cover change in Indiana, with an emphasis on forest-cover change.

At the time of the first federal government surveys of Indiana in the early 1800s, 86% of the state's 22.9 million acres was forested (Lindsey, et al., 1965). During the next century, settlers cleared the forests to create homes, farmland, pastures, businesses, towns, and cities. By 1920, forested land had shrunk to 1.4 million acres, or only 6% of the land base. This deforestation process was followed by a period of gradual reforestation that still seems to be in progress (Schmidt, et al., 2000) (see Figure 1). Reforestation in Indiana has been cyclic and spatially nonuniform (Schweik, 1998). Much of the reforestation occurred in the early 1900s and was primarily due to localized processes like agricultural abandonment (Sieber and Munson, 1992), while deforestation due to metropolitan growth and urban sprawl continues to contribute to deforestation today (LeMaster, 1993). Currently, 19% of Indiana is forested, or approximately 4.2 million acres, and much of this land is private nonindustrial forest (Schmidt, et al., 2000). The complex dynamic interactions of people and forests are not unique to Indiana. Similar patterns have occurred in multiple eastern states and in some countries of Europe.

The effort to explain changes in forest cover over time is directly related to many of the major environmental issues of the day — how to maintain vital ecosystem services, protect

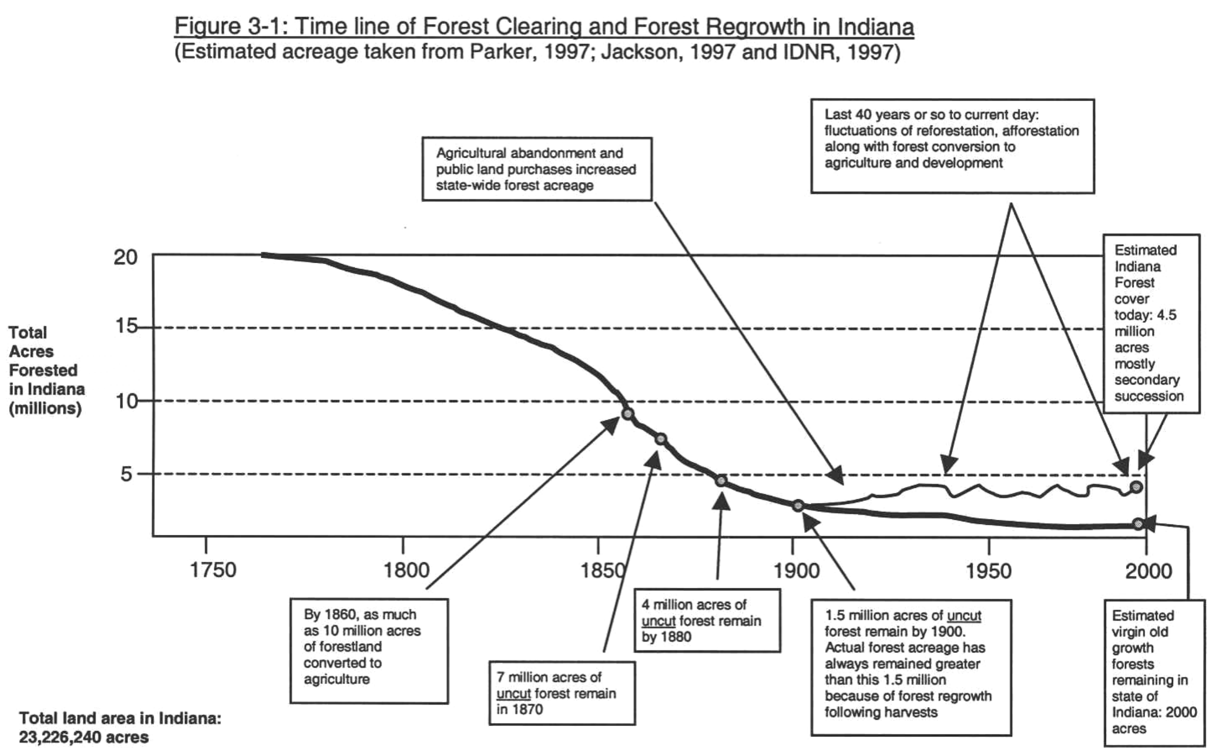

FIGURE 1 Indiana Land-cover Change (Source: Schweik, 1998, Chap. 3, p. 102, Figure 3-1)

biodiversity, and increase carbon sequestration so as to reduce global warming. The history of land cover in Indiana provides a good setting for developing ABM, as similar cyclic patterns of forest growth have occurred elsewhere and are desirable in many tropical countries currently undergoing massive deforestation. Further, relatively good historical data exist even though these data are located in scattered sources and have not previously been brought together as the foundation for a single project. If it is possible to understand the complex interactions among biophysical, social, and institutional factors affecting individual land-use and land-cover decisions in Indiana, many applications can be made to other locations.

THE INDIANA BIOCOMPLEXITY AGENT-BASED MODEL OF LAND-USE AND LAND-COVER CHANGE

Currently, our team is transforming a prototype model developed early in our project to provide an initial, very general ABM of land-use decision making without locating the agents in a "real" location. We are also creating a more realistic model in which the matrix of land-use characteristics is based on extensive acquisition and processing of data representing actual land cover in southern Monroe County from 1939 to 1997.[1] The agents in our model are private landowners who have the potential to "grow" forest on their lands or to use their land for agriculture or other purposes. In addition to the behavior of individual landowners, heterogeneity among biophysical (topography and soil quality) (Evans, et al., 2001) and socioeconomic (Koontz, 2001) factors influence the current spatial pattern of forest cover in Indiana. Thus, in our basic modular structure (Figure 2), individual landowners (agents) are in the center and interact with other modules that could potentially affect decision making.

The four basic modules are the social, economic, political, and the biophysical modules. Each module provides some constraints and opportunities that affect the decision space of the agents. Institutions can be a source of information, incentives, or sanctions that agents use in their land-use decision-making process. While our colleagues are developing the biophysical, economic, and social modules, we are attempting to obtain sufficient, detailed knowledge about government-sponsored inducements or sanctions that could potentially affect an agent's decision to form the basis for the political module. As we try to understand each module's impact on the agent's decision-making process, we are aware that the agent's actions or inactions may modify the structures of the modules, such as erosion of soil due to poor farming practices, which cause changes in the biophysical module and perhaps lead to the creation of new land-use laws. In the early stages of our work, we can only investigate how programs and policies may affect the agent and not the agent's impact on the programs and policies, other institutions within the political module, or other modules and their components.

Agents make decisions based on various characteristics of their household (e.g., size, age and gender distribution, income) within a biophysical, social, economic, and political setting. Institutions, which make up the political setting, are considered to be the *de jure* and *de facto*

[1] In this project, several models will be developed to address different questions. As data availability for other variables (e.g., economic, demographic, agricultural) varies greatly over time, some models will encompass the entire time series of land-cover data, whereas others will cover smaller periods, such as from 1972 to 1997.

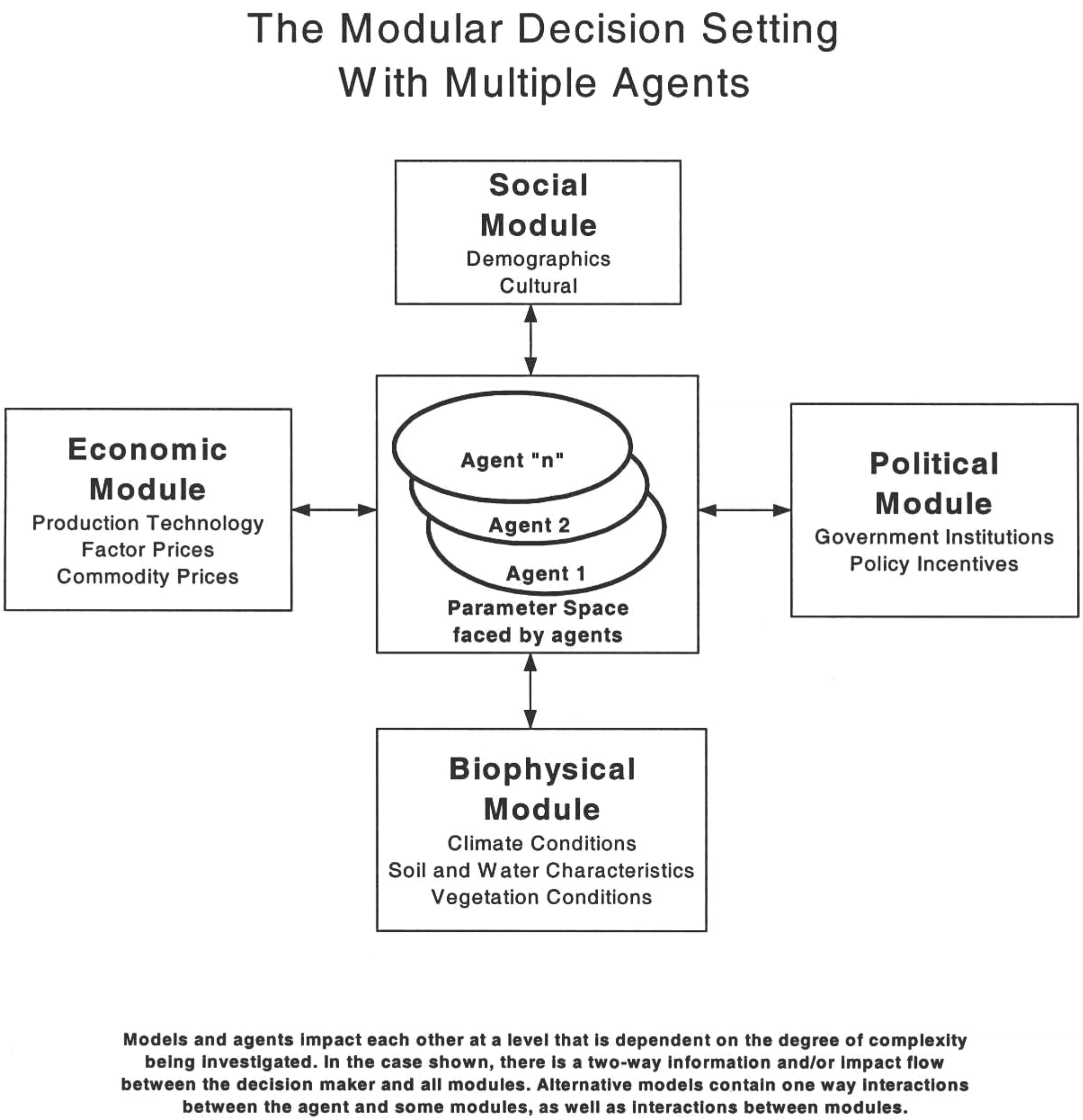

FIGURE 2 The Basic Model of the Biocomplexity Project (Source: Hoffmann, et al., 2002)

rules created by multiple levels of governments that attempt to establish incentives and sanctions for land-use management decision making. Rules as used here are actions and/or outcomes that are required, prohibited, or permitted, as well as the sanctions that are authorized if the rules are broken (Crawford and Ostrom, 1995). Many organizations in the public and private realm have programs that could potentially affect the de/reforestation process in Indiana. These include professional organizations, such as farming cooperatives; nongovernmental organizations, such as stewardship and conservation groups; and government programs, both state and federal.

At this time, we are trying to learn about and understand the potential connections between policy initiatives of diverse governmental and nongovernmental programs and the agent in the biocomplexity model. Initially, we focused on governmental programs because information about these programs is more easily obtained, and the rules are highly formalized, simplifying the effort to use them in our model. We have started to gather information on about 100 or more such programs and have acquired more detailed information about 30 state and federal governmental programs related to land use (see Appendix 1). In this paper, we begin with a conceptual model of potential interactions between these programs and agents. We then investigate in some depth one governmental program, The Indiana Classified Forest Program, in an effort to understand how these programs affect landowner decisions.

AN UNDERLYING CONCEPTUAL MODEL

It is useful to sketch a broad conceptual model that eventually may be implemented in the Indiana Biocomplexity ABM. The aspiration adaptation framework of Selten (1998) is particularly helpful. In this framework, an agent has multiple aspirations but does not have a complete preference order for them. In fact, the agent makes slight adjustments to increase progress toward different goals. Which action an agent takes depends on the feasibility and the urgency of the actions. The interesting element of this theory is that different types of goals do not have to be translated in one aggregated utility function, since the agents move in a landscape of different goal variable changes and try to make local improvements.

These concepts are well captured in *belief-desire-intention (BDI) agents*, in which decision making depends on the manipulation of data structures representing the BDIs of the agent. The BDI architecture is based on the concept of practical reasoning (Bratman, et al., 1988). By practical reasoning, we mean reasoning that is directed toward actions. Practical reasoning agents weigh conflicting options. Considerations of their options are affected by the BDIs of the agent. A contrasting approach is deductive reasoning, where agents use purely logical reasoning (Woolridge, 2002). BDI architecture involves two key processes: deciding *what* goals an agent wants to achieve (deliberation) and deciding *how* an agent is going to achieve these goals (means-ends reasoning). The main idea is that an agent has limited resources to make decisions, in terms of time and knowledge. The beliefs represent information about the agent's current environment. Beliefs, together with desires, filter in a deliberation process the range of possible options to a set of intentions. The intentions may lead the agent to take various actions. Because of changes in the environment (affecting beliefs and/or desires), both the intentions and the actions that flow out of them may change. Thus far, BDI agents have been mainly applied for agents doing real-time activities, which differ greatly from the long-term dynamics of the land-use change we are attempting to capture. Nevertheless, the BDI framework provides a basic structure to implement the aspiration adaptation framework for agents in our project.

As previously stated, institutions can affect a landowner's decision-making process through a wide variety of incentives, sanctions, and information resources. A simple diagram helps to explain the potential role of a governmental program on the decision making of an agent (Figure 3). For various reasons, an agent may choose not to participate in the program after learning more about it. The agent may decide that the expected benefits (both financial

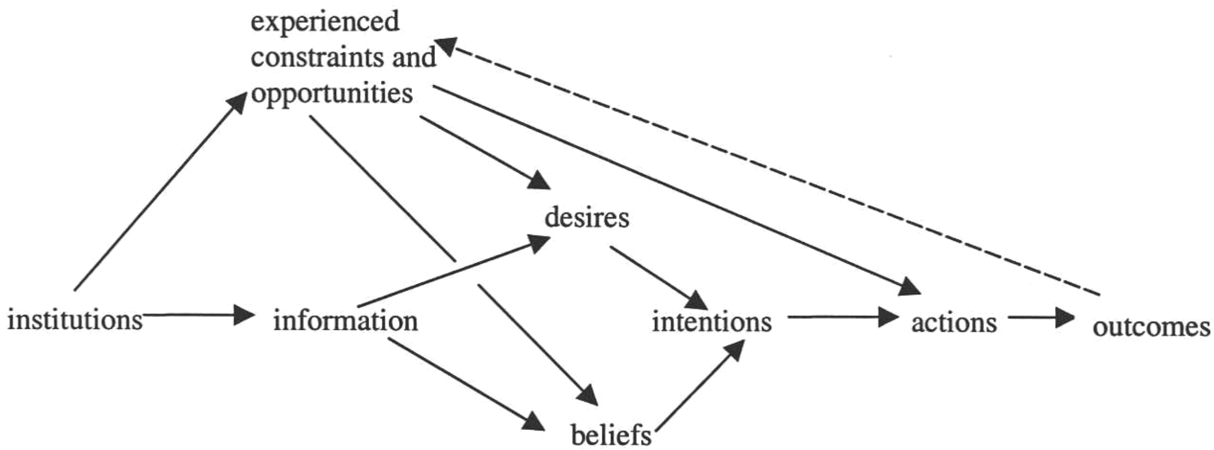

FIGURE 3 A General Conceptual Model of the Decision Process of a Landowner

and otherwise) are less than the expected costs of participation. We do not expect many agents to calculate this net benefit in the form of a utility function, either consciously or subconsciously, but rather as Selten (1998) theorizes, the decision may come from the evaluation of multiple aspirations. Agents may place little trust in the government and therefore decide not to participate in a government-sponsored program. Others may perceive that joining is too much of a hassle (i.e., potentially with surveys, registration fees, and other bureaucratic "hoops to jump through"). Agents may also evaluate the program information and determine that certain actions, such as clear-cutting, building homes, or subdividing the parcel, may not be allowed while participating. Perhaps the easiest explanation for nonparticipation is that some agents are not eligible.

Even a program participant does not necessarily change his/her behavior (intentions, actions and outcomes) with regard to land use. For example, some agents' existing management practices may have been in line with a particular program's guidelines. These agents may join the program to obtain additional benefits from an action that they would have already taken. Some agents, however, may modify their behavior once participating in order to meet the guidelines. Others may participate in the program but continue or begin unauthorized land use as defined by the program guidelines or rules. The level of compliance with program rules may depend on sanctioning and enforcement.

Even in the absence of governmental programs, landowners will have desires that affect what they do with their property. The desires are partly innate; that is, they depend on the personality and attitudes of the landowner. Private landowners value their property in different ways (including various economic, environmental, and amenity values) (Birch, 1996; Koontz, et al., 1998; Baughman, 2002). In a stratified, random sample of landowners in Monroe County, considerable variation was found among landowners in their evaluation of land attributes (Kauneckis and Novac, 2000; Koontz, 2001). We therefore expect that landowners differ in their initial desires and beliefs. These are in turn continually being shaped by experienced constraints and opportunities. For example, a developer approaches the agent with an offer to develop the property for a good price or the agent witnesses the neighbors selling timber for a hefty sum. In line with the framework of Selten, agents' desires adapt over time with these experiences (Selten, 1998). Like desires, beliefs can be affected by new information. These beliefs and desires ultimately result in intentions. These intentions may eventually lead to actions, depending on other factors such as physical and financial restrictions, lack of sufficient time to realize intentions, or the emergence of new opportunities. Agents may or may not be aware of a particular institution, which further complicates analysis. Once an institution becomes engaged with an agent, it may provide new information (perhaps through rules) that affects beliefs, desires, or intentions. Therefore, the agent's information, BDIs, and actions may all be affected directly or indirectly through the institution.

It is difficult to separate the effect of an institution from that of original BDIs, as well as the mitigating factors mentioned above. It is clear then that it is extremely difficult to relate land-use change directly to activities sponsored by institutional programs. Therefore, ABM, in combination with empirical data regarding participation rates and attitudes, enables us to explore potential impacts on land use in Indiana.

It is useful to examine the structure of various land-use programs to evaluate their potential impact. From this array of available government programs, we hope to acquire sufficient information to help create a set of institutional variables that potentially will impact landowner beliefs, desires, intentions, and actions in our model.

GOVERNMENTAL PROGRAMS IN INDIANA

A vast array of Indiana and federal governmental land-use programs have been in place throughout the state's history. In the late 1800s, clearing of forest land in Indiana did not occur without concerns regarding the conservation of forest. Historical records from the late 1800s identify various organizations such as the Indiana Horticultural Society, debating the needs for conservation of timber resources (Clark, 1987). The concern for the continued loss of forest land and erosion of soil due to land clearing eventually led to the passage in 1899 of House Bill 436, Indiana's first forest classification, which gave participants a tax reduction on one-eighth the area of their woodland, with the following restrictions: cut no more than 20% of their timber, limit grazing in the woodland, and replant every tree that was cut (Clark 1987). A revised Forest Tax Classification Act passed in 1921 required a forest management agreement with the State Forestry Department and allowed unlimited woodland acreage to be assessed at $1 per acre (Clark, 1987).

Today, more than 101 programs exist in Indiana. Of these, 76 state and federal governmental programs and 25 nongovernmental programs may affect a landowner's decisions. These programs offer a variety of services — from information and ideas to funding — to help landowners manage natural resources. From these 101 programs, we have collected detailed information from 30 state and federal governmental programs that are directly and indirectly targeted at creation and maintenance of forest cover on private lands in Indiana (Appendix 1).

Information about governmental programs is spread through various channels of knowledge diffusion. Among the 30 identified programs, most advertise through various media, such as news bulletins and newsletters, although frequently this information is targeted to landowners already participating in the program. Many have their own Web sites with annual reports discussing missions, participation rates, and funding, as well as links to other information resources for potential participants. For landowners, the most utilized and trusted source of information about assistance programs is often word-of-mouth.[2]

Many of the 30 programs are educational programs (e.g., the Lake & River Enhancement Program) or are not focused on individual landowners (e.g., the Arbor Day Grant). The effect of these programs is difficult to evaluate because the focus is mostly on altering agents' beliefs and desires through information with no sanctions and few incentives. In comparison, a few voluntary programs are directly related to private landowners' land-use decision making. Some of these institutions offer a property tax assessment reduction as a financial incentive for participation. One such program is the Indiana Department of Natural Resources (IDNR) Division of Forestry's Classified Forest Program (CFP). The following section examines this program in detail in relation to our conceptual model.

THE CLASSIFIED FOREST PROGRAM

The CFP was established by the Indiana Classified Forest Act 6-1.1-6 in 1921. This program was developed to encourage people to keep areas in forest land or create forest lands, by planting trees, for the purpose of forest conservation. The IDNR Division of Forestry sums up the Classification Act as (IDNR, 2002):

1. Both native timberland and land planted to acceptable tree species are eligible for classification.

2. A Classified Forest must be protected from domestic livestock and fire.

3. Timber may be cut at any time and sold or used as the owner desires, provided that such cuttings or sales of timber are not so severe that they will destroy or seriously set back the timber-producing values of the forest.

4. No dwellings are permitted in a Classified Forest, but owners may maintain a sawmill or operate a sugar camp.

5. The land must be posted with signs provided by the Division of Forestry.

6. An annual report must be made to the state forester regarding the condition of each Classified Forest.

7. Once classified, the forest must remain in the program indefinitely unless withdrawn. If withdrawn, the landowner could be subject to paying back taxes and a 10% penalty.

CFP landowners receive a property tax assessment of $1 per acre for general property taxation purposes. Woodland that is not in a Classified Forest is assessed at 20% of value determined by the soil productivity map (State Board of Tax Commissioners, 1992). Since the

---

[2] Many program officials mentioned that word-of-mouth is the best publicity. Landowners responding in the Monroe County Landowner Survey frequently stated that positive information from neighbors and friends regarding governmental programs led to their participation.

1960s, agricultural land in Indiana (which includes any land parcel of 10 acres or more with no commercial or industrial use) has been assessed for tax purposes at $495/acre (Kelly and Wuensch 2000). This amount is adjusted according to a soil productivity factor[3] and reduced by 80% if the land is wooded, so the greatest assessment reduction for CFP landowners is approximately $126/acre. Owners with Classified Forest in agriculturally productive soil receive a greater reduction in tax through participation than those with poorer soils.

CFP landowners file a written Forest Stewardship Management Plan (FSMP) created by their district forester and signed by the owner. The plan must adequately describe the present condition of the forest and prescribe a plan of action meeting the objectives of the owner, while following the guidelines for inclusion in the classified forest land program. Timber extraction is allowed on CFP land and is, in fact, often encouraged by the management plan. The Classified Forest Act requires the Classified Forest owner to follow minimum standards of good timber management as prescribed by the FSMP. In addition to property tax breaks, landowners receive forestry literature and periodic free inspections by their district forester while the forest is enrolled in the program. The FSMP may be revised periodically to meet changing landowner objectives and forest conditions. Therefore, upon joining the CFP, landowners receive a flexible management plan designed around and potentially changing the current set of BDIs through information and resources (see Figure 3).

The only sanction that the CFP authorizes is IDNR removal of the property from the program and collection of back taxes with 10% interest. According to IDNR officials, this has rarely been done. Overall, the limited rules and sanctioning, as well as the limited amount of eligible land, may decrease the statewide impact of the CFP on landowners' decision making. Alternatively, the lack of restrictions, beyond the eligibility requirements, may increase the participation levels for owners with 10 continuous acres of forest, as it may already fit with their current BDIs.

Currently, more than 8,300 pieces of property, covering nearly 410,000 acres, are enrolled in this voluntary program, with an average growth rate of approximately 10,000 acres per year (IDNR, 2002). A glance at Figure 4 shows that the number of acres of Classified Forest has increased steadily since the beginning of the CFP. However, it is unclear if the success of the CFP is a cause or a consequence of the general reforestation trend in Indiana (Figure 1).

As mentioned earlier, landowners that have decided to maintain forest cover may join the CFP for the tax benefit after making their land-use decision. Eligible landowners that participate may or may not follow through with the management practices outlined in their plan. These owners may not want to actively manage their land or may decide to cut their forests without direction from the plan. On the other hand, a landowner may start participating in the program attracted by the tax relief, but due to increased information after developing a management plan, the landowner may become inspired and may perform more or less active management of the property than previously intended. Thus, joining the program may or may not affect the intentions, actions, and outcomes of an agent's land-use decisions.

---

[3] The highest soil productivity factor in Indiana is 1.28 (Wuensch, et al., 2000).

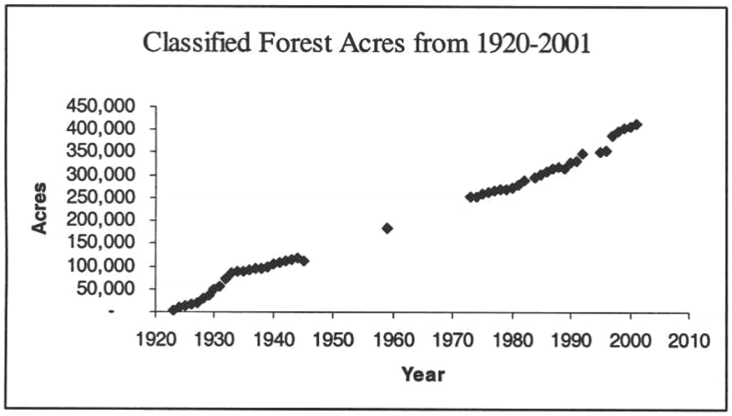

FIGURE 4 Classified Forest Acreage over Time (Source: Data adapted from IDNR Division of Forestry information)

FINAL THOUGHTS

We have started to obtain key information about various national and state policies and programs that may impact landowner decisions about land cover in diverse ways. One of our first findings is that a plethora of programs could potentially impact landowner decisions. It was rather challenging to find consistent information about each program by searching Web pages, published reports, and eventually contacting public officials to gain more information. If researchers who are well equipped with access to libraries, the Web, and email find it difficult to obtain information, we can imagine that citizens without such equipment find it even more challenging. This assumption is supported by information obtained from a 1998 survey of landowners that shows many of these programs are unknown to Indiana landowners (Summers, 1998). If unknown, an institution is not likely to generate information affecting desires and beliefs, as these affect intentions and actions. Thus, our immediate task is to focus on a smaller set of programs that have a higher chance of affecting desires, beliefs, intentions, and actions.

In a closer investigation of one program (the CFP) that has substantial participation, we have shown through the use of our conceptual model its potential to influence some landowners to change their land-use decisions. However, because of the eligibility requirement for the program of 10 continuous acres of forest land, an increasing number of landowners with smaller parcels or discontinuous coverage (biophysical constraints) are ineligible, and the impact of the program is minimized. Conversely, the flexibility of the program may attract a group of agents with a more varied set of beliefs and desires, which may or may not eventually be affected by the opportunities and constraints of the program. Likewise, changes in land prices due to nearby development or decreased agricultural commodity prices change the economic constraints that agents face and impact the program’s effect on land use. Thus, intentions, actions, and outcomes may be altered or only facilitated by the institution itself. To understand the impact of the institution on the agents’ behavior, we must understand the condition without the institution (e.g., the property tax assessment without the reduction) — another aspect of the political module or the initial set of, or the initial set of beliefs and desires created by an agent’s experienced constraints and opportunities. Institutions have the potential to intervene and alter the beliefs and desires of an agent through provision of information and incentives. We have described our initial efforts to create a model structure by which we can examine institutional impacts on individual land-use decisions. Our model is based on the concept that an initial condition endows an agent with a particular set of beliefs and desires that may lead to any number of intentions, actions, and outcomes. We plan to extend this model to study the impact of other political institutions, such as taxation and zoning, as well as utilize the conceptual model to facilitate implementation of institutions in the agent-based model.

ACKNOWLEDGMENTS

This paper is part of the Biocomplexity Project of the CIPEC at Indiana University, Bloomington. The project is funded by the U.S. National Science Foundation (Grant Number SES 0083511).

In addition to the authors, the project team includes the following members:

Jerome Busemeyer, Department of Psychology, Indiana University
Rich Caldanaro, CIPEC, Indiana University
Juan-Camilo Cardenas, Environmental and Rural Studies, Universidad Javeriana, Bogota, Colombia
Peter Deadman, Department of Geography, University of Waterloo
Shanon Donnelly, CIPEC, Indiana University
Tom Evans, Department of Geography, and CIPEC Indiana University
Min Gong, CIPEC, Indiana University
Matt Hoffman, Department of Political and International Relations, University of Delaware
Hugh E. Kelley, Department of Economics, Indiana University
Tei Laine, CIPEC, Indiana University
Vicky Meretsky, School of Public and Environmental Affairs (SPEA), Indiana University
Darla Munroe, Department of Geography, University of North Carolina, Chapel Hill
Dawn Parker, CIPEC, Indiana University
Jörg Rieskamp, Department of Psychology, Indiana University
Jim Walker, Workshop in Political Theory and Policy Analysis and Department of Economics, Indiana University
David Welch, CIPEC, Indiana University

REFERENCES

Baughman, M., 2002, *Landowner Survey of Forest Stewardship Plan Implementation*, University of Minnesota, Minneapolis, MN.

Birch, T., 1996, *Private Forest-Landowners of the Northern United States*, U.S. Department of Agriculture, Northeastern Experiment Station, Randor, PA.

Bratman, M.E., D.J. Israel, and M.E. Pollack, 1988, "Plans and Resource-bounded Practical Reasoning, *Computational Intelligence* **4**:349–355.

Clark, D.L., 1987, *Entrepreneurs in Hardwoods: A Study of Small Business Strategies*, Indianapolis, IN: White Arts, Inc.

Crawford, S., and E. Ostrom, 1995, "A Grammar of Institutions," *American Political Science Review* **89**:582–600.

Evans, T.P., G.M. Green, and L. Carlson, 2001, "Multi-scale Analysis of Landcover Composition and Landscape Management of Public and Private Lands in Indiana," pp. 271–287 in A. Millington, S.J. Walsh, and P. Osborne, eds. *GIS and Remote Sensing Applications in Biogeography and Ecology*, Boston, MA: Kluwer Academic Publications.

Hoffmann, M., H. Kelley, D. Parker, and J.M. Walker, 2002, *Agent Based Models: Alternative Decision Making Strategies*, Center for the Study of Institutions, Populations, and Environmental Change, Bloomington, IN.

IDNR: See Indiana Department of Natural Resources.

Indiana Department of Natural Resources, Division of Forestry, 2002, *Private Landowner Assistance*, Sept. 10. Available at http://www.in.gov/dnr/forestry/.

Janssen, M.A., ed., 2002, *Complexity and Ecosystem Management*, Cheltenham, UK: Edward Elgar Publishers, Dec.

Kauneckis, D., and C. Novac, 2000, "Preferences on the Landscape: How Much Do Individual Values Matter?" *Constituting the Commons, the 8th Annual Conference of the International Association for the Study of Common Property*, Bloomington, IN.

Kelly, F., and J. Wuensch, 2000, "Assessment Reform in Indiana: One Step Forward, Two Steps Back," *Land Lines: Newsletter of the Lincoln Institute of Land Policy* **12**:1–3.

Koontz, T.M., 2001, Money Talks — But to Whom? Financial versus Nonmonetary Motivations in Land Use Decisions," *Society and Natural Resources* **14**:51–65.

Koontz, T.M., L. Carlson, and C. Schweik, 1998, *The Role of Institutions in Shaping Land Use: An Exploratory Study of Southern Indiana Non-Industrial Private Forests*, Center for the Study of Institutions, Populations, and Environmental Change, Bloomington, IN.

LeMaster, D., 1993, *Forest Policy Issues in Indiana.* Department of Forestry and Natural Resources, Purdue University, Lafayette, IN.

Lindsey, A.A., W.B. Crankshaw, and S.A. Qadir, 1965, "Soil Relations and Distribution Map of the Vegetation of Presettlement Indiana," *Botanical Gazette* **126**:155–163.

Parker, D., S.M. Manson, M.A. Janssen, M. Hoffmann, and P. Deadman, 2003, "Multi-agent Systems for Simulation of Land-use and Land-cover Change: A Review," *Annals of the Association of American Geographers*.

Schmidt, T., M.H. Hansen, and J.A. Solomakos, 2000, *Indiana's Forests in 1998*, U.S. Department of Agriculture, Forest Service.

Schweik, C., 1998, *The Spatial and Temporal Analysis of Forest Resources and Institutions*, Center for the Study of Institutions, Population, and Environmental Change, Bloomington, IN.

Selten, R., 1998, "Aspiration Adaptation Theory," *Journal of Mathematical Psychology* **42**:191–214.

Sieber, E., and C.A. Munson, 1992, *Looking at History: Indiana's Hoosier National Forest Region: 1600 to 1950*, Bloomington, IN: Indiana University Press.

State Board of Tax Commissioners, 1992, *Indiana Real Property Assessment Manual*, Indianapolis, IN: State of Indiana.

Summers, K., 1998, *Joint Management of Nonindustrial Private Forests: An Alternative Method of Encouraging Stewardship in Indiana.* Center for the Study of Institutions, Populations, and Environmental Change, Bloomington, IN.

Woolridge, M., 2002, *Introduction to Multiagent Systems*, Hoboken, NJ: John Wiley and Sons, Inc.

Wuensch, J., F. Kelly, and T. Hamilton, 2000, *Land Value Taxation in Indiana: Challenges and Issues*, Lincoln Institute of Land Policy, Cambridge, MA.

APPENDIX 1: Land-use–related Programs in Indiana[a]

Program	Purpose	Members	Funding Source	Information Diffusion
Agricultural Conservation	To protect erosion on land	100 annual	A portion of state cigarette tax and $5.00 lake and river enhancement fee on boats	Bulletins and publication through USDA county offices
Arbor Day Grant	To encourage urban forest	1,000 schools annual	State funded	Letters sent to schools by December or January
Best Management Practice Cost Share	To help manage logging practice	NA	Environmental Protection Agency (EPA) grants	Web site and media publications
Classified Forest Program	To keep Indiana's forest	8,339 landowners with nearly 410,000 acres	From Division of Forestry (86% Timber Sale Tax and Seedling)	Web sites and brochures
Clean Water Indiana	To reduce water pollution from NPS	Counties	1999 State Assembly designed for 3 years with $1 million	County offices
DNR Clean Water Indiana	To protect erosion and water resources	92 counties (farmers and land owners)	State budget ($3 million)	Web site, publications, radio programs
Conservation of Private Grazing Land Initiative	To help manage grazing land	1,590 (2001), 854 (2000), 725 (1999)	USDA Technical Assistance Allocated Fund	News articles, TV, and radio programs
Conservation Reserve Enhancement Program	To remove land from agricultural production (land retirement)	16,000 participants with 330,000 acres	Commodity Credit Cooperation Acre ceiling (federal)	County-level office, some national, farm publication, agricultural newspapers
Cooperative Forestry Assistance/Management Program	To provide forest stewardship	NA (county offices maintain)	State (22%), dedicated fund (78%), t-sale, mail tax, federal programs	Sister agencies (network) recommendation, word-of-mouth, articles in local papers, no budget for ads
Emergency Conservation Program (ECP)	To help farmers in natural disasters	NA (county offices maintain)	Funds appropriated from Congress/ Community Credit Cooperation funding	County offices

Program	Purpose	Members	Funding Source	Information Diffusion
Environmental Education Grants Program	Financial support for environmental education to landowners	7,001 in three nongovernmental organizations maintained by program	EPA grants $190,000 per year from Congressional Appropriation for all of Region 5 states.	Mailing list and Web site
Environmental Quality Incentives Program (EQIP)	To provide technical, financial, and educational assistance	NA	Federal (USDA)	Form publication, county office, some federal information sources
Farmland Protection Program	To keep land in agricultural use	NA	USDA fund	Web site and newsletters, county affair events
Farm Loan Program	To provide financial help to farmers in land management	Vary by county	USDA general fund	Web site, state/county USDA offices
Farm Mediation/Farm Counseling	To provide financial advice to farmers	500–700 per year	Grants from Office of Commission of Agriculture and federal sources	Farm bureau, Purdue extension, community agriculture association, county extension office
Five Star Restoration Program	To restore streams and wetland	NA	$500,000 annual federal funds through EPA	Participants' network
Flood Mitigation Assistance Program	To eliminate long term risk of flood damage	NA	$160,000 annually from FEMA	County-level office, national flood insurance program
Flood Hazard Mitigation and Riverine Ecosystem Restoration Program	To conserve wetlands and to restore flood plains	NA	1999 Water Resources Development Act designated fund	Web sites
Forest Legacy Program (FLP)	Congress 1990 Farm Bill to identify and protect environmentally important forest lands	Six legacy areas in IN. FLP buys development rights from landowners.	Federal funding can be used for up to 74% of the purchase price (no exact dollar amount)	Web site, state and county DNR-Forestry offices, newsletter
Forestry Incentives Program also known as Forest Improvement Program	To support forest management practices	200 private land and forest owners (32 granted in 2000; 20 cannot be funded for lack of funds)	USDA fund	NRCS Web site

Program	Purpose	Members	Funding Source	Information Diffusion
Forest Stewardship Incentive Program	To encourage stewardship for privately owned woodlands	NA	USDA fund	Web site, grant proposal announcements, county offices
Hoosier Homestead Program	To encourage keeping farms in family	4,500 Indiana Fund	Department of Agriculture	Web site
Indiana's River Friendly Farmer Program	To decrease water pollution	20 in 1999, 2000, 60 in 2001	Farm bureau ($4,000 annually)	County-provided promotional items
Lake & River Enhancement Program	To reduce sediment and nutrient pollution in Indiana's watersheds	NA	$1.1 million per fiscal year from $5 cigarette tax; some cost share	Promotional letters sent to lake association and county officials/posters
Resources Conservation and Development Program	To accelerate conservation and development of natural and historic resources	NA	USDA fund	Web sites, USDA – NRCS offices
State Wetland Protection Grant	To protect wetlands in Indiana	NA	EPA Region 5	EPA offices and Web site
Tree Steward Program Grants	To provide educational training for tree care	NA county manage	Equal match of $500–1,000 is available for a grant proposal from state.	Web site, county offices, application announcements
Urban Forest Management	To help communities manage urban forests	NA	U.S. Forest Services ($2,000 to 20,000 grants)	Web site and DNR newsletters
Watershed Protection and Flood Prevention Program	To prevent floods and to increase proper utilization of land in watershed areas	NA	USDA – NRCS fund	Web sites
Wildlife Habitat Incentive Program	To provide financial incentives for fish and wildlife on private lands	NA	USDA – NRCS fund	Web sites

[a] Information about these programs was collected through Web sites, telephone calls, and e-mail communication. First, we collected all possible information about these programs through Web sites. Second, if needed, we called program offices and asked for background information missing in the Web sites. Finally, if we were unable to reach a person by telephone, we sent them an e-mail. For federal programs in Indiana, we often called the offices in DC to seek information about programs in Indiana. We then called or e-mailed the state contacts provided by the DC offices. We used a uniform template of background information sheet to collect information about these programs.

THE COMPLEX INTERACTION OF AGENTS AND ENVIRONMENTS: AN EXAMPLE IN URBAN SPRAWL

W. RAND,* M. ZELLNER, S.E. PAGE, R. RIOLO,
D.G. BROWN, and L.E. FERNANDEZ
University of Michigan, Ann Arbor, MI

ABSTRACT

We present and evaluate a foundational agent-based model of land-use change at the rural-urban fringe within the context of a larger project that will link surveys of the environmental and community preferences of residents with historical data on patterns of development. This paper focuses on the dynamics arising from a model of residential location resulting from preferences for services, density, and aesthetics, in particular on the relationship between micro-level preferences and policy-relevant macro phenomena such as scattered development, largest city size, and the number of residential clusters. We consider two representations of agents' utility functions — one additively separable and one multiplicative — to see if functional form has any impact on the dynamics of the system, and we find that they produce similar results. Our analysis includes both representative agent runs, in which all agents have identical preferences, as well as runs in which the agents have diverse preferences. We find that diversity can increase sprawl through feedbacks associated with the spatial locations of services and agents. In addition, we examine cases in which the agents' location decisions affect the aesthetic quality of neighboring sites and find that these feedbacks further exacerbate the sprawl effect.

INTRODUCTION

The goal of our project is to use an agent-based model to evaluate the ecological effects of alternative plans and designs for urban development. Development at the urban-rural fringe has been linked to a variety of negative ecosystem impacts, including habitat and migration corridor destruction (Johnson, 2001). In the modeling portion of our research agenda, we focus on how agent level preferences alter land-use change in this fringe region. Our immediate goal is to understand how residential agents make decisions on where to live and which dimensions of that decision-making process influence settlement patterns on the urban rural fringe. If we have a better understanding of this process, we can design policy instruments to control the patterns of urban development and thereby improve ecological performance.

Our project encompasses a suite of models that extends from very simple analytical models to full-blown, agent-based models with heterogeneous types and spatio-temporal feedbacks. This paper presents some of the results from our agent-based models. While our study focuses on modeling land-use changes at the urban-rural fringe in the Detroit Metropolitan Area, USA, our analysis here involves a hypothetical area. In our experiments, we use preference distributions designed to test archetypal or extreme cases. This approach allows us to analyze the

* *Corresponding author address*: William Rand, Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109; e-mail: wrand@umich.edu.

verisimilitude of our assumptions in a controlled manner, thus testing the limits of our model. In future experiments, the assumptions driving our model will be linked to empirical data.

Our agent-based model extends the traditional land-use and settlement models in several directions. First, we include feedback on several dimensions simultaneously. Residential choices affect services, density, and aesthetic quality, all of which influence the decisions of other agents. Second, our model includes heterogeneous agents. We find that heterogeneity matters. Knowing only the mean preferences of the agent population is insufficient to predict land-use patterns. These changes — multiple feedbacks and heterogeneity — would be difficult if not impossible to include in an analytic model. Our agent-based approach enables us to include in our model both of these important, empirically validated features (Ewing, 1997; Irwin and Bockstael, 2001; Chin, 2002).

In this paper, we explain our approach, its various parameters, and its potential. We then present some results from this model and discuss how those results compare with other models. We conclude with a discussion of future explorations. Our long-range goal is to extend this model to address normative and descriptive questions.

EXPLANATION OF MODEL

The simplified model presented here was developed in Swarm using agents who have locational preferences. These agents exist on a heterogeneous two-dimensional landscape, which can be defined by using data stored in a geographic information system or set to a hypothetical landscape. The model generates dozens of outcome variables of both a spatial and quantitative nature, and even more can be added. We call this model SLUCE (**S**patial **L**and **U**se **C**hange and **E**cological effects). The model is composed of three primary parts: the environment, the agents, and the agents' interaction with the environment. The following sections describe each of these parts.

Environment

We represent geographic space with a two-dimensional square lattice. The results presented below take place on an 80×80 lattice. For purposes of calibration, each site can be interpreted as 0.5×0.5 mile so that the entire lattice is 40×40 miles. Each location on the landscape has two exogenous characteristics: a natural beauty score in the interval [0,1] and the presence or absence of an initial service center. The model presented has only one initial service center in the center of the lattice. We then compute the distance to services for location (x,y), sd_{xy} by taking the sum of the inverse Euclidean distances (for simplicity) to the nearest eight service center locations from that cell. Thus, a cell surrounded by service centers would receive a score of 8. Because it seems reasonable that the residents of a cell would not receive additional benefit from more than about 2 immediately adjacent service centers, we set the service center score to a maximum of 2 and normalize the value. Thus,

$$sd_{xy} = \max\left[1, \frac{\left(\frac{1}{\| sc_1 \|} + \ldots + \frac{1}{\| sc_8 \|}\right)}{2}\right], \tag{1}$$

where $\|sc_i\|$ is the Euclidean distance to the i'th nearest service center from x,y. Service center distance changes over time as new service centers arise. The other variables that affect agents' choices are endogenous, such as density, and also change as time progresses.

Agents

The basic agent types are residents and service centers (e.g., retail firms). Residents and service centers enter the world at each time step, and each takes up one cell in the lattice. Both agents and service centers have the capacity to include heterogeneous attributes and behaviors, but at present, service centers do not have any attributes themselves. They might be more accurately called proto-agents. However, their presence greatly affects how residents determine where to live.

In the most basic model, residents have two attributes:

- *Beauty Preference* ($\alpha_{nb} \in [0,1]$), the weight that an agent gives to the natural beauty of an area. The natural beauty of an area can be generated from a distribution or set to a particular value exogenously. The beauty value of a cell x,y to an agent i is $nb_{xy} \times \alpha_{nb,i}$.

- *Service Center Preference* ($\alpha_{sd} \in [0,1]$), the weight that an agent gives to the nearness of an area to service centers. The service value of a cell x,y to an agent i is $sd_{xy} \times \alpha_{sd,i}$.

The distribution of the attributes across agents can be set at normal, uniform, or homogeneous. With the normal distribution, the variance must also be chosen.

Agent Behavior

The agents choose locations on the lattice, which in turn influences how other agents choose locations, resulting in a settlement pattern. For each iteration of the model, a group of new residents enters the map. The rate at which residents move into the landscape is determined exogenously. For the experiments below, we set the rate to 10 per step. Residents use a hedonic utility calculation to decide where to live, which currently takes into account some or all of the landscape variables. We endow the service centers with much less intelligence. Every time some number of residents (arbitrarily set to 100) is created, a service center is created in an empty cell near the last resident to enter the model. An initial service center is located in the middle of the map.

To select a cell, a new resident looks at some number of randomly selected cells (10 for all runs presented here) and moves into the cell that has the highest utility for r or selects randomly among tied cells. In the results presented, we calculate the utility in two different ways. The first is an additive model, which assumes that the preferences are separable from each other:

$$u_{xy} = \alpha_{nb} \times nb_{xy} + \alpha_{sd} \times sd_{xy}\ . \tag{2}$$

The second is a multiplicative model, which assumes that the preferences are dependent; i.e., being near a service is irrelevant if there is no natural beauty:

$$u_{xy} = nb_{xy}^{\alpha_{nb}} \times sd_{xy}^{\alpha_{sd}}\ . \tag{3}$$

ALTERNATIVE MODEL STRUCTURES

We start with our standard model (described above) and alter it to examine alternative structural assumptions. Our goal is to determine the effect of different modeling choices on the output and results of the system. By doing so, we intend to make recommendations about issues that must be taken into account when looking at questions of urban sprawl and those features that can be ignored in order to gain a simpler understanding of the general dynamics of the system. We start by looking at the effect of modifying the distribution and use of preferences in the system on the outcome. We then add feedbacks between the agents and the environment.

Preferences and Utility Functions

We first examine whether the form of the utility function affects the outcome of the model and whether representative agent models are sufficient to capture settlement patterns or heterogeneous agent models are necessary. We find that both matter with regard to quantitative measures. When agents have diverse preferences, we see much more sprawl. Moreover, when we change agents' preferences from separable to nonseparable, we see less extreme residential choices.

Heterogeneous Preferences

We begin with a model where all agents have the same preferences (i.e., all agents value distance to services and natural beauty equally). However, our model also allows for agent preferences to be drawn from a distribution of preferences. In most of our experiments, this distribution was a normal distribution with the same mean and different levels of variance. This capability allows us to examine how having heterogeneous agents can change the outcome of the model. The variances are increased equally for all preferences. This allows us to characterize how varying levels of heterogeneity affect the system, as opposed to looking at the separate question of how the positive or negative correlation of preferences affects the system.

Feedbacks

Another area of interest is how the agents' locational choices influence future agents' decisions. For instance, an agent's development decision can change the natural beauty of the location around it and make that area less desirable for future agents. In this case, the interaction of the agents and the environment determines the course of development in a run of the model. Thus, we examine two ways that this interaction might occur:

- *Neighborhood Density*, where agents can prefer to live in areas of higher or lower density, and
- *Land Use Affects Beauty*, where agents moving into a location decrease the natural beauty of the area around them.

Neighborhood Density

The first feedback that we added to the model gave agents a preference for a neighborhood density. In other words, an agent determines what the density of a neighborhood will be if it moves into it, and the agent then compares that to its ideal value for the density of the area in which it wants to live. This value is then used as part of the utility function similar to the distance to services and natural beauty.

The neighborhood density function around cell *C* is a weighted average of the fraction of Moore (8) neighbors that are inhabited and whether *C* is also inhabited (when this is being used in the locational decision, *C* is always considered to be inhabited). The neighbors' fraction is worth one-half the value, and *C* is worth one-half the value. For example for a nonedge cell *C*, if two neighbors are developed in a square and *C* is developed, the result is (2/8 + 1)/2 = 5/8. Thus,

$$nd_{xy} = \frac{\dfrac{Developed\ Neighbors_{xy}}{Neighbors_{xy}}}{2} + \frac{Developed_{xy}}{2}\ , \tag{4}$$

where

$Developed\ Neighbors_{xy}$ = number of developed neighbors of cell *x*,*y*,

$Neighbors_{xy}$ = number of possible neighbors of *x*,*y*, and

$Developed_{xy}$ = 1 if *x*,*y* is developed and 0 otherwise.

Since agents always calculate what the value of the neighborhood density would be if they moved in, the minimum value is 0.5 and the maximum value is 1.

This adds two more attributes to the agents:

- *Neighborhood Density Ideal Value* ($\beta_{nd} \in [0,1]$), the density that agents prefer in their neighborhood, and
- *Neighborhood Density Preference* ($\alpha_{nd} \in [0,1]$), the weight that agents give to living in locations that have a density near its ideal density value.

The density value of a cell x,y to an agent i is $(1 - |nd_{xy} - \beta_{nd,i}|) \times \alpha_{nd,i}$, where the nd_{xy} is calculated as if the agent was already living in the cell. Thus, the greater the difference between the ideal and the actual value of neighborhood density, the lower the utility derived from locating at x,y.

This results in two modified utility functions. First the additive,

$$u_{xy} = \alpha_{nb} \times nb_{xy} + \alpha_{sd} \times sd_{xy} + (1 - |\, nd_{xy} - \beta_{nd,i} \,|) \times \alpha_{nd,i} \ . \tag{5}$$

Second, the multiplicative,

$$u_{xy} = nb_{xy}^{\alpha_{nb}} \times sd_{xy}^{\alpha_{sd}} \times (1 - |\, nd_{xy} - \beta_{nd,i} \,|)^{\alpha_{nd,i}} \ . \tag{6}$$

Land Use Affects Beauty

The second feedback incorporated was to have land use decrease natural beauty of surrounding locations. Specifically, as an area becomes more and more developed, we decrease the natural beauty of neighboring cells by a proportional amount. The modified natural beauty landscape simply takes the neighborhood density, subtracts that value from 1, and multiplies the result by the original natural beauty to arrive at a new natural beauty measure. Thus,

$$mnb_{xy} = (1 - nd_{xy}) \times nb_{xy} \, , \tag{7}$$

where nb_{xy} is the original natural beauty at cell x,y.

In determining a location, the residents calculate what the *mnb* of a cell will be if they move into it and then use that value in the utility function. This results in new utility functions. First the additive,

$$u_{xy} = \alpha_{nb} \times mnb_{xy} + \alpha_{sd} \times sd_{xy} \, . \tag{8}$$

Second, the multiplicative,

$$u_{xy} = mnb_{xy}^{\alpha_{nb}} \times sd_{xy}^{\alpha_{sd}} \ . \tag{9}$$

Of course, this interaction can be used with or without the neighborhood density preferences and ideal values described above.

Separable versus Nonseparable Preferences

Our final comparison involves changing the functional form of our utility function and rerunning our entire suite of experiments. In our original model, preferences are additively separable. This implies that agents can choose locations that have extremely high values on one attribute but low values on another. If we assume instead that preferences are multiplicative, so that the utility from natural beauty and distance to services equals their product not their sum, then agents will choose locations that better balance beauty with distance to services. For example, suppose that one location has a natural beauty value of 1.0 but a distance to services value of 0.1, and that another location has a natural beauty value of 0.4 and a distance to services value of 0.5. Moreover, assume that the agent has preferences such that $\alpha_{nb} = \alpha_{sd} = 1.0$. Using an additive utility function, the first location is preferred since 1.1 > 0.9. But, using a multiplicative utility function, the latter is preferred because 0.2 > 0.1. This tendency for choosing locations with more extreme attribute values can cause more sprawl, as agents choose locations of high natural beauty that are far from service centers and can at the same time cause bigger central clusters as agents who do not care about natural beauty choose locations near service centers.

EXPERIMENTS AND RESULTS

Experiments were run starting with the two base models, corresponding to the additive and multiplicative value functions described above. The natural beauty of each cell was derived from a normal distribution ($\mu = 0.5$; $\sigma^2 = 0.5$) between 0 and 1, and then a local filter was applied to create spatial autocorrelation.[1] Results for each experiment were averaged over a minimum of 30 runs. These models were subsequently modified to add heterogeneous preferences, the effect of neighborhood density preferences, and the link between land use and natural beauty.

Our model generates dozens of output measures. Two that closely track sprawl are the size of the largest cluster (*LRGECLUS*) and development beyond a 30-cell radius of the center (*DEV* + 30). The first measure is used to estimate the size of the main development. Clustered development is generally considered to result in less ecological impact because it reduces the amount of area directly affected by development. In addition, such a development pattern increases accessibility to urban amenities, reducing commuting distances, and consequently energy consumption and pollution (Ewing, 1994; Beatley and Manning, 1997). A larger central cluster, compared among patterns with the same amount of area developed, usually means less scattered development in outlying areas and, therefore, less impact. *DEV* + 30 provides a good measure of how much leapfrogging or scattered development we see. All measures were averaged over time steps 95 to 105 to give an average measure at time step 100. After 100 time steps, 1,000 agents and 10 service centers have located in the world.

Heterogeneous Preferences

The first set of experiments was run keeping the mean value of all α parameters at 0.5, while sweeping the variance between 0.0 and 0.4, at 0.05 intervals. Land Use Affects Beauty and Neighborhood Density were not included. The results for this experiment and others are listed in Tables 1 and 2. This set of experiments allowed us to examine the effect of heterogeneous preferences. These results were averaged over 100 runs. Including heterogeneity ($\sigma^2 = 0.25$) increases the amount of sprawl according to the (*DEV* + 30) measure by 31 units. This should have been expected given that they are now agents who care relatively more about natural beauty and are willing to move away from service centers. The effect on the size of the largest cluster was harder to predict prior to running the model. Those agents that want to be relatively close to

---

1 The samples in all normal distributions in this paper are drawn with the given μ and σ^2, but if the result is outside the bounds, a new sample is drawn until the sample is within bounds.

TABLE 1 Observed Results — Additive Case (mean and standard deviation of the mean)

	***DEV* + 30**		***LRGCLUS***	
Basic model	123	(8)	800	(3)
Diversity	131	(7)	761	(5)
Land Use Affects Beauty	251	(9)	475	(12)
Density	96	(10)	771	(9)
Land Use Affects Beauty and Diversity	303	(8)	371	(10)
Density and Diversity	124	(11)	760	(9)
Land Use Affects Beauty and Density	65	(5)	814	(7)
Land Use Affects Beauty, Density, and Diversity	109	(9)	740	(9)

TABLE 2 Independent and Interaction Effects — Additive Case

	***DEV* + 30**	***LRGCLUS***
Basic model	123	800
Diversity	+8	–39
Land Use Affects Beauty	+128	–325
Density	–27	–29
Land Use Affects Beauty and Diversity	+44	–65
Density and Diversity	+20	+28
Land Use Affects Beauty and Density	–159	+368
Land Use Affects Beauty, Density, and Diversity	–123	+349

services should make the cluster larger, but what was less clear was whether these agents could fill in all of the gaps created by the agents who want high natural beauty. In the model, we found that the largest cluster (*LRGCLUS*) decreased by 33 units when we included heterogeneity ($\sigma^2 = 0.25$). This indicates a larger amount of sprawl, which shows that these agents were not able to fill in all of the gaps.

The inclusion of diversity created more sprawl and clustering behavior. These two effects can be seen in screen captures of the model in Figure 1 (black indicates development).

With high levels of diversity, we see dark patches in the center where the early service centers locate. We also see isolated agents jumping far from the service centers to locations of great natural beauty fairly early on in the run of the model. Interestingly, while we found that diversity mattered, we also found that the level of diversity was less important. Our measures did not vary much once the variance was increased to 0.25. Thus, when we state that we are using heterogeneous agents, we mean that they had preferences with a variance of 0.25 and a mean of 0.5. This regularity proves useful in running other scenarios, as we were able to work with only two diversity levels when exploring other changes in the model.

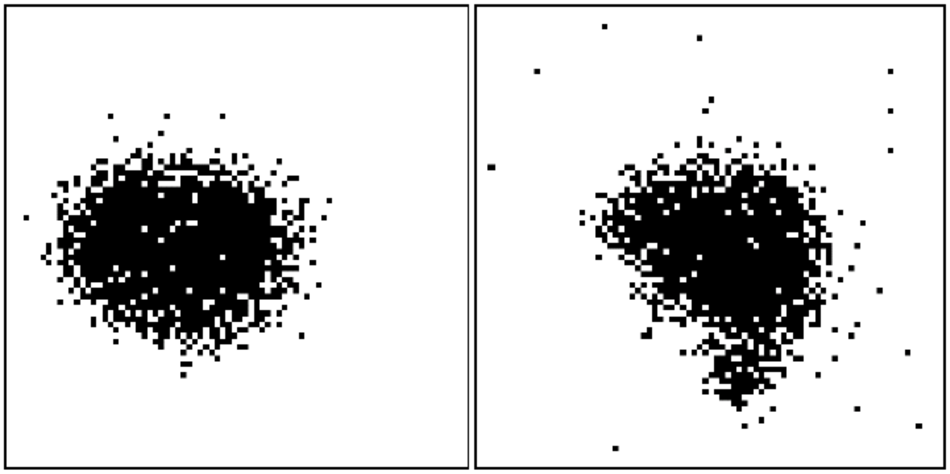

FIGURE 1 Homogeneous Agents (left) and Heterogeneous Agents (right)

Land Use Affects Beauty

The next set of experiments included feedbacks on the natural beauty, which had a much greater effect than including diversity (Tables 1 and 2). The results were averaged over 100 runs. We found huge increases in the (*DEV* + 30) parameter and much larger decreases in the size of the largest cluster (*LRGCLUS*). The reason for this disparity is straightforward. The decrease in natural beauty caused subsequent agents in search of natural beauty to leapfrog out into undeveloped territory, increasing *DEV* + 30. The destruction of beauty decreased *LRGCLUS* by making locations near the largest cluster less attractive.

Neighborhood Density

Our next set of experiments included preferences for density in the agents' utility functions (Tables 1 and 2). The results are averaged over 30 runs, and agents are sampling 10 cells at each time step. The mean value of the parameter for ideal density (β_{nd}) was swept from 0.5 to 1.0 at 0.1 intervals. As a crude benchmark we first ran a model where agents only cared about density, turning off preferences for service centers and beauty. When density preference was set to 0.0, the development pattern resembled a perfectly spaced splattering of points. When the parameter was set to 1.0, we obtained a collection of randomly placed clusters of agents. These effects played out when the density preference was also a part of preference. We present the results here for agents whose β_{nd} is set to 1.0, which means that they prefer densely inhabited areas. Also, when we discuss heterogeneity in preferences, we mean that the variance of the α_{nd} is set to 0.25. As expected, a high-density preference causes sprawl to decrease. However, the change is not as great as the "land use affects beauty" change is in the other direction.

Interactions of Feedbacks and Diversity

In addition to providing the raw numbers, we unpack the data to show the individual and linear effects of each variation by showing the marginal contribution to each output variable from each variation. For example, *DEV* + 30 has a value of 123 in the base model, but a value of 131 in the model with diversity. Therefore, the marginal contribution of diversity equals +8. Similarly, the marginal contribution of Land Use Affects Beauty equals +128. To approximate the interaction effect, we compute the total expected contribution of each change separately and assign to the interactive term the difference between that value and the data. For example, with Diversity and Land Use Affects Beauty without any interactive effect, we would obtain

$$DEV + 30 = 259 = 123 + 8 + 128\ .$$

Instead, we get a value of 303. Therefore, we assign a value of +44 to the interactive effect. We perform similar calculations for the other interactive effects. To capture the effect of all three variations, we calculate the expected value using all three pairs of interactive effects plus the individual effects and allocate the difference between that variable and the data to the interactive effect. When we included both the effect of land use on natural beauty and preference heterogeneity (variance set to 0.25), we found that the two effects reinforced one another, and we got even more sprawl than we would have had just by summing the two effects. This reinforcement effect occurs because with diverse preferences, we have some agents who care greatly about natural beauty. When development destroys natural beauty, these agents have to go even further out than they would have were this effect not in place. Thus, diversity plus natural beauty feedback creates an even larger effect than the two would separately.

Whereas diversity tends to increase sprawl, a preference for density should decrease it. Moreover, it appears that the preference for density decreases the effect that diversity in preferences has on the system. This is because once agents want to live near the center, it causes more service centers to choose central locations. This process decreases the effect that a strong preference for natural beauty has on agents who still have some preference for service centers.

Combinations of Feedbacks

In our final set of experiments, we see the effect of the two feedbacks when they interact with each other (Tables 1 and 2). The interesting feedback here is that density has a greater effect on the results than land use and diversity of preferences. In fact as shown in the last case, the density preference can turn the diversity of preferences from a positive sprawl feedback to a negative sprawl feedback. Land Use Affects Beauty can only serve to increase sprawl, but apparently its effect is not great enough to overcome the preference for higher density.

Separable versus Nonseparable Preferences

Tables 3 and 4 show our findings from the same scenarios as before, but with multiplicative utility functions. For the most part, the results are similar. The linear effects are directionally nearly identical: Diversity increases sprawl as does Land Use Affects Beauty, and

TABLE 3 Observed Values — Multiplicative Case

	DEV + 30		*LRGCLUS*	
Basic model	84	(5)	814	(3)
Diversity	111	(7)	799	(3)
Land Use Affects Beauty	218	(8)	600	(6)
Land Use Affects Beauty and Diversity	227	(6)	514	(8)
Density	67	(9)	828	(6)
Density and Diversity	82	(7)	822	(5)
Land Use Affects Beauty and Density	139	(15)	771	(7)
Land Use Affects Beauty, Density, and Diversity	131	(12)	695	(8)

TABLE 4 Independent and Interaction Effects — Multiplicative Case

	DEV + 30	*LRGCLUS*
Basic Model	84	814
Diversity	+27	+15
Land Use Affects Beauty	+134	–214
Density	–17	+14
Land Use Affects Beauty and Diversity	–18	–71
Density and Diversity	–12	+9
Land Use Affects Beauty and Density	–62	+157
Land Use Affects Beauty, Density, and Diversity	–97	+96

preferences for density tend to decrease sprawl. The only difference is that density preferences appear to lead to a smaller largest cluster with linear preferences but not with multiplicative preference. The only difference in magnitude of note relates to the effect of diversity. It is more pronounced in the multiplicative case. This is expected given that to get people to locate far from service centers, preferences must be diverse in the multiplicative case but not in the linear case. This is verified by the fact that with diversity there is less sprawl under multiplicative preference than there is with homogeneous linear preferences.

When we turn to the interaction effects, we begin to observe differences between the models. The interaction effect between diversity and land use affects beauty is positive in the separable case but not significant or even smaller in the multiplicative case. Again, this results from people who care more about natural beauty being willing to move far from service centers with linear preferences but being held closer with the more moderating multiplicative preferences. This same effect is even more pronounced when we look at the interaction between the density and land use affects beauty preferences. In the multiplicative case, the interactive term is negative, but moderately so. In the linear case, the interactive term is so negative that development beyond the radius of 30 is lower than the base case. This paradoxical result occurs because the agents with linear preferences are willing to sacrifice one attribute for another. As agents move to regions of high natural beauty, they destroy the natural beauty around them. Therefore, agents who locate later choose to move into dense central regions that have almost no natural beauty but that have high density. In a cursory investigation of screen shots, we found that agents with additive preferences left very few holes in the largest cluster, while agents with multiplicative preferences left holes in locations that had high density but almost no natural beauty. Further, agents with additive preferences who located outside the central region were predominantly isolated, whereas agents with multiplicative preferences tended to form little population clusters around areas that initially had high natural beauty.

DISCUSSION

In this paper, we have shown some preliminary results from an agent-based model that can be used to explore development patterns on the urban rural fringe and their ecological impact. Our findings suggest that diversity and feedbacks matter independently and jointly, but that the separability of the utility function is qualitatively unimportant for the parameters we are examining. The scientific and policy implications of these findings are provocative. On the scientific end, the fact that heterogeneity matters so much means that the empirical focus on means of variables may be misplaced. Perhaps variances are as important or more important than means. Similarly, the importance of feedbacks on natural beauty for sprawl suggests that their extent should be measured empirically, but doing so can be difficult. Third, the feedbacks support the Schelling-inspired possibility that preferences can be inconsistent in that people’s micro-level preferences may lead to macro-level development patterns that generate low levels of utility (Schelling, 1978).

From a policy standpoint, the implications are obvious. The interaction between higher preferences for density and natural beauty feedbacks can lead to significant effects on reducing sprawl, even when heterogeneous preferences are played out. Changing the ecological and aesthetic quality of development can modify the perception that dense development precludes access to privacy, quiet, and open areas away from congestion. Education on the environmental impacts of development, and supporting alternative modes of transportation, can also change the attitudes toward density. In addition, facilitating the location of urban amenities in central areas can reinforce the clustering effect of density preferences. More generally, the policy instruments can be seen as “motivators,” (i.e., as tools which change the agents’ preferences). Agent-based models, like the one presented here, can then show how those new preferences will aggregate, sometimes in unintended ways.

REFERENCES

Beatley, T., and K. Manning, 1997, *The Ecology of Place: Planning for the Environment, Economy and Community*, Cavelo, CA: Island Press.

Chin, N., 2002, *Unearthing the Roots of Sprawl: A Critical Analysis of Form, Function and Methodology*, Paper 47, Center for Advanced Spatial Analysis, University College London, United Kingdom.

Ewing, R.H., 1994, "Characteristics, Causes, and Effects of Sprawl: A Literature Review," *Environmental and Urban Issues* **21**(2):1–15.

Ewing, R.H., 1997, "Is Los Angeles-style Sprawl Desirable?" *Journal of American Planning Association* **63**(1):107–126.

Irwin, E.G., and N.E. Bockstael, 2001, *Interacting Agents, Spatial Externalities and the Evolution of Residential Land Use Patterns*, Working Paper AEDE-WP-0010-01, Department of Agricultural, Environmental, and Development Economics, The Ohio State University.

Johnson, M.P., 2001, "Environmental Impacts of Urban Sprawl: A Survey of the Literature and Proposed Research Agenda," *Environment and Planning A* **33**:717–735.

Schelling, T., 1978, *Micromotives and Macrobehavior*, New York: Norton.

DISCUSSION: ECOLOGICAL SIMULATION

P. SYDELKO, Argonne National Laboratory, Moderator

Pamela Sydelko: I have some overall thoughts regarding the similarities among the talks and topics that came up during the day. In this field of combining ecological and sociological modeling, what strikes me is that the concept of heterogeneity is three-, five-, and tenfold. We are trying to put together several disciplines. The speakers on this panel said that they are part of large interdisciplinary teams. Any modules in these very complex “put-together” systems have their own challenges as to the heterogeneity that comes into play between the disciplines and understanding what this ecosystem is, even if the human being isn’t in the picture.

We need to look at many feedback loops. In the ecosystem, feedback loops are what it’s all about. It is very tempting to want to isolate certain parts of the ecosystem and model while holding everything else equal. We’re just starting to ask about what would happen if we do hydrologic modeling and have feedback loops with the atmosphere. That is challenging enough. When we actually start to put in feedback loops, however, which Bill [Rand] stated quite clearly, that becomes the key point. The reason is that often the environment causes changes in society, and the changes in society’s reaction to the changes in the environment then cause different patterns to emerge.

As a result, feedback loops are very important, but very hard to tease out and determine which are the most important and where to start. Should we start with a very complex model, or should we start with a very simplistic model and move toward the more complex? What is the approach? It strikes me that when everyone talks about a simple model questions invariably turn toward, what if you do this? Everyone immediately wants to make the model complex. People who have very good, simple models are never happy with them. They say, much as Kathleen [Carley] would say, that they understand it, so they know you could make it more complex by doing this, this, and this. We may all have that tendency, and perhaps a good approach is to take a simple model and move to the complex.

We also can find ways — and this is similar to some research we’re doing at Argonne — to find architectures that can treat these agents, these objects of societal objects, either individually or within institutions. We can make them interact in a system that can be more dynamic with the environment. At Argonne, we’ve developed a system called DIAS — the Dynamic Information Architecture System. John Christiansen from Argonne is the inventor of DIAS. I would be interested in discussing with some of the panel members ways in which that architecture might help, by plugging in some of your systems, to make a common system where the objects can be more interactive and cause feedback loops. Those are some concepts for discussion.

Robert Reynolds: Bob Reynolds from Wayne State University. In light of the talks by Nigel [Gilbert] and Kathleen [Carley], it is interesting that all of these models had configurations of model components, and each of the model components could be of different granularity. It is possible to have, for example, hydrological and social models, and they produce input in different levels of granularity and detail. One of the key issues is model integration. We are trying to effectively link up these models at the same level of detail to produce consistent results that can then be checked and validated.

One of the challenges for the "tools people" is that as we build prefabricated model components, eventually these components will need to be integrated quickly, easily, and painlessly. One of the key issues, for example, in game design is that a model is only as good as its weakest component. Say that you have a very strong component that is generating real values to 10 or 15 precision points, and then say that these other lower-level models are being integrated, and one is binary — one or zero. All of that detail or knowledge is effectively being lost because of the way that it's transitioning into the other components.

Making these adjustments is very important. For example, you mentioned the agents going into the area. Once agents locate, can they then relocate or are they basically fixed? In other words, once you locate in a position, the environment around you starts to change. You learn that it's changing, and you move. If you want to add a learning component to your model, for example, how would that affect how it relates to the other components? Would you need additional detail from GIS to facilitate that upgrade? In these adjustments, then, it's not just one model that is affected. Rather, those effects ripple through all of the other models, and you have to deal with that integration all the time.

Harko Verhagen: Harko Verhagen from Stockholm University. I agree. In our model, agents stay where they're placed. But in reference to the modularity question, we deal with this issue not weekly, but at least on a monthly basis. Although this model has only been under development for about a year, it's already to the point where we have many different modules that we can turn on, turn off, plug in, and plug out.

I like the idea of having a general-purpose modeling system by which you can plug in different components and pull them out. However, if you try to build a system that answers every possible question, that is, has a module to address every possible situation, I worry about two things. First, are you going to answer any questions, and second, is that model actually simplifying real life in any way that's useful for telling you anything?

I'm not sure if there's a good way to address either of those concerns. It seems important even within the smaller framework, however, to keep our systems modular so that we can check the interdependent effects. The overview mentioned going from simple to complex, and this is a good way to do it. We start with the simplest model possible, come up with a small module that can be turned on, see what effect it has on the system, see if it is needed within the system, and continue to add things.

Part of the problem with models that start out at a high level of complexity is that they might have pieces that are completely irrelevant to the model dynamics. It might have no effect whatsoever on what's going on and almost no effect on the truthfulness of the predictions that the model produces for you. Therefore, it seems irrelevant to equate complexity with necessarily true empirical data in those ways. I think it's important to check modules and to have those within your system; however, I'm not sure it's always necessarily good to have as many modules as possible.

Sydelko: As far as the integrated system approach, the one that we [referring to Argonne] have is flexible, and we built it in that way to get away from what you are talking about. That is, we start with a couple of components, and then we add a third one and a fourth one if needed, and we end up having to kind of “glom” these things together to make it work. We started out doing it that way and realized that we had an *n*-squared problem, because as soon as we pulled one out, we had all these connections.

The approach we took is very generic. It’s a framework for doing model integration called DIAS, and it provides a very generic framework. It’s not discipline-specific. It’s only discipline-specific when you add your objects — we call them entity objects, which could be agents, or your watershed, or whatever objects you want — but it only becomes discipline-specific when you populate those objects. There is a way that you can then connect the models to those objects so that if one’s pulled out, you only have that one connection. The models themselves — the modules — don’t ever talk to each other.

That’s one of the approaches that we’ve used. For me, the exciting thing is that we’ve been using this approach to integrate the environmental models, but we haven’t yet done it with the agent-based models. It seems like a perfect environment for us to experiment with this, and that’s why I’d love to talk about how we might collaborate because I think we do have an environment that could make integration easier.

Another area that I want to explore with the panel is the use of GIS. It’s something that gets across this concept of scale. We’re talking about the micro and the macro levels, so we’re talking about the heterogeneity of the agents, but also the heterogeneity of the environment. Some things fall out with different levels of complexity, which are also very scale dependent. When you’re looking at the micro level, things that are happening in your environment — the GIS environment that you might have with these different layers — may be very important at the individual level. However, when you start looking at institutions and governments and you’re coming up at the macro level, many heterogeneity issues of the landscape fall out, because they aren’t as important as the regional- or national-scale level. And there’s also this concept of needing to deal with the micro and macro level geographically because some concepts of scale cause problems. I don’t know if anyone has comments about that problem. I’m interested in Abigail’s [York’s] talk with regard to looking at more macro and government scales.

Abigail York: Abigail York from Indiana University. We’re actually looking at a fairly micro scale. We’re looking at how government programs impact individual behaviors, so in our GIS, we’re covering one county, Currently, our model’s working with two townships. I definitely agree that if we were modeling things at a regional or a national scale, it would be quite complicated. With our current model, we use parcel — individual parcel — data, and it’s fairly detailed. We’re interested in individual decision making and basically local processes. We haven’t had to deal with a larger-scale issue, although in the sense of dealing with individual processes, we’ve had to. We’re working on a number of experiments, for example, to validate how individuals make decisions. We are very interested in that type of validation because it might be possible to make statements about a population distribution at an aggregate level. But when you’re talking about individuals making actual decisions, we’re trying to get experimental data to really understand what’s going on.

Jesse Voss: Jesse Voss from University of Wisconsin, Milwaukee. This question is a two-part question directed at Abigail. The moderator suggested that issues of heterogeneity tend to drop out when we start moving up in scales to …

Sydelko: Just different parts of it. It’s not the same set of heterogeneous things. It’s still heterogeneous, but it may not be the same ones.

Voss: I question that based on an observation that came out in this other paper. It was observed that individuals had differential perceptions of what the meaning of these tax incentives or disincentives were, and these perceptions varied across individuals. Those things both feed upward into the macro-level observables, and there are also downward kinds…[inaudible on tape]. Interaction takes place between those two, and I think that it might be very, very important not to overlook the effects of heterogeneity that exist at the multiple levels within complex systems, no matter what level you're working at. Also, there could be great danger in not doing so because that's actually one of the greatest benefits offered by complex adaptive systems, that is, to show the undeniable importance of those effects at all levels. Would you disagree with that?

Sydelko: Actually, I was focusing more on the ecological side than I was on the agent-modeling side. I was saying, for example, that if we're modeling at a watershed level, I might be using certain kinds of data. If I'm modeling at a more regional setting, I'm not going to be so worried about evapotransporation at the vegetation level, but maybe just one value for an entire watershed. There's this concept of when you go up and down in the scales, you have to prioritize which of these numerous parts of the data sets that you're using in a GIS situation. That's been a real frustration, and I don't have a good framework for knowing how to do that.

I didn't know how much of that is also on the social side. If we're modeling at the township level, is there a certain set of things we're using for our agents that we would not necessarily be using at a higher level? I don't know that much about that side. It would be interesting to find out if people had some thoughts in that area.

Lisa Brouwers: Lisa Brouwers from Stockholm University. In talking about different scales of granularity, I think that a model dealing with these issues needs to be as finely scaled as necessary, but not more. For dealing with flood management issues, we need a hydrological model and that by necessity needs to have small cells in the geographic representation. We're working with 10- × 10-meter cells, but if you're working with other things, say cyclones, you can use larger cells. As you said, maybe you can skip — you didn't say skip, but talking about different ….

I wish I knew the answer to your question. I wish I knew what I wanted in those toolkits, how we could solve the question of how to visualize things at different levels, but I don't. We've been working very hard in this area to find — for even the small region simulated — how to visualize 2,500 households, give the results for those, and show the variability of the outcomes. Some houses are much worse off than others, and you want to point that out. At the same time, we have 11 municipalities in the region, and we want to give an overview of the differences among the municipalities and the aggregate entities. However, we also want to look at the entire region and be able to compare different policies when we are assimilating those. I don't have an answer. It's a very hard question, and I don't think it's an easy fix to plug in a toolkit or something. It's a bigger problem than that.

Rick Riolo: Rick Riolo from the University of Michigan. I'd prefer to see less energy spent on building tools in the toolkit. I'd rather see more energy spent on making the toolkits easy to integrate with all the other packages out there where people are building lots and lots of these tools.

Sydelko: That makes a lot of sense, definitely. Were there other comments? I agree with actually trying to export facilities, particularly to be able to move things out.

Bill Rand: Bill Rand, University of Michigan. Rick, by the way, is one of the co-authors on this paper, so if he has anything to add to any of the other comments that have been made, we're looking to hear something.

York: I don't have a strong background in the computer end of this. I'm one of the empirical social scientists coming to a group. We have a number of computer scientists working with us, and I don't know how a toolkit would necessarily help them, but I agree that it definitely depends on the context of the problem. One of the interesting things that I know they're working on for our model is being able to input a starting point of ownership, parcel boundaries onto a landscape, and use a GIS framework for that instead of using a grid. You've got similar interesting dynamics with that.

Sydelko: That makes sense and that's a good point. Being able to export or connect to existing tools is probably the most important answer to that question.

Claudio Cioffi-Revilla: Claudio Cioffi, George Mason University. The issue of a graphic rendering of complex processes and visualization is extremely important, and I'd like to emphasize the salience that you attach to this issue. It's more than coming up with ideas for interesting graphics though. It requires the development of specialized notation and standards in a way that we're not even familiar with in this area and social sciences.

Keep in mind that notation, for instance in weather maps in meteorology, was not invented overnight, and it also took a substantial organizational foundation to maintain that type of visualization and allow it to evolve to its current form. It's not an arbitrary notation; it has evolved over time for proven things. It corresponds closely between theory and data, so it's a theoretically grounded empirical system of notation or portrayal of empirical data. These are features that we must take into consideration.

It's not sufficient to read a book. It's also important to understand that the most successful graphic notational systems in any specific domain have always been related to theory and to basic understanding. That's something that we need to look at closely. Perhaps in the future, as this community develops methodology and theory components, something like that will come about.

Finally, from an institutional point of view, it would be helpful — perhaps by a future organization of this community, for example — to encourage the development of standards and their validation, maintenance, dissemination, and so on. Even musicians do this better than we do. If you put cello players who have never seen each other together with a quartet, after a few minutes they'll be playing something that is cohesive and pleasant to listen to. They can do this because they have a uniform system of notation that is interpersonal, but that took centuries to evolve. If musicians have done it, I think we should be able to do this as well.

Costas Alexandridas: Costas Alexandridas from Michigan State. I agree. I am working with GIS and agent-based models and see that scale is important, as is notation. When we assume a level two GIS representation and go to level three or start to use notation that sends us into secondary land uses, the dynamics change. We ran such a simulation, and the dynamics were completely different. We then had to agree on understanding what we were talking about, especially with GIS.

Nigel Gilbert: We need standards in notation, but we don't even agree on the concepts and the theories. So without those, I think a standard notation is quite impossible to reach.

Unidentified Speaker: We've been giving a lot of attention to visualization. It's very critical to what we do in intelligence analysis. In reference to the notation that musicians use, it's a language just like any programming language, and that's very different than visualization. We're starting a project to figure out what is required to have an effective visualization.

Visualization has many purposes. If I'm presenting a visualization to a very high level policy maker, say the governor of a state, I may need to use something different than the visualization I show to this community, because the degree of information — the data, the detail, all of this — is very different. Your knowledge allows you to understand more complex visualizations than perhaps I could present to the governor of a state or, for us, to the president of the United States. We have had a lot of trouble with that part of the visualization. We know that in the modeling arena, visualization is going to be very critical for us, and we don't know what makes for effective visualizations. Kathleen talked about some of the socio-cognitive things and the effect of learning.

We do have some experience with visualizations. We've got some very high tech visualization information retrieval systems that are causing analysts to think differently about how they retrieve information compared with the old methods, and they're very uncomfortable with it. We've got to understand that, and it's also part of the agent-based models. Who is our audience? Because identifying the audience is critical, we may be looking at things like who is our audience and what are we trying to get across. As we get into this international society, we might even have to be looking at methods and toolkits and things like that to have research individualization for modeling. What types of visualizations could be most effective in presenting information from modeling?

Brouwers: I think you are right, and that's our experience from our modeling work and communication with policy makers from different disciplines and different professionals. It's very important to have the visualization or the communication of the results within the toolkits and to be able to use it interactively in decision making or policy processes so we can look at output, discuss it, change view, and change perspective. From our point of view, this is good, but how would it look from the government's point of view or from the insurer's point of view? Then we can change some parameters and rerun it, so we have it as a tool instead. I think that, in general for us, it would be very good to have it integrated in the tool.

Sydelko: I think integration was what we meant. When Mike [North] was talking about that he meant still integrated, but not necessarily that we had to build a new one within the system. We could borrow an existing one and make it integrated and dynamic.

Gilbert: To emphasize a separation there, I would argue that we are looking at huge interdisciplinary groups. To expect everyone within those groups to not only be experts in their own subdomain, but also to be experts within the entire realm of visualization techniques, is quite impossible. We need to consider consulting with outside experts and others in this area and see what the best visualization tools are for use within these areas, not necessarily do it ourselves and reinvent the wheel.

Rand: We could borrow from these other disciplines that have people with good experience.

Unidentified Speaker: Yes, I also sense that you need to have, at a minimum, basic visualization capabilities in the models, so you can run them, test them, and get feedback very quickly. At the same time, you probably want to have, as Rick [Riolo] mentioned, more advanced features to either export to or directly integrate with some of the more advanced packages for people who want to do other things.

I think you're right, though, in saying that you have to have something in the model, because as you were saying, as Bill [Rand] was saying, he needed to look at it to see what happened and have it integrated at least at some level. Having a basic capability that's integrated would be very useful.

Fernando Oliveira: Fernando Oliveira from London Business School. I have one question and two comments for everyone on the panel. In listening to your talks, I noticed that all had something in common, that is, the way you define behavior and how your agents behave. You start by assuming that people are not rational and that they have some patterns of behavior. How did you come up with those patterns? How can you validate those patterns?

The comment involves the sensitivity analysis. None of the models, even those that run simulations, had any sensitivity analysis. One of the speakers, perhaps Dr. Verhagen, said that the parameters don't matter. Is that true? Don't we have to test our parameters? This comment also seems to apply to other presentations.

Rand: I disagree with the last statement [about sensitivity analysis]. The entire point of our paper is in some sense sensitivity analysis. Our aim was to find out exactly what system parameters greatly affect the model outcomes. Maybe I'm misunderstanding what you're saying. We developed much more data than we presented here. We presented the things that we thought made the most significant sensitivity changes.

Oliveira: In your model, it is my understanding that you were only having one parameter toward the variance, and …

Rand: No, that's not true. First of all, there's one parameter …

Oliveira: But for the parameter for the variance, I think you only tested with two cases.

Rand: No, that's not true either. We tested multiple cases. We presented one particular case because the other cases didn't matter. It turned out that those cases were statistically insignificant when compared to each other — no variance versus variance. It's a bullying question at that point. As to other parameters within the model, you only have the parameters of whether or not you turn various aspects on within the model. Are you talking about some other parameter testing?

Oliveira: No, it was not in your presentation. It was not obvious that you did statistical testing. It was not obvious that you tested all the other parameters.

Rand: You're right; I didn't cover those details.

Oliveira: I liked your comment about the error in the programming. That's something I think everyone should take into account when talking about either simple or big models. The bigger the model, the greater the opportunity there is to have errors in your parameters.

I also have another comment, this one about the visualization. Everybody is talking about how to visualize or not visualize, and the most important thing, from my experience in doing agent-based work, is that if you have an analytical model, you need to justify the work of your agent-based approach. It's more important to make it critical than to have a visualization toolkit.

York: Your first question was with regard to human behavior. While this area isn't my field, we have psychologists on our team who are doing experiments and looking at learning behavior — reinforcement learning versus hill-climbing learning. These are experiments with people, not agent-based experiments. We're also planning to take these into the field and do more contextual field experiments with actual landowners and how they make decisions.

To give a brief background of what the experiments are about, an agent or the participants are allocating among three different resources. There's also a local maximum and a global maximum, and we're seeing how many agents get to each of those. That's how we're trying to validate our behavior assumptions.

Rand: We also have a sociological group that does a comprehensive survey of the Detroit area, and most of our data will eventually be tied to that. This particular paper is not based on that data because that information dealt more with the model's theoretic questions rather than with how to exactly validate or verify the model against the sociological data. In general, we do look into those questions.

York: In our case, patterns of behavior of individuals are partly based on the interviews we had with people living in the actual area and with well-verified psychological theories that we used with the model. As to the sensitivity analysis and my remark that parameters do not really matter, parts of our parameters were based on real-world data, and the others were educated guesses.

Verhagen: I was referring to the fact that the neighbors get to fight the neighborhoods and that was the guess part. That's where you can make the model more critical if you test it with different types of parameters because you're guessing people will believe more of what you're saying.

Rand: We did use tests, and they didn't really matter. If I had to make a well-educated guess about the parameters we had, I'd say that it didn't really matter if we turned them a bit up or a bit down.

Verhagen: I'd like to add something. In the last experiments presented in our talk, we added a social network. These experiments weren't the ones presented and used in a policy setting earlier. That's why we felt it didn't matter if we used five or four or seven or ten neighbors. Our point was to see if it made a difference to include a social network. Of course, it would be interesting to investigate the size of the social network and the number of neighbors and so on. You can incorporate and change many things in the model and find it very interesting. But that wasn't our purpose now; it was just the first step.

Sydelko: I would like to thank all who participated in the discussion and invite you to stay to listen to Lisa Brouwers provide some additional comments on the model used in their presentation. [Lisa Brouwers then demonstrated the model used in her work with Dr. Verhagen.]

Brouwers: The interface used in our model during the final stakeholder workshop in Hungary is implemented in Met Lab, which perhaps isn’t the most natural choice of language for programming agent-based models, but it was quite efficient when using GIS data. The model, or the simulation, is the one that we presented to the different stakeholders at a recent workshop. The stakeholders were from the insurance companies, the majors, the water management bureaus, environmentalists, NGOs, average people, property owners, and probably others that I can’t remember right now. There were about 20 or 30 people from the region.

People were given the choice of experimenting freely or using the tool to analyze a number of predefined policy scenarios. The analysis permitted people to compare the three policy scenarios that were taken out during the entire project. They could change the time period to five years, the number of times to iterate it, and the flood frequency. For example, if I increase it here, more flood failures will occur, which makes it more interesting. We’ll try to solve it this way to show information without making it too messy. Also, people could compare how many scenarios they want to run simultaneously. Some assumptions can still be made during the predefined scenario, for example, the insurance rate. How many people choose to have insurance? We say 68% to 70%.

I won’t go into details, but I’ll give you a glimpse of how it works. It will take a little while, but you can see the region while the simulation is running. The city was near Aszhasha, where the workshop was held.

Unidentified Speaker: What are the plans now that you’ve had the stakeholder meeting? What’s the next step?

Brouwers: That project is finished, but we’re trying, of course, to start new, similar projects in different countries. The Hungarian Financial Ministry is very interested in the results from this project. We will not leave it, but I don’t know exactly in which direction we will move.

We used this visualization because we wanted to be able to choose perspective. That is the economic outcome from the government for the entire pilot base, and that’s an aggregate for 11 municipalities for insurer companies or for an example individual property owner in the region.

We can stop looking at only one scenario from the insurer perspective, for example. It takes a while because we’ve got some music to it.

We had a lot of information to present, and it’s difficult to do it easily. We chose to split our information into two different areas, with the lower one showing the frequency of the outcomes. During the 1,000 simulated five-year periods, we had 12 different outcomes. Some were repeated many times. In 92% of the times, this happened, and that means no failure occurred because the probabilities that a flood failure would occur were very low. So in 92% of the cases, the insurer gained 2.5 million Hungarian florins. During a five-year period, that would be the premiums they gain, and they don’t pay any compensation. Then you had flood failures or combinations of flood failures during five-year periods of varying degrees of severity.

The most frequent outcomes are here, and the less frequent outcomes are here [referring to the demonstration]. Of course, you can say you would like to add these together or something to present it because it will yield many different outcomes. Still, when looking at simulations of catastrophes, it is the extremes that are interesting; you can't reduce those.

We are also interested in the individual. The number of outcomes has decreased. It's just one property. Most of the time, this property paid premiums but did not receive compensation. When flood failures occurred, this property owner was lucky because he received more than 100% compensation: 50% from the government and 80% from the insurance company, that is, 20% deductible on insurance. You can see different perspectives.

We can compare two or more scenarios to see which is better from an individual point of view. You would choose scenario one. If you look at the government, though, you would choose number two, of course. It doesn't give the policymaker a direct answer, which is overall the best scenario. Further, we know if it is what they want. They might say that this is very nice; it gives me a lot of information, but please, which is the best scenario? That's why we couple this with Decision 3-2, which is probably not installed [so we can't demonstrate it today]. You can see in the first part that you can choose the person who is the most important stakeholder. We could say that the government is equally important as the pilot base and the insured. The individual is not included because it was hard to weight one property owner with regard to the others. So you can change this interactively and see the results. You can then generate the decision tree. Once you have the decision tree, you can evaluate it and perform sensitivity analysis with it afterward. This is what we did and presented during the project.

Unidentified Speaker: When you say you presented it, did you do what you've done now? Did you give the individuals a chance to play with it?

Brouwers: We did, and in fact it was the main use of the tool. At the meeting, the different stakeholders agreed that they preferred the third predefined scenario, but we wanted to modify it. Someone asked if we could get together in three groups and discuss proposing a new policy design that they could all agree on at this time, and they did. Following two hours of very intense debate among them, all in Hungarian, they used this tool where they were able to freely design their own policy strategies and where there is more freedom in setting the number of variables. You can also run the simulations and compare them, and the output will be presented in the same way.

Of course, it's a choice. How many variables or parameters would you want to include in the interface? We could have added more, but it would be too complicated, too complex. So they used this, and they tried, and they increased the levels here, they decreased there, they discussed, and they looked, and they looked from different perspectives, and then they were able to agree on a reasonable policy that was nothing we would have imagined, that is, that they would agree or how the policy structure would look.

[Demonstration ends]

Saturday, October 12, 2002

Invited Speaker:

Lars-Erik Cederman

LEVELS OF COMPLEXITY: ENDOGENIZING AGENT-BASED MODELING

LARS-ERIK CEDERMAN, Harvard University*

ABSTRACT

Agent-based modeling is a computational methodology that allows scientists to create, analyze, and experiment with artificial worlds populated by agents that interact in non-trivial ways and that constitute their own environment. In these “complex adaptive systems,” computation is used to simulate agents’ cognitive processes and behavior in order to explore emergent macro phenomena (i.e., structural patterns that are not reducible to, or even understandable in terms of, properties of the micro-level agents). Such models typically feature local and dispersed interaction rather than centralized control (Resnick, 1994). Moreover, as opposed to conventional rational-choice models that assume either a small number of dissimilar or numerous identical actors, agent-based models normally include large numbers of heterogeneous agents. Rather than studying equilibrium behavior, the focus is often on dynamics and transient trajectories far from equilibrium. Finally, instead of assuming the environment to be fixed, many agent-based models let the agents constitute their own endogenous environment.

INTRODUCTION

The agent-based methodology enables the analyst to explore how social forms are generated (for introductions, see Axelrod, 1997a; Casti, 1997; Epstein and Axtell, 1996). For the present purposes, social forms are defined as configurations of social interactions and actors that together constitute the structures in which they are embedded (cf., Wolff, 1950; Barth, 1981). Computational models differ with respect to what kinds of configurations are endogenized. Whereas some models are limited to the generation of behavioral patterns, others feature a much more profound “rewiring” of social reality, including the context and constitution of the main actors. Based on these principles, four levels of complexity can be identified in ascending order of ontological depth.[1] According to this scheme, social forms can be modeled as follows:

- *Behavioral interaction configurations* constituting patterns of behavioral choices of the micro-level agents, usually in space.
- *Property configurations* constituting arrangements of the agents’ micro-level properties, such as their identities.
- *Interactive networks* constituting dynamic configurations of relations that modify the agents’ ability or inclination to interact with other actors.

* *Corresponding author address*: Lars-Erik Cederman, Harvard University, Weatherhead Center for International Affairs, 1033 Massachusetts Avenue, Room 318A, Cambridge, MA 02138; e-mail: cederman@wcfia.harvard.edu.

[1] Cederman (2001a) offers a similar taxonomy of models where the two last categories are merged into one.

- *Actor structures* constituting social actors' boundaries and internal organization.

Rather than being mutually exclusive, these types of social forms usually appear recursively. For example, models that generate property configurations frequently endogenize behavioral interaction configurations as well. By the same token, models featuring dynamic boundaries often require endogenization of the relevant interaction networks.

As shown by Macy and Willer's (2002) comprehensive review of recent sociological applications, most existing agent-based models treat social forms as behavioral interaction or property configurations, while keeping the interaction topology and the actors' corporate identities fixed. These two first categories correspond to what Macy and Willer refer to as models of "emergent order" and "emergent structure," respectively.

BEHAVIORAL INTERACTION CONFIGURATIONS

Studies explaining behavioral aspects of social systems remain the most active research area in agent-based modeling. Since the path-breaking work by Axelrod (1984), the literature on behavioral interaction patterns has centered on explaining the emergence of cooperation in anarchic settings. As is well known, Axelrod's main result indicates that cooperation is possible in anarchic situations provided that the actors' interactions are iterated. More important, however, the book introduces and popularizes evolutionary thinking in the context of the social sciences. Subsequently, this work has spawned a whole literature on the dynamic conditions of cooperation (for reviews, see Axelrod and Dion, 1988; Macy, 1998; Axelrod, 2000; Hoffmann, 2000; Macy and Willer, 2002).

It should be noted that these cooperation-theoretic models serve other purposes than generating social forms. Instead, the outcome dimension is typically limited to a one-dimensional statistic measuring the level of cooperation in the system (Macy and Willer, 2002). If social forms do emerge, they are usually seen as side effects of this primary goal. Illustrating this point, Axelrod (1984, Chap. 8) devoted an often overlooked chapter of his celebrated book to the "structure of cooperation." Noting that cooperative strategies would stand a better chance to invade noncooperative populations if they appeared in clusters, Axelrod studied the geographic distribution of collaborating agents (see also Cohen, et al., 2001). Although Axelrod's book primarily focuses on growing cooperation in general, the spatial configuration of cooperative strategies exemplifies social forms seen as behavioral interaction configurations. Applying the same logic to democratic cooperation in world politics, Cederman (2001b) illustrates that behavioral interaction configurations can emerge in models that contain the other three categories of social forms.[2]

[2] If one generalizes the spatial logic from geographic locations to strategic space, Lomborg's (1996) study of strategic configurations involving strategy mixes serving as "nuclei" and "shields" could also be classified as one that generates behavioral interaction configurations.

PROPERTY CONFIGURATIONS

The generation of social forms appears to be a more central goal in studies that explore property configurations, such as cultural traits and attitudinal dispositions. Schelling's (1978) classical model of segregation remains the best known example. As we have already seen, this model generates a spatial configuration of actors possessing dichotomous "ethnic" traits. By providing numerous references to similar models, Macy and Willer (2002) demonstrate that this field of research is still very lively. Many models omit Schelling's focus on movement in favor of permanently located social actors influencing each other locally (e.g., Carley, 1991; Mark, 1998). As for behavioral interaction models, researchers typically focus on clustering. Axelrod's (1997b, Chap. 7) well-known culture model, for example, generates distinct spatial clusters of agents with identical traits despite the presence of a homogenizing micro-level mechanism.

Although these two categories of social forms dominate the computational literature, it would be a mistake to believe that they exhaust it. Based on the process-theoretic critique of "substantialism," it is clear that configurative explanations often require a more flexible ontology than that evidenced by the two first categories of social forms. In this sense, one could fault Macy and Willer's (2002) otherwise excellent review for adopting too shallow a notion of "social structure." On the other hand, the limited scope of the review is understandable given the scarcity of attempts to generate deeper social forms. Fortunately, however, computational modeling in no way precludes endogenization of such sociational configurations, although the costs in terms of complexity may discourage some researchers from venturing into this domain. Indeed, endogenous networks and actors are comparatively understudied by computational researchers, but as we will see, they are far from empty.

INTERACTIVE NETWORKS

What does it mean to generate interaction networks as social forms? First, it is necessary to differentiate this research agenda from the behavioral interaction configuration that I listed as the first type of social forms. The distinction hinges on the endogenization of interaction opportunities, or what game theorists refer to as the "game form". Whereas behavioral configurations assume the underlying interaction topology to remain fixed, researchers exploring dynamic networks vary the possibilities of interacting.

Here we limit the scope to studies that explore social forms explicitly, something that normally requires a direct representation of the dynamic network in question even if its generation is not the main purpose of the model (see Skyrms and Pemantle, 2000, for such an analytical network model).[3] Because the first generation of agent-based models stems from cellular automata (Troitzsch, 1997), it is not surprising that the interaction topologies have tended to be predominantly grid-based rather than expressed in terms of networks.

In mathematical sociology, there is a very rich literature on social networks dating back several decades (see e.g., Wellman, 1983). Yet, partly because of the analytical tools employed, much of this scholarly activity has focused almost entirely on static characteristics of network

[3] In an indirect sense, it could be argued that Schelling's (1978) segregation model features this type of entities, for in his framework, the actors' mobility implies that their interaction topology evolves over time. Likewise, some behaviorist interaction models feature exit and other aspects of voluntary partner selection that implicitly produce network structures (see examples in Macy, 1998; Axelrod, 2000; Hoffman, 2000).

structures. To the extent that dynamics are explored at all, tractability constraints force the analysts working with models to assume stationarity and homogeneity (Zeggelink, 1994; see also Sawyer, 2003). However, computational modeling has added new, powerful tools that facilitate the rendering of networks as emergent social forms. Zeggelink (1994) was among the first to rely on agent-based modeling as a way to analyze dynamic friendship networks (see also Conte, et al., 1998). Today, computational toolkits, such as Swarm and RePast, provide standard library modules for modeling of networks, thus reducing the implementational effort.

The dramatic improvements in computational power have also enabled researchers to study larger real-world networks than has previously been the case. These advances have triggered a surge of interest in networks across a great variety of disciplines. In particular, theoretical physicists have made significant progress in developing both analytical results and computational models of large-scale networks (Albert and Barabási, 2002). These explorations have revealed that most empirical cases deviate significantly from classical random graphs, whose links are added entirely randomly and thus do not contain any clusters. In an important paper, Watts and Strogatz (1998) generated a family of realistic “small-world” networks that is characterized by a very high connectivity and clustering thanks to the addition of a small number of long-range links to an otherwise locally connected topology (see also Watts, 1999; for a popular introduction, see Buchanan, 2002).

Although physicists construct small-world configurations by adding links successively, such a process is not meant to model the actual creation of these networks. Rather, the main purpose is to reproduce and explore the static topology of existing, empirical networks. In that respect, the research cannot be characterized as generative even though social forms are replicated. Nevertheless, the puzzling structure of the Internet has given rise to a series of analytical and computational studies that subscribe to a generative logic. To account for the highly skewed connectivity of Web sites, Albert and Barabási (2002) were forced to address the question: “what is the mechanism responsible for the emergence of scale-free networks” (p. 71). It turns out that a very simple model can generate a power-law distributed configuration. Barabási and Albert (1999) created such a network by continuously adding nodes and connecting the new nodes to the previous ones such that the new links exhibit “preferential attachment.” The latter means that new links are added in proportion to each node’s popularity measured as its current number of links.

Obviously far from all empirical networks exhibit power-law distributed connectivity. This property appears to follow from topologies where there is little or no cost in adding new nodes. Where there are constraints due to geography or other factors, the dynamic process of network formation and evolution may well generate emergent configurations characterized by exponential distributions rather than by power laws (Jin, et al., 2001).

ACTOR STRUCTURES

After having considered dynamic interactive networks, it is natural to turn to models that generate the structure of actors as emergent social forms. Though inevitably increasing the complexity of the modeling task, this deepest level of endogenization has been a key element in sociological process theory since Simmel (e.g., Giddens, 1979; Fararo, 1987; Abbott, 1995; Archer, 1995). Unsurprisingly, this is also the least explored type of social form in agent-based studies.

Conventional "substantialist" theories take the constitution of agents as given (Emirbayer, 1997). This assumption is especially common in methodologically individualist approaches, that assume there to be an unchanging number of presocially fixed actors. To follow the principles of sociological process theory, it would be necessary to deviate from these ontological assumptions by generating actors as social forms. This in turn requires that the actors be problematized in terms of their external boundaries and internal structure (see also Axelrod, 1997b, Chap. 6).

How could agent-based modeling circumvent the need to postulate a set of reified actors at the outset of the modeling process? Drawing explicitly on sociological process theory, Abbott (1995) shows how to avoid the trap of anthropomorphic extrapolation from biological individuals. The trick is to focus on potential boundary elements before assuming the presence of the actors to be generated. Typically, such "sites of difference" are formed in a "soup" of micro-level actors: "Previously-constituted actors enter interaction but have no ability to traverse the interaction inviolable. They ford it with difficulty and in it many disappear. What comes out are new actors, new entities, new relations among old parts" (p. 863).

Abbott's scenario dovetails with Simmel's theory of how collective actors emerge in reaction to external threats. Simmel distinguishes between cases in which the cohesion of an already existing group increases as it enters into an antagonistic relationship with another group, and those cases where there was no preexisting group consciousness before conflictual interaction: "Conflict may not only heighten the concentration of an existing unit, radically eliminating all elements which might blur the distinctness of its boundaries against the enemy; it may also bring persons and groups together which have otherwise nothing to do with each other" (Simmel [1908], 1955, pp. 98–99).[4]

Fortunately, these ideas can be translated readily into computational language. In a generic model of ecological morphogenesis called ECHO, John Holland (1995) lets "primitive agents" amalgamate into "multi-agents" through a process of boundary formation where lower-level agents are able to merge into composite entities, thus assuming the status of "agent-compartments" (see also Cederman, 2002). Such an organizational transformation requires that explicit rules of action scope and resource transfer be specified. This is why some type of internal organizational structure is needed to hold together, and coordinate the activities of, the agent-compartments.

Axtell's firm-size model provides a particularly concise and instructive example of this approach to the formation of actors. Drawing on work by Simon and Stanley, Axtell (1999) notes that the size of real-world companies is power-law distributed. Treating this phenomenon as an emergent statistic reflecting a specific social form, his generative approach accounts for the mechanisms producing it. By postulating straightforward rules of joining and leaving firms, Axtell generates scale-free aggregate behavior.

Featuring a similar, though inherently more complicated, logic of organizational genesis, another line of research focuses on properties of state systems in world politics. Already in 1977, Bremer and Mihalka (1977) introduced a model featuring conquest in a hexagonal grid, which was later extended and further explored by Cusack and Stoll (1990). Cederman (1997) introduces

[4] Coser (1964) offers an influential, functionalist interpretation of Simmel's conflict hypothesis that reduces it to a behavioral phenomenon applying to relations among exogenous actors. For a critique of this perspective, see Sylvan and Glassner (1985).

a new generation of models in the same tradition. These models share a common architecture that starts with a territorial grid of fixed and indivisible primitive agents that can be thought of as villages or counties. The states that survive grow and their boundaries expand endogenously through a repeated process of conquest. The resulting states become hierarchical organizations linking capitals to their respective provinces through direct, asymmetric relations of domination. It should be noted that although the agents reside in a grid-based space, the underlying organizing principles presuppose that the actors be organized as a dynamic network, for conquest inevitably changes the list of neighbors. Models of this type can be used to explore dynamic features of competitive geopolitical systems, such as the duration of balance of power (Cusack and Stoll, 1990) and war-size distributions (Cederman, 2003).

The examples covered so far comprise only two-level organizations. However, computational organization theorists have gone beyond this limitation by analyzing multilevel networks explicitly. Whereas most of the literature investigates the properties of fixed social forms, some studies set out to grow organizational structures. In a prominent example, Carley and Svoboda (1996) apply simulated annealing as a way to represent organizational adaptation in terms of restructuring and learning (see also Dittrich, et al., 2000; Carley, 2002).

Do actor configurations generated in these illustrative models exhibit emergence? The answer to this question hinges on which type of emergence is aspired to. If the goal is to generate emergent patterns in the "bottom-up" sense, the question can be answered in the affirmative, at least when it comes to specific instances of the social forms. For example, in Bremer and Mihalka's (1977) model, it is impossible to predict what the states' boundaries are going to look like based on inspection of the micro-level rules. As regards entirely new organizational forms, however, it is less obvious that existing research generates even individualist emergence. Axelrod (1997b, Chap. 6) proposes a "tribute" model of new political actors that may be at least a partial exception to this observation. According to Axelrod's algorithm, collective actors emerge if there is a pattern of interactions that confirms a number of properties, seen to be constitutive of agency. These include effective control over subordinates, collective action, and recognition by third parties that an actor has been formed.

Nevertheless, "intrinsic emergence" of this type calls for explicit representations of organizational forms inside the actors' "internal models," i.e., their cognitive maps of themselves and their environments (Crutchfield, 1994). Because of their relative simplicity, the agents in all of the aforementioned cases fail to live up to this standard. This does not mean that there are no examples of downward causation, however. In a dynamic model of a geopolitical state system, Cederman (2001c; 2002) introduces nations as actor configuration as categorical networks that bring together primitive actors in culturally coded objects. The identity of these nations is implemented with the help of Holland's (1995) notion of schema that enables the actors to highlight specific traits as politically relevant while suppressing others. If belonging to a nation, a state agent modifies its behavior to incorporate nationalist secession, unification, and irredentism. Thus, there can be no doubt that the nations exhibit downward causation. To simplify the computational effort, however, the list of nations is kept outside the internal models of the agents, and thus in principle constitute a primitive "blackboard" system, which is not compatible with intrinsic emergence (Sawyer, 1998, p. 59). Nor is nationalism in general, viewed as a new type of organizational configuration, truly emergent in this particular framework.

Whereas there are models that exhibit downward causation, in these cases the causal pattern is not emergent in itself. To date, the computational literature lacks models that "are both micro-to-macro and macro-to-micro modeled simultaneously" (Sawyer, 2003). To close this gap, model building must confront the challenge of developing agents with internal models that recognize emergent effects while at the same time predicating their actions on them. All this has to be done without resorting to reification of the emergent patterns themselves. As explained by Sawyer (2002), philosophers refer to such phenomena as "supervenient" in that these "higher-level entities and properties [are] grounded in and determined by the more basic properties of physical matter" without being reducible to the latter. We are still quite far from realizing this ideal in contemporary agent-based frameworks, but richer cognitive models employing "agent communication languages" and some developments in Artificial Life should give us hope that such a project is indeed feasible. For example, Fontana and Buss (1994) synthesize the emergence of life as self-replicating LISP functions that are capable of acting on themselves.

BIBLIOGRAPHY

Abbott, A., 1995, "Things of Boundaries," *Social Research* **62**:857–882.

Albert, R., and A.-L. Barabási, 2002, "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics* **74**:47–97.

Archer, M.S., 1995, *Realist Social Theory: The Morphogenetic Approach*, Cambridge, MA: Cambridge University Press.

Axelrod, R., 1984, *The Evolution of Cooperation*, New York: Basic Books.

Axelrod, R., 1997a, "Advancing the Art of Simulation in the Social Sciences," in *Simulating Social Phenomena*, R. Conte, R. Hegselmann, and P. Terna, eds., Berlin: Springer-Verlag.

Axelrod, R., 1997b, *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration*, Princeton, NJ: Princeton University Press.

Axelrod, R., 2000, "On Six Advances in Cooperation Theory," *Analyse & Kritik* **22**:130–151.

Axelrod, R., and D. Dion, 1988, "The Further Evolution of Cooperation," *Science* **242**:1385–1390.

Axtell, R., 1999, "The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distributions," Brookings Institution CSED Working Paper No. 3; available at http://www.brook.edu/es/dynamics/papers/firms/firms.pdf.

Barabási, A.-L., and R. Albert, 1999, "Emergence of Scaling in Random Networks," *Science* **286**:509–512.

Barth, F., 1981, *Process and Form in Social Life: Selected Essays of Fredrik Barth*, London: Routledge & Kegan Paul.

Bremer, S.A., and M. Mihalka, 1977, "Machiavelli in Machina: Or Politics among Hexagons," in *Problems of World Modeling*, Karl W. Deutsch, ed., Boston, MA: Ballinger.

Buchanan, M., 2002, *Nexus: Small Worlds and the Groundbreaking Science of Networks*, New York: W.W. Norton.

Carley, K.M., 1991, "A Theory of Group Stability," *American Sociological Review* **56**:331–354.

Carley, K.M., 2002, "Computational Organization Science: A New Frontier," *Proceedings of the National Academy of Sciences* **99**(3):7257–7262.

Carley, K.M., and D.M. Svoboda, 1996, "Modeling Organizational Adaptation as a Simulated Annealing Process," *Sociological Methods and Research* **25**:138–168.

Casti, J.L., 1997, *Would-be Worlds: How Simulation Is Changing the Frontiers of Science*, New York: Wiley.

Cederman, L.-E., 1997, *Emergent Actors in World Politics: How States and Nations Develop and Dissolve*, Princeton, NJ: Princeton University Press.

Cederman, L.-E., 2001a, "Agent-based Modeling in Political Science," *The Political Methodologist* **10**:16–22.

Cederman, L.-E., 2001b, "Modeling the Democratic Peace as a Kantian Selection Process," *Journal of Conflict Resolution* **45**:470–502.

Cederman, L.-E., 2001c, "Modeling the Co-evolution of States and Nations," in D. Sallach and T. Wolsko, eds. *Proceedings of the Workshop on Simulation of Social Agents: Architectures and Institutions*, Chicago: Argonne National Laboratory.

Cederman, L.-E., 2002, "Endogenizing Geopolitical Boundaries with Agent-based Modeling," *Proceedings of the National Academy* **99**(3):7796–7303.

Cederman, L.-E., 2003, "Modeling the Size of Wars: From Billiard Balls to Sandpiles," *American Political Science Review* (forthcoming in March).

Cohen, M.D., R.L. Riolo, and R. Axelrod, 2001, "The Role of Social Structure in the Maintenance of Cooperative Regimes," *Rationality and Society* **13**:5–32.

Conte, R., C. Castelfranchi, and V. Veneziano, 1998, "The Computer Simulation of Partnership Formation," *Computational and Mathematical Organization Theory* **4**:293–315.

Coser, L.A., 1964, *The Functions of Social Conflict*, New York: The Free Press.

Crutchfield, J.P., 1994, "Is Anything Ever New? Considering Emergence," Santa Fe Institute, Working Paper 94-03-011.

Cusack, T.R., and R. Stoll, 1990, *Exploring Realpolitik: Probing International Relations Theory with Computer Simulation*, Boulder, CO: Lynnie Rienner.

Dittrich, P., F. Liljeros, A. Soulier, and W. Banzaf, 2000, "Spontaneous Group Formation in the Seceder Model," *Physical Review Letters* **84**:3205–3208.

Emirbayer, M., 1997, “Manifesto for a Relational Sociology,” *American Journal of Sociology* **103**:281–317.

Epstein, J., 1999, “Agent-based Computational Models and Generative Social Science,” *Complexity* **4**:41–60.

Epstein, J.M., and R. Axtell, 1996, *Growing Artificial Societies: Social Science from the Bottom Up*, Cambridge, MA: MIT Press.

Fararo, T.J., 1989, *The Meaning of General Theoretical Sociology: Tradition and Formalization*, Cambridge, MA: Cambridge University Press.

Fontana, W., and L. Buss, 1994, “What Would Be Conserved If 'the Tape Were Played Twice?',” *Proceedings of the National Academy of Sciences* **91**:751–761.

Friedman, M., 1953, *Essays in Positive Economics*, Chicago, IL: Chicago University Press.

Giddens, A., 1979, *Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis*, Berkeley, CA: University of California Press.

Hoffmann, R., 2000, “Twenty Years on: The Evolution of Cooperation Revisited,” *Journal of Artificial Societies and Social Simulation* **3**; available at http://www.soc.surrey.ac.uk/JASSS/3/2/forum/1.html.

Holland, J.H., 1995, *Hidden Order: How Adaptation Builds Complexity*, Reading, MA: Addison-Wesley.

Jin, E.M., M. Girvan, and M.E.J. Newman, 2001, “The Structure of Growing Social Networks,” *Phys. Rev. E* **64**:046132.

Lomborg, B., 1996, “Nucleus and Shield: The Evolution of Social Structure in the Iterated Prisoner's Dilemma,” *American Sociological Review* **61**:278–307.

Macy, M.W., 1998, “Social Order in Artificial Worlds,” *Journal of Artificial Societies and Social Simulation* **1**; available at http://www.soc.surrey.ac.uk/JASSS/1/1/r.html.

Macy, M.W., and Willer, R., 2002, “From Factors to Actors: Computational Sociology and Agent-based Modeling,” *Annual Review of Sociology*.

Mark, N., 1998, “Beyond Individual Differences: Social Differentiation from First Principles,” *American Sociological Review* **63**:309–330.

McCauley, E., W.G. Wilson, and A.M. de Roos, 1993, “Dynamics of Age-structured and Spatially Structured Predator-prey Interactions: Individual-based Models and Population-level Formulations,” *American Naturalist* **142**:412–442.

McMullin, E., 1964, “Two Ideals of Explanation in Natural Science,” in *Midwest Studies in Philosophy*, P.A. French, T.E. Uehling, Jr., and H.K. Wettstein, eds., Minneapolis, MN: University of Minnesota Press.

Miller, R.W., 1987, *Fact and Method: Explanation, Confirmation and Reality in the Natural and the Social Sciences*, Princeton, NJ: Princeton University Press.

Müller, H.J., T. Malsch, and I. Schulz-Schaeffer, 1998, "SOCIONICS: Introduction and Potential," *Journal of Artificial Societies and Social Simulation*; available at http://www.soc.surrey.ac.uk/JASSS/1/3/5.html.

Resnick, M., 1994, *Turtles, Termites, and Traffic Jams*, Cambridge, MA: MIT Press.

Sawyer, R.K., 1998, "Simulating Emergence and Downward Causation in Small Groups," in *Mutli-agent-based Simulation*, S. Moss and P. Davidson, eds., Berlin: Springer-Verlag.

Sawyer, R.K., 2002, "Emergence in Sociology: Contemporary Philosophy of Mind and Some Implications for Sociological Theory," *American Journal of Sociology* **108** (forthcoming in May).

Sawyer, R.K., 2003, "Artificial Societies: Multi Agent Systems and the Micro-macro Link in Sociological Theory," *Sociological Methods and Research* (in press).

Scharpf, F.W., 1997, *Games Real Actors Play: Actor-centered Institutionalism in Policy Research*, Boulder, CO: Westview Press.

Schelling, T.C., 1978, *Micromotives and Macrobehavior*, New York: W.W. Norton.

Simmel, G., 1955, *Conflict and the Web of Group-affiliations*, ed. and transl. Wolff, New York: Free Press.

Skyrms, B., and R. Pemantle, 2000, "A Dynamic Model of Social Network Formation," *Proceedings of the National Academy of Sciences* **97**:9340–9346.

Sylvan, D., and B. Glassner, 1985, *A Rationalist Methodology for the Social Sciences*, Oxford: Basil Blackwell.

Troitzsch, K.G., 1997, "Social Science Simulation: Origins, Prospects, Purposes," in *Simulating Social Phenomena*, R. Conte, R. Hegselmann, and P. Terna, eds., Berlin: Springer-Verlag.

Watts, D.J., 1999, *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton, NJ: Princeton University Press.

Watts, D.J., and S.H. Strogatz, 1998, "Collective Dynamics of 'Small-world' Networks," *Nature* **393**:440–442.

Wellman, B., 1983, "Network Analysis: Some Basic Principles," *Sociological Theory* **1**:155–200.

Wolff, K.H., ed., 1950, *The Sociology of Georg Simmel*, New York: The Free Press.

Zeggelink, E., 1994, "Dynamics of Structure: An Individual Oriented Approach," *Social Networks* **16**:295–333.

DISCUSSION:
LEVELS OF COMPLEXITY:
ENDOGENIZING AGENT-BASED MODELING

A. WENDT, The University of Chicago, Moderator

Alexander Wendt: It's my pleasure to introduce Lars-Erik Cederman who is one of — or perhaps the leading — computational social scientists in my own subpolitical science, which is international relations. It's a pleasure to be here, and we're looking forward to hearing your talk.

Lars-Erik Cederman: Thank you very much for the invitation to give this talk. I want to direct my thanks, in particular, to David Sallach who proposed this idea. It's an honor to be here at the University of Chicago, especially because this paper draws quite heavily on an intellectual tradition that has its American home right here in Chicago. I'm going to be talking about computational models of social forms, advancing generative macro theory, or of added macro to be a bit more precise.

[Presentation by Cederman]

Wendt: I'm an outsider to computational social science, having just become interested in this area about a year ago. I want to thank David Sallach for including me in the discussion, although I feel like a first-year graduate student again, which is a little unsettling. On the other hand, I have some background in social theory, and so in that sense, Lars-Erik's paper was very congenial for me for a couple of reasons.

First, I very much support the effort to link sociological process theory to agent-based modeling. I believe, as does Lars-Erik, it would benefit both fields. My sense of the agent-based modeling literature is that, with a few exceptions, notably recent work by Keith Sawyer, it's been relatively unreflexive about its social-theoretic foundations. As a result, it has tended to fall, almost by default, into an individualistic or bottom-up way of thinking. As Lars-Erik shows, a process theory approach "problematizes" that kind of thinking without completely rejecting its insights.

Second, and partly in virtue of this connection to process theory, the paper's view of agent-based modeling is quite expansive and points toward underworked areas in the literature. I'm referring to Lars-Erik's emphasis on the need to address the construction of actors rather than taking them as given and also raising the question of downward causation, which I'll revisit later. That said, I want to raise six questions or concerns that I had after reading the paper, and then we can take a few minutes for discussion.

The first concern, which is brief, has to do less with agent-based modeling than with Lars-Erik's picture of process theory, which is a very big "tent" in his characterization and includes people from me to Simmel, Whitehead, Giddens, Harré, and so on. I'm in favor of big tents, and these theorists do have important similarities. On the other hand, I think specialists may have doubts about some of their process theory credentials, especially when we start including scientific realists in the tent. After all, scientific realism is very much associated with substantialism — trying to ground explanations at the end of the day on material substances. This is one of the doctrines that process theorists have sought to critique. I don't understand agent-based modeling well enough to know if it too is critical of substantialism. I have my doubts, since in the end, it seems to require positing agentic substances, but I do think the relationship between realism and process theory is more complicated than Lars-Erik suggests.

The second concern is whether process theory offers adequate conceptual resources to capture all the things that Lars-Erik wants to capture. I think process theory is quite good on bottom-up causation and also on the problem of andogenizing actors, but it's less clear to me that it can ground a robust idea of downward causation, which is an important part of the paper's agenda. After all, process theories tend to be very micro oriented and are often criticized for being insufficiently structural. To my mind, a robust idea of downward causation fits much more naturally with a more structuralist view of social theory in which social structure has an ontological status that is less tied to the processes by which it is reproduced. In that context, Keith Sawyer's work on downward causation is instructive, because in that work Durkheim rather than Mead is more the touchstone, and Durkheim is hard to classify as a process theorist.

This comment relates to my third question, which is whether process theory, specifically, agent-based modeling, can capture what I see as one of the most fundamental features of human social life, which differentiates it from the behavior of particles, insects, or animals. That is, that many of the properties of agents are irreducibly relational rather than intrinsic. The classic example here is Hegel's master and slave, neither of which can be what it is without a relationship to the other. This is not a causal relationship. Masters don't *cause* slaves to be — well, maybe masters do cause slaves to be slaves — but slaves don't cause masters to be masters. This is a constitutive relationship. The property of being a master is constituted or defined by the relationship to a slave. So it's a downward constitution rather than a downward causation. To that extent, it's the relationship or the social whole that is ontologically primitive and not agents, so it's a more holistic principle than an individualistic one.

So I guess this third question is, can such a holistic idea be reconciled with process theory or agent-based modeling? Perhaps it can with process theory, although most process theory tends to be more causal than constitutive, but it's less clear to me that a social holism fits at all with an agent-based modeling perspective. After all, agent-based modeling is *agent*-based, not relationship-based or structure-based, and that would seem to lead inevitably to a bracketing off of the relational sources of some agent properties.

I don't take that to be a criticism of agent-based modeling, since no approach can do everything, but more as a possible limit of the approach as a basis for a general social theory. However, it does point to an ambiguity in the title of this conference, Social Agents. Namely, does the word "social" refer to the fact that agents interact socially, but are not inherently social, or to something deeper, namely, that the agents are intrinsically social creatures?

This in turn relates to a fourth question, which has to do with the issue of emergence. It seems that the relational constitution of agent properties suggests that there is a further kind of emergence beyond those that Nigel Gilbert and Lars-Erik talked about — an emergence in which agents lose some or all of their individuality by virtue of the relationship in which they're embedded. In contrast, as I understand it, agent-based modeling seems to assume that even when we're talking about emergence, the agents from which emergence emerges have an individuality or an identity that is independent of the emergent whole.

There was a neat article a few years back by Paul Humphreys, arguing that this idea of agents is somehow prior to the whole. It does not involve emergence at all but is something more like supervenience, according to which macro-level properties are not reducible to micro properties. So that's sort of the sense in which Lars-Erik was talking about emergence. But nonetheless, the micro-level properties are assumed to exist prior to or independent of the larger social whole.

Now, whether or not Humphreys is correct about the definition of emergence, I think it may be useful to think about the question of the relationship between supervenience and emergence in discussing these issues. Parenthetically, Humphreys argues that one only gets genuine emergence — namely, where the units lose their individuality — in quantum mechanics. I'm not sure that's right, because I suggested earlier that a lot of agent properties exist only in relation to other agents. So in a sense, in holistic social relationships agents also have lost some of their individuality, and perhaps something like Humphreys' idea is going on in the social realm as well.

A fifth question comes to mind, which didn't really come up in Lars-Erik's talk, but I think is interesting as I read the paper. This question relates to a possible tension between process theory and agent-based modeling, and that concerns the place of consciousness or subjectivity in agent-based modeling.

Much of what I understand to be process theory — and I'm thinking especially of people like Mead or Whitehead, although I'm not so sure about Simmel — much of this literature is very much concerned with consciousness. By this they mean, and I mean, not just the ability to perform certain cognitive functions, like conceptualize, recognize, and communicate, such as in Nigel Gilbert's talk, but the subjective experience of what it's like to be a particular kind of agent. In other words, many process theorists take a phenomenological view of consciousness and subjectivity as opposed to a functionalist view.

Agent-based modeling seems, though, to be more functionalist in its treatment of subjectivity or agency. If an entity can conceptualize, recognize, and communicate, then it is an agent. Okay? Perhaps that's all we need from our agents, but I wonder. After all, a machine could, in principle, perform these functions, as could what philosophers have taken to calling zombies, which are beings just like ourselves, but they lack subjective experience. I'm not sure that machines or zombies are agents. So my question is, are the agents in agent-based modeling zombies and does that matter? I'm not sure it does matter, but given that process theorists are talking about people rather than zombies, if we're going to connect these two literatures it may be worthwhile to think about this issue.

Finally, I want to raise a question that Lars-Erik does not address but relates to the larger issue of the relationship between social theory and agent-based modeling, and which has been on my mind lately. One of the things that strikes me as I work to familiarize myself with this literature is that many of these models seem to be end-directed, in the sense that their dynamics converge on stable end states or attractors. Moreover, as we saw in Nigel Gilbert's talk yesterday and again in Lars-Erik's talk, this idea of downward causation and emergence are invoked to … [inaudible on tape] and in a variety of fields.

So my concluding question is whether, or to what extent, an explanation in agent-based modeling is teleological. I don't know the answer to that question, but I'd welcome any suggestions that you might have.

That's the end of my comments. We have about 10 minutes for questions, if anybody wants to weigh in on any of this.

Kathleen Carley: Kathleen Carley, Carnegie Mellon University. I was struck by your comment about agent systems focusing on agents and not dealing with structure or process. It seems that one of the midpoints of multi-agent systems is in part the fact that you've got agents that have internal processes, and it is those processes that lead them to interact. And through that interaction, structure emerges in terms of the structure between the groups. This approach doesn't leave that out, and, in fact, many of the new models are precisely dynamic network models because of what's going on inside the agents. I'm also curious about the distinction.

Wendt: I didn't mean to suggest that this literature is not concerned with structure. It's more the way that it thinks about structure such as the character of the agents, their identity, and their properties seem to be something that is not intrinsically tied to structure. Structure is created by their interactions, and then structure shapes their behavior. The qualities that those agents have are not defined in relation to structure. That may not be correct, but that's how I see this literature.

Unidentified Speaker: In some new models, they're calling these agents multi-level agents. So you have individual agents and organizational agents — the same thing — and you get levels of structure.

John Padgett: John Padgett from the University of Chicago. I want to follow up on Kathleen Carley's question, because, in effect, I'm extremely sympathetic with your critique. I've always been nervous at the name *agent-based* and what's implied by that, precisely because I, like you and Kathleen, am very interested in images of human action that see the interactions or the networks themselves as the generative sources of behavior rather than the agents being the generative sources of behavior.

My version of Kathleen's answer is that, at the level of technology, at the level of object-oriented code writing, I think it is possible, as Kathleen says, to think of agents as nothing more than holding bins for series of rules. The series of rules — the program that produces the action — is not necessarily located within an agent but could be distributed over agents. Agents, in the sense of this holding bin, could simply be one little component of the action-producing program. In the technology sense, there might be room for responding to your type of concerns by having distributed programming rather than centrally located programming. That requires agents as a bin rather than agents as sources of action. That's at the technology level. I think it's possible to answer that.

The problem is that we still have this word "agent." The linguistic label agent rubs directly against that sort of reasoning, and the word agent implies *agency*, which implies that the bin is the source of the action, is the agent, and so forth. You're basically picking up on a contradiction that needs some highlighting by the fact that our linguistic labels are very much buying into the problems that you're talking about, even though with the technology it is possible to incorporate some of your ideas more generally.

I see how that works at the level of behavior, the level of programs, that produces action, behaviorism. Of course, you're also interested in linguistic consciousness issues. I'm not speaking to that, but in principle, although I have no idea how it could work, you could imagine a communicative theory that is distributed rather than located in the brain cells, in the skull, and so forth. So technologically I can see how it could move in your direction, but there is a relatively deep linguistic connotation to the community of agent-based modeling that creates a bit of a blind spot against what you're talking about.

Wendt: Your comment suggests that within the social theory literature, polists and individualists have very different conceptions of agency in human social life. Given that much of this literature seems to be quite explicitly individualistic, there seems to be some tension in the community regarding how we should think about what agents are.

Unidentified Speaker: I really enjoyed both of the talks, and I agree that in sociological theory, there is more of a debate between individualism and collectivism, or individualism and holism. In the modeling community, there is less of that debate, and so when you come from the perspective of the sociological theory debate and enter the multi-agent community, everyone seems to be on the individualist side. That's my impression as well.

What often is not mentioned is the history of agent modeling. People did modeling in simulation long before we had agents, and it was system dynamics modeling or macro-sociological or macro-economic modeling. Many of us are nervous when people talk about modeling wholes or modeling collectives. It starts to look like the bad old stuff, the macro modeling or multi-level modeling where you're actually explicitly representing macro variables as well as some lower-level agents or individuals. Something like that might be necessary, however, to have the system dynamics and the agent-based modeling. I think there's a resistance to going above the level of the agent, although you see it in some areas.

Unidentified Speaker: I'd like to make a comment to help clarify things, based on 25 years of experience in classical mathematical and statistical work. What clarifies it for me is that with cases, you do statistics; with variables you do classical mathematical modeling. It is through objects that you engage in the third way of science. If we think in terms of objects and of agents as an instantiation, of a specification of objects, to me that clarifies a great deal, because the world of objects is very different from the world of variables or cases in doing scientific work. It's not only technically different; it's a different way of looking at how the world works and moving from a state space that is built on variables — for example, in way number two to a world where variables also play a role — but it's a subservient role to the basic building blocks, which are objects, and these objects are constituted by attributes and relations and methods among them and so forth. Once you get past that point, the landscape looks more logical, cohesive, reassuring, and also presents a much more natural way of modeling social interactions and complex processes. It keeps coming back to this basic notion of objects in the third way of looking at things.

David Sallach: David Sallach, University of Chicago. I want to respond specifically to the point that Lars-Erik made about the necessity of having internal models. Keith [Sawyer] made a similar point in another context. I want to suggest a cautionary note: a lot of insight has come out in recent years about community of practice — deeply situated community of practice — where the concepts or models that arise reflexively are something that's achieved; something occurs sometimes. But it's something that actually has to arise out of the community of practice, a lot of which is tacit, a lot of which is not under continual examination. I think there's a danger if we make that the definition of emergence or a major component of the definition of emergence, so that we may move again toward a straightforward "cognitivism" and lose some of the insights that have come out of the situated definitions of the interaction framework.

Nigel Gilbert: Nigel Gilbert from University of Surrey. I think we all agree that Lars-Erik's paper is an important paper for moving us on. I want to make a point about where we might move on *to*, in a sense, because I think there is a danger, there could be a danger, with this type of paper, that it is seen as counter-posing, on the hand, social theory, and on the other hand, agent-based modeling, and then saying they're doing two things that are slightly different or that they are in a sense in competition with each other and things of this kind.

I'd like to see what Lars-Erik is pointing out: that social theorists, or the generative social theorists, are doing that is the same thing as agent-based modeling. If we think about that and its implications, it follows that agent-based modeling *is* social theory. That's not a big point, but it's a matter, as it were, of presentation of what we are doing. It is, of course, all sorts of other things as well. I'm not saying it's *only* social theory, theorizing, but I strongly believe that agent-based modeling is a way of doing social theory. It isn't something separate from or informing social theory; it *is* doing social theory.

Cederman: Obviously, I wholeheartedly agree with Nigel that the ultimate goal is a merger of the two, but that's more an aspiration than a reality. I'm working in that direction. In fact, I don't think of agent-based modeling as merely a method. I have a certain idea of how the social world is constituted. That was the fact and that was the situation when I started writing my dissertation. At that time, I was working with rational choice theory and then the Wall came down, and the Cold War ended, and suddenly everything was in flux, and boundaries were changing, and I couldn't make headway. That's why I turned to agent-based modeling. But that's a secondary move reflecting a certain worldview that I seem to have, as opposed to many of my colleagues, in the U.S. especially. So I agree with you entirely.

Just briefly on Alex's very thoughtful, and partly provocative, comments. I can't respond to them all; I don't want to take up that much time. But certainly the tent is probably a bit too big in the paper, and I need the help of Alex and other experts on social theory and philosophers of science to make sure it's not too spacious. The paper looks a little like a smorgasbord right now on that side, possibly reflecting my national identity.

Now I think you can combine things. Certainly, you don't have to buy into all of the substantialist assumptions of scientific realism to pick up some of the really good things about social realism. I don't think they are completely hardwired. I need to do a better job showing how that's possible because it's not obvious.

Concerning the internal models, David's point is well taken, and I didn't necessarily want to move in the direction of cognitive psychology. We're talking about shared internal models. Exactly how that's going to be done technically is a tricky problem. Ultimately, this notion of intrinsic emergence depends fundamentally on the pattern-recognizing capability that we are endowed with as human beings, something that emerged in evolution. It turned out to be very good to connect all those signals we get from the neurons to attacking tigers or whatever it may be. By the same token, the same kind of hardware can be used to detect social movements or groups of people, and you name it. As long as our agents are not equipped with that kind of hardware, we're not going to make much progress toward modeling downward causation.

At the same time, I think possibly Keith [Sawyer] is a bit too harsh in the criticism, calling agent-based modeling "reductionist." That seems to gloss over a very important distinction between the truly reductionist paradigms, like rational choice theory, for instance, and the computational paradigm because we have — downward causation very often — path dependence, and the aggregate outcomes feed back into the reality of the agents, even if it's not processed at the symbolic level by the agents. In that sense, there is a distinction that we need to insist on. We have made some progress compared to the truly reductionist paradigms, but it seems to be that either we have downward causation, but not really emergence, or we have emergence phenomena, but they don't really exert downward causation. What I'm trying to push for is the combination of both at the same time and that would be very nice. But we're not there yet, clearly.

Therefore, I agree with Alex that this paper is not saying that we are doing what Simmel did 100 years ago, but just better and more precisely. This paper is pointing to gaps and challenges, and Alex did a wonderful job pointing out precisely —much better than I did in the paper — where those challenges are. I think the tone of your comments may be too pessimistic. Ultimately, all that you're asking for is going to be possible, and so I end on that note of optimism.

Electricity Applications

SIMULATING ENERGY MARKETS AND INFRASTRUCTURE INTERDEPENDENCIES WITH AGENT-BASED MODELS

C.M. MACAL, Argonne National Laboratory, Argonne, IL*
M.J. NORTH, Argonne National Laboratory, Argonne, IL

ABSTRACT

National infrastructure systems are becoming more complex and interdependent. Markets and industries for electric power, natural gas, petroleum, and telecommunications are examples of physical network infrastructures and markets that are undergoing rapid evolution. For example, electric power markets, which have pioneered the transition from a regulated monopolistic system to decentralized open markets, have faced many challenges. The continuing restructuring of the natural gas industry is another example. This paper explores the use of agent-based modeling methodologies to simulate interactions among the interdependent infrastructures, focusing on the electric power and natural gas systems. Aspects of modeling infrastructure agent behaviors include agents' selection of objectives, pricing and bidding strategies, learning and adaptation regarding market evolution, and capacity expansion decisions. Modeling the decision processes and actions of the individual agents (e.g., natural gas suppliers, transmission companies, and independent power producers) is informed by approaches to modeling agent behavior that are being taken in the computational social sciences.

1 INTRODUCTION

National infrastructure systems are becoming more complex and interdependent. Electric power, natural gas, petroleum, and telecommunications are examples of physical infrastructures and markets that are undergoing rapid evolution. For example, electric power markets that have attempted the transition from a regulated monopolistic system to decentralized open markets have faced many challenges. The restructuring of the natural gas industry is another example.

As the national infrastructures become more competitive and are squeezed to maximize efficiencies, as safety margins narrow, and as systems approach their design limitations, infrastructures are becoming more physically and economically interdependent. Recently, breakdowns in the infrastructure markets and systems have become the object of the public's attention. The California electricity crisis and the natural gas price spike of December 2000 are examples. These incidents have the potential to create ripple effects in other infrastructures and raise important questions concerning the extent of infrastructure interdependencies, such as:

- Is it possible to quantify the physical and economic interdependencies among the infrastructures?
- How long does it take for disruptions, whether physical or economic, in one infrastructure to propagate through another infrastructure?

* *Corresponding author address*: Charles M. Macal, Decision and Information Sciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439; email: macal@anl.gov.

- Under what conditions or system states could amplification occur in which disruptions in one infrastructure propagate through other infrastructures, thereby leading to unstable behavior?

- How will the infrastructures adapt or adjust to shocks and disruptions, both physical and economic?

- More specifically, will the electric power and natural gas industries co-evolve in such a way as to increase interdependencies? What are the implications for the stability of both systems?

These questions are extremely difficult to address using traditional modeling and simulation approaches. However, agent-based simulation is a natural approach to simulating the dynamics and diversity of agents within interdependent infrastructures. In this regard, there is a natural connection to the social sciences through the representation of behaviors of individuals and organizational structures. As Thomas, et al. (2002) observe:

> The rules of business are at least as important as the rules of physics when it comes to the generation, sale, and delivery of electrical power, for example, as well as the other infrastructures. The decision-making behavior of firms in an industry and the financial vehicles that allow a utility to exist and conduct business are crucial to gaining an understanding of the system evolution.

Modeling the decision processes and actions of the individual agents (electric power generation companies, natural gas suppliers, transmission companies, independent power traders, and others) involved in the operation and use of the commodities provided by the infrastructures presents a formidable challenge. This paper explores the use of agent-based modeling methodologies to simulate interactions among the interdependent infrastructures, focusing on the electric power and natural gas systems. Section 2 describes the salient features of the electric power and natural gas systems for modeling these industries in an agent simulation framework. Applicable notions from agent simulation and computational social sciences are discussed in Section 3. Section 4 presents an agent-based simulation approach to the analysis of infrastructure interdependencies.

2 THE ELECTRIC POWER AND NATURAL GAS SYSTEMS

2.1 Electric Power

The physical infrastructure of the electric power system comprises several components that generate, transmit, distribute, and utilize electricity (Figure 1) (Sadaat, 1999). Electric power plants consist of one or more generating units of varying sizes that use various fuels. In Illinois, for example, generating units are fueled primarily by uranium (nuclear), coal, natural gas, or oil. Other portions of the United States, especially the Northwest, rely on hydroelectric generation

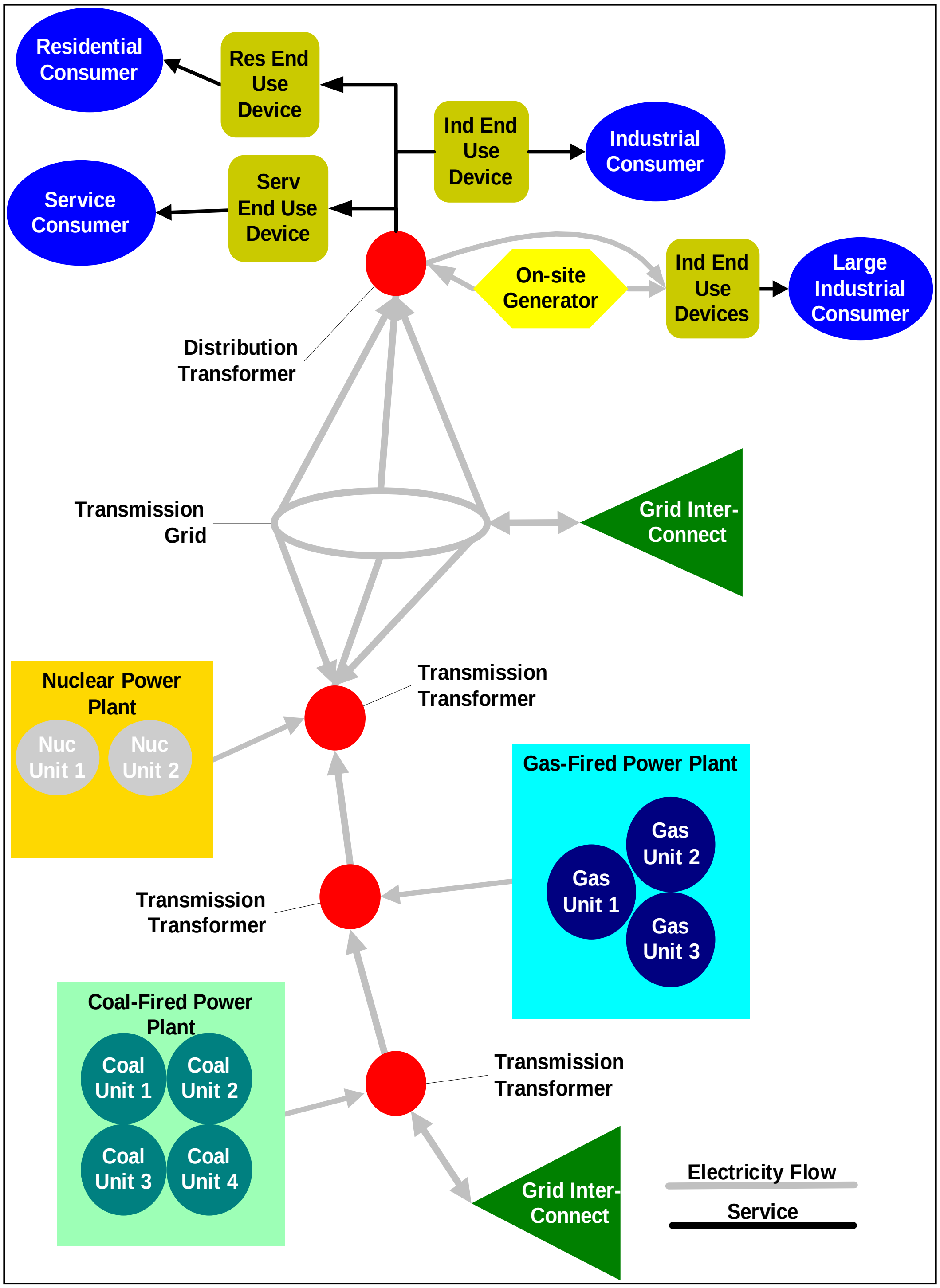

FIGURE 1 Electric Power System

where hydropower resources are abundant. Electricity is transmitted over an electric power network or grid. Generators and distribution subnetworks are connected to the grid at points called buses.

Generators create voltage potential, which is necessary for electric current to flow. Electricity is transmitted over long distances at high voltages to minimize losses. Transmission lines have capacity limits that are largely based on the physical properties of the lines. Transformers increase or decrease voltage levels at various points on the grid, which is necessary for efficient transmission and conversion to voltage levels for final use. The response of the electric grid is almost immediate. When a generator increases its power output, the effects are felt instantly throughout the network.

Electricity demand or load has a characteristic time-dependent profile that varies by sector (e.g., household, service industry, heavy industry). Load varies by hour of the day, day of the week, and season, due to societal and commercial consumption patterns and the weather. A typical electricity load pattern has load that is fairly low throughout the night, increasing from early morning and continuing throughout the day to a peak in the afternoon before declining toward the evening. It is typical for the system load to vary by a factor of 2 between base and peak hours during the course of a single day during the peak production season. Generating units, depending on fuel type, take more or less time to start up or shut down from full production capability, at which point the unit functions most efficiently. Start-up and shutdown costs may be sizable, so it may be desirable to continue to operate units at a loss during periods of reduced load to avoid the added costs of shutting down and restarting the unit. To generate enough electricity to meet the load, generating units can be cycled into production during base, peak, or intermediate demand periods based on the relative operating and fuel costs and response rates of the individual units.

The laws of physics endow the electric power system with some unique properties from an infrastructure point of view. Electricity cannot be stored, generally, but can be converted to other energy forms on a temporary basis; this is not currently done for electric power production on a large scale. Practically all electric power generation is in the form of alternating current. As a generator spins, it creates a voltage level and a corresponding current flow that fluctuates according to a sinusoidal wave. Such a wave is characterized by amplitude and frequency, and frequencies must be synchronized throughout the network to ensure the maximum flow of electricity and minimize energy losses. Increased electrical generation at a node (bus) has the effect of increasing the voltage at the node, which in turn affects the current flowing on *all* links (transmission lines) connected to the node. Electricity cannot be sent from point A to point B in the same way that most other discrete physical commodities are shipped from A to B, or the way that packets are routed over the Internet. The analogy often used for electricity transmission is increasing the flow of water through an interconnected set of pipes by increasing the pressure at any single point.

The physical aspects of the real electric power system are much more detailed and complex than described here. There exist very sophisticated physics-based models that consider all of the salient features of electric power generation and transmission, at least in enough physical detail to plan and operate the electric power grid successfully. At least one research program has been initiated on modeling the physical components of the electric power system in an agent modeling framework (Amin, 2000; Wildberger, 1997).

Figure 2 shows the typical decision-making agents in the electric power industry. Decision making operates in various time scales or decision levels that include everything from hourly unit dispatch to day-ahead, week-ahead, month-ahead, year-ahead, and multiyear time frames. At each decision level, supply agents make decisions regarding the operation of the generating resources they manage and formulate marketing strategies. Different types of markets are available to players at each time scale. For example, these could include markets for bilateral contracts, spot market pool, and ancillary services. Decision-making behavior includes decisions regarding bid pricing for day-ahead power generation and ancillary services markets, bilateral contracts, generating unit scheduling, and long-term capacity expansion. The decision process may be segmented. For example, one type of strategic decision made every day in the electric power industry is on what hourly generation prices to bid into the day-ahead market and each generation unit's schedule for the following day (Wen and David, 2001); another decision is to coordinate these generation bids for the ancillary services (such as reserve) markets (Wen and David, 2002).

2.2 Natural Gas

The physical infrastructure of the natural gas system comprises several components that produce, process, transport, distribute, and use natural gas (Figure 3). Natural gas is extracted from fields by wells, processed to separate gas constituents and remove moisture and impurities, and transported through the interstate pipeline system. Natural gas imports in the form of liquefied natural gas (LNG) are sizable in some parts of the United States, and the processing of LNG is part of the natural gas infrastructure. Natural gas is transported long distances through transmission pipelines. Compressor stations are distributed along pipelines at regular intervals to boost pressure and regulate the flow of gas. Gas is transported cross-country at high pressures and moves at high velocities. Pipelines have capacity limits based on the diameter of the pipe segments. Natural gas is a compressible fluid and, within limits, more gas can be moved through or stored in a pipeline with corresponding pressure increases in a process called line packing. It may take days for gas that is injected into the interstate pipeline system to reach its cross-country destination.

At regional gas markets called hubs, gas is traded and physically routed or wheeled to final regional destinations. When gas reaches a service area, it can be stored in large quantities, typically in underground storage fields (aquifers or abandoned gas fields) for future use. Gas from a transmission pipeline or withdrawn from storage field enters the distribution system through a city gate station that regulates pressure and flow. The gas is then sent through a distribution network that includes a series of regulators that reduce pressure to standard levels appropriate for final consumption. Unlike electricity, natural gas can be readily stored. Gas demand or load has a characteristic time profile or shape that varies by sector (e.g., residential, commercial, and industrial) and mix of loads. Although load varies by hour of the day, and day of the week as in the case of electricity, these fluctuations can be buffered by storage capability and line pack. The main problem facing natural gas supply is that load varies by season and highly depends on weather, which makes forecasting natural gas consumption difficult.

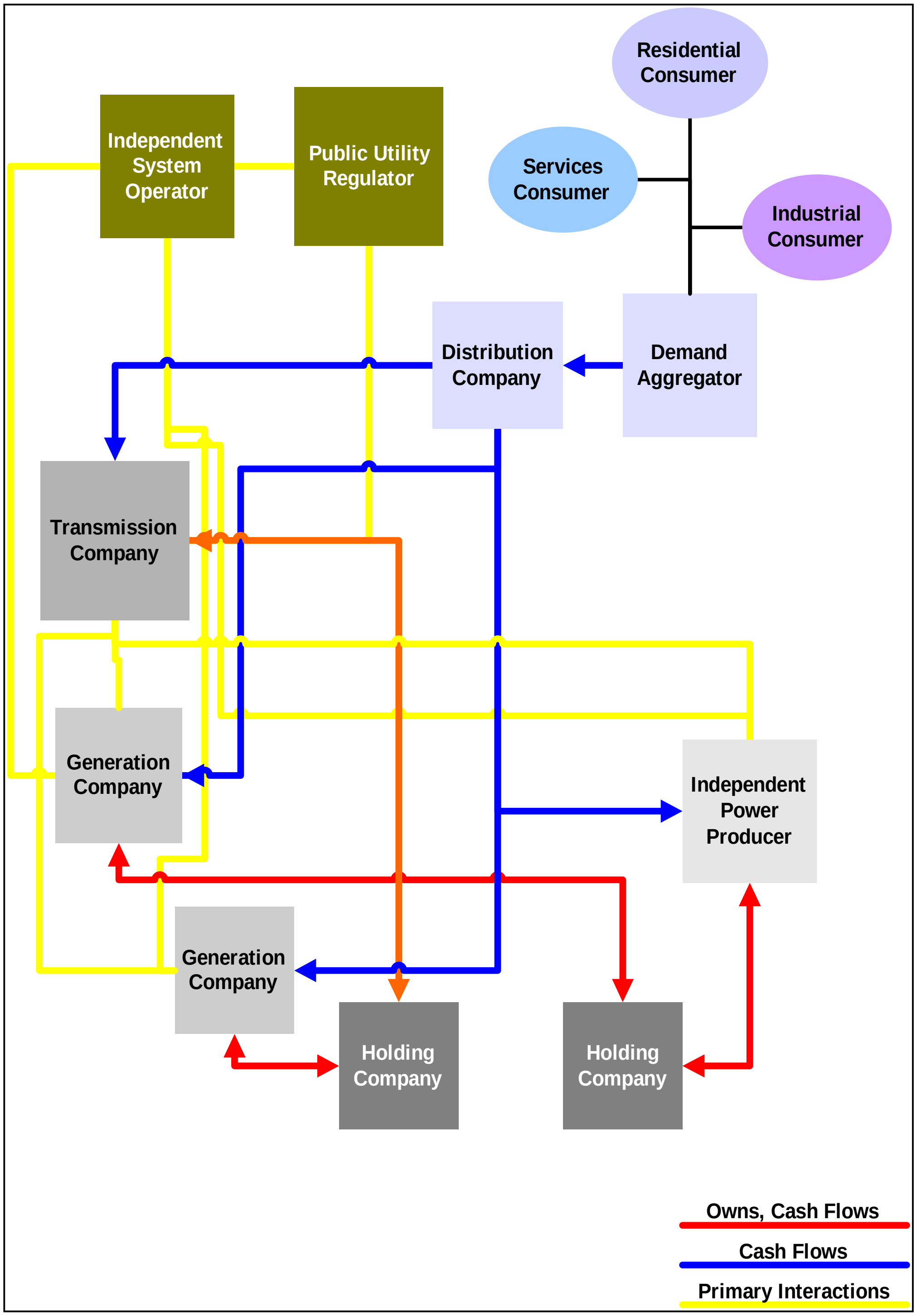

FIGURE 2 Electric Power Industry Decision-making Entities

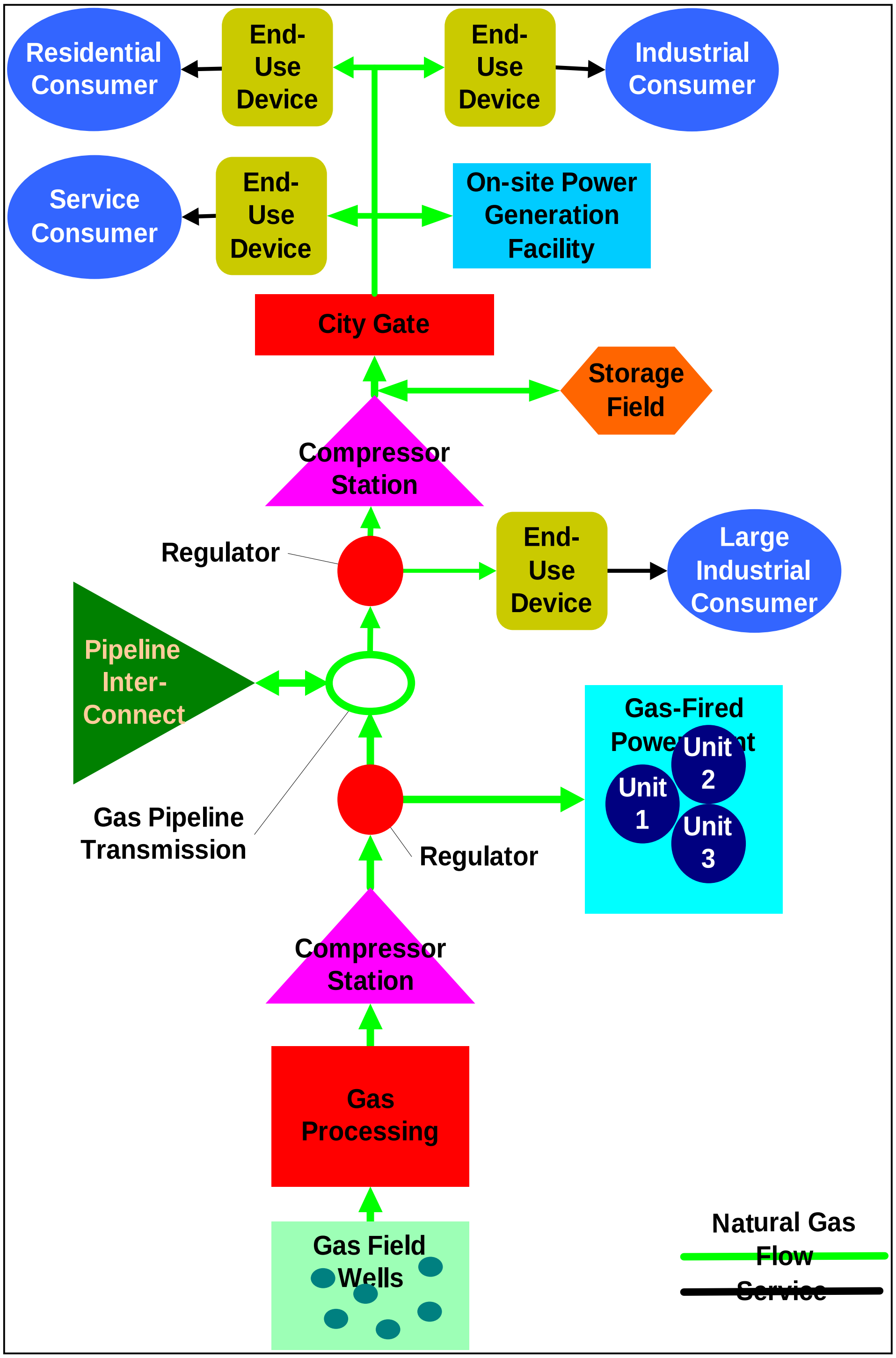

FIGURE 3 Natural Gas System

The physical aspects of the real gas network are much more detailed and complex than described here. Very sophisticated physics-based models consider all of the salient features of gas transport and distribution, at least in sufficient physical detail to plan and operate the system successfully.

Figure 4 shows the typical decision-making agents that make up the natural gas industry. Strategic decisions in the natural gas industry depend on such factors as portfolio of gas supply, storage contracts, and weather effects, primarily seasonal effects (Knowles and Wirick, 1998; Rosenkranz, 1989). The typical decision problem faced by the natural gas supplier consists of maximizing earnings per share, maximizing rate of return on investment, and finding co-rerelated business opportunities that generate a high rate of return. The local distribution company's (LDC's) decision problem centers on gas storage in regions where a large portion of gas supply in the winter months may come from storage. The typical LDC decision problem consists of deciding on the quantity of gas to go into storage in advance of winter, deciding on an acceptable level of risk regarding the severity of next winter's weather, deciding on a storage fill schedule, and deciding on capital investments to improve the situation. For example, capital investments could consider improving the deliverability system, adding compressors or more storage capacity, and possibly extending the pipeline system.

2.3 Market Mechanisms

Electricity markets are undergoing major restructuring in response to deregulation. Electricity market trading is becoming much less tied to the traditional operations-based goals of reliability maximization and cost minimization. Buying and selling of electric power are beginning to resemble the trading of many commodities in both the spot and future markets (Stephenson and Paun, 2001). The natural gas industry also continues to undergo a major restructuring (Leitzinger and Collette, 2002; Economides, et al., 2001). Both markets are expected to be in transition for some time to come.

The selection of particular market rules that will be applied in the deregulation process is a key concern of companies, organizations charged with industry oversight, and industry analysts. Even competitive markets can be set up with significantly different market rules that lead to quite different outcomes. The types of markets that are available and the specific rules under which each market operates influence the decisions made by the market players and the evolution of the industry. For example, Bower and Bunn (2000) compared markets in which all suppliers bid into a common pool to markets in which bilateral mechanisms predominated. Different market rules can create different degrees of market power. A general issue of concern is how to appropriately mix regulation and competition for the restructured energy industry (David, 2001) in a manner that minimizes the potential market power of market participants and minimizes costs to the consumer.

2.4 Interdependencies

The electric power and natural gas markets are undergoing fundamental transformations in the sources of fuel for electric power generators. Large electric generators that use natural gas as a fuel source are rapidly gaining market share, due to their relatively capital construction costs and relatively short construction lead times (1 to 2 years). Many types of gas-fired electric

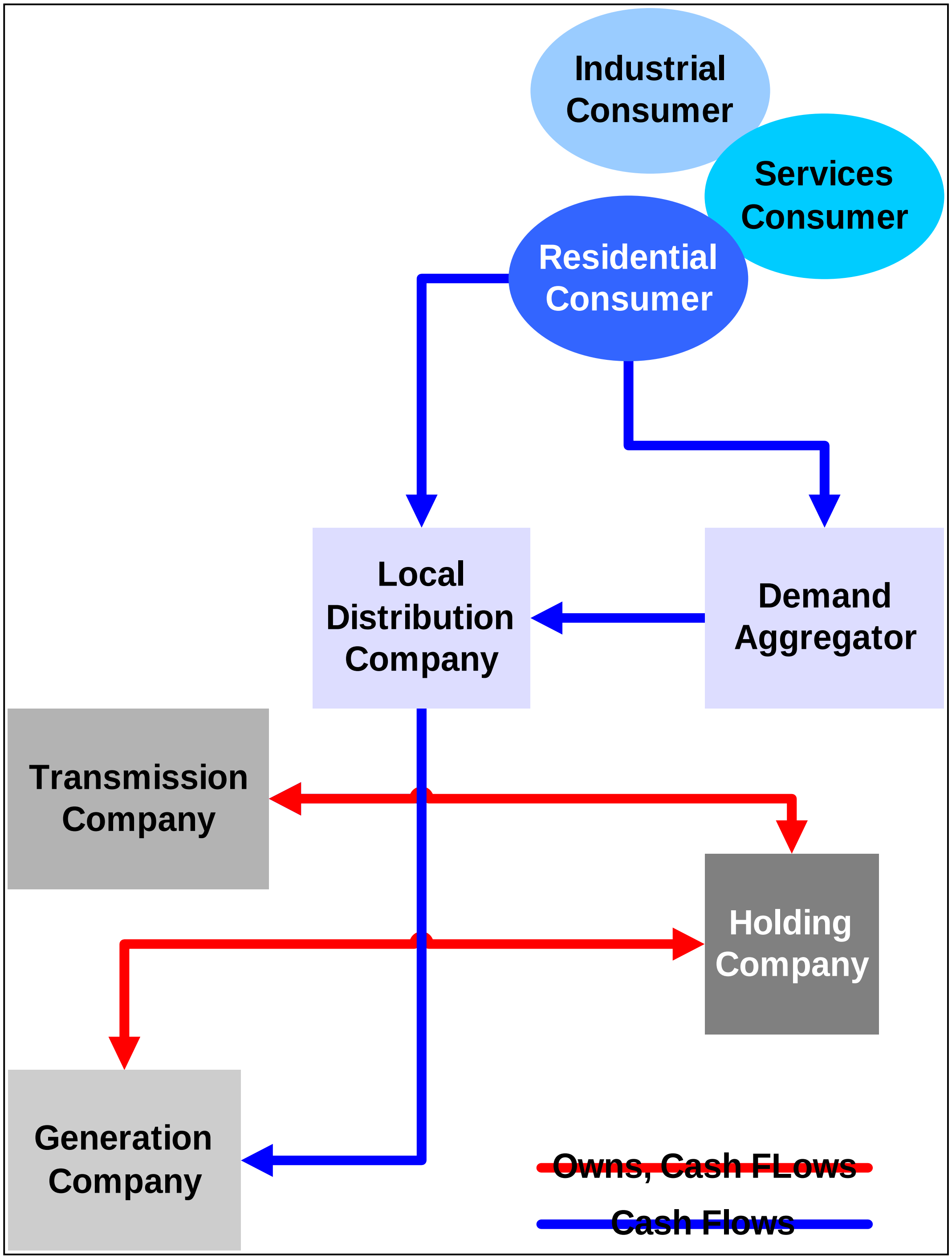

FIGURE 4 Natural Gas Industry Decision-making Entities

generating units can be started up or shut down in a very short time with minimal cost. Small gas-fired units are commonly used to respond to short-term fluctuations in peak electricity load. The recent expansion in the construction and use of gas turbine peaking units for electric power generation, due to technology improvements, favorable economics, and readily available gas supply, has introduced an interdependency between the electric power and natural gas infrastructures of potentially sizable proportions (Anderson, 2000). In addition, the increased use of natural gas for electric power production has led to the observed correlation of electricity and natural gas futures prices over diverse markets (Emery and Liu, 2002).

3 SIMULATION OF THE INFRASTRUCTURE

3.1 Agent Simulation

Several models have been developed for the electric power industry. These systemwide models include economic factors and physical constraints unique to the electric power industry. Fewer system-wide models have been developed for the natural gas industry (Avery, et al., 1992; Bopp and Kanan, 1996; Guldmann and Wang, 1999; MacAvoy and Moshkin, 2000). In both the electric power and natural gas cases, these models are formulated and solved using traditional techniques, such as optimization, in which an organizational objective function is specified and maximized, or discrete event simulation, in which the steps in a dynamic process are modeled. Market outcomes for electric power markets have been modeled using traditional game theoretic equilibrium frameworks (Moitre, 2002; Nguyen and Wong, 2002). Models using traditional simulation and optimization techniques have limitations in addressing questions about the stability, robustness, and interdependent evolution of infrastructure industries, as they lack the capability to model agent adaptation in response to changing economic and physical factors.

The sheer complexity of the types of decisions that need to be made in the electric power and natural gas industries, uncertainty regarding the data upon which the decisions are based, and the short time frames for decisions are all factors that preclude the practicality of formulating or solving in real-time “optimal” decision problems for these industries as a whole. To a large extent, these systems have always been too complex to model adequately. For example, modeling economic markets has often relied on the notions of perfect markets, homogeneous agents, and long-run equilibrium. The need to capture transitory behaviors of the infrastructure in response to disruptions is a key issue in infrastructure interdependency analysis, as evolutionary paths may affect the long-run configuration of the infrastructure.

Agent-based simulation (ABS) offers a promising modeling alternative to capturing and discovering realistic infrastructure behavior as compared to traditional simulation approaches. An ABS consists of a set of agents and a framework for simulating their decisions and interactions. An agent is a self-directed software representation of a decision-making unit. The complexity of an ABS arises from the interaction patterns among the agents. Emergent system behavior is a common result from agent simulations and occurs when the behavior of a system is more complicated than the simple sum of the behavior of its components (Bonabeau, et al., 1999). ABS is related to a variety of other simulation techniques, including discrete event simulation (Law and Kelton, 2000) and distributed artificial intelligence or multiagent systems. Although many traits are shared, ABS is differentiated from these approaches by its focus on achieving “clarity through simplicity” as opposed to maximizing representation detail (Sallach and Macal, 2001). Agents typically are modeled as having bounded rationality, meaning that they make decisions using limited internal decision rules that depend only on imperfect local information. Agent simulation is more amenable to modeling the segmented decision processes as they exist in real infrastructure industries.

Infrastructures lend themselves to structuring agent interaction patterns as networks, which can be readily defined and represented. Modelers are beginning to realize that the topology of networks of interacting agents influences the dynamic behaviors of the network as a whole and therefore the emergent properties of the system. For example, Watts (1999) characterized the topology of electric power system networks for the State of New York and found that it exhibited a scale-free distribution of link-node connectivity, which has considerable implications for reliability and economic features of the infrastructure in general.

3.2 Computational Social Science

New developments in computational environments and modeling toolkits have opened up the possibility of and even created the demand for integrating diverse fields of knowledge and investigation into practical frameworks for modeling real-world problems. Tesfatsion (2002) notes:

> Advances in modeling tools have been enlarging the possibility set for economists…. Researchers can now quantitatively model a wide variety of complex phenomena associated with decentralized market economies, such as inductive learning, imperfect competition, endogenous trade networks formation, and open-ended co-evolution of individual behaviors and economic institutions.

The complex interactions and interdependencies among electricity market participants are much like those studied in game theory (Picker, 1997). Unfortunately, the strategies used by many electric power and natural gas market participants are often too complex to be modeled using standard game theoretic techniques. In particular, the ability of market participants to repeatedly probe markets and adapt their strategies adds complexity.

Computational social science (Epstein and Axtell, 1996), which involves the use of agent simulations to study complex social systems, offers appealing extensions to traditional game theory. Social agents have a behavioral “repertoire” — behaviors they are capable of acting upon. For example, such a behavioral repertoire may consist of reproduction (the ability to form new firms and create larger organization structures, for example, through mergers), resource gathering (revenue generation), vision and perception of the behaviors of other agents (visibility), credit, trade, and cognitive complexity (decision-making sophistication). Behavioral experiments (Erev and Roth, 1998) can motivate candidate heuristics (Sterman, 1987) for modeling complex but generally applicable decision-making behaviors; these heursitics can be tested in agent simulations.

4 AN APPROACH TO AGENT-BASED SIMULATION OF THE INFRASTRUCTURE

Agent-based simulation applications to modeling infrastructure industries are very recent. Special-purpose agent-based simulation tools such as Swarm (Burkhart, et al., 2000), the Recursive Agent Simulation Toolkit (Repast) (Collier and Sallach, 2001), StarLogo, and Ascape are among the most widely used options for implementation of ABS models. A few electricity market ABSs have been constructed, including those created by Bunn and Oliveira (2000), Petrov and Sheblé (2000), and Veselka, et al. (2002). ABS has been applied to analyzing the new electricity trading arrangements for England and Wales (Bunn and Oliveira, 2001). North (2001a) applied ABS to identify infrastructure factors in electric power generation and transmission leading to local price spikes. North (2001b) demonstrated the feasibility of applying agent simulation to quantify the extent of interdependencies between the electric power and natural gas infrastructures. Thomas, et al. (2002) present a conceptual modeling framework for examining infrastructure interdependencies. These models have demonstrated the potential of agent simulations to act as electronic laboratories, or "e-laboratories," suitable for repeated experimentation under controlled conditions.

The ABS approach to infrastructure interdependency analysis consists of representing the physical and behavioral aspects of the infrastructures as a system of highly connected, interacting agents (Figure 5). Agents interact in terms of physical and financial flows and by exchanging information on system performance; key economic parameters are essential to model realistic system operation and adaptation. As an agent is a representation of a decision-making unit, the emphasis on modeling the behavioral components within the infrastructure translates into identifying the primary decision-making processes that are carried out. Each agent has rules of behavior and a decision-making capability that broadly considers salient aspects of the immediate environment and other agents' behaviors. Ideally, organizations could be modeled explicitly as collections of agents that form spontaneously in response to the physical and economic variables in the simulation. The goal of developing the simulation is to monitor and understand the behavior of various system properties (such as reliability and stability), and market issues (e.g., pricing, market share patterns, company profitability, cost recovery).

The physical properties of the electric power and natural gas systems have several implications as to how to structure an agent simulation that includes behavior agents in conjunction with its physical representation. In effect, these physical properties define a topology and create a "landscape" of constraints within which an agent simulation must operate.

4.1 Modeling Issues

Several modeling issues are relevant to the practical application of agent simulation to infrastructure analysis.

4.1.1 Aggregation

Aggregation of the physical and behavioral components of the infrastructure for representation in an agent simulation is necessary to some degree and is especially so in representing multiple infrastructures and their interactions. Aggregation represents a trade-off

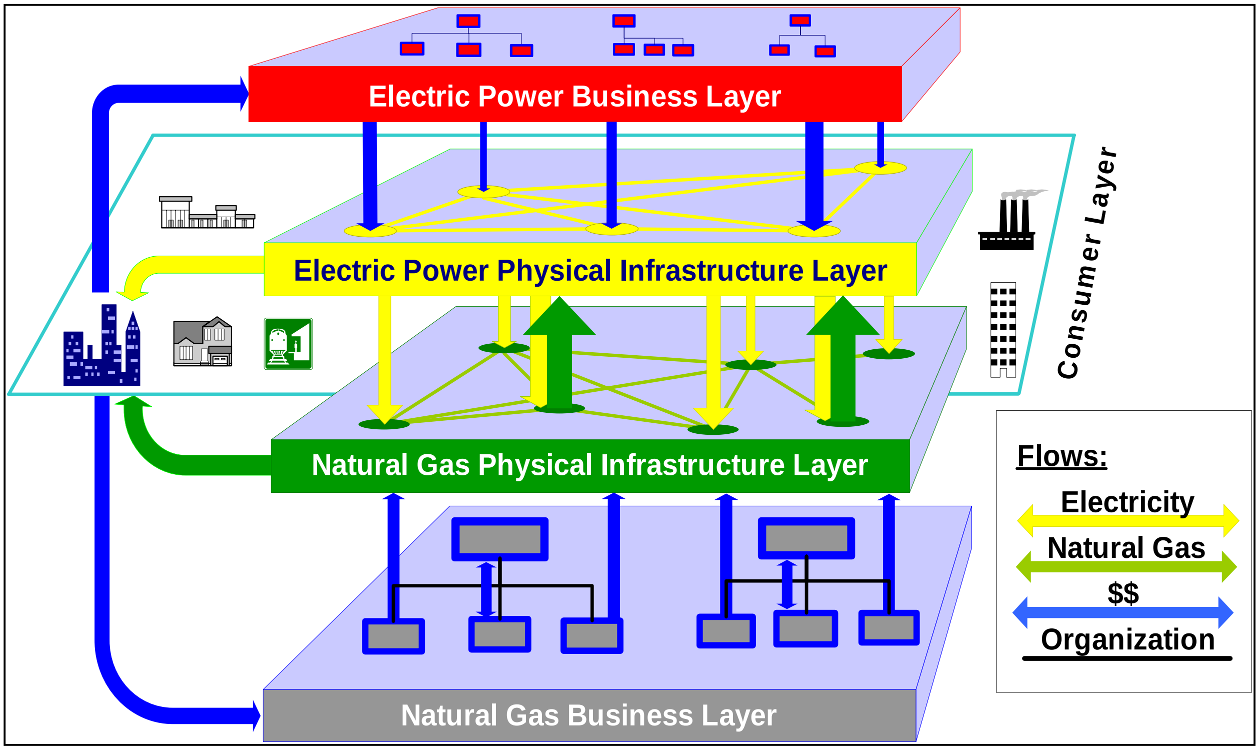

FIGURE 5 Interdependencies between Electric Power and Natural Gas Systems

among various factors such as computational performance and data availability on the one hand and "correspondence credibility" and accuracy on the other hand. Modeling agents at more detail establishes credibility through more direct, one-to-one correspondence between the real system and their representation in the modeled system. Accuracy improves to a point as more detail is included and then levels off. For example, an aggregate model of consumer behavior may be nearly as accurate as the results obtained from a model in which each consumer's behavior is simulated individually. As agents are modeled at greater levels of detail, computational performance degrades, and data requirements become infeasible to satisfy with available and accurate data. The level of aggregation needed for modeling the infrastructure while reasonably satisfying these trade-offs appears to be consistent with the representations shown in Figures 1–4.

4.1.2 Model Collaboration and Consistency

Even if one is able to model the physical infrastructure to the level of detail of individual components in an agent simulation, it is only an approximation to the dynamics of the actual infrastructure for a limited range of operational parameters. Very detailed physical models of regional electric power grids are used to operate the grid, plan transmission, and consider generation expansion. To include these complex physical models within an agent framework along with the decision-making agents is not feasible. Alternatively, the detailed physical models can be used to derive approximations to transmission network transfer capabilities for local neighborhoods in which the agent simulation operates. This entails establishing a close and ongoing "collaboration" between the agent simulation and the physical system model to ensure consistency between physical variables in each of the models. A similar situation exists for the natural gas industry. Detailed physical models for natural gas are based on mass balance relations and pressure and temperature variables.

4.1.3 Data

Of paramount importance is the issue of whether data are available to support the development of an agent simulation infrastructure model at the level of detailed complexity needed for credibility. Much of the data on the physical infrastructure are available from public sources, but compilation and integration of the data are a formidable task. The economic data and financial data on infrastructure markets also appear to be available. Data on the decision-making processes used by individual company agents included in a simulation may, however, be very difficult to acquire. The data problem is largely one of verifying the accuracy of the data and maintaining its currency. Figure 6 shows a preliminary network representation of regional infrastructure data for the natural gas and electric power systems and ownership relations.

4.1.4 Agent Decision Making

Aspects of decision-making behavior included within the scope of modeling infrastructure agent behaviors include agents' objectives and risk preferences, future price expectations, and learning and adaptation in response to simulated market conditions. Each agent has a set of objectives such as maximizing profits, maximizing market share, maximizing capacity utilization, minimizing unserved energy, etc. Objectives may conflict with each other in that improvement in one objective may negate improvement in other objectives. For example, if a generation company agent tries to maximize the capacity factor of a unit at times of low market clearing prices, maximum profits may not be achieved. Each objective of an agent is represented by a minimum expected value, a maximum expected value, and a risk preference. An agent's risk preference is broadly classified as risk-averse, risk-neutral, and risk-seeking and could be modeled using, for example, a von Neumann-Morgenstern expected multiobjective utility function. The overall utility is then computed as the weighted summation of all single-objective utilities. On the demand side, an objective of a demand agent could be minimizing the unserved energy to its customers.

Agents develop price expectations for the markets in which they participate. These expectations are based on a combination of public information available to all market participants, and private information available only to the specific agent. The differing private information available to the agents results in a diversity of price expectations. Initially, the agents have prior price expectations based only on public information (i.e., information on pool prices, system load, reserve margin). Agents may also have differing skills in forecasting the future markets and differ in the historical information available to them on the acceptance and rejection of their own bids. On the basis of results from the simulation, agents update their price expectations using private information on bids that are accepted and rejected and public information that is available to all participants.

An agent learns about market behavior and the actions of other agents based on an exploration process. Agents explore various marketing and bidding strategies and observe the results of their actions. Once a strategy is found that performs well, it is exercised and fine-tuned

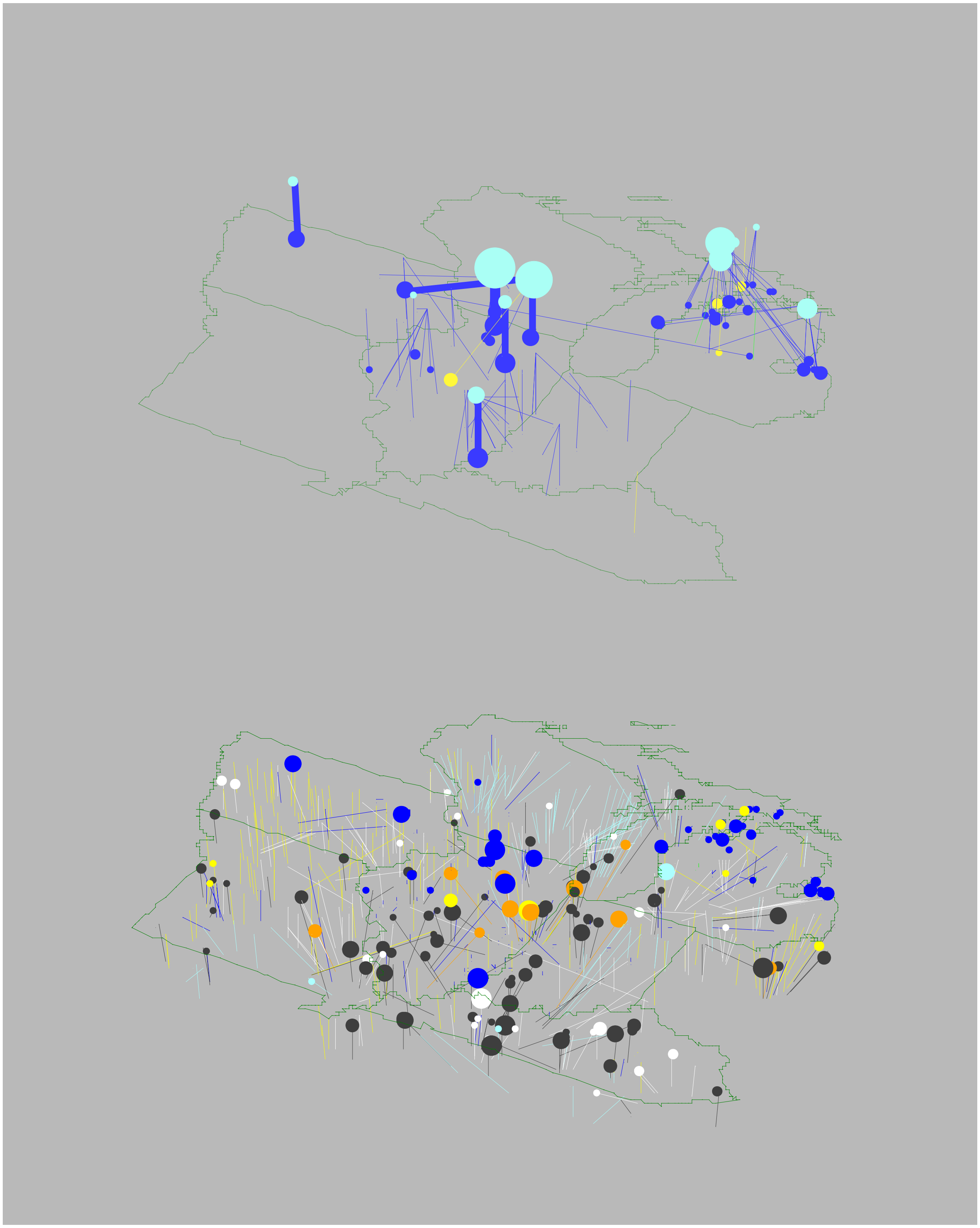

FIGURE 6 Regional Natural Gas (top) and Electric Power Plants and Facilities (bottom) and Ownership Relationships

as subtle changes occur in the marketplace. When more dramatic market changes take place and a strategy begins to fail, an agent more frequently explores new strategies in an attempt to adapt to the dynamic and evolving supply and demand forces in the marketplace. Even when a strategy continues to perform well, an agent periodically explores and evaluates other strategies in its search for one that performs better. Through this process, agents engage in a price discovery process and learn how they may potentially influence the market through their own actions to incrementally increase their utility.

4.2 Prototype Agent Simulation

A preliminary prototype model SMART II+ has been developed to explore infrastructure interdependencies (North, 2001a). The model includes an integrated set of agents and interconnections representing (1) the electric power marketing and transmission infrastructure, (2) the natural gas marketing and transmission infrastructure, and (3) the interconnections between the two infrastructures in the form of natural-gas–fired electric generators. SMART II+ includes two different kinds of market agents — producers and consumers. Agents are connected by a complex physical network of links representing electric power transmission and natural gas transmission systems and nodes representing their transformation and interconnection. Each transmission link and pipeline link has a capacity which limits flow over the network.

Agents are also connected with the physical infrastructure network through ownership and financial relationships. Economic variables in the model consist of investment capital and generation capacity expansion for profitable producers, bankruptcy for noncompetitive organizations, and demand growth for successful consumers. Link capacity constraints and transmission losses have the effect of creating spatially separated regional markets for electricity and natural gas. The electric power infrastructure includes the gas-fired electric generators that buy fuel from the natural gas market. The resulting electricity is then sold in the electric power market.

Key market indicators derived from SMART II+ are market prices, unserved energy (UE) and gas-fired electrical generator market share (MS), as shown in Figure 7. Unserved energy is the energy demand that was not met by the market and represents a form of market failure. (UE is given as a percentage of total energy demand.) Natural-gas–fired electric generator market share is a measure of the electric generation capacity that is supplied by natural gas units and is a key indicator of infrastructure interdependency. Investigation of the interdependencies between the electric power and natural gas markets indicates that natural-gas–fired electrical generators are highly competitive, which causes their market share to rise rapidly. In turn, rising natural-gas–fired electrical generator market share radically increasing market interdependence. Finally, increasing market interdependence pits the electric power and natural gas markets against each other during simultaneous disruptions, since both markets are fighting for the same underlying resource — natural gas — driving up prices for both commodities.

5 SUMMARY AND CONCLUSIONS

Agent-based simulation offers promise for modeling the complexities of interdependent infrastructures, their co-evolution, and response to changing market conditions and physical disruptions. Detailed physical models can be used in collaboration with ABSs that include

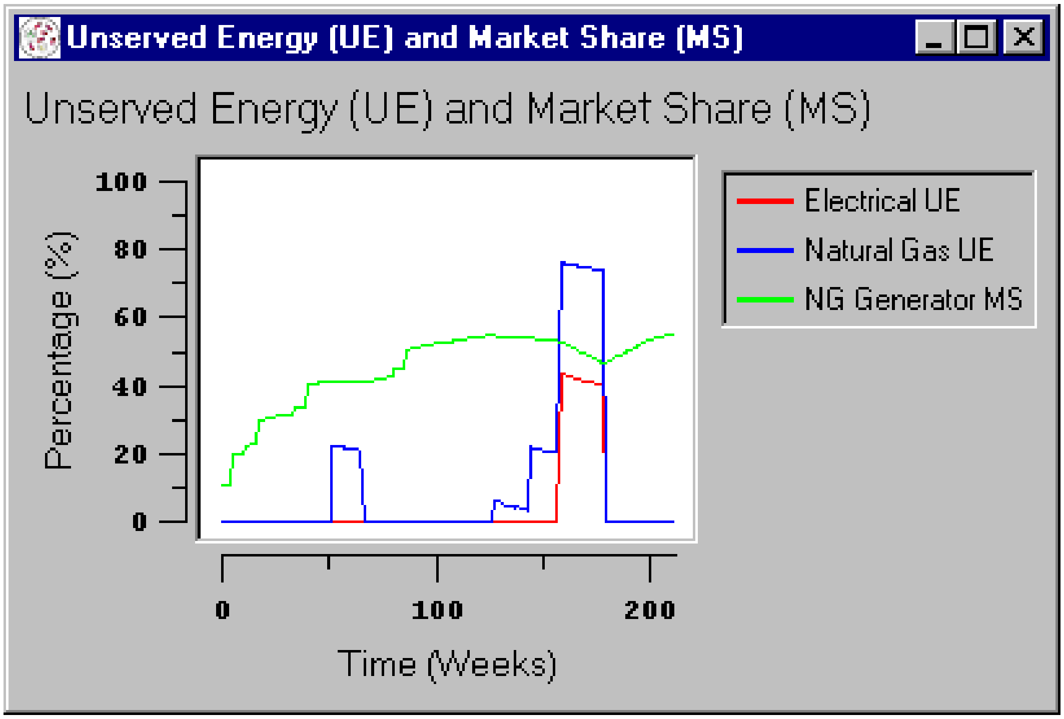

FIGURE 7 Unserved Energy and Natural Gas Generator Market Share in Response to Price Spike

behavioral components and less-detailed representations of the physical systems. This approach provides more information about infrastructure interdependencies than can be produced by traditional simulations.

6 ACKNOWLEDGMENTS

The authors thank Vladimir Koritarov, Richard Cirillo, and Michael McLamore for helpful discussions on various aspects of the infrastructure systems. This work is sponsored by the U.S. Department of Energy under contract W-31-109-ENG-38.

7 REFERENCES

Amin, M., 2000, "Toward Self-healing Infrastructure Systems," *IEEE Computer*, pp. 44–52, Aug.

Anderson, J.A., 2000, "What's Electrifying Natural Gas," *Business Week*, Sep. 19.

Avery, W., G. Brown, J. Rosenkranz, and R.K. Wood, 1992, "Optimization of Purchase, Storage, and Transmission Contracts for Natural Gas Utilities," *Operations Research* **403**:446–462, May-June.

Bonabeau, E.M. Dorigo, and G. Theraulaz, 1999, *Swarm Intelligence: From Natural to Artificial Systems*, Oxford, U.K.: Oxford University Press.

Bopp, A.E., and V.R. Kanan, 1996, “An Optimization Model for Planning Natural Gas Purchases, Transportation, Storage and Delivery,” *Omega International Journal of Management Science* **245**:511–522.

Bower, J., and D. Bunn, 2000, “Model-based Comparison of Pool and Bilateral Markets for Electricity,” *Energy Journal* **21**(3):1–29.

Bunn, D.W., and F.S. Oliveira, 2001, “Agent-based Simulation: An Application to the New Electricity Trading Arrangements of England and Wales,” *IEEE Transactions on Evolutionary Computation* **55**:493–503, Oct.

Burkhart, R., M. Askenazi, and N. Minar, 2000, *Swarm Release Documentation*. Available at http://www.santafe.edu/projects/swarm/swarmdocs/set/set.html.

Collier, N., and D. Sallach, 2001, *RePast*. Available at http://repast.sourceforge.net/.

David, A.K., 2001, “Market Power in Electricity Supply,” *IEEE Transactions on Energy Conversion* **16**(4):352–360, Dec.

Economides, M.J., R.E. Oligney, and A.S. Demarchos, 2001, “Natural Gas: The Revolution Is Coming,” *Journal of Petroleum Technology* **535**:64–72, May.

Emery, G.W., and Quinfeng Liu, 2002, “An Analysis of the Relationship between Electricity and Natural-gas Futures Prices,” *The Journal of Futures Markets* **222**:95–122.

Epstein, J.M., and R. Axtell, 1996, *Growing Artificial Societies: Social Science from the Bottom Up*, Washington, D.C.: Brookings Institution Press.

Erev, I., and A. Roth, 1998, “Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique Mixed Strategy Equilibria,” *American Economic Review* **8**:848–881.

Guldmann, J.-M., and F. Wang, 1999, “Theory and Methodology, Optimizing the Natural Gas Supply Mix of Local Distribution Utilities,” *European Journal of Operational Research* **112**:598–612.

Knowles, T.W., and J.P. Wirick, Jr., 1998, “The Peoples Gas Light and Coke Company Plans Gas Supply,” *Interfaces*, Sept.-Oct.

Law, A.M., and W.D. Kelton, 2000, *Simulation Modeling and Analysis*, 3rd ed., New York, NY: McGraw-Hill Book Company.

Leitzinger, J., and M. Collette, 2002, “A Retrospective Look at Wholesale Gas Industry Restructuring,” *Journal of Regulatory Economics* **211**:79–101.

MacAvoy, P.W., and N.V. Moshkin, 2000, “A New Trend in the Long-term Price of Natural Gas,” *Resource and Energy Economics* **22**:315–338.

Moitre, D., 2002, "Nash Equilibria in Competitive Electric Energy Markets," *Electric Power Systems Research* **60**:153–160.

Nguyen, D.H.M., and K.P. Wong, 2002, "Natural Dynamic Equilibrium and Multiple Equilibria of Competitive Power Markets," *IEEE Proc. Generation, Transmission, Distribution,* **149**(2):133–138, March.

North, M.J., 2001a, "Toward Strength and Stability: Agent-based Modeling of Infrastructure Markets," in *Social Science Computer Review*, pp. 307–323, Thousand Oaks, CA: Sage Publications, Fall.

North, M.J., 2001b, "Technical Note: Multi-agent Social and Organizational Modeling of Electric Power and Natural Gas Markets," *Computational and Mathematical Organization Theory Journal*, pp. 331–337, Norwell, MA: Kluwer Academic Publishers, Dec.

Petrov V., and G.B. Sheblé, 2000, "Power Auctions Bid Generation with Adaptive Agents Using Genetic Programming," *Proceedings of the 2000 North American Power Symposium*, Waterloo, Ontario, Canada: Institute of Electrical and Electronic Engineers, Oct.

Picker R.C., 1997, "Simple Games in a Complex World: A Generative Approach to the Adoption of Norms," *University of Chicago Law Review*, Chicago, IL: University of Chicago.

Rosenkranz, J.S., 1989, "Strategic Gas Supply Decisions," *Pipeline and Gas Journal*, Dec.

Sadaat, H., 1999, *Power System Analysis,* New York: NY: McGraw-Hill Book Company.

Sallach, D.L., and C.M. Macal, 2001, "Introduction: The Simulation of Social Agents," in *Social Science Computer Review*, Thousand Oaks, CA: Sage Publications, Fall.

Stephenson, P., and M. Paun, 2001, "Electricity Market Trading," *Power Engineering Journal* 277–288, Dec.

Sterman, J.D., 1987, "Testing Behavioral Simulation Models by Direct Experiment," *Management Science* **33**(12):1572–1592, Dec.

Tesfatsion, L., 2002, "Agent-based Computational Economics: Growing Economies from the Bottom Up," Artificial Life **81**:55–82, Cambridge, MA: MIT Press, Winter.

Thomas, W.H., M.J. North, C.M. Macal, and J.P. Peerenboom, 2002, "From Physics to Finances: Complex Adaptive Systems Representation of Infrastructure Interdependencies," *Naval Surface Warfare Center Technical Digest*, Naval Surface Warfare Center, Dahlgren, VA, (in press).

Veselka, T., G. Boyd, G. Conzelmann, V. Koritarov, C. Macal, M. North, and P. Thimmapuram, 2002, "Simulating the Behavior of Electricity Markets with an Agent-based Methodology: The Electricity Market Complex Adaptive System EMCAS Model," *International Association for Energy Economists IAEE North American Conference*, Vancouver, B.C., Oct. 6–8.

Watts, D.J., 1999, *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton, NJ: Princeton University Press.

Wen, F.S., and A.K. David, 2001, "Strategic Bidding for Electricity Supply in a Day-ahead Energy Market," *Electric Power Systems Research* **59**:197–206.

Wen, F.S., and A.K. David, 2002, "Coordination of Bidding Strategies in Day-ahead Energy and Spinning Reserve Markets," *Electric Power and Energy Systems* **24**:251–261.

Wildberger, M.W., 1997, "Complex Adaptive Systems: Concepts and Power Industry Applications," *IEEE Control Systems*, pp. 77–88, Dec.

MULTIAGENT MODELING OF ELECTRICITY MARKETS

M. NORTH,* C. MACAL, R. CIRILLO, G. CONZELMANN,
V. KORITAROV, P. THIMMAPURAM, and T. VESELKA
Argonne National Laboratory, Argonne, IL

ABSTRACT

Electric utility systems around the world continue to evolve from regulated, vertically integrated monopoly structures to open markets that promote competition among suppliers and provide consumers with a choice of services. Decentralized decision making is one of the key features of the new deregulated markets. Many of the modeling tools for power systems analysis that were developed over the last two decades are based on the implicit assumption of a centralized decision-making process. Although these tools are very detailed and complex and should continue to provide many useful insights into power systems operation, they are limited in their ability to adequately analyze the intricate web of interactions among all the forces prevalent in the new markets. Driven by these observations, Argonne National Laboratory's Center for Energy, Environmental, and Economic Systems Analysis (CEEESA) has started to develop a new deregulated market analysis tool, the Electricity Market Complex Adaptive Systems (EMCAS) model. Unlike conventional electrical system models, the EMCAS agent-based modeling and simulation system does not postulate a single decision maker with a single objective for the entire system. Rather, agents are allowed to establish their own objectives and apply their own decision rules. Genetic algorithms are used to provide a learning capability for certain agents. With its agent-based approach, EMCAS is specifically designed to analyze multiagent markets and allow testing of regulatory structures before they are applied to real systems.

INTRODUCTION

Electric utility systems around the world continue to evolve from regulated, vertically integrated monopoly structures to open markets that promote competition among suppliers and provide consumers with a choice of services. The unbundling of the generation, transmission, and distribution functions that is part of this evolution creates opportunities for many new players or agents to enter the market. It even creates new types of industries, including power brokers, marketers, and load aggregators or consolidators. As a result, fully functioning markets are distinguished by the presence of a large number of companies and players that are in direct competition. Economic theory holds that this new market will lead to increased economic efficiency expressed in higher quality services and products at lower retail prices. Each market participant has its own, unique business strategy, risk preference, and decision model. Decentralized decision making is one of the key features of the new deregulated markets.

Many of the modeling tools for power systems analysis that were developed over the last two decades are based on the implicit assumption of a centralized decision-making process.

* *Corresponding author address*: Michael North, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439; e-mail: north@anl.gov.

Although these tools are very detailed and complex and should continue to provide many useful insights into power systems operation (Conzelmann, et al., 1999; Koritarov, et al., 1999; Harza, 2001), they are limited in their ability to adequately analyze the intricate web of interactions among all the market forces prevalent in the new markets. Driven by these observations, Argonne National Laboratory's Center for Energy, Environmental, and Economic Systems Analysis (CEEESA) has started to develop a new deregulated market analysis tool, the Electricity Market Complex Adaptive Systems (EMCAS) model. Unlike conventional electrical system models, the EMCAS agent-based modeling and simulation (ABMS) model uses techniques that do not postulate a single decision maker with a single objective for the entire system. Rather, agents are allowed to establish their own objectives and apply their own decision rules. Genetic algorithms are used to provide a learning capability for certain agents. With its agent-based approach, EMCAS is specifically designed to analyze multiagent markets and allow testing of regulatory structures before they are applied to real systems. Computational social science offers potential solutions.

AGENT-BASED MODELING AND SIMULATION

Computational social science involves the use of agent-based models (ABMs) to study complex social systems (Epstein and Axtell, 1996). An ABM consists of a set of agents and a framework for simulating their decisions and interactions. ABM is related to other simulation techniques, including discrete event simulation and distributed artificial intelligence or multiagent systems (Pritsker, 1986; Law and Kelton, 2000). Although many traits are shared, ABM is differentiated from these approaches by its focus on achieving "clarity through simplicity" as opposed to deprecating "simplicity in favor of inferential and communicative depth and verisimilitude" (Sallach and Macal, 2001).

An agent is a software representation of a decision-making unit. Agents are self-directed objects with specific traits. Agents typically exhibit bounded rationality, meaning that they make decisions using limited internal decision rules that depend only on imperfect local information. Emergent behavior is a key feature of ABMS. Emergent behavior occurs when the behavior of a system is more complicated than the simple sum of the behavior of its components (Bonabeau, et al., 1999).

A wide variety of ABM implementation approaches exist. Live simulation, where people play the role of individual agents, is an approach used successfully by economists studying complex market behavior. General-purpose tools such as spreadsheets, mathematics packages, or traditional programming languages can also be used. However, special-purpose tools such as Swarm, the Recursive Agent Simulation Toolkit, StarLogo, and Ascape are among the most widely used options (Burkhart et al., 2000; Collier and Sallach, 2001).

Several electricity market ABMs have been constructed, including those created by Bower and Bunn (2000), Petrov and Sheblé (2000), and North (2000a,b, 2001). These models have hinted at the potential of ABMs to act as electronic laboratories, or "e-laboratories," suitable for repeated experimentation under controlled conditions.

EMCAS

EMCAS is a Recursive Porous Agent Simulation Toolkit (Repast) ABMS with agents that represent generation companies, demand aggregation companies, transmission companies, consumers, system operators, and government regulators. These agents use a variety of computer learning techniques to improve their individual competitiveness as the market within which they are embedded evolves. EMCAS is related to several earlier models by VanKuiken, et al. (1994) and Veselka, et al. (1994).

The underlying structure of EMCAS is that of a time continuum ranging from hours to decades. Modeling over this range of time scales is necessary to understand the complex operation of electricity marketplaces.

On the scale of decades, the focus is long-term human decisions constrained by economics as shown in Figure 1. On the scale of years, the focus is short-term human economic decisions constrained by economics. On the scale of months, days, and hours, the focus is short-term human economic decisions constrained by economics and physical laws. On the scale of minutes or less, the focus is physical laws that govern energy distribution systems. In EMCAS, time scales equate to decision levels. Six decision levels are implemented in the model, with decision level 1 representing the smallest time resolution, that is, the hourly or real-time dispatch. Decision level 6, on the other hand, is where agents perform their long-term, multiyear planning.

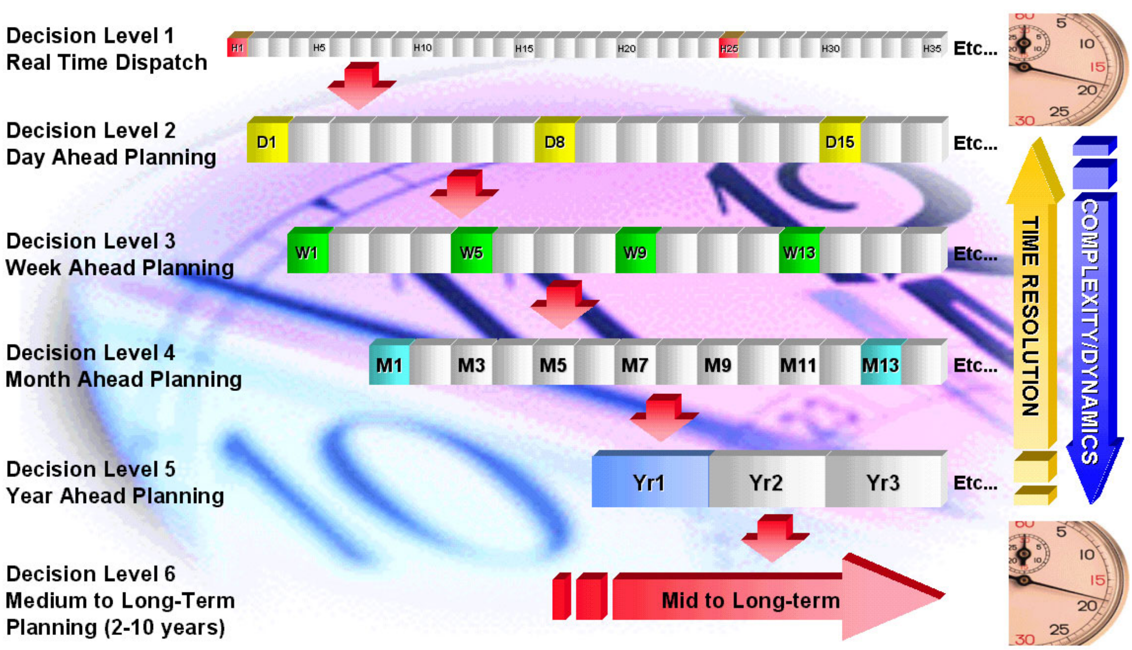

FIGURE 1 EMCAS Time Scales and Decision Levels

EMCAS includes many different agents to model the full range of time scales as shown in Figure 2. The focus of agent rules in EMCAS varies to match the time continuum. Over longer time scales, human economic decisions dominate. Over shorter time scales, physical laws dominate. Many EMCAS agents are relatively complex or “thick” compared with typical agents. EMCAS agents are highly specialized to perform diverse tasks ranging from acting as generation companies to modeling transmission lines. To support specialization, EMCAS agents include numerous highly specific rules. EMCAS agent strategies are highly programmable. Users can easily define new strategies to be used for EMCAS agents and then examine the marketplace consequences of these strategies. EMCAS and its component agents are currently being subjected to rigorous quantitative validation and calibration.

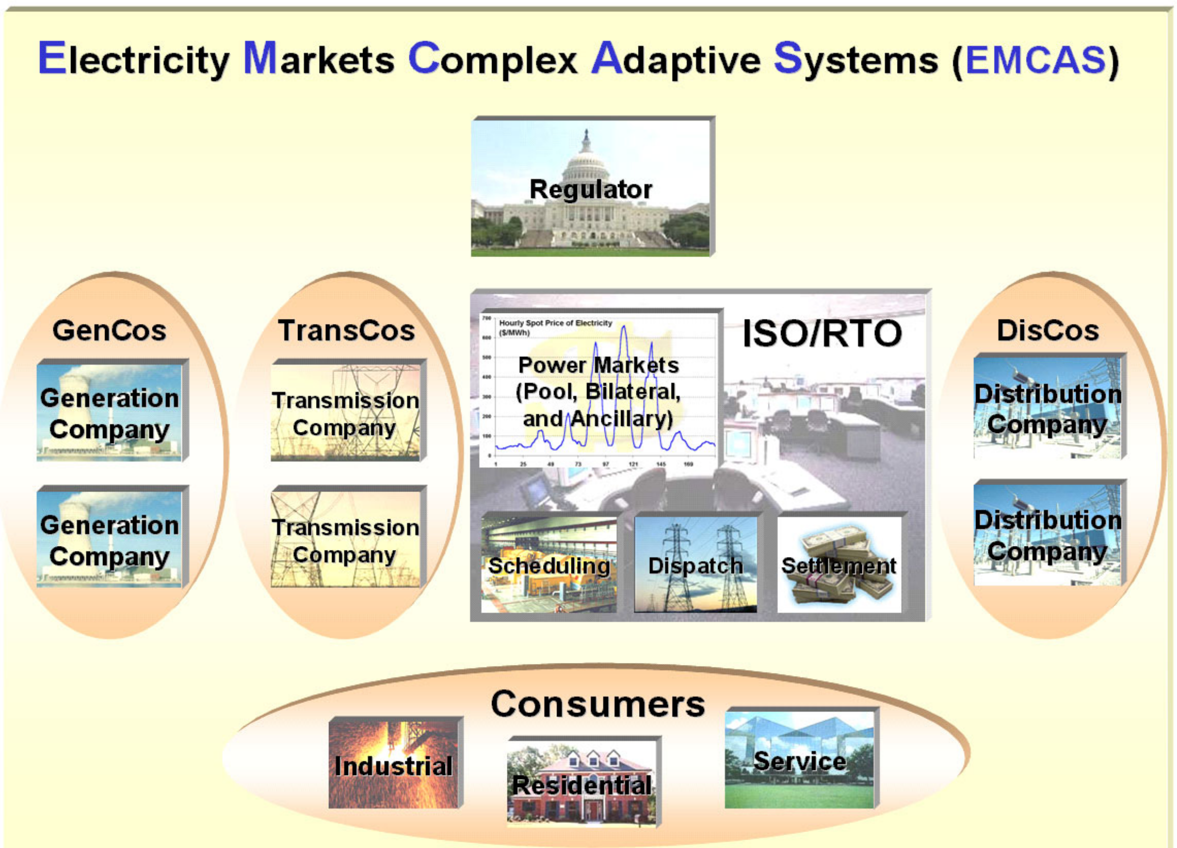

FIGURE 2 EMCAS Structure and Agents

THE EMCAS ARCHITECTURE

EMCAS agents make decisions based on past experiences and anticipated conditions in the future as shown in Figure 3. They also make decisions in the context of current market rules and the potential impact that other players have on markets.

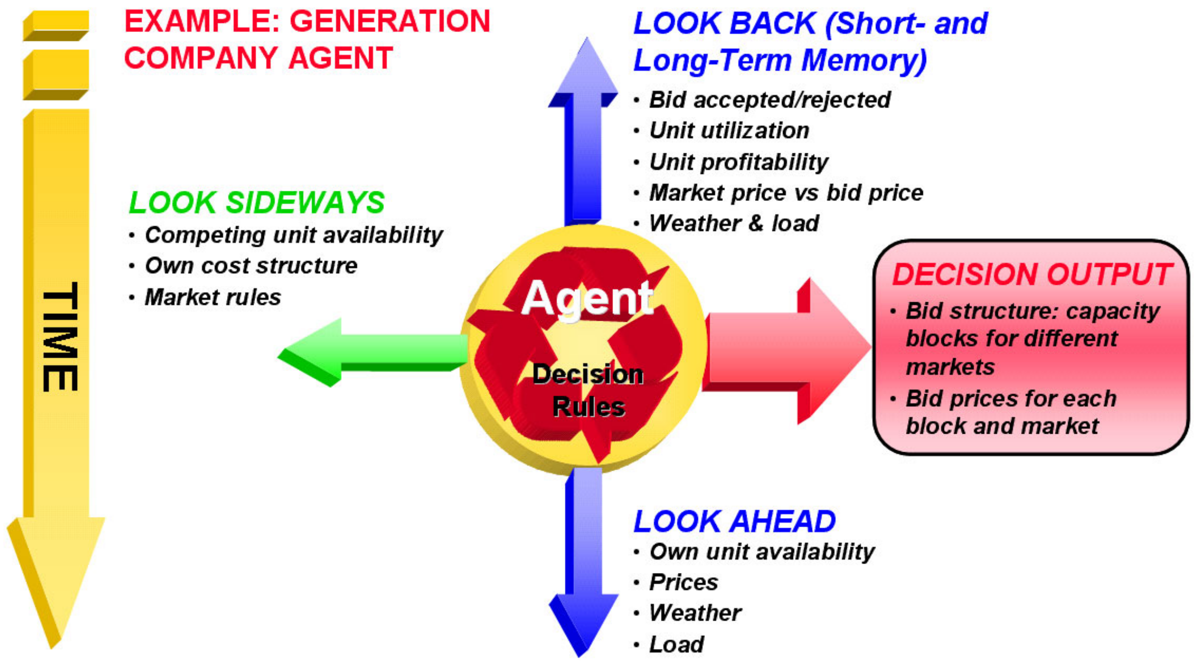

FIGURE 3 Generation Company Agent

The EMCAS model consists of two components, a simulation server and an interface client, both of which are currently under development. The EMCAS simulation server uses the new ABM approach to simulate deregulated electricity marketplaces. The EMCAS interface client uses a Web-based approach to permit shared universal access to the EMCAS model.

The EMCAS simulation server is written in Java. Java directly supports object-oriented implementation, allowing the EMCAS simulation server to be easily extended. Java also supports complex multithreading, allowing the EMCAS simulation server to maximize concurrent execution. The simulation server is designed to use Java Remote Method Invocation (RMI) for distributed computing. Java RMI allows distributed simulation runs across all major platforms, including large computing clusters. The simulation server uses extensible markup language (XML) for data storage. XML is an open, worldwide standard supported by virtually all major software vendors. Because XML is highly portable, EMCAS can be easily interconnected with external data sources, models, and tools.

The EMCAS interface client uses Dynamic Hypertext Markup Language (DHTML) and Scalable Vector Graphics (SVG), allowing it to be displayed in all major Web browsers. The interface client can be used anywhere in the world that a server is available via the Internet or on portable computers without a network connection but with a local server.

A POWER MARKET SIMULATION GAME

To better understand the requirements of an electricity market structure testing tool, a live electricity market simulation was created. In this market simulation game, individuals played the role of generation companies. One additional person played the role of the ISO/RTO.

Each generation company in the market simulation game had three identical generators. The generators included a small natural-gas–fired turbine generator, a medium-sized natural-gas–fired, combined-cycle unit, and a large coal-fired power plant. Players were allowed up to five bid blocks for each unit. Players submitted bids electronically, basing their bids on public information posted by the system operator. This information included historical and projected prices, demands, supply, and weather.

The system operator collected the players' bids on a periodic basis and used to them to simulate the operation of an electricity spot market. The simulation calculated MCPs and player profits based on internally derived demands, supplies, and weather. The actual simulation demands, supply, and weather differed from the publicly posted projections by small random amounts. Generating units also suffered from unannounced random outages.

An initial market simulation game was run with 6 players. The price results from this run are shown in Figure 4. Subsequently, a second market game with 10 players was run. Experience from these market simulation games suggested that the development of an electricity market ABMS might be extremely beneficial. This experience helped to shape the development of EMCAS.

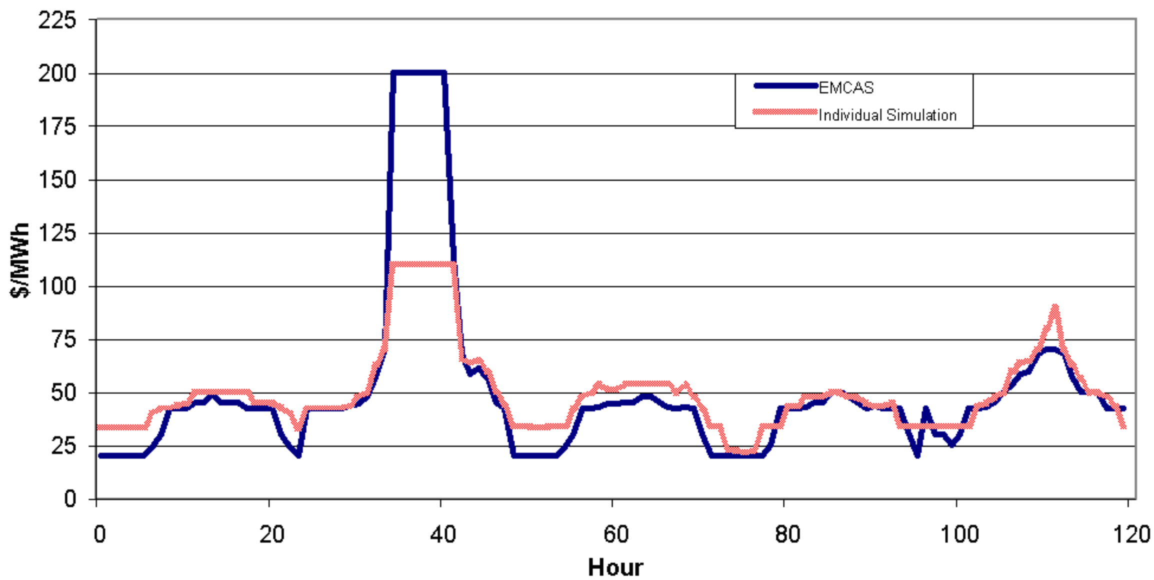

FIGURE 4 Market Clearing Prices — EMCAS vs. Market Game

THE GAME AND EMCAS

An EMCAS case was created based on the previously described market game. Specific agents representing individual market game players were implemented by using the EMCAS agent architecture. The strategies of the individual players were determined by asking them to write short descriptions of their approaches after completing the game; follow-up consisted of a series of focused interviews. Once the strategies were determined, agents implementing each of the strategies were programmed.

The individual agents developed to emulate the market game players were run using the game's original data. The resulting prices are similar to those found in the individual market game (Figure 4). The main difference is that the prices near hour 40 are higher in the EMCAS case because the EMCAS agents were programmed to use the evolved final strategies of the players. Many of the market game players had begun the game using a relatively cautious approach to bidding. As the game progressed, they learned to become much more aggressive. For example, several players developed "hockey stick" strategies that have low prices for the majority of each generator's capacity followed by extremely high prices for the last few megawatts. This approach can be effective because players have little to risk and much to gain. The risk is minimal because the vast majority of their generation bids are likely to be accepted. The gain is potentially high because MCP pricing will assign the last few megawatts high prices to all generation during times of shortage. The result lends new meaning to the hockey term "high sticking."

The EMCAS agents were programmed with the final, more aggressive strategies of the human players. Thus, EMCAS tended to have higher prices throughout the simulation. Once EMCAS was able to replicate the original market game, it was used to explore its suitability as an electricity market structure testing tool.

CHANGING THE RULES

To explore the potential of EMCAS, several variations of the original market game case were created and simulated. These variations probed the effects of changing power plant outages and price setting rules on electricity market prices. As previously mentioned, EMCAS and its component agents are currently being subjected to rigorous quantitative validation and calibration. All of the EMCAS results presented here are intended to explore the potential use of EMCAS as an electricity market structure testing tool. As such, these results are not intended to represent complete analyses of the issues described.

Figure 5 shows the results for the baseline case. This EMCAS run assumes a Pay-MCP market without power plant outages with prices closely following the assumed daily load pattern. The first variation to the base case that was tested was the effect of power plant outages in a Pay-MCP market. The hourly prices are shown in Figure 6. In this example, the overall effect of power plant outages is to greatly increase market prices during periods of peak demand. This suggests that an important concern for regulators setting pricing rules is the relative balance between system supply and demand. In particular, systems that have demands that approach the maximum generation supply might experience significant price spikes under Pay-MCP. Such systems might fare better under Pay-as-Bid because they could potentially be victimized by strategies such as high sticking.

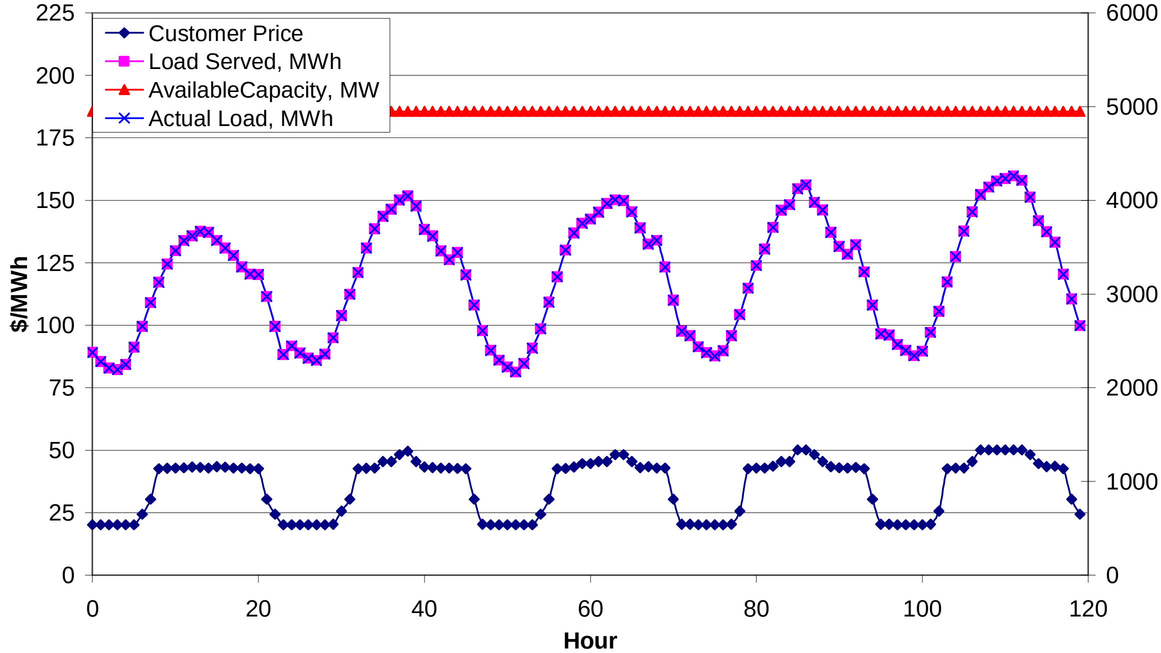

FIGURE 5 Pay-MCP without Outages

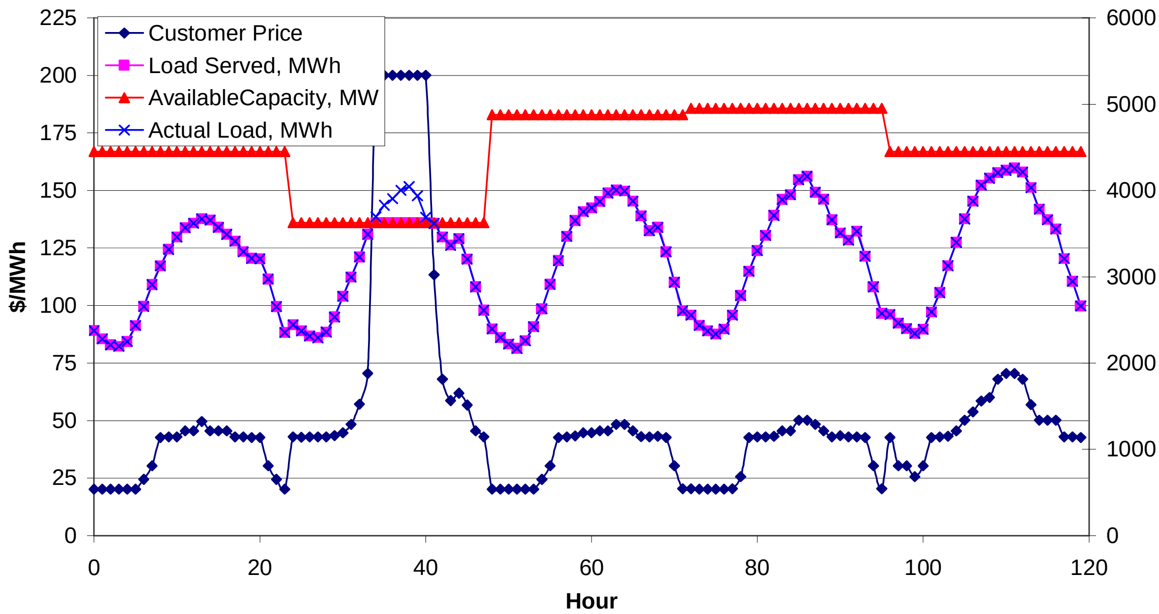

FIGURE 6 Pay MCP with Outages

In the second variation, the market was set up as Pay-as-Bid. Agent pricing strategies were suitably modified to reflect the new price setting rule. The actual hourly loads, the hourly loads served, the available generation capacity, and the resulting hourly prices are shown in Figure 7. In this case, all of the loads were served, so the actual hourly loads and the hourly loads served are the same. In this example, the overall effect of Pay-as-Bid is to noticeably reduce price fluctuations. This observation suggested a third experiment.

The third variation looked at the effect of Pay-as-Bid price setting with power plant outages. As before, agent pricing strategies were suitably modified to reflect the price setting rule. The hourly prices are shown in Figure 8. As with the previous Pay-as-Bid example, in this run, the overall effect is to substantially reduce price volatility compared to Pay-MCP, particularly during times when high demands intersect with reduced supplies.

THE PROFIT MOTIVE

Considering the lower and more stable prices found under Pay-as-Bid, it appears that this form of pricing is better for consumers under this simplified model run. Producers, however, might have a different view. While prices are lower and more stable under Pay-as-Bid, producers lose money under this approach, as shown in Figure 9.

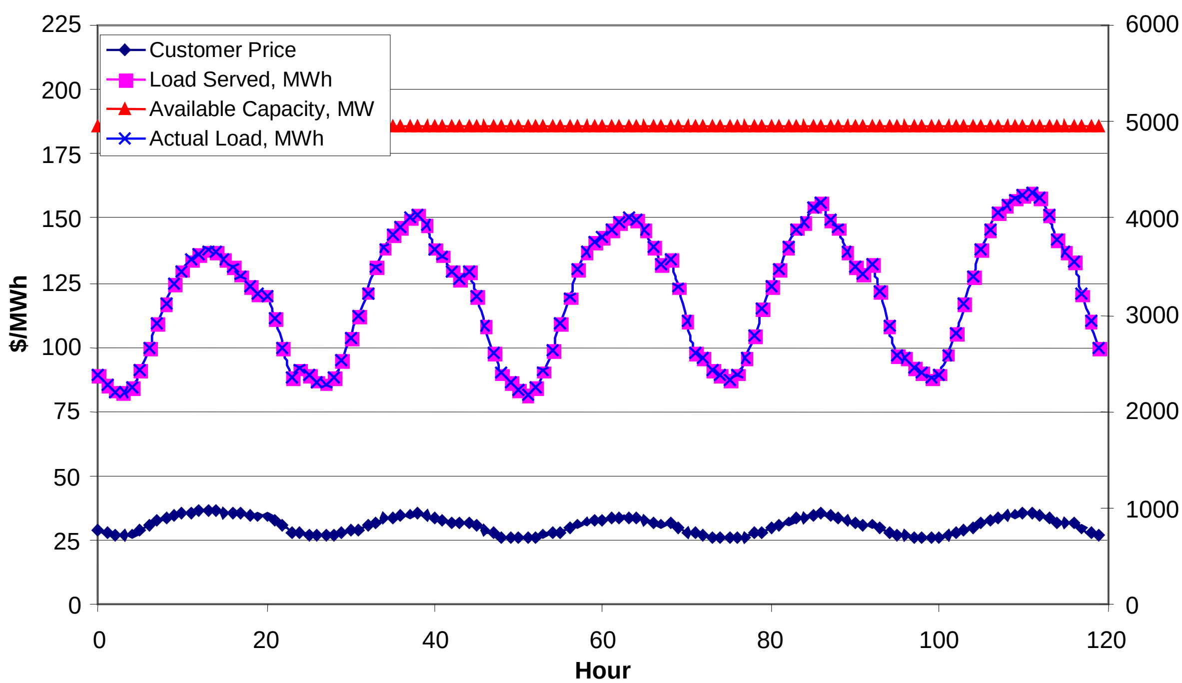

FIGURE 7 Pay-as-Bid without Outages

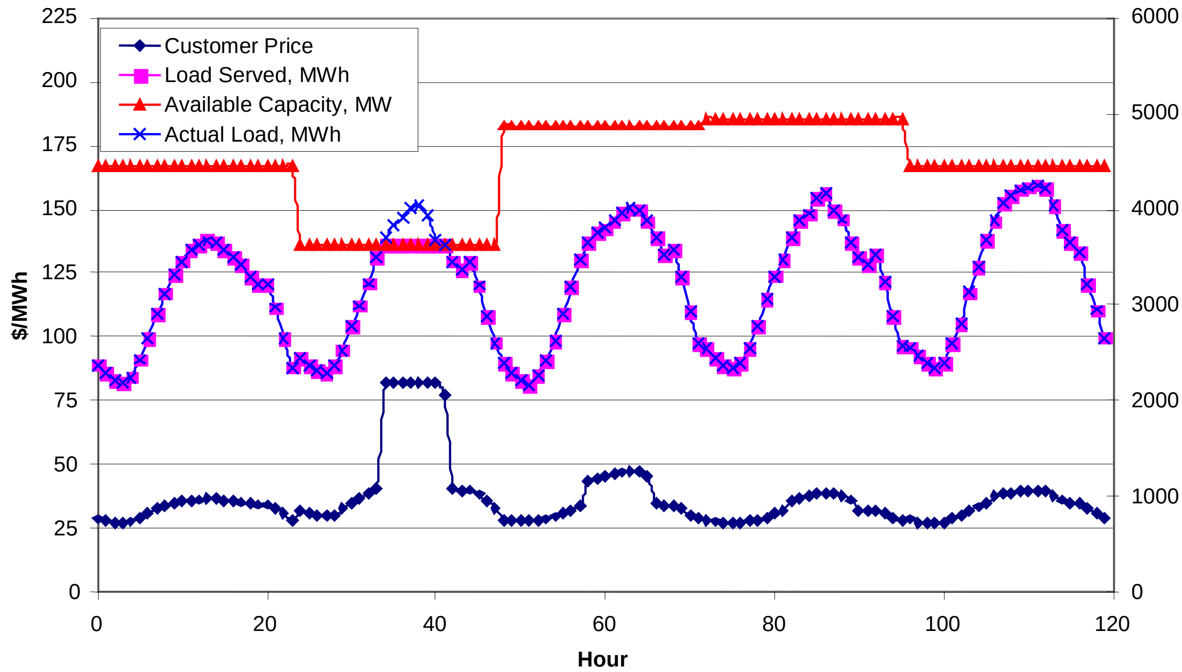

FIGURE 8 Pay-as-Bid with Outages

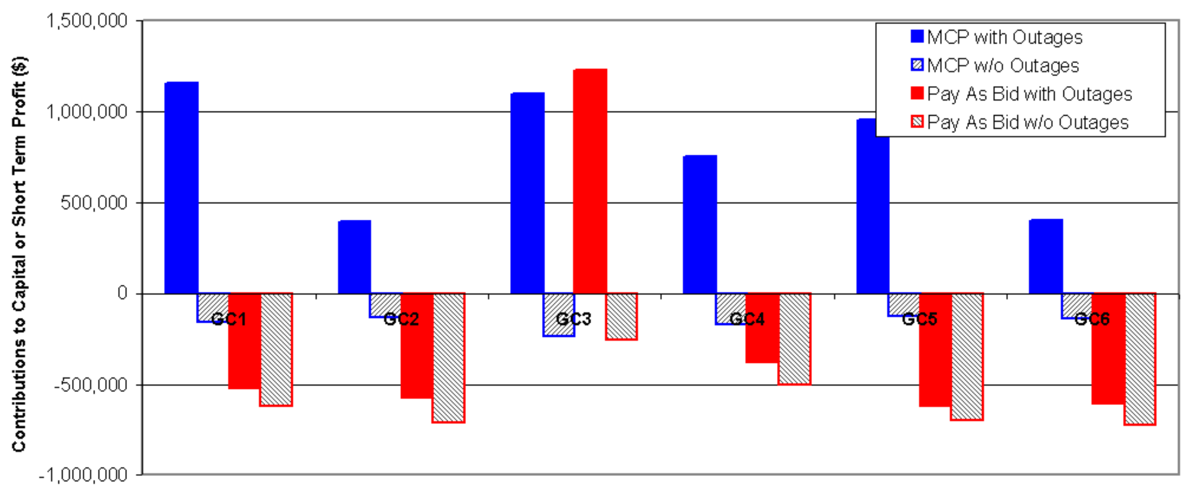

FIGURE 9 Generation Company Profits under Various Market Rules and Outages Regimes

CONCLUSIONS

As electric utility systems around the world continue to move toward open, competitive markets, the need for new modeling techniques will become more obvious. Although traditional optimization and simulation tools will continue to provide many useful insights into market operations, they are typically limited in their ability to adequately reflect the diversity of agents participating in the new markets, each with unique business strategies, risk preferences, and decision processes. Rather than relying on an implicit single decision maker, ABMS techniques, such as EMCAS, make it possible to represent power markets with multiple agents, each with their own objectives and decision rules. The complex adaptive systems approach allows analysis of the effects of agent learning and adaptation. The simple test runs presented in this paper clearly demonstrate the value of using EMCAS as an electricity market structure testing tool, where regulatory structures can be tested before they are applied to real systems.

ACKNOWLEDGMENT

This work was supported by the U.S. Department of Energy, Assistant Secretary for Policy, Planning, and Program Evaluation, under Contract W-31-109-Eng-38.

BIBLIOGRAPHY

Bonabeau, E., M. Dorigo, and G. Theraulaz, 1999, *Swan Intelligence: From Natural to Artificial Systems*, Oxford, U.K.: Oxford University Press.

Bower, J., and Bunn, D.W., 2000, "A Model-based Comparison of Pool and Bilateral Market Mechanisms for Electricity Trading," *Energy Journal* **21**(3), July.

Burkhart, R.M., M. Askenazi, and N. Minar, 2000, *Swarm Release Documentation*. Available at http://www.santafe.edu/projects/swarm/swarmdocs/set/set.html.

Carley, K.M., M.J. Prietula, and Z. Lin, 1998, "Design Versus Cognition: The Interaction of Agent Cognition and Organizational Design on Organizational Performance," *Journal of Artificial Societies and Social Simulation* **1**(3), June.

Collier, N., and D. Sallach, 2001, *Repast*. Available at http://repast.sourceforge.net/.

Conzelmann, G., V.S. Koritarov, K. Guziel, and T.D. Veselka, 1999, *Final Audit Report — New Generating Capacity Tenders 97/1 and 97/2*, report submitted to the Hungarian Power Companies Ltd., Budapest, Hungary, by Argonne National Laboratory, Argonne, IL, Feb.

Epstein, J.M., and R. Axtell, 1996, *Growing Artificial Societies: Social Science from the Bottom Up*, Washington, DC: Brookings Institution Press.

Harza Engineering Company in association with Argonne National Laboratory, 2001, *Trans-Balkan Power Line Project*, Final Report, May.

Koritarov V., G. Conzelmann, T. Veselka, W. Buehring, and R. Cirillo, 1999, "Incorporating Environmental Concerns into Electric System Expansion Planning Using a Multi-criteria Decision Support System," *Int. Journal of Global Energy Issues* **12**(1-6):6–67.

Law, A.M., and W.D. Kelton, *Simulation Modeling and Analysis*, 3rd ed., New York, NY: McGraw-Hill Book Company.

North, M.J., 2000a, "SMART II+: The Spot Market Agent Research Tool Version 2.0 Plus Natural Gas," *Proceedings of the Computational Analysis of Social and Organizational Science Conference*, Carnegie Mellon University, Pittsburgh, PA.

North, M.J., 2000b, "SMART II: The Spot Market Agent Research Tool Version 2.0," *Proceedings of SwarmFest 2000*, Swarm Development Group, Logan, UT.

North, M.J., 2001, Agent-based Infrastructure Modeling, *Social Science Computer Review*, Thousand Oaks, CA: Sage Publications, Fall.

Petrov, V., and G.B. Sheblé, 2000, "Power Auctions Bid Generation with Adaptive Agents Using Genetic Programming," *Proceedings of the 2000 North American Power Symposium*, Institute of Electrical and Electronic Engineers, Waterloo-Ontario, Canada, Oct.

Picker, R.C., 1997, "Simple Games in a Complex World: A Generative Approach to the Adoption of Norms," *University of Chicago Law Review*, Chicago, IL: University of Chicago.

Pritsker, A.A.B., 1986, *Introduction to Simulation and SLAM II*, New York, NY: Wiley.

Sallach, D.L., and C.M. Macal, 2001, "Introduction: The Simulation of Social Agents," *Social Science Computer Review*, Thousand Oaks, CA: Sage Publications, Fall.

VanKuiken, J.C., T.D. Veselka, K.A. Guziel, D.W. Blodgett, S. Hamilton, J.A. Kavicky, V.S. Koritarov, M.J. North, A.A. Novickas, K.R. Paprockas, E.C. Portante, and D.L. Willing, 1994, *APEX User's Guide (Argonne Production, Expansion, and Exchange Model for Electrical Systems) Version 3.0*, Argonne National Laboratory, Argonne, IL.

Veselka, T.D., E.C. Portante, V.S. Koritarov, S. Hamilton, J.C. VanKuiken, K.R. Paprockas, M.J. North, J.A. Kavicky, K.A. Guziel, L.A. Poch, S. Folga, M.M. Tompkins, and A.A. Novickas, 1994, *Impacts of Western Area Power Administration's Power Marketing Alternatives on Electric Utility Systems*, Argonne National Laboratory, Argonne, IL.

NOTICE

VALUATION OF ELECTRICITY COMPANIES IN EVOLUTIONARY OLIGOPOLISTIC MARKETS: A METHODOLOGY

F.S. OLIVEIRA, London Business School*
D.W. BUNN, London Business School

ABSTRACT

We develop a methodology to analyze dynamic aspects of market design and asset valuation. This model has two main components: an "electricity assets market" (EAM) game and a "wholesale electricity market" (WEM) game. This game simulates how bounded-rational agents would trade assets to maximize their expected long-run profit, given the initial conditions. Mathematically, this corresponds to a search in the space of the possible market structures for the one that maximizes the value of each player. We analyze the relation between the EAM and WEM games, identifying the relation between capacity withholding and the portfolio structure of a player. We show that capacity withholding is not an optimal strategy for technological diversified Cournot players. Moreover, if this player acquires extra capacity of a certain technology, it reduces the output of the other technologies in this technology *natural market*. The methodology developed in this paper is motivated by the ultimate aim of analyzing basic questions related to (1) industry structure and the regulatory objectives regarding price and market shares (e.g., how should an industry be restructured; what evolution can be expected regarding technology diversification and vertical integration); and (2) the portfolio of an electricity company (asset mix and vertical integration) and its value (e.g., when does it pay to integrate vertically; when does it pay to diversify or to be specialized).

INTRODUCTION

The logic behind strategic asset ownership in electricity markets changed with the restructuring and deregulation processes. Larsen and Bunn (1999) summarized the changes in the industry that resulted from privatization. The new industry is characterized by unstable and volatile prices, strategic behavior, regulatory uncertainty, and information opacity (the price signaling effects on investment may be misleading). At the corporate level, the new market is characterized by a focus on shareholder value, which replaces the social optimum, and new methods of linking strategic thinking, uncertainty, and limited information, which replaces the classic operations research deterministic planning.

Restructuring and regulatory actions have an important impact on shareholder value. In the electricity industry, market structuring and design have been widely analyzed (e.g., Borenstein, et al. [1995], Elmaghraby and Oren [1999], Bower and Bunn [2000], Bunn and Oliveira [2001]), but this research has only looked at the impact of market structuring on short-term pricing strategies. However, the regulatory body has a longer-term obligation. It must choose how to balance controls on prices with investment incentives, defending both the consumers' interest and efficient entry (and exit). Cox (1999) informs us that mergers,

* *Corresponding author address*: Fernando S. Oliveira, London Business School, Sussex Place, Regent's Park, London NW1 4SA, United Kingdom; e-mail: foliveira@london.edu.

acquisitions, and divestures are the main problems faced by a regulator and the companies operating in a market. Electricity companies may use mergers and acquisitions to adapt to the new environment (e.g., risk management) or to gain market power, whereas divestments by incumbent generation companies may be needed to ensure that the market is competitive. This divestment issue has been addressed by Green and Newbery (1992) and Day and Bunn (2001), by analyzing the impact of the divestment actions in the England and Wales (E&W) electricity market.

In this paper, we develop a methodology aimed at supporting the analysis of market structure evolution as an endogenous variable. Our objective is to model how companies learn to trade generation and supply assets between themselves, but not the investment planning issue, which has been addressed by Skantze, et al. (2000) and Visudhiphan, et al. (2001).

The evidence from the E&W electricity market suggests that, in a restructured electricity market, companies are active in trading assets between themselves in their quest for the "optimal portfolio"; see Appendixes 1 and 2. Thus, in this paper, we present a model formulation that could simulate the possible market structure evolution and the possible equilibria to which this structure may converge. This model has two main components: an "electricity assets market" (EAM) game and a "wholesale electricity market" (WEM) game.

The EAM game simulates the interaction between electricity companies that trade generation and supply assets to maximize their expected profit, taking into account regulatory targets and the underlying fuel prices. Furthermore, this new methodology will enable analysis of the adaptation process, studying the trajectory by which the model convergences toward equilibrium and how the asset value evolves with the industry structure.

The WEM game is an extension of the Cournot game with capacity constraints, taking into account, at the same time, the supply business. The Cournot model has been widely used in the electricity markets game theoretical literature: Allaz and Vila (1993) analyze Cournot competition in forward markets; Borenstein and Bushnell (1999) have used it to analyze market power, and divestment in the California electricity market; Jing-Yuan and Smeers (1999) and Hobbs (2001) have analyzed spatial competition in restructured electricity markets assuming Cournot behavior.

Next, we give a brief summary of the Finite Automata Dynamic Game (FADG) tool used in the EAM game. We then describe the EAM and the WEM games and analyze the interactions between them.

THE FINITE AUTOMATA DYNAMIC GAME FRAMEWORK

The players in this game have rules of behavior (automata) that associate a possible state of the world to a decision. The core of this theory has been developed by Rubinstein (1986), Abreu and Rubinstein (1988), Gilboa (1988), Banks and Sundaram (1990), and Piccione (1992).

In an automata game *G*, each player *i* has a finite automaton,

$$A^i = (Q^i, q_0^i, \Sigma^i, \delta^i, \lambda^i),$$

where

Q^i = finite nonempty set of internal states,

q_0^i = initial internal state,

Σ^i = set of all the possible actions,

δ^i = transition function ($\delta^i : Q^i \times Z \to Q^i$), and

λ^i = behavioral function ($\lambda^i : Q^i \to \Sigma^i$) associating an action to each possible internal state.

At stage 1, each player *i* plays $\lambda^i(q_0^i)$. At stage $t \geq 1$, after each player executing its actions with an outcome $\mathrm{z}_t = \lambda^i(q_t)$, where q_t is the state of the environment, each automaton A^i moves from the state q_t^i to the state $\delta^i(q_t^i, \mathrm{z}_t)$. Each player *i* then chooses a new move $\lambda^i(q_{t+1}^i)$. Player *i* uses the automation A^i.

A *Finite Automata Dynamic Game,* presented in Oliveira (2002), is a game played by automata where a player is allowed to change automaton during a game. In addition, the player does not have a model of the game: it has to infer it by using an identification algorithm. More formally, an FADG is an *N* player discrete-time game with incomplete information, where each player *i* faces a *sequential decision process* $\Pi^i = (A^i, P^i, u^i, \xi^i)$ such that:

1. $A^i = (Q^i, q_0^i, \Sigma^i, \delta^i, \lambda^i)$ represents player *i*'s automaton.

2. $P^i = (Q^{pi}, q_0^{pi}, \Sigma^{pi}, \delta^{pi}, \lambda^{pi})$ represents agent *i*'s perception of the residual product automaton of all $N - i$ players. $M^i = (Q^{mi}, q_0^{mi}, \Sigma^{mi}, \delta^{mi}, \lambda^{mi})$ is the *true* residual product automaton of the game, representing the residual product automaton of all $N - i$ players.

3. *W is* the *true* product automaton of the game. At stage *t*, *W* is a 5-tuple $W_t = (Q_t, q_0, \Sigma_t, \delta_t, \lambda_t)$ that defines how the environment behaves, where $Q_t = Q_t^1 \times Q_t^2 \times ... \times Q_t^N$, $q_0 = q_0^1 \times q_0^2 \times ... \times q_0^N$, $\Sigma_t = \Sigma_t^1 \times \Sigma_t^2 \times ... \times \Sigma_t^N$, $\delta_t : Q_t \times \Sigma_t \to Q_t$, and $\lambda_t : Q_t \to \Sigma_t$.

4. The objective of each player is to maximize ξ^i by choosing a behavioral function λ^i and a transition δ^i.

5. The information set of each player, at each stage, contains the perceived residual current automaton P^i and the new data arriving from the interactions with the environment D^i.

6. The player's perceived residual product automaton P^i is updated by an identification algorithm (quasi-perfect rationality) that transforms the history of the game into the new P^i.

7. Adaptation algorithm (adaptive best-response) transforms the history of the game into a new automaton A^i.

The pseudo-code for the FADG algorithm is represented, in a stylized way, in Table 1. During the game, each agent collects information and updates its perception of the system's behavior. Each agent then revises its automata A^i by using the adaptive best-response algorithm.

TABLE 1 The Finite Automata Dynamic Game Algorithm

While last iteration is not reached, each agent *i* in the game:

1. Makes a move given its current automaton
2. Computes:
   a. the new state and the reward,.
   b. the value of its current automaton,
   c. a model of the environment (P^i) using the identification algorithm, and
   d. Its new automaton A^i.

Each agent uses the identification algorithm, quasi-perfect rationality, to rebuild the model of the system behavior. This identification process obeys two rules necessary for the rational behavior of a certain agent. The first rule is consistency: the model identified by each agent must be consistent (i.e., the same action in a certain state of the automaton it always leads to the same new state). The second rule is closure: an agent must build a model that has a forecast for the outcome of every possible action.

Finally, the agent must adapt to its new perception of the environment, using the adaptive best-response (ABR) algorithm. The ABR algorithm applies three principles to model rational behavior: inertia, tyranny of the weak, and best-response behavior. Inertia reflects the cost of changing. Tyranny of the weak represents the conduct of a player that imposes its behavior on others. Finally, the best response behavior is the attitude of a player maximizing its behavior in stable environments (assuming that the behavior of the other players is stable). The ABR algorithm is presented in Table 2.

Thus, FADG enables the modeling of how players learn and adapt in a dynamic environment: players learn the automaton on the environment behavior and then adapt their behavior to maximize their long-term profit.

TABLE 2 ABR (A^i, P^i) Algorithm

ABR (A^i, P^i) Algorithm
1. Each player *i* computes the volatility measure.
2. Each player *i* computes the internal change measure.
If environment is stable and the player did not change its behavior recently,
$A^i := BR(P^i)$.
Otherwise,
$A^i := A^i$.
Algorithm Best-response $A^i := BR(P^i)$:
1. Compute the optimal policy play against P^i.
2. Compute A^i from the optimal policy.

THE ELECTRICITY ASSET MARKET GAME

The EAM game is developed using the FADG framework. The EAM can be used to analyze the evolution of the value of electricity plants and customers in a market where players might decide to buy or sell some assets given the current market value of these assets and their expectations regarding the future evolution of the wholesale electricity market. The EAM game represents a search mechanism in the space of possible market structures. The search dynamics is guided by players' strategic decision making. The main issues that this model is intended to support relate to this evolution. More specifically, we ultimately seek to analyze the following questions: (1) Under which conditions does the game converge to equilibrium? (2) Which type of equilibria emerged — vertical integration or diversification? (3) How does the trajectory along the adaptation process evolve?

Market structure evolves as a function of several factors: fuel prices, technological innovation, regulatory intervention, threat of new entrance, and the players' expectations and behavior. The focus of this paper is the analysis of the industry's optimal structure, assuming no new entrants and no technological innovation (however, these issues can be straightforwardly incorporated into the model).

Thus, we can define the EAM game as a two-stage dynamic game, where in the first stage, the players choose the amount of capacity they want to hold from each different technology, and in the second stage, they define the amount of generation they want to sell in the market. It is noteworthy that this game is more than the repetition of a single-stage game: the structure of the market changes as players buy and sell assets; thus, the *payoff mapping* of the single-stage game changes as well.

Let us now look more closely at the EAM game dynamics. For a game with *P* players, let a vector $V = \left[(W_1, K_1, C_1),...,(W_i, K_i, C_i),...,(W_{M+S}, D_{M+S}, C_{M+S})\right]$ describe the state of the game in an industry with *M* plants and *S* suppliers (i.e., companies that purchase from the wholesale market and sell to end-use customers). In vector *V* the triples (W_i, K_i, C_i) and (W_i, D_i, C_i) represent the owner (W_i); the capacity (K_i) of plant *i*; the demand of supplier *i* (D_i); and the cost (C_i) of plant or supplier *i*. This vector *V* describes the ownership structure of the industry.

Given the description V and the initial strategies of the other players, each player builds a model that associates each possible succession of actions with different possible outcomes. The automaton used by each player associates the possible states of the industry structure (defined by V and by the expected behavior of others) to its possible actions.

The possible actions of each player are to buy or sell a certain plant or supplier or to exercise no action at all. The outcome of this game is the combination of actions of all the agents. A reward is associated with each one of these outcomes. Moreover, a transition function maps the outcomes into new states, and a behavior function specifies which action to take in each state of the industry structure. Finally, the system's behavior is defined by the product automaton of the automata used by the players in the game.

Therefore, each player, at each stage of the game, can play $M+S+1$ different actions. Thus, there are $(M+S+1)^P$ possible combinations of actions, and $P^{(M+S)}$ possible states of the industry. These two last figures are an exponential function, respectively, of the number of players and of the number of plants and suppliers. The implication of this is striking: for a player to analyze all possible events of this game with a depth D, it has to analyze $(M+S+1)^{DP}$ possible combinations. Besides, as the trading of plant and suppliers implies a bilateral agreement, it requires that at the same time there is a buyer and a seller for the same asset, which makes this game a very hard coordination problem.

Thus, the complexity of this problem makes it hard to assume that the "optimal market" structure can be computed and implemented by any regulator or player in this market. This is where the FADG algorithm might help us. Given the complexity of the problem, each agent only plans D steps ahead which assets to buy or sell, if any. Furthermore, to guide its behavior, a player computes a rule of behavior (an automaton) that associates a certain action to the perceived possible states. Moreover, at the same time, each player infers a model of how the industry is going to evolve and evaluates if it should adapt to this new perception.

Overall, two main features of the EAM simulator differ from the FADG approach: (1) each player assumes a stage game played in a finite number of stages D; (2) each player defines a rule of behavior and infers a model for the industry evolution that can be represented as tree-automata. The EAM algorithm pseudo-code is represented in Table 3.

TABLE 3 EAM Algorithm

At every stage a player:

1. Infers a model of the industry structure evolution (only if it has not been inferred recently to minimize the cost of information gathering).
2. Chooses to change behavior (or to keep the same one) to maximize the present value of its portfolio and to minimize the cost of change.
3. Trades assets if bilateral agreement is possible.
4. Computes the value of the new portfolio.

Finally, it is noteworthy that there are two main economic drivers of industry evolution: the average electricity prices and the average generation costs. The average price is a consequence of the possible market structures. Therefore, when an agent simulates the possible evolution of the industry, it computes the average price in a perceived future industry evolution. A major part of this model is the calculation of this price, which is explained in the next section.

THE WHOLESALE ELECTRICITY MARKET GAME

To simulate the long-term evolution of a market structure, a crucial variable that has to be endogenous is average market price. The WEM simulator is modeled as a Cournot game where each player decides, in each possible stage of the game, how much to generate from each plant it owns. The output of this Cournot game is a market price, and ultimately the profit of each player.

In this section, we present the algorithm to compute the expected average prices and loads, for each plant, given a certain market structure.

To tackle this problem, define a theoretical model of prices and loads in electricity markets. This model relates the industry structure and the asset portfolio of each generator to its strategic behavior and captures the following stylized facts:

- A generator’s supply function is step-shaped.
- A generator may price its plants differently, even if they are identical.
- Generators may price the same type of plant differently.
- A generator aims at maximizing the value of its asset portfolio as a whole.

Several models are available to compute this equilibrium price. The Bertrand model, which assumes price competition, is nonlinear and difficult to compute. The Cournot model, which assumes quantity competition, is the simplest to compute. Finally, the supply-function equilibrium model, which assumes that generators offer a supply-function in the market, allows players to modify at the same time the quantities and prices bid into the market. All of them have been used to describe the behavior of players in the electricity industry. Different researchers have various reasons to use each of them. For example, researchers focusing on the nonlinearity of competition and on price competition between players tend to prefer the Bertrand model. However, while a solution for the Bertrand model can be found when each player owns one technology only, no solution has been found for a game where the same player owns different generation technologies. On the other hand, the supply function game has been the favorite of researchers analyzing the behavior in pool markets where generators offer a supply function. Finally, the oldest of these models, the Cournot, has the main advantage of being the easiest to compute, and a solution may be found even when different players own different generation assets. Thus, we decided initially to model the WEM as a Cournot model.

In the case of bilateral electricity markets, each generator has the possibility of selling the electricity of its different plants in different markets. On the basis of evidence in the E&W electricity market, the baseload, mid-merit, and peak plants tend to be sold over different time scales with different prices. The evidence seems to suggest that different technologies sell into different market segments. Thus, nuclear and combined-cycle gas turbine (CCGT) seem to behave as baseload and mid-merit plants; coal seems to behave as a mid-merit plant; and finally, oil, open-cycle gas turbine (OCGT), and pumped storage behave as peak plants (Power UK, 2002, p. 20–21). Furthermore, the very high spread between high and low prices in the UKPX, and between the System Buy Price (in the Balancing Mechanism) and the UKPX prices, are indicators that in the new bilateral markets flexibility has a price, and that technologies may achieve different rents (Power UK, 2002, p. 42–44). It is also noteworthy that this evidence does not take into account the "forward market" effect where some price "discounting" may take place due to quantity trading and risk aversion.

Therefore, it seems that trades in the baseload market tend to happen in bulk, in the forward market, a long time ahead of the trading day. Second, the mid-merit market needs plants with high flexibility (i.e., capable of reacting with a very short period to demand fluctuations. Finally, a new type of market has emerged, which in the E&W market is called "balancing mechanism," where the generators and suppliers may sell (buy) directly to (from) the system operator. The balancing mechanism is the place where the more flexible technologies can sell their electricity with a high premium (as well as in the power exchange).

A Cournot Pricing Model

In this paper, to capture this flexibility effect, a Cournot model with three different markets is used. Each player is modeled as a Cournot agent *i*, choosing the output of a plant *g* in a baseload market (q_{ig}^{b}), in a mid-merit market (q_{ig}^{m}), and in a peak market (q_{ig}^{κ}). It receives the clearing prices P_b, P_m, and P_k, respectively, for the quantities sold in each of these markets. The capacity constraints in the baseload market are defined by the total capacity available in each technology (and peak plants cannot sell in the baseload market). In the mid-merit market, the capacity constraints are the total capacity available deduced from the baseload plants and from the mid-merit capacity sold in the baseload market. Finally, in the peak market only the peak plant generation that was not sold in the mid-merit market may be offered. This procedure follows the model proposed by Elmaghraby and Oren (1999) and aims at capturing the interaction between different markets.

The WEM game is more complex than a straightforward Cournot model with market separation. Owing to the effect of vertical integration, our players may own, simultaneously, generation and supply assets. In this paper, the suppliers are modeled as having the same type of clients and as having price-taking behavior in the wholesale market. Therefore, we assume that the quantities sold (and the prices in the wholesale market) are defined by the generators and that the suppliers charge a profit margin large enough to cover their long-term average cost.

Thus, for player *i*, the profit (π_i) maximization problem is represented by Equations 1 where

C_{ig} = marginal cost of plant *g*;

r = profit margin charged by the supply businesses;

D_{is}^{j} = market share of supplier *s* (owned by player *i*) in market *j* (where *j* stands for baseload, mid-merit, and peak);

D_j, α_j = intercept and slope of the inverse demand curve of the j market, respectively;

K_{ig} = plant g's total capacity; and

G_i = number of plants of player i.

In Equations 1, as a simplification that does not affect the results, it is assumed that there are only three types of generation technologies: baseload (*b*), mid-merit (*m*), and peak (*k*).

$$
\begin{aligned}
&\max \pi_i = \sum_g q_{ig}^b.\left(P_b - C_{ig}\right) + \sum_{g\neq b} q_{ig}^m.\left(P_m - C_{ig}\right) + \sum_{g\neq b,m} q_{ig}^{pk}.\left(P_k - C_{ig}\right) \\
&\quad + P_b.r.\sum_s D_{is}^b + P_m.r.\sum_s D_{is}^m + P_k.r.\sum_s D_{is}^k \\
&st. \\
&P_b = D_b - \alpha_b.\sum_i \sum_g q_{ig}^b \\
&P_m = D_m - \alpha_m.\sum_i \sum_{g\neq b} q_{ig}^m \\
&P_k = D_k - \alpha_k.\sum_i \sum_{g\neq b,m} q_{ig}^k \\
&q_{ig}^b + q_{ig}^m + q_{ig}^k \leq K_{ig}, \quad \text{for } g = 1,\dots,G_i \\
&q_{ig}^b \geq 0,\ q_{ig}^m \geq 0,\ q_{ig}^k \ \geq 0, \quad \text{for } g = 1,\dots,G_i \ .
\end{aligned}
\tag{1}
$$

Does this model respect the stylized facts? The generation of every mid-merit (peak) plant may be sold in the baseload (mid-merit) or in the mid-merit (peak) markets, therefore being paid different prices even in equilibrium. On the other hand, the baseload plants always receive the same price. In addition, our model is very rich, and the interaction between different technologies is even more subtle: the quantities offered in each market, and the capacity owned of each technology, may influence the profitability of different assets and strategies of different players. How much should a generator offer in each market? Does this depend on the plants owned? We analyze the properties of our model with respect to these questions in the next section.

THE TWO-STAGE GAME: CAPACITY AND GENERATION COMPETITION

It follows from the presentation in the two previous sections that we have defined a two-stage game. In the first stage, a player defines its goals for capacity and the type of technology it wants to own. In the second stage, a player defines how much it wants to sell from each of its plants given the generation technologies and suppliers owned. Therefore, we analyze the relation between these two stages of the game and how the competition in the second stage bounds the choices of capacity and technology in the first stage.

Before proceeding, we analyze the optimality conditions of each one of the players in our model. By computing the partial derivatives of the profit function with respect to each one of the decision variables (the quantities sold from each one of the technologies in each market), we obtain the necessary conditions for optimal behavior in the wholesale electricity market. This procedure follows the complementarity problems (see Ferris and Pang [1997]), and has been extensively used in game theory models in electricity markets (see Jing-Yuan and Smeers [1997] and Hobbs [2001]). Equations 2 represent these optimality conditions. In these equations, *g* equals *b*, *m*, or *k*, and λ_{ij} stands for the shadow price of the *j* (*b*, *m*, *k*) technology for player *i*.

$$D_b - 2.\alpha_b.\sum_{g=b}^{m} q_{ig}^b - \alpha_b.\sum_{\substack{g=b,\\ j\neq i}}^{m} q_{jg}^b - \alpha_b.r.\sum_{s} D_{is}^b - C_{ib} = \lambda_{ib} \quad , \tag{2.1.a}$$

and

$$D_b - 2.\alpha_b.\sum_{g=b}^{m} q_{ig}^b - \alpha_b.\sum_{\substack{g=b,\\ j\neq i}}^{m} q_{jg}^b - \alpha_b.r.\sum_{s} D_{is}^b - C_{im} = \lambda_{im} \quad . \tag{2.1.b}$$

For the mid-merit market, the agent can offer its mid-merit and peak technologies, but it cannot use its baseload technology. The necessary conditions are given by Equations 2.2:

$$D_m - 2.\alpha_m.\sum_{g=m}^{k} q_{ig}^m - \alpha_m.\sum_{\substack{g=m,\\ j\neq i}}^{k} q_{jg}^m - \alpha_m.r.\sum_{s} D_{is}^m - C_{im} = \lambda_{im} \quad , \tag{2.2.a}$$

and

$$D_m - 2.\alpha_m.\sum_{g=m}^{k} q_{ig}^m - \alpha_m.\sum_{\substack{g=m,\\ j\neq i}}^{k} q_{jg}^m - \alpha_m.r.\sum_{s} D_{is}^m - C_{ik} = \lambda_{ik} \quad . \tag{2.2.b}$$

The optimality conditions for the technologies being offered in the peak market are represented by Equation 2.3.

$$D_k - 2.\alpha_k.q_{ik}^k - \alpha_k.\sum_{j\neq i} q_{jk}^k - \alpha_k.r.\sum_{s} D_{is}^k - C_{ik} = \lambda_{ik} \, . \tag{2.3}$$

Finally, to complete the Karush-Kuhn-Tucker necessary conditions, we need the complementarity constraints, as shown in Equations 3, for each of the capacity constraints.

$$\lambda_{ib} = 0 \vee q_{ib}^b = K_{ib} \quad , \tag{3.1}$$

$$\lambda_{im} = 0 \vee q_{im}^b + q_{im}^m = K_{im} \quad , \tag{3.2}$$

and

$$\lambda_{ik} = 0 \vee q_{ik}^m + q_{ik}^k = K_{ik} \quad . \tag{3.3}$$

Given the optimality and complementarity conditions (Equations 2 and 3) we can analyze the players' behavior and the interactions between the two stages of the game.

Proposition 1: *A generator owning two different technologies — baseload and mid-merit plants (or mid-merit and peak plants) — offers the mid-merit (peak) technology in the baseload (mid-merit) market if and only if it does not withhold capacity from the baseload (mid-merit) technology.*

Proof: By replacing Equation 2.1.a into Equation 2.1.b, we get $\lambda_{ib} = \lambda_{im} + C_{im} - C_{ib}$. As $\lambda_{im} \geq 0$ and $C_{im} - C_{ib} > 0$, it follows that $\lambda_{ib} > 0$. Then, by Equation 3.1, it results that this player offers its full baseload capacity (i.e., $q^b_{ib} = K_{ib}$). Similar proof follows for the mid-merit technology.

Proposition 2: *A Cournot player may profit, under certain parameters, from capacity withholding in the baseload (mid-merit) market if it does not sell any electricity generated by mid-merit (peak) plants in this market.*

Proof: By splitting the markets into separate entities, Equations 2.1.b and 2.2.b are eliminated from the problem, with $q^b_{im} = 0$ and $q^m_{ik} = 0$. The proof follows by contradiction. Assume that a player cannot profit from capacity withholding. Then, in the baseload market $q^b_{ib} = K_{ib}$ and $\lambda_{ib} > 0$ by Equation 3.1. This implies that, by Equation 2.1.a,

$$D_b - 2.\alpha_b.K_{ib} - \alpha_b.\sum_{\substack{g=b,\\ j\neq i}}^{m} q^b_{jg} - \alpha_b.r.\sum_s D^b_{is} - C_{ib} > 0\,.$$

This inequality is *not* true for every possible parameter. Hence, under certain demand conditions and competition behavior, it may be profitable for a player to generate less than its full capacity.

Theorem 1: *In the Cournot model specified by Equations 1, generation transfer from baseload to mid-merit plants (or from mid-merit to peak plants) via capacity withholding is not an optimal strategy.*

Proof: From Proposition 1 it follows that generation from mid-merit (peak) technology is offered in a baseload (mid-merit) market if baseload (mid-merit) technology is not withheld. Further, Proposition 2 implies that capacity withholding of baseload (mid-merit) plants may occur, but this capacity is not replaced by mid-merit (peak) generation. Hence, capacity transfer from cheaper to more expensive technologies it is not optimal under Cournot behavior.

Theorem 2 (Crowding-out principle): *In the Cournot model specified by Equations 1, a technologically diversified player may expel the mid-merit (peak) plants from the baseload (mid-merit) market by increasing the capacity it owns of baseload (mid-merit) plants.*

Proof: Proposition 1 implies that a diversified player selling some of its mid-merit (peak) electricity in the baseload (mid-merit) market offers its full baseload (available mid-merit) capacity in the baseload (mid-merit) market. Therefore, if a player increases its baseload (mid-merit) generation capacity, it may reduce the quantity sold from mid-merit (peak) technology in these markets. Moreover, by Proposition 2, if the capacity increase is such that some baseload (mid-merit) capacity is withheld from the baseload (mid-merit) market, then no mid-merit (baseload) capacity is sold in this market.

The crowding-out principle completes Theorem 1 by saying that not only is capacity transfer not an optimal strategy for Cournot players, but the capacity increase (reduction) of cheaper technologies may also drive out (drive in) the more expensive generation ones. Thus, the crowding-out principle establishes a bridge between the two stages of the game — capacity and generation competition. Whenever a player decides to acquire or sell baseload (mid-merit) capacity, it takes into account the crowding-out effects and also the effects of that decision on the other players' behavior.

SUMMARY

Modeling of the evolution of electricity market structure is an open research issue. A better understanding of this evolutionary process will have important implications on both regulatory and ownership policies. We are starting to study this issue by developing a new methodology, the EAM-WEM game, which enables the analysis of market structure evolution as an endogenous variable. The aim of this model is to provide a framework that enables an endogenous search for the possible market structure equilibria and, at the same time, gives better insights into the trajectory to equilibrium.

Further, the search space of this problem is an exponential function of the number of plants and supply businesses in an industry, which means that the optimal structure of the industry and the equilibrium of this game are extremely hard to compute. Moreover, the possible number of different "combinations of actions" is an exponential of the number of players, which means that the EAM game represents a hard coordination game, where each player has to search for bilateral trading opportunities with other players.

Hence, in developing the EAM-WEM game, a number of behavioral properties of the players were identified, some of which challenge conventional wisdom regarding generation capacity manipulation:

- A generator owning two different technologies — a baseload and a mid-merit plant (or a mid-merit and a peak plant) — only offers the mid-merit (peak) technology in the baseload (mid-merit) market if and only if it does not withhold capacity from the baseload (mid-merit) technology.

- A Cournot player may profit, under certain parameters, from capacity withholding in the baseload (mid-merit) market if it does not sell any electricity generated by its mid-merit (peak) plants in this market.

- It is not an optimal strategy for a Cournot player to transfer generation from a baseload to a mid-merit plant (or from a mid-merit to a peak plant) via capacity withholding.

- A Cournot player that owns different types of technologies expels the mid-merit (peak) technologies from the baseload (mid-merit) market if it increases the capacity of the baseload (mid-merit) plants.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support from Fundação para a Ciência e a Tecnologia, Portugal, under Grant Praxis XXI/BD/21769/99 and would like to thank A.V. De Miguel for his helpful comments.

REFERENCES

Abreu, D., and A. Rubinstein, 1988, "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica* **56**(6):1259–1281.

Allaz, B., and J.-L. Vila, 1993, "Cournot Competition, Futures Markets and Efficiency," *Journal of Economic Theory* **59** (1):1–16.

Banks, J.S., and R.K. Sundaram, 1990, "Repeated Games, Finite Automata, and Complexity," *Games and Economic Behavior* **2**:97–117.

Borenstein, S., J. Bushnell, E. Kahn, and S. Stoft, 1995, "Market Power in California Electricity Markets," *Utilities Policy* **5**(3-4):219–236.

Borenstein, S., and J. Bushnell, 1999, "An Empirical Analysis of the Potential for Market Power in California's Electricity Industry," *The Journal of Industrial Economics* **XLVII**(3):285–323.

Bower, J., and D.W. Bunn, 2000, "A Model-based Comparison of Pool and Bilateral Market Mechanisms for Electricity Trading," *The Energy Journal* **21**(3):1–29.

Bunn, D.W., and F.S. Oliveira, 2001, "Agent-based Simulation — An Application to the New Electricity Trading Arrangements of England and Wales," *IEEE Transactions on Evolutionary Computation* **5**(5):493–503.

Cox, A.J., 1999, "Mergers, Acquisitions, Divestures, and Applications for Market-based Rates in a Deregulated Electric Utility Industry," *The Electricity Journal* **12**(4):27–36.

Day, C.J., and D.W. Bunn, 2001, "Divestiture of Generation Assets in the Electricity Pool of England and Wales: A Computational Approach to Analysing Market Power," *Journal of Regulatory Economics* **19**(2):123–141.

Elmaghraby, W., and S.S. Oren, 1999, "The Efficiency of Multi-unit Electricity Auctions," *The Energy Journal* **20**(4):89–116.

Ferris, M.C., and J.S. Pang, 1997, "Engineering and Economic Applications of Complementarity Problems," *SIAM Review* **39**(4):669–713.

Gilboa, I., 1988, "The Complexity of Computing Best-response Automata in Repeated Games," *Journal of Economic Theory* **45**:342–352.

Green, R.G., and D. Newbery, 1992, "Competition in the British Electricity Spot Market," *Journal of Political Economy* **100**(5):929–953.

Hobbs, B.F., 2001, "Linear Complementarity Models of Nash-Cournot Competition in Bilateral and POOLCO Power Markets," *IEEE Transactions on Power Systems* **16**(2):194–202.

Jing-Yuan, W., and Y. Smeers, 1999, "Spatial Oligopolistic Electricity Models with Cournot Generators and Regulated Transmission Prices," *Operations Research* **47**(1):102–112.

Larsen, E.R., and D.W. Bunn, 1999, "Deregulation in Electricity: Understanding Strategic and Regulatory Risk," *Journal of the Operational Research Society* **50**:337–344.

Oliveira, F.S., 2002, "Quasi-perfect Rationality, Playing Finite Automata Dynamic Games," Working Paper, *8th Conference of the Society for Computational Economics*, June, Aix-en-Provence, France.

Piccione, M., 1992, "Finite Automata Equilibria with Discounting," *Journal of Economic Theory* **56**:180–193.

Power UK, 2002, "News and Analysis of the UK Power Industry," 97, Mar.

Rubinstein, A., 1986, "Finite Automata Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory* **39**:83–96.

Skantze, P., P. Visudhiphan, and M. Ilic, 2000, "Valuation of Generation Assets with Unit Commitment Constraints under Uncertain Fuel Prices," Mimeo MIT_EL 00-006, Nov.

Visudhiphan, P., P. Skantze, and M. Ilic, 2001. "Dynamic Investment in Electricity Markets and Its Impact on System Reliability," Mimeo MIT_EL 01-012WP, Massachusetts Institute of Technology, Energy Laboratory Report, July.

APPENDIX 1:

TRADING OF POWER STATIONS OWNERSHIP 1994–2002

Plant	Fuel[a]	Installed Capacity (MW)	Commission or Generation Began	Seller	Acquirer
Brigg	CCGT	240	1993	European Gas Turbine	Regional Power Generators Ltd.
West Burton Rugeley Ironbridge	Coal	1,932 976 970	1967 1972 1970	Eastern Group	National Power
Drakelow High Marnham	Coal	976 945	1965 1959	Eastern Group	PowerGen
Keadby	CCGT	720	1994	Norweb	Scottish and Southern Energy
Brigg	CCGT	240	1993	Yorkshire Electricity	IVO Energy
Ferrybridge C	Coal	1,955	1966	PowerGen	Edison Mission Energy
Fiddler's Ferry	Coal	1,961	1971	PowerGen	Edison Mission Energy
Drax	Coal	3,870	1974	National Power	AES
Eggborough	Coal	1,960	1968	National Power	British Energy
Killingholme	CCGT	650	1994	National Power	NRG Energy
Rye House	CCGT	715	1993	PowerGen	Scottish Power
Sutton Bridge	CCGT	803	1999	Enron	London Electricity
Corby	CCGT	401	1993	Dominion Energy	PowerGen (50%) ESBI (50%)
CoGen	CHP	176	NA[b]	Yorkshire Electricity	PowerGen
Cottam	Coal	2,008	1969	PowerGen	London Electricity
Fawley	Oil	518	1969	IPUK	Innogy Plc
Stallingborough	CCGT	1,250	1996 &1998	Humber Power	Centrica (60%) TotalFinaElf (40%)
Rugeley	Coal	976	1972	TXU	International Power
Saltend	CCGT	1,200	2000	Entergy	Calpine
Kings Lynn Peterborough	CCGT	350 380	1996 1993	Centrica	TXU
West Burton	Coal	1,932	1967	TXU	London Electricity

[a] CCGT = combined-cycle gas turbine; CHP = combined heat and power.

[b] NA = not available.

APPENDIX 2:

ACQUISITIONS OF SUPPLY BUSINESSES

Target	Owner	Acquirer
SWALEC	SWALEC	Welsh Water
London Electricity	London Electricity	Entergy
East Midlands	East Midlands	Dominion Resources
East Midlands	Dominion Resources	PowerGen
London Electricity	Entergy	Electricité de France
Midands	Midlands	National Power
SWEB	Southern Eastern	London Electricity
SWALEC	Hyder	British Energy
Norweb Energi	United Utilities	TXU
SWALEC	British Energy	Scottish & Southern Energy
Supply business	Independent Energy	Innogy
Yorkshire Power	Yorkshire Power	Innogy
YED	Innogy	Northern Electric
Northern Electric	Northern Electric	Innogy
Eastern Electric	TXU	London Electricity
Amerada Hess	Amerada Hess	TXU
Innogy	Innogy	RWE

DISCUSSION:
ELECTRICITY APPLICATIONS

Randy Picker, The University of Chicago, Moderator

Time did not allow a discussion of the Electricity Applications papers.

MacroEvolution

THE IMPACT OF RESOURCE ALLOCATION STRATEGIES ON PRE-HOMINID GROUP SURVIVAL

S.J. GOODHALL, Compuware Corporation, Farmington Hills, MI
and Department of Computer Science, Wayne State University, Detroit, MI
R. WHALLON, Museum of Anthropology, University of Michigan, Ann Arbor, MI
R.G. REYNOLDS, Department of Computer Science, Wayne State University, Detroit, MI*

ABSTRACT

This paper investigates the effectiveness of a simple vector voting model of Proto-hunter-gatherer decision making in exploiting the mobile and stationary resources patterns in a complex environment. The formal vector voting model is based on the observation of chimpanzee group decision making. The cognitive limitations of the decision makers on the performance of the system over time are investigated using an agent-based Swarm simulation model. It is suggested that the resource-sharing differences between hominid and pre-hominid groups reflect the type of decision making that they each can perform.

It is shown in the Proto-hunter-gatherer model that the allocation of the resultant resources to those most responsible for making decisions is an important aspect of group survivability. This is because individuals cannot directly communicate the reasons for their decisions to the other group members. Resource-sharing protocols that regulate the distribution in terms of relative decision-making success or status perform much better than altruistic approaches that distribute resources equally. This is in contrast to many observations made with regard to hominid hunter-gatherer resource-sharing practices where hunting and collecting knowledge can be shared among individuals in the group.

INTRODUCTION

The ultimate goal of our research is to assess the impact of culture on group decision-making behavior. Specifically, we are concerned whether the emergence of human culture provided humans with an adaptive advantage over nonhuman primate counterparts in terms of hunting-gathering capabilities. Reynolds (1978) has proposed several mathematical models of hunter-gatherer and primate decision making based on differences in human and primate cultural traits. These are called the cultural algorithm and vector voting models, respectively.

The context in which these decision-theoretic models were compared was a two-dimensional cellular space divided into R discrete subregions or cells each of unit area. The model groups computed the answers to various spatial predicates or queries about the region based on the agents' current knowledge. The models were analyzed theoretically, and it was shown that the ability to form a collective intelligence through the pooling of knowledge with cultural algorithms had some distinct advantages over the vector voting model. In particular,

* *Corresponding author address*: Dr. Robert Reynolds, Department of Computer Science, Wayne State University, Detroit, MI; e-mail: Reynolds@cs.wayne.edu. Dr. Reynolds' work on this paper was supported by a National Science Foundation Intelligent Systems Grant #9907257.

predicates such as the best direction within a region in which to forage (i.e., the direction containing the most resources) were limited in the vector voting model by the maximum area over which each individual had knowledge. On the other hand, pooling of that knowledge theoretically in the cultural model allowed a group to make these decisions over the entire region.

MODELS OF RESOURCE ALLOCATION

Although the above results identified theoretical differences between the models, the theoretical models did not address the impact of resource sharing among participants. It seemed important to identify the impact that various resource-sharing strategies had on the actual differences in model performance in environments of varying resource complexity. In this paper, the effect that the distribution of food resources within the foraging group has on the outcome of the pre-hominid vector voting model is tested using an agent-based implementation of the model using the Swarm simulation system. Learning takes place in the model by emulation, and the performance of the model is compared with that of a random walk model.

In previous work, we explored the performance of the vector voting model in combination with several different strategies for allocating acquired resources within the foraging band. These strategies included Equal Allocation, Round Robin, Satisfied First, Neediest First, and Fixed Rank Order. The results suggest that the vector voting model performs best when certain resource allocation strategies are used (e.g., Satisfied First, Fixed Rank Order) rather than, for example, Neediest First and Equal Allocation. In the latter case, the vector voting model behaves equivalently to a random walk model. Thus, when knowledge is not shared equally, it also is not adaptive to share resources equally. We recently extended our consideration of the effects of resource allocation strategies on the performance of the vector voting model foraging to a new strategy involving periodic reordering of precedence.

The model simulates groups of hunter-gatherers traveling and foraging in a physical space. Food is distributed in varying quantities throughout the space, and visible landmarks are distributed randomly throughout the space. Each member of a group is modeled as an independent agent participating in group decision making, movement, and consumption of food. The basic temporal unit used in the model is days.

Movement Model

Each band moves independently within the defined model space. At the beginning of each successive day, each band uses the vector voting model to select a direction of movement; it then moves in that direction for the remainder of the day. During the day, each band moves one or more squares. The model space consists of a rectangular grid. Movement is from square to square along one of eight compass points, with foraging performed in each square. Movement ceases when the maximum number of squares has been traversed for the day or the caloric needs of the group have been met.

Decision Model

Each band chooses a direction of travel based on a vote by its members. Each member of the band selects a preferred direction based on memories of past consumption, choosing, for example, to move in a direction associated with being satisfied or away from a direction associated with not receiving enough to eat. The band moves in a consensus direction. If it encounters another band or an edge of the model space along its path, it uses the same process to select a different direction. Bands are constrained against reversing their path from the previous day.

The decision model reflects the vector voting process by which primate groups make consensus-based decisions. The motivation for the vector model is captured in the following quote by Kummer (1968, p. 66):

> In choosing the direction of their departure, Hamadryas baboons have to rely on information gathered on previous trips. We do not know how or by whom these sites are explored and remembered, but we know that on different mornings different males of the troop strive in different directions. The decision is made by a long process in which most of the males in the troop participate while the rest seem unconcerned. I have already described how a Hamadryas troop prepares for departure during its morning rest. The troop performs slow, on-the-spot movement, changing its shape like an undecided ameba. Here and there, males move a few yards away from the troop and sit down, facing in a particular direction away from the center. Pseudopods are generally formed by the younger males and their groups, until one of the older males in the center rises and struts toward one of the pseudopods. At this, the entire troop is alerted and begins to depart in the indicated direction. Detailed observation reveals that two male roles are involved in leading the Hamadryas troop: the younger initiators who 'propose' certain directions and the deciders who choose among the proposed directions.

We have abstracted this process into a model where each band member chooses a preferred direction based on remembered experiences of resource acquisition success and failure when the icon was in sight. The consensus process is emulated by calculating the consensus direction of the individual preferences.

Consumption Model

The overall consumption model is based on movement and foraging. After the band has selected a direction of travel for the day, it moves and forages in that direction until the nutritional needs of all members have been fully satisfied or until the band has covered the maximum distance for one day's movement. In each square, the members of the band consume food in sequence. There are several options for the sequence and quantity consumed by each member on each movement.

Each individual has a minimum daily requirement for consumption. If an individual does not receive this amount in the course of a day's movement, a deficit develops. If an individual's deficit reaches a specified level, the individual dies. Individuals can consume more than the daily requirement, permitting them to reduce the deficit in areas with abundant food and to build a reserve for future privation. If all of the members of a band die, the band becomes inactive.

Some of the consumption strategies can be characterized as selfish, whereas others are more altruistic. A total of five strategies have been investigated here; each is described below:

- *Fixed order.* The order of consumption is established when the band is created and does not change. If there is not enough food to meet the needs of individuals at the end of the order, they receive nothing. On each successive move, consumption resumes at the top of the order.

- *Equal Shares.* The order of consumption is established when the band is created and does not change. Each member of the band is allocated a proportionate share of the available food in each square visited.

- *Round Robin.* The order of consumption is established when the band is created and does not change. Given the consumption order, each individual has a chance to be the first in the order that they are on the consumption list starting with the first member, then the second member and so on. The individual who is first on one day will be last on the next as the pointer scrolls through the list from top to bottom.

- *Neediest First.* The order of consumption is redetermined each day. The members of the band are ordered by their total food deficit. The member with the largest deficit has the first turn.

- *Satisfied First.* The order of consumption is redetermined each day. Members are assigned a “dissatisfaction” score based on whether they voted for the direction of travel that was chosen. The greater the difference between the chosen direction and their score, the greater the dissatisfaction. Consumption proceeds in ascending order of dissatisfaction (descending order of satisfaction) with the decision.

Memory Model

Individuals make decisions regarding movement based on their own memories. Memories are recorded for each individual at the end of each model day. If the individual received at least the daily requirement for nutrition, the memory is positive in proportion. If the individual received less than this amount, the memory is negative in proportion to the deficit. Memories are not directly related to the point in space where they are formed. Instead, a system of icons or landmarks is used. Here, landmarks are distributed randomly across the model space and are visible for a specified distance. Individuals can also see for a specified distance. Each of these constraints can be either fixed or variable.

When a memory is formed, each individual selects one or more landmarks in approximately the same direction from the current location. These landmarks are associated with the newly formed memory and are used to access the memory for decision making. Each individual looks in a direction chosen at random, so that members of the same band may have quite different associations. Recent memories are given greater weight than older memories. Memories expire after a specified number of days.

Meritocracy

Our most recent variation is a modified consumption strategy that involves a rank order that is periodically reordered based on individual performance. At predetermined intervals, the members of the band are reranked based on their leadership performance. On each day, their decision is assessed against the consensus decision of the group.

Reproduction Model

If all members of a band are fully nourished for a specified time, reproduction becomes possible. One individual is added to the band, and the interval is restarted. If a band reaches a certain threshold size, it is divided into two independent bands, each consisting of approximately one-half the members of the original band.

Resource Distribution Model

Food resources are distributed in a rectangular space according to some algorithm. During model development, a number of algorithms were explored, including a random distribution; several variations of DeJong's formula, which produces a bowl-shaped gradient; and inverses of DeJong that produce a dome-shaped gradient. Figure 1 shows a modified DeJong distribution.

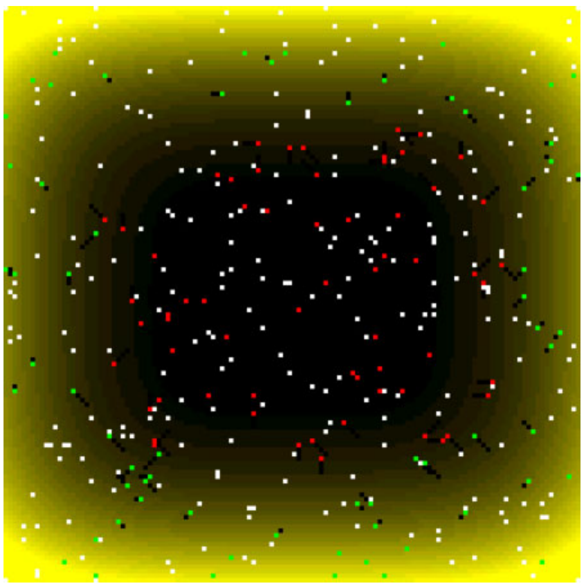

FIGURE 1 Modified DeJong Distribution

In Figure 1, the food resource is represented by the graduated shading at the edge of the image. The individual dots represent landmarks and foraging groups. Landmarks are bright white; groups are green or red, depending on whether they are receiving sufficient resources or not.

Our recent work uses a "patchwork" distribution used in population ecology. The patchwork distribution can be used to emulate the attributes of a semi-arid environment. This distribution is generated by creating uniform areas of food availability around each landmark on the landscape. This distribution is created according to the following procedure:

- Landmarks are distributed randomly on the landscape.
- Each landmark is assigned a random food value between zero and the maximum permitted quantity of food in a cell.
- An area of fixed dimensions around the landmark is populated with food at that value.
- If the areas around two landmarks overlap, the food level is at the value of the last landmark processed.

Figure 2 shows a typical patchwork distribution.

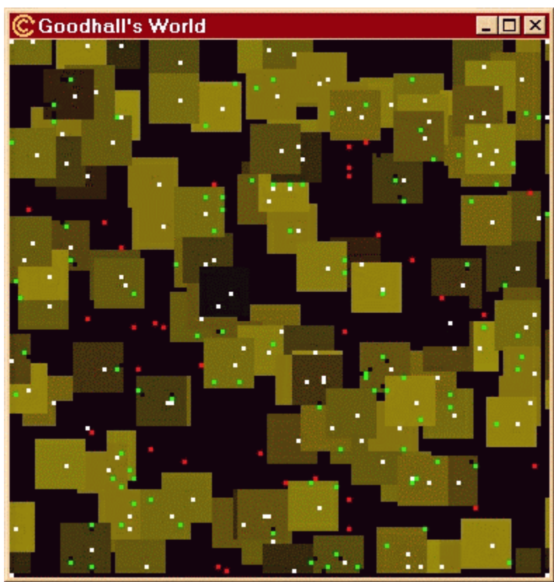

FIGURE 2 Patchwork Distribution

Mobile Resources

Mobile resources in the model represent the existence of mobile herds of animals. Mobile resources are organized into groups or herds. Herds wander the landscape according to a simple pattern. In the currently implemented pattern, herds move in a diamond-shaped, counterclockwise path and occupy a range of cells. If a herd encounters the edge of the space, it rebounds at a 90° angle to its line of travel. If a band moves into the area occupied by a herd, adult members of the band can capture and consume a herd animal. Figure 3 shows a food distribution with mobile resources. Herds of mobile resources are represented as purple patches in the display. The range of a herd is shown by the boundaries of the patch. Each herd moves in a straight line for a specified number of days and then turns 90° to the left, a basic nonlinear pattern. If unhindered, it produces a polygon whose size can be adjusted in each experiment.

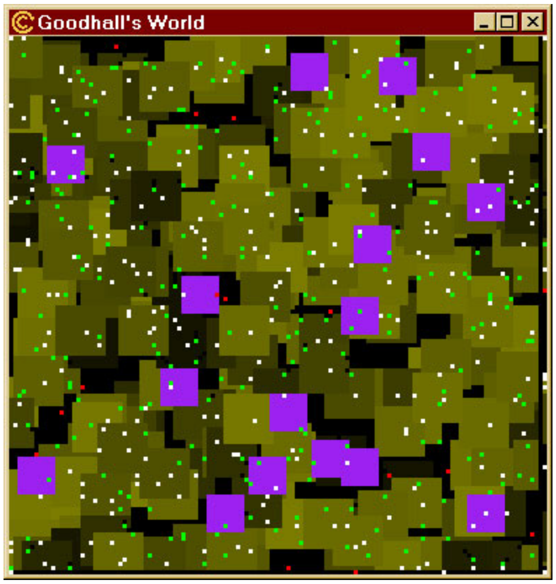

FIGURE 3 Patchwork Distribution with Herds of Mobile Resources

When a band encounters mobile resources, the memory formation algorithm for the band is altered. While the band is in the domain of a herd, all memories are formed using landmarks that appear in the direction of travel of the herd. The purpose is to induce a tendency for the band to move after the herd.

Memory Exchange

Another feature recently introduced into the model is the ability for individuals in different bands to communicate with one another. This capability is accomplished by a selective exchange of memories between individuals. Whenever two bands move into adjacent cells, each adult member of each band has the opportunity to exchange one or more memories with a member of the other band.

MODEL IMPLEMENTATION

The model is implemented using a multi-agent simulation model using Swarm and Objective C. Swarm is an object-oriented, multi-agent simulation framework developed by the Santa Fe Institute.[1] Figure 4 shows the object model.

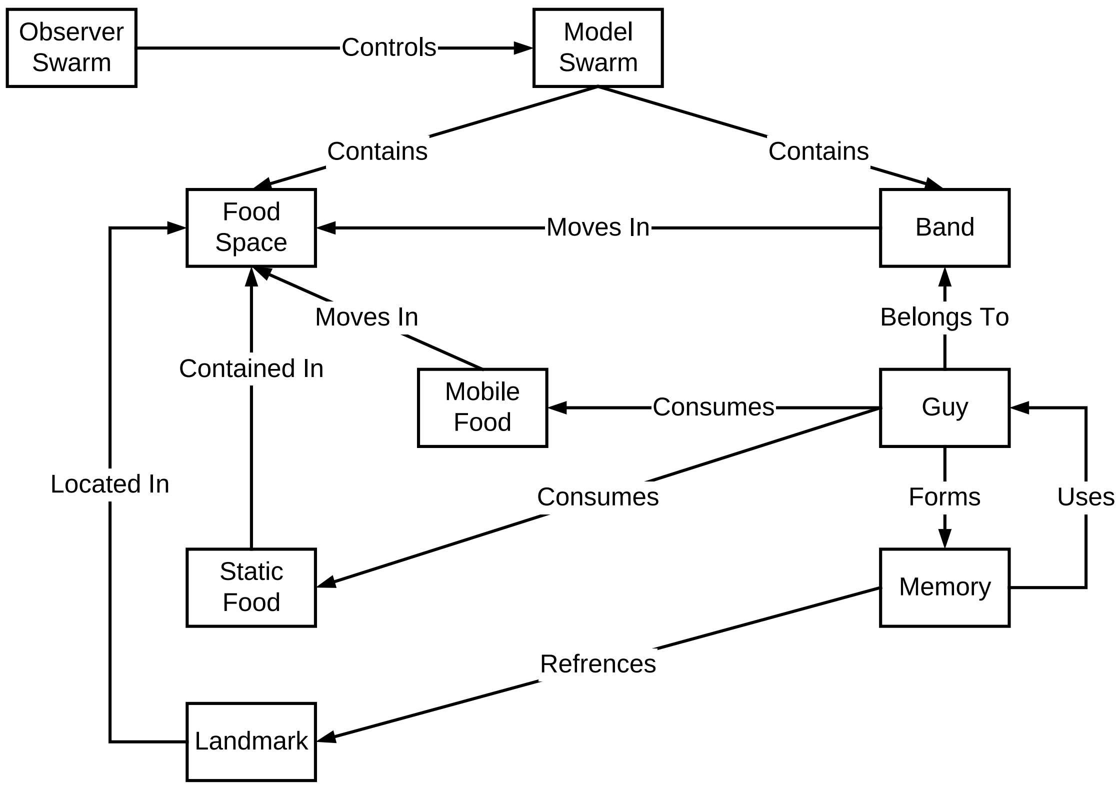

FIGURE 4 Object-relationship Model

The model objects are as follows:

- The Observer Swarm is a Swarm Control structure.
- The Model Swarm is a Swarm Control structure.
- A Band is a group of individuals (Guys) who move and forage together.
- A Guy is a member of a Band.

---

[1] See the Santa Fe Institute Web site (www.santafe.edu) for information about the model framework. Information about Swarm can be obtained at www.swarm.org.

- An Individual Memory is formed by a Guy at the end of each day of movement and foraging.

- The Food Space is a rectangular grid in which Bands move and forage. Food is distributed across the food space according to one of several distribution algorithms. Separate Food Spaces exist for static resources and mobile resources.

- Landmarks are navigation objects scattered across the Food Space. Bands use Landmarks for navigation. When a Memory is formed, it is associated with one or more Landmarks that were visible to the Band Member when the Memory was formed.

The model relationships are as follows:

- The Observer Swarm controls and monitors the Model Swarm. The model event loop is controlled by the Observer Swarm object.

- The Model Swarm is a container for other model objects. It directly contains both the Band objects and the Food Space objects. The external parameters and the file input-output are handled by the Model Swarm.

- Each Band contains a specified number of members (Guys).

- Each Band moves in the Food Space according to distributed decision-making algorithms.

- Each (living) Guy in the Band consumes food as the band moves through the Food Space. Static resources are consumed according to the selected sharing algorithm. Mobile resources are consumed opportunistically.

- At the end of each day, each living member forms one memory of the food satisfaction experienced during the day. Each memory references one or more landmarks that are visible from the location of the Band at the end of the day.

- At the beginning of each day, each living member forms a preference of direction using previously captured memories.

- Landmarks reside at a static location in the Food Space.

Swarm provides a cyclical control element that operates the model. Each cycle represents a single day of movement and foraging. Action methods associated with objects perform the model calculations.

ANALYSIS OF RESULTS

At fixed intervals during the model execution, as measured in model days, status information is written to a set of log files. This information includes:

- The number of Bands that have at least one living individual.
- The number of living individuals.
- The number of individuals that received the minimum required nutrition on the current model day.
- The average age of each Band taken on the current day for Bands with at least one living member or at the time the last member died for Bands with no active members.
- The distribution of the number of members in each Band on the current model day.
- The distribution of distances between Bands on the current model day.
- The distribution of range of Band movement since the last log event.
- The distribution of memories based on calories received relative to need.

These data are loaded into a relational database and analyzed using custom software.

EXPERIMENTAL DESIGN

Experiments are conducted to compare specific alternatives. Using a common food distribution and set of model parameters, a single factor is varied, and the impacts on individual, Band, and population survival are assessed. To ensure consistency of results, a series of simulations is performed with a common set of parameters, and the average performance is produced by the analysis software. Model runs must be long enough to achieve a steady state for measurement. In general, we found that 1,000 cycles (days) are sufficient to achieve stability.

EXPERIMENTAL RESULTS

We have documented comparisons of decision strategies and consumption models (Reynolds, et al., 2000, 2001; Goodhall, 2002). Some of the comparisons to date are summarized in Table 1.

TABLE 1 Summary of Results

Superior Performance	Inferior Performance
Vector Voting	Random Movement
Fixed Consumption Order	Equal Allocation
Fixed Consumption Order	Round Robin Allocation
Satisfied First	Fixed Consumption Order
Fixed Consumption Order	Neediest First
Neediest First	Equal Allocation

It was no surprise that a decision process based on the memory of past successes and failures performs better than random movement. Two exceptions to this occur in the model: (1) when there are highly abundant food resources and (2) and when there are “harsh” environments where the resources are randomly distributed. Since such an environment is information free, the vector voting model acts in a random fashion and performs the same as the random movement model. However, in any harsh environment with any local concentrations of resources, the vector voting process is more successful.

Somewhat more surprising were the results from comparing consumption models. “Selfish” resource allocation strategies, such as the Fixed Consumption Order and Satisfied First patterns, significantly outperformed “altruistic” strategies, such as Equal Allocation. The Neediest First strategy performs somewhere in between but never stabilizes and continues to decline toward extinction.

Figure 5 shows the number of individuals surviving more than 1,000 cycles for each of the consumption modes. These data are drawn from a series of 10 trials each for an environment configuration with a carrying capacity of around 300 individuals.

Postanalysis of the model behavior reveals several findings:

- Total population spikes upward as some bands grow in size before those in sparse resource areas die out. This finding is more an artifact of the model than a useful result and represents the initial colonization activity in a region. Meaningful comparisons become possible once the model stabilizes.

- Under the altruistic patterns, there is a tendency for entire bands to die off in a single cycle. This finding suggests that more individuals are being kept alive than the local environment will sustain. When starvation finally occurs, there is no time for any member to recover from the food deficit.

- Under the selfish patterns, membership in individual bands tends to stabilize relatively quickly and remains at that level for a long time. Some oscillation occurs as a band flourishes long enough to add a member that pushes the local

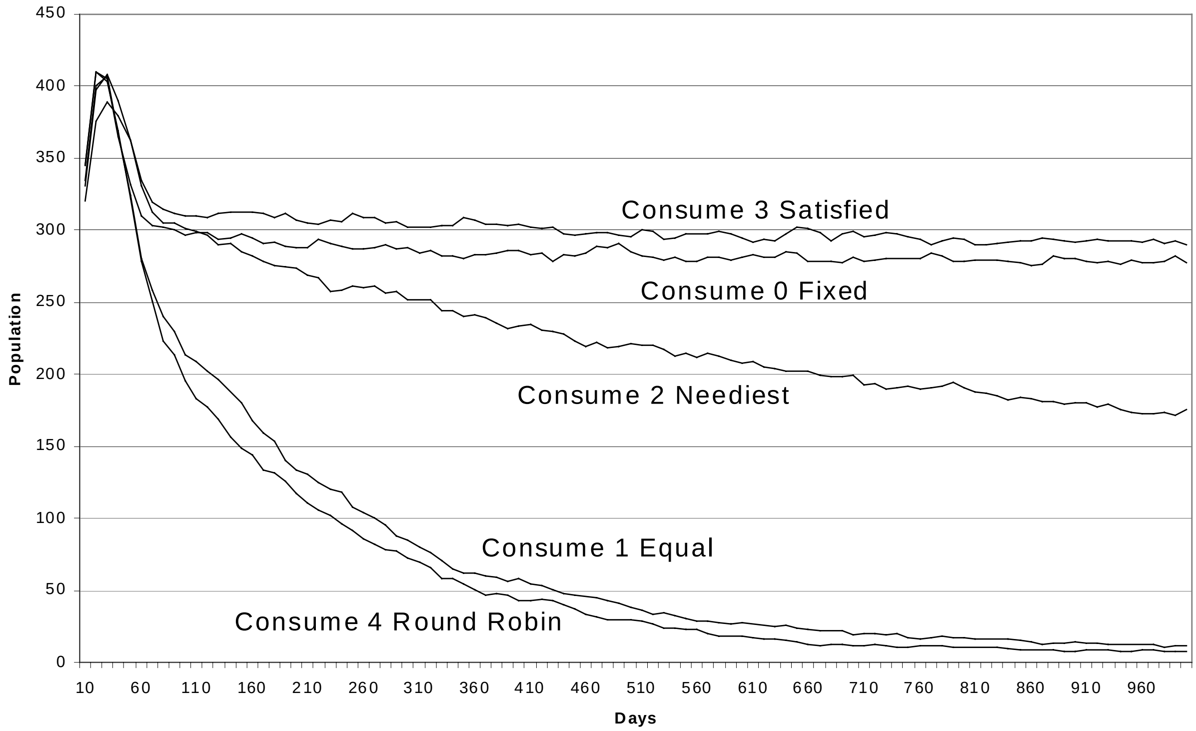

FIGURE 5 Comparison of Resource-sharing Strategies in Terms of Surviving Individuals over Time

population beyond the capacity of the local environment. The group then struggles to survive until the last member in the rotation starves, permitting the cycle to repeat.

- Under successful strategies, the population eventually concentrates into the areas with the most resources and competes with one another for resources.

Our most recent trials have used sharing based on a fixed order that is recalculated periodically, as mentioned earlier. Between recalculations, members of the group consume resources in fixed order. At periodic intervals, the order is recalculated based on the cumulative "satisfaction" score, which in turn is based on each individual's agreement with navigational decisions. Figure 6 shows that periodic reordering is somewhat more successful than a permanent, fixed order.

We recently explored another area: the impact of different cognitive abilities among members of the group on the performance of various resource-sharing strategies. Each landmark has a maximum distance, measured in cells, at which it is visible. Each individual has the ability to see for a particular distance, also measured in cells. In most trials to date, all individuals have had same vision ability. Here, we provide each individual with a different vision ability, which can vary from one cell to some maximum (Figure 7). The addition of these different cognitive abilities produces a small but significant improvement in population survival. This result may in part be due to the additional information provided to the system via the different measurements.

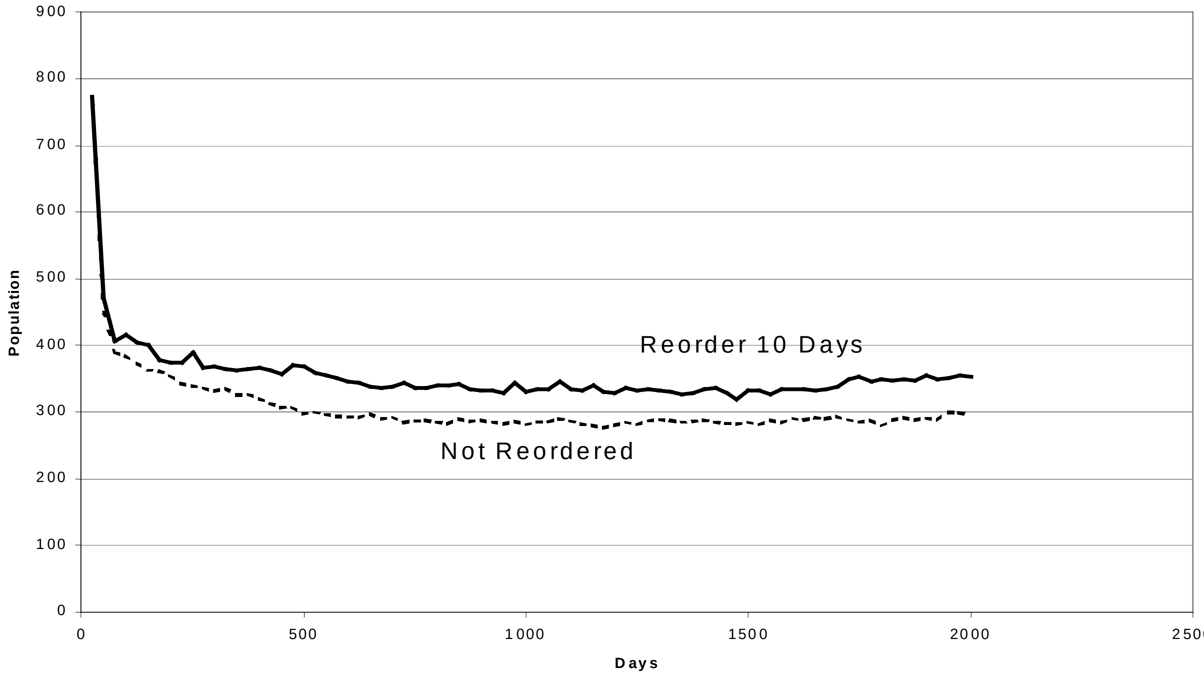

FIGURE 6 Fixed Order vs. Periodic Reordering

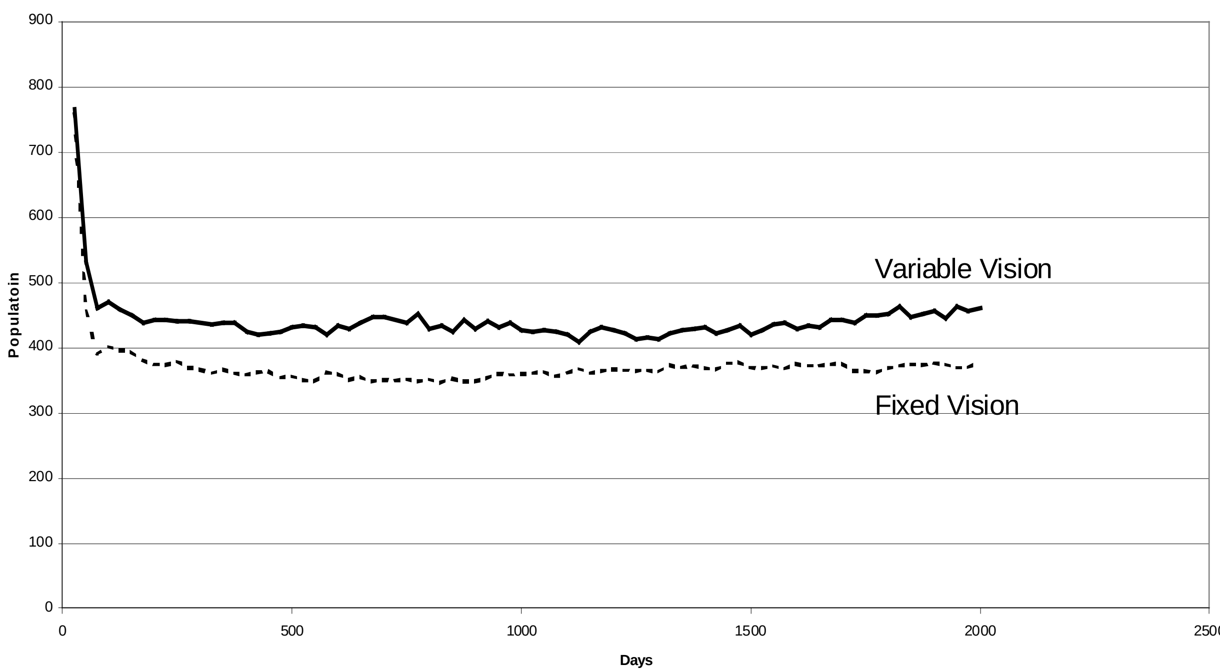

FIGURE 7 Variable Vision vs. Fixed Vision

MOBILE RESOURCES

Figure 8 shows the survival rate of individuals over time with or without mobile resources. Figure 9 shows the survival rate of bands over time with or without mobile resources. Curiously, the addition of mobile resources produces a smaller population of individuals. The fact that there are more bands indicates that the environment is supporting a slightly larger number of slightly smaller bands. Evaluation of the detailed results indicates that mobile resources represent only 2 to 3% of the total resources consumed here. This suggests that the decision making of proto-hominid groups does not allow extensive exploitation of mobile resources with complex, large-scale patterns of mobility and may distract the bands from the more reliable activity of gathering static (and more predictable) resources. Further experiments will be conducted with mobile resources that possess simple linear and very local patterns of movement. These patterns will be more likely to be exploited by the linear vector voting model than the current nonlinear patterns.

SYMBOLIC COMMUNICATION

Figure 10 shows the results of runs in which simple iconic information is exchanged compared with runs in which iconic information is not exchanged. In this situation, a simple icon representing a landscape feature and associated performance (e.g., good, bad) can be exchanged between individuals. In this instance, up to three memories were exchanged by up to three individuals on each encounter. Diagnostic outputs from the modeling software recorded dozens of encounters on each model day. It was observed that there is no significant difference between

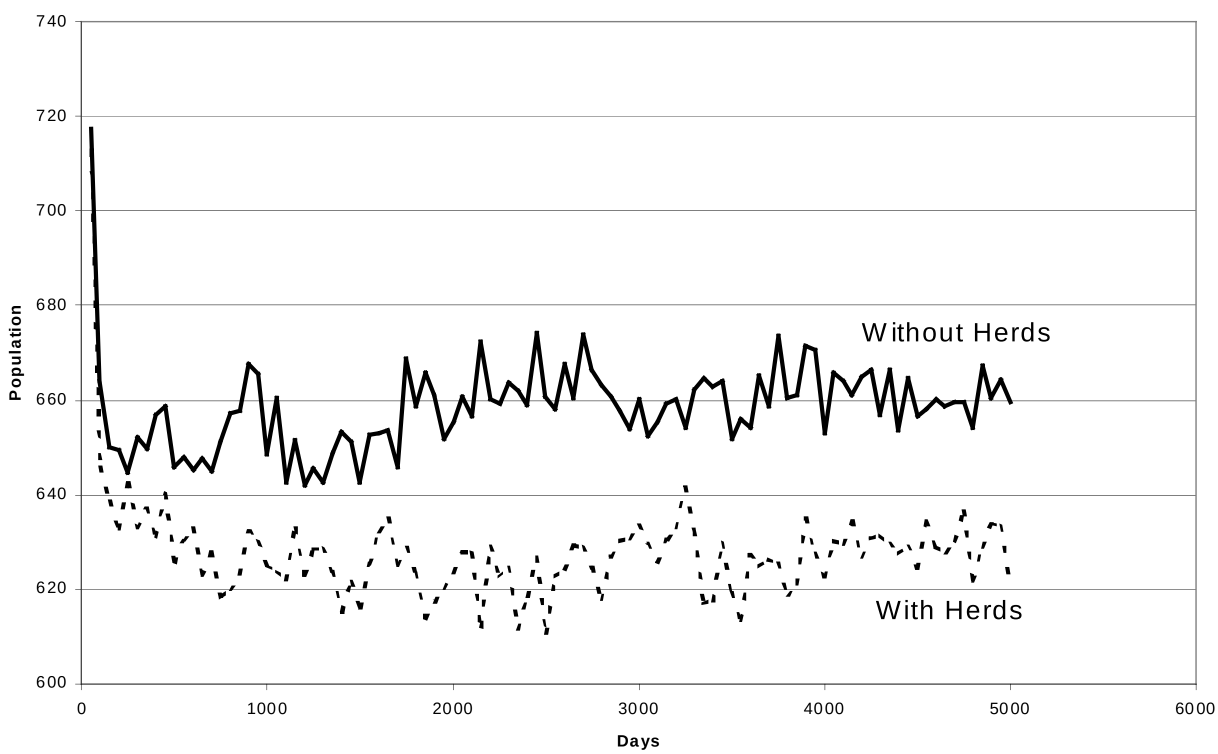

FIGURE 8 Individual Survival Rate, with and without Mobile Resources

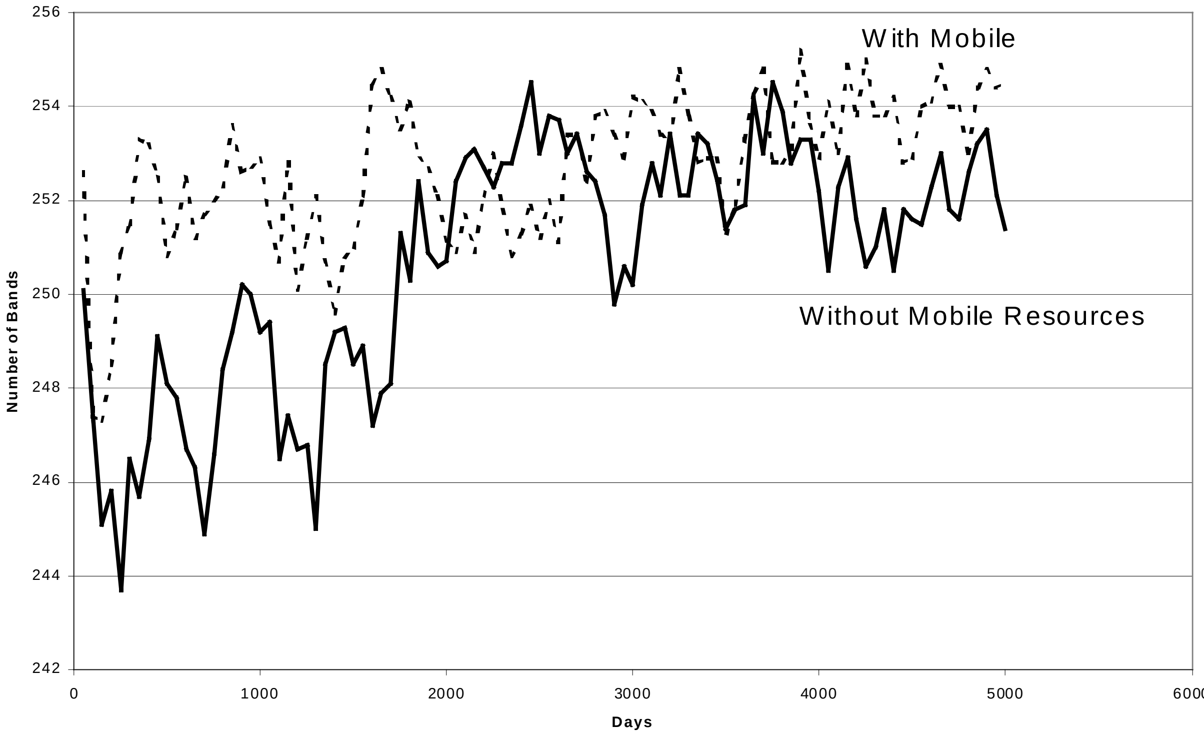

FIGURE 9 Band Survival Rate, with and without Mobile Resources

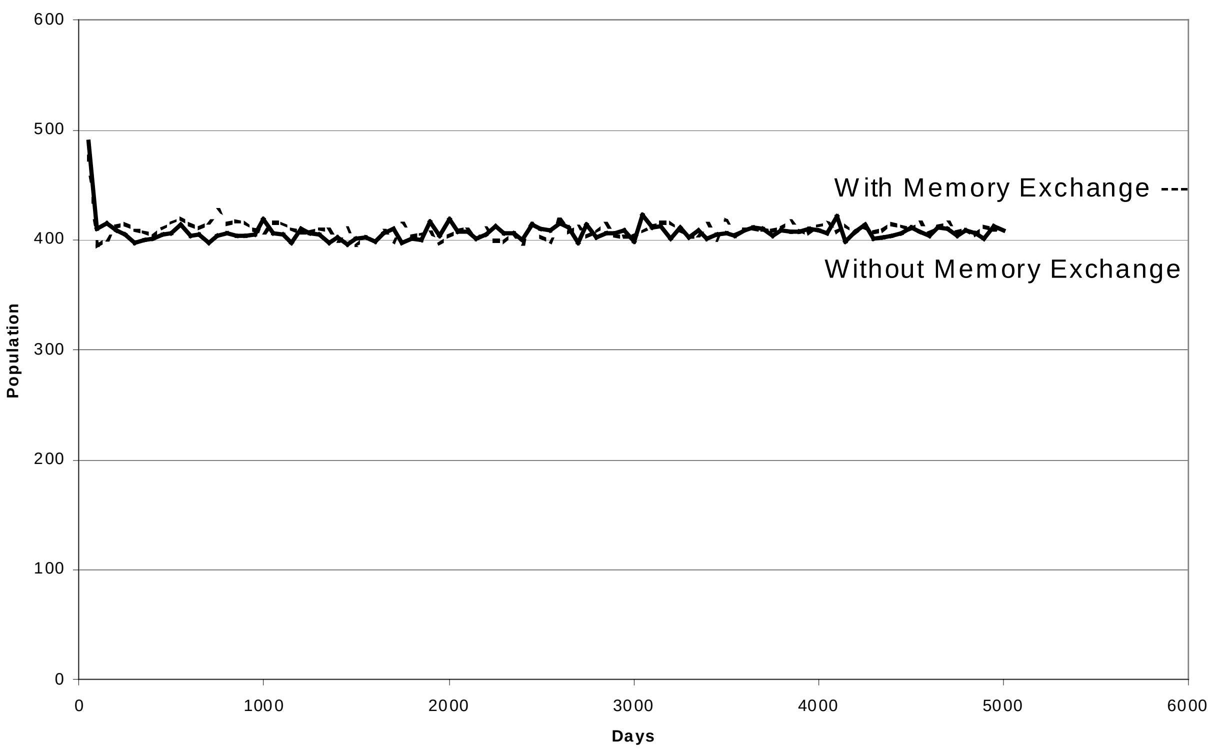

FIGURE 10 Individual Survival Rate with Information Exchange

the options. This suggests that having the ability to symbolize or abstract information is not sufficient to guarantee any improvement in decision-making performance. The abstraction process does not guarantee that all individuals will share access to the information during decision making. To do this would require the sharing of all recorded icons among members. But in the absence of any generalization activities by individuals, the shear bulk of knowledge needed for exchange would be infeasible here. What is needed is a cultural framework in which individual experiences can be integrated and generalized to allow more efficient storage, retrieval, and decision making.

One interesting pattern emerges across all trials. Under every resource-sharing strategy and variation on information exchange, groups exhibit territoriality. As a group gathers information about resources, it tends to develop an area in which it moves. Sometimes a group crosses into the territories of other individual groups, but eventually a stable pattern emerges in which single groups, or a small number of groups, establish control of a territory. This is best understood by viewing the dynamic model display. Figure 11 shows a portion of the display containing three territories: two of the territories are occupied by a single band, and the third is occupied by two bands.

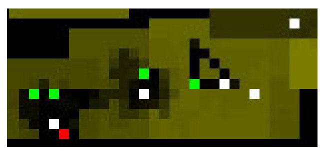

FIGURE 11 Territories Containing One or Two Bands Each

CONCLUSIONS

The goal of these experiments was to assess the advantages and disadvantages of the pre-hominid decision-making model (vector voting) in a complex, patchwork environment. Some of the positive results are:

- The pre-hominid group attains a level of population that is close to carrying capacity for patchwork environments. It outperforms random walk decision making in environments with nonrandom concentrations of resources.

- The pre-hominid group exploits both static resources and dynamic resources — the latter to a much lesser degree.

- Standardized spacing of groups and their associated territories tends to emerge in environments with nonrandom distribution of resources.

- Static resource-sharing strategies that reward individual decision makers for their successes outperform altruistic approaches when little or no sharing of memories occurs between individuals.

- A dynamic version of the best static resource-sharing strategy is an improvement over its static counterpart.

- Adding cognitive differences to individual members improves group performance, perhaps because it adds new information that can be used to discriminate between choices.

On the other hand, the pre-hominid group was less successful in exploiting certain situations:

- Limited communication of memories had little impact on decision-making performance. This suggests that larger, more substantial changes in conceptualization are required to support large gains in performance.

- It is difficult for these groups to exploit mobile resources that exhibit relatively large-scale, complex patterns of movement. These patterns require cooperative strategies and perhaps some division of labor, neither of which is possible here. Exploitation of mobile resources locally is a more suitable alternative here.

- Performance of the vector voting model and associated resource-sharing strategies is sensitive to changes in resource distribution.

In summary, this paper has extended the agent-based pre-hominid model to incorporate three additional elements of real-world hunter-gatherer populations: hunting of mobile resources, dynamic reranking of individuals, and symbolic communication. Our next step will be to allow for the emergence of shared culture among the group members and to investigate its impact on resource acquisition performance.

REFERENCES

Goodhall, S., R. Reynolds, and R. Whallon, “Altruism, Selfishness, and Survival: An Agent-based Model of Sharing Behavior,” *2002 World Congress on Computational Intelligence.*

Kummer, H., 1968, *Primate Societies*, New York, NY: Aldine Press.

Reynolds, R.G., 1978, “On Modeling the Evolution of Hunter-Gatherer Decision-making Systems,” *Geographical Analysis* **X**:31–46.

Reynolds, R., R. Whallon, and S. Goodhall, 2000, “Transmission of Cultural Traits by Emulation: A Vector Voting Model,” *Proceedings of the Workshop on Simulation of Social Agents: Architectures and Institutions*.

Reynolds, R., R. Whallon, and S. Goodhall, 2001, “Transmission of Cultural Traits by Emulation: An Agent-based Model of Group Foraging Behavior,” *Journal of Memetics — Evolutionary Models of Information Transmission*, **4**. Available at http://www.cpm.mmu.ac.uk/jom-emit/2001/vol4/reynolds_r&al.html.

THE AGENT-BASED SIMULATION OF THE EVOLUTION OF ARCHAIC STATES

R.G. REYNOLDS, Department of Computer Science, Wayne State University
A. LAZAR, Department of Computer Science, Youngstown State University
S. KIM, Department of Computer Science, Wayne State University*

ABSTRACT

We investigate the role that warfare played in the formation of the network of alliances between sites associated with the formation of the state in the Valley of Oaxaca, Mexico. A model of state formation proposed by Marcus and Flannery is used as the basis for an agent-based simulation model. Agents reside in sites, and their actions are constrained by knowledge extracted from the Oaxaca Surface Archaeological Survey. The simulation is run with two different sets of constraint rules for the agents. The first set is based on the raw data collected in the surface survey. This set represents 79 sites and constitutes a minimal level of warfare (raiding) in the valley. The other site represents the generalization of these constraints to sites with similar locational characteristics. This set corresponds to 987 sites and represents a much more active role for warfare in the valley. The rules were produced by a data mining technique, decision trees, guided by genetic algorithms. Simulations were run using the two different rule sets and compared with each other and the archaeological data for the valley. The results strongly suggest that warfare was a necessary process in the aggregations of resources needed to support the emergence of the state in the valley.

INTRODUCTION

We are interested in simulating the emergence of the archaic state in the Valley of Oaxaca, Mexico. A state is among the most sophisticated and powerful structures that has emerged from the social evolution process. In the modern world, these states are termed “nation states,” with a government composed of a hierarchical decision-making structure where the decision makers are either elected or appointed. States are supported by various economies and can interact with each other via warfare, trade, etc. Most states in the ancient world — often called archaic states — were ruled by hereditary royal families. These archaic states exhibited great internal diversity, with populations numbering from tens of thousands to millions. They had a bureaucracy, organized religion, a military presence, large urban centers, public buildings, public works, and services provided by various professional specialists. The state itself could enter into warfare and trade-based relationships with other states and less complex neighbors.

First, a particular theory of the state formation was selected to be the focus of the implementation. Marcus and Flannery (1996) had proposed a process-based model, “Evolution without Stages,” of state formation in the valley based on the long-term interaction of “actors.” The goal of the model was to focus on how the interactions of the various actors shaped the emergent social structures over time in a manner that was independent of the archaeological phases in the valley. The prehistory of the valley is commonly divided into phases on the basis of

* *Corresponding author address*: Soon won Kim, Wayne State University, Department of Computer Science, Detroit, MI 48202; e-mail: swleok@cs.wayne.edu.

the characteristics of the pottery produced in each phase. Table 1 gives the basic phases and all of the relevant periods of social evolution in the valley.

Tierras Largas marks the beginning of early village settlement. The state emerged at Monte Alban in period Monte Alban Ia. The valley came under control of the state by Monte Alban II, and Monte Alban IIIa signaled the decline of the state and its succession by a collection of city states localized in different parts of the valley. The phases represent uneven slices through time. Traditional models of state formation in the valley have focused on modeling the changes exhibited from one occupational stage to the other. Each occupational stage represents a stable phase of occupation in a given site.

However, each stage may have formed over a different length of time and reflects the accumulated result of an underlying continuous process. The model developed here views these periods as temporal snapshots that frame an underlying continuous process. Marcus and Flannery characterize this underlying process

> ...in terms of the changing relations between the actors and the environment. In such an analysis it is the social and political institutions, not the stages, that provide the milestones along the way. Transitional periods — those brief phases of rapid evolution during which the system changed, or the actors deliberately changed it — become more crucial to our analysis than the long stable periods which gave rise to the typology of stages (Marcus and Flannery, 1996, p. 236).

The basic stages of the Marcus-Flannery model are:

1. The onset of sedentary agriculture around the time of the Tierras Largas phase led to steady population growth in the valley. This population growth was “an unintended consequence of sedentism and agriculture [that] produced a whole new set of problems for the actors to solve” (Marcus and Flannery, 1996, p. 238).

TABLE 4 The Basic Occupational Phases of the Valley

Period	Approximate Date
Tierras Largas	1400 – 1150 BC
San Jose	1150 – 850 BC
Guadalupe	850 – 700 BC
Rosario	700 – 500 BC
Monte Alban Ia	500 – 300 BC
Monte Alban Ic	300 – 150/100 BC
Monte Alban II	150/100 BC – AD 200
Monte Alban IIIa	AD 200 – 500
Monte Alban IIIb	AD 500 – 700/750
Monte Alban IV	AD 700/750
Monte Alban V	AD 1000 – 1521

2. The actors developed ritual practices focused on men's houses, fraternal orders, and large descent groups to provide a basis for the integration of this larger population.

    > Stable villages integrated by ritual, fraternal orders, and large descent groups had spread throughout the valley, displacing other settlement types. Some villages grew larger than others because self-selected leaders worked hard, accumulated valuables, led raids, and build men's houses were able to attract a lot of followers. Their leadership, however, was not transferable to their offspring; it died with them (Marcus and Flannery, 1996, p. 239).

3. As suggested above, the scale of aggregation was restricted by the lifetime of leaders in terms of their ability to attract resources. To produce larger degrees of aggregation, the span of leadership needed to be extended. This level of aggregation was achieved again "by changing the meaning of existing relations" (Marcus and Flannery, 1996, p. 240). Successful leaders were said to be descended from supernatural forces such as the earth and sky and other more aggressive forms such as earthquakes, lightning, and thunder. This belief implied that only the immediate offspring of such individuals had the right and duty to lead. As a result, leaders aggregated wealth and resources over generations and extended alliances over larger numbers of surrounding villages.

4. The changing relationships described above produced a system in which the actors were relentlessly competing, resulting in periodic outbursts of violence. In this system,

    > ...the culturally defined goals of a leader were to have as many farmers, craftspeople, and warriors under his control as possible. The two main strategies for reaching those goals were: (1) alliance building — through feasting, gift-giving, and bride exchange; and (2) warfare, mostly at the level of raiding and burning rival villages" (Marcus and Flannery, 1996, p. 240).

5. Escalation of warfare led to a major shift in emphasis on site location — from access to high-quality agricultural land to the need for defensible locations. One result was the emergence (between 500 and 200 BC) of Monte Alban, a large urban center located on a mountaintop in the center of the valley. The leaders at Monte Alban then attempted to control the entire valley. Again, this involved an even larger spatial and temporal aggregation than before, which resulted in another "change in the meaning of social relations" to support aggregation at that level. This change supported the shift from a ranked society to a stratified one by restricting the ability of those of lower rank to marry people from upper ranks. Over time, two basic strata resulted — the commoners and the elites. The pragmatics behind this shift in the meaning of social relationships was engendered by the need to incorporate other conquered, highly ranked elites into the fold via intermarriage valleywide.

6. With increased aggregation, it was no longer possible for a single individual to monitor the entire system. Marcus and Flannery state that "in order to control thousands of farmers, laborers, and warriors required that many tasks be delegated to administrative, scribal, architectural, craft, and military specialists" (Marcus and Flannery, 1996, pp. 242–243). This resulted in the formation of the state.

The state formation process described above in terms of actors and their interactions lends itself readily to agent-based simulation of social evolutionary processes using cultural algorithms. The key is that the pressures toward increased aggregation resulted in a sequence of “changes in the meaning of social relations” among the agents or actors. Each change in relations allowed for opportunities to aggregate. The basic processes involved in the model were agricultural production and exchange, craft production, trade, and warfare. In this implementation, only agricultural productivity and warfare are used explicitly in the model. Alliances are generated as a result of warfare only. Excess productivity can be exchanged between agents if an alliance or relationship exists between them.

We use rules to constrain how agents interact based on their environmental location. One rule set only allows those sites for which evidence for raiding exists directly on the surface. We call this the minimal raiding hypothesis. The other set of rules represents generalizations of the site location characteristics to other sites with similar situations, which results in a much larger set of sites that can engage in raiding and warfare. The main question here concerns the role of warfare and raiding in the formation of the complex social networks required to sustain a state. If warfare is at a minimal level, does the network of alliances that is formed and the corresponding distribution of sites have the complexity needed to support state formation? If not, what complexities are added to the association network with warfare and when do these additional complexities emerge? The results for each approach are assessed in terms of the archaeological information for the region. The results strongly suggest the presence of warfare as an important process in the production of alliance networks needed to support the emergence of the state in the valley.

USING DATA MINING FROM THE OAXACA DATABASE TO EXTRACT AGENT KNOWLEDGE

The goal of this project is to produce a large-scale, knowledge-based computational model of the origins of the Zapotec State, centered at Monte Alban, in the Valley of Oaxaca, Mexico. The state was formed between 1400 and 300 BC. While archaic states have emerged in various parts of the world, the relative isolation of the valley made the processes of social evolution more visible. In the 1970s and 1980s, the Oaxaca Settlement Pattern Project conducted extensive surveys of the valley. The location and features of nearly 3,000 sites dating from the archaic period (8000 BC) to Late Monte Alban V (just prior to the arrival of the Spaniards) were documented (Kowalewski, et. al, 1989). Several hundred variables were recorded for each site. These data are the basis for generating the knowledge used in the model.

The knowledge used to constrain the behavior of agents in the model corresponds to their abilities to conduct warfare, establish trade relations, and support specialized craft production. The presence of each of these factors was supported by the presence at a site of one or more diagnostics variables for the factor. In this section, we use the warfare factor to illustrate the general approach. The diagnostic variables for determining whether a site is a target for raiding and warfare are the presence of defensive walls, the presence of burned daub, and other evidence of burning. Sites that have one or more of those variables present in a given period were said to be positive examples for the factor. In the case of warfare, only 79 of more than 3,000 sites exhibited these variables; however, these are just the sites that exhibit this evidence on the surface. We take that as the minimum number of sites that are targets for raiding in the valley. This is equivalent to assuming a minimal amount of raiding in the valley.

On the other hand, it is reasonable to assume that if a site was excavated, much more information would be available concerning its cultural relations than is present on the surface. Therefore, we used inductive learning techniques to generalize the properties of positive sites. In other words, the goal was to produce generalized rule sets that would classify all sites associated with one or more of these rules as positive examples of the concept. For warfare, this would reflect a more active pattern of raiding between sites. The approach described here is decision tree induction.

Decision tree induction is a very good method for high-dimensional applications. It is a fast nonlinear prediction method and employs dynamic feature selection. The solution complexity is expressed in terms of the number of terminal nodes. The most complex tree covers all cases in the training data. Pruning the tree and measuring the errors in progressively smaller trees find less complex solutions. Any decision tree can be rewritten in a form of decision rule set. An implied decision rule in a tree is a complete path to a terminal node. Because these rules are not mutually exclusive, the size of the decision rule set can be much larger than the logic needed for overlapping rules. The decision tree attempts to classify all of the training examples using potentially all of the variables. Genetic algorithms (Holland, 1975) and cultural algorithms (Reynolds and Al-Shehri, 1998) were used to guide the decision tree induction process performed by Utgoff's Incremental Tree Induction (ITI) system. Here, the decision tree approach generalized the collection of sites that can actively initiate raiding activities to 987 sites.

Figure 1 gives an example of a decision tree for warfare in the Etla region for sites dated to the Rosario phase. Each path in the tree can be viewed as a rule whose conclusion is the category that labels its leaf node. If the category is labeled "1," it is a target for raiding; if labeled "0," it is not based on archaeological evidence. Both of these categories are called homogeneous or pure decisions. Also, leaf nodes are labeled by both 1 and 0. This represents the fact that no further distinction can be made between them based on the available data. Table 1 gives the rules produced by extracting each possible path through the decision tree.

Thus, there were two different measures of raiding activity in terms of predicting sites involved with raiding. A minimal degree of raiding was represented by the surface collection with 75 sites. The rules produced by the decision tree approach generalized the number of targeted sites for raiding to 987 based on the similarity in environmental location. Table 2 gives the extent to which rules based on the environmental variables predicted the original set of examples, as well as the accuracy of the rule prediction for sites in the training set. Each row represents a particular phase and associated region in the valley. The number of sites found in that region, the number of incorrectly classified sites, and the overall percentage of incorrectly classified sites follows. Clearly, the generated rules are good predictors of warfare and raiding as measured in terms of the diagnostic variables used here.

In the following section, the simulation model is run using the two different knowledge sources. One effectively represents minimal raiding, whereas the other represents raiding as a much larger component. The question of interest is the extent to which the degree of warfare

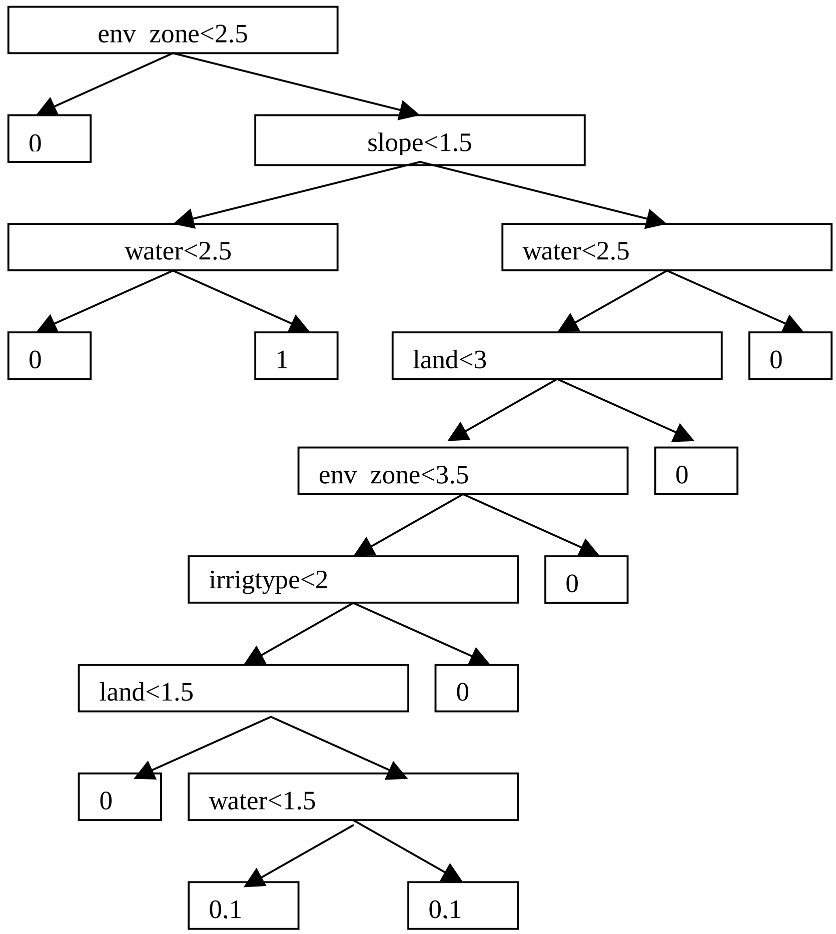

FIGURE 1 Decision Tree for the Locations of Sites with Evidence of Raiding in the Rosario Phase in the Etla Region of the Valley

TABLE 1 Decision Rule Set Induced from the Decision Tree

	Rules
1	env_zone<2.5 => decision(0)
2	env_zone>=2.5 and slope<1.5 and water<2.5 => decision(0)
3	env_zone>=2.5 and slope<1.5 and water>=2.5 => decision(1)
4	env_zone>=2.5 and slope>=1.5 and water<2.5 and land<3 and env_zone<3.5 and irrigtype<2 and land<1.5 => decision(0)
5	env_zone>=2.5 and slope>=1.5 and water<2.5 and land<3 and env_zone<3.5 and irrigtype<2 and land>=1.5 and water<1.5 => decision(0) or decision(1)
6	env_zone>=2.5 and slope>=1.5 and water<2.5 and land<3 and env_zone<3.5 and irrigtype<2 and land>=1.5 and water>=1.5 => decision(0) or decision(1)
7	env_zone>=2.5 and slope>=1.5 and water<2.5 and land<3 and env_zone<3.5 and irrigtype>=2 => decision(0)
8	env_zone>=2.5 and slope>=1.5 and water<2.5 and land<3 and env_zone>=3.5 => decision(0)
9	env_zone>=2.5 and slope>=1.5 and water<2.5 and land>3 => decision(0)
10	env_zone>=2.5 and slope>=1.5 and water>=2.5 => decision(0)

TABLE 2 Accuracy of Warfare and Raiding Predictions for Each Phase and Region during the State Formation Process

Phase[a]	Area[a]	No. of Sites	No. of Errors	Error (%)
Tierras Largas (1)	Etla (1)	15	1	6.7
Tierras Largas (1)	Central Valley (2)	2	0	0
Tierras Largas (1)	Valle Grande (3)	6	0	0
San Jose (2)	Etla (1)	19	1	5.3
San Jose (2)	Central Valley (2)	4	0	0
San Jose (2)	Valle Grande (3)	7	0	0
San Jose (2)	Tlacolula (4)	4	0	0
San Jose (2)	Ocotlan (5)	2	0	0
Guadalupe (3)	Etla (1)	24	1	4.2
Guadalupe (3)	Central Valley (2)	6	0	0
Guadalupe (3)	Valle Grande (3)	8	0	0
Guadalupe (3)	Tlacolula (4)	3	0	0
Guadalupe (3)	Ocotlan (5)	2	0	0
Rosario (4)	Etla (1)	35	2	5.7
Rosario (4)	Central Valley (2)	5	0	0
Rosario (4)	Valle Grande (3)	18	1	5.6
Rosario (4)	Tlacolula (4)	18	0	0
Rosario (4)	Ocotlan (5)	4	0	0
Monte Alban Early I (5)	Etla (1)	68	2	2.9
Monte Alban Early I (5)	Central Valley (2)	37	0	0
Monte Alban Early I (5)	Valle Grande (3)	63	2	3.2
Monte Alban Early I (5)	Tlacolula (4)	65	4	6.2
Monte Alban Early I (5)	Ocotlan (5)	13	0	0
Monte Alban Late I (6)	Etla (1)	212	2	0.9
Monte Alban Late I (6)	Central Valley (2)	153	3	2.0
Monte Alban Late I (6)	Valle Grande (3)	150	2	1.3
Monte Alban Late I (6)	Tlacolula (4)	164	3	1.8
Monte Alban Late I (6)	Ocotlan (5)	43	0	0
Monte Alban II (7)	Etla (1)	139	1	0.7
Monte Alban II (7)	Central Valley (2)	23	0	0
Monte Alban II (7)	Valle Grande (3)	104	1	1.0
Monte Alban II (7)	Tlacolula (4)	210	6	2.9
Monte Alban II (7)	Ocotlan (5)	24	0	0
Monte Alban IIIA (8)	Etla (1)	65	0	0
Monte Alban IIIA (8)	Central Valley (2)	37	0	0
Monte Alban IIIA (8)	Valle Grande (3)	313	1	0.3
Monte Alban IIIA (8)	Tlacolula (4)	490	5	1.0
Monte Alban IIIA (8)	Ocotlan (5)	133	0	0

[a] The numbers in parentheses refer to the index of the phase or period.

impacts the complexity of the network of sites formed in each simulation. If a difference exists, in which phases do the differences emerge and what properties do they have? Finally, which of the network structures produced by the model exhibits the best fit with the archaeological data?

RESULTS

The simulation was run using only warfare knowledge. Warfare knowledge was used to determine what sites could form alliances with other sites where agents resided. At this level of granularity, a site is modeled as an agent. If a site is a target for warfare or raiding, it is assumed that agents at that site can participate in alliance formation. Each of the two rule sets represents a different degree of raiding activity. In both cases, each site has the same productivity potential. However, it is hypothesized that differences in agricultural productivity throughout the valley provide the potential for the aggregation of population and resources at certain sites. The presence of warfare may allow agents to exploit that potential by forming alliance networks. While the actual networks are no longer visible, the simulation shows how the archaeological distribution of sites (their sizes and locations) reflects the underlying networks that produced them.

The model was implemented in Java and run in the Swarm simulation environment. The details of the model and its Swarm implementation can be found in Reynolds and Lazar (2002). In the experiments described here, runs were made using each of the two sets of warfare constraints for agents, one derived directly from the surface survey and other inferred using decision tree induction. The model was run starting from the earliest phase, Tierras Largas, for a period of years that lasted until phase 8. Data were extracted at points in time that corresponded to the projected middle of each archaeological phase. The first three phases of the valley exhibited a similar low level of warfare in each of the two models. At this point, we pick up the simulations in the Rosario phase. The Rosario phase immediately preceded the emergence of Monte Alban, a site located on a hilltop in a “no man’s land” between the Etla arm in the northern part of the Valley and the Valle Grande to its south. The state that was to form was located at the Monte Alban site.

The Rosario phase was characterized by an increase in raiding activity in both data sets. Figures 2 and 3 give the generated network of alliances produced with the original survey data and the inferred set, respectively, for the 35 largest sites in the valley. Several differences are notable. In the minimal warfare case, several centers in the network exhibit a relatively high degree of the number of alliances in which they participate. Although the Etla arm in the upper left exhibits slightly more complexity than the remainder, they are generally similar in complexity. Figure 3 shows a marked increase in the complexity of the alliances produced within the Etla arm relative to the rest of valley. The Etla arm is able to acquire a substantially larger population and resources aggregate with the presence of warfare. This fact is significant because it is hypothesized that the Monte Alban site was colonized primarily from the Etla arm and represents an extension of this area’s network.

In the next phase, Monte Alban Ia, the Monte Alban site emerged in the archaeological record. Figures 4 and 5 show the locations of all active sites in the valley at that time period as produced by each model, respectively. Each site is represented as a circle. Figure 6 gives the archaeological distribution of site collections observed in the valley. The site distributions

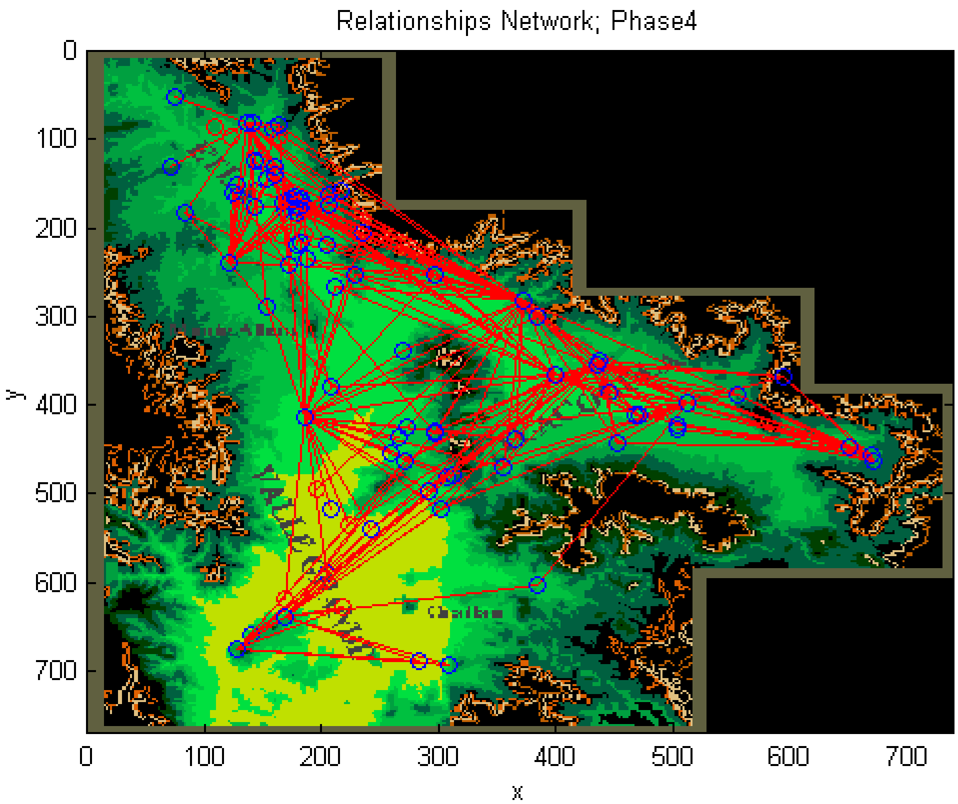

FIGURE 2 Alliance Network Formed among the Top 35 Largest Sites in the Valley in Terms of Population for the Rosario Phase Using the Minimal Warfare Rule Set

produced by both simulations (see Figures 4 and 5) exhibit a good correspondence with the site distribution shown in Figure 6. The main difference in the simulations is again the intensity of settlement in the Etla arm and adjacent areas. It is much more intensely settled with warfare than without. Again, we presume that this reflects the ability of warfare to produce alliances that attract larger aggregations of agents and resources than without warfare. As such, it is a much better fit with the archaeological data.

In both simulations, the future site of the archaic state, Monte Alban, emerged; it was unoccupied in the previous time step. Figures 7 and 8 give the alliances associated with Monte Alban in the two scenarios, respectively. What is interesting is that in the minimal warfare model, Figure 7, Monte Alban has connections in all directions up to a certain distance away. The distance may reflect limits on the ability to sustain alliances beyond that point. In the warfare case, Figure 8, it is interesting to note that few connections are made to the Valle Grande region that lies to its south. This is consistent with archaeological thought that Monte Alban was previously in a no man's land beyond the two regions of relatively high productivity. Thus, in the presence of warfare, the two regions do not exhibit the interactions shown in the nonwarfare case. This observation tends to confirm the archaeological hypothesis that intense friction occurred between the regions.

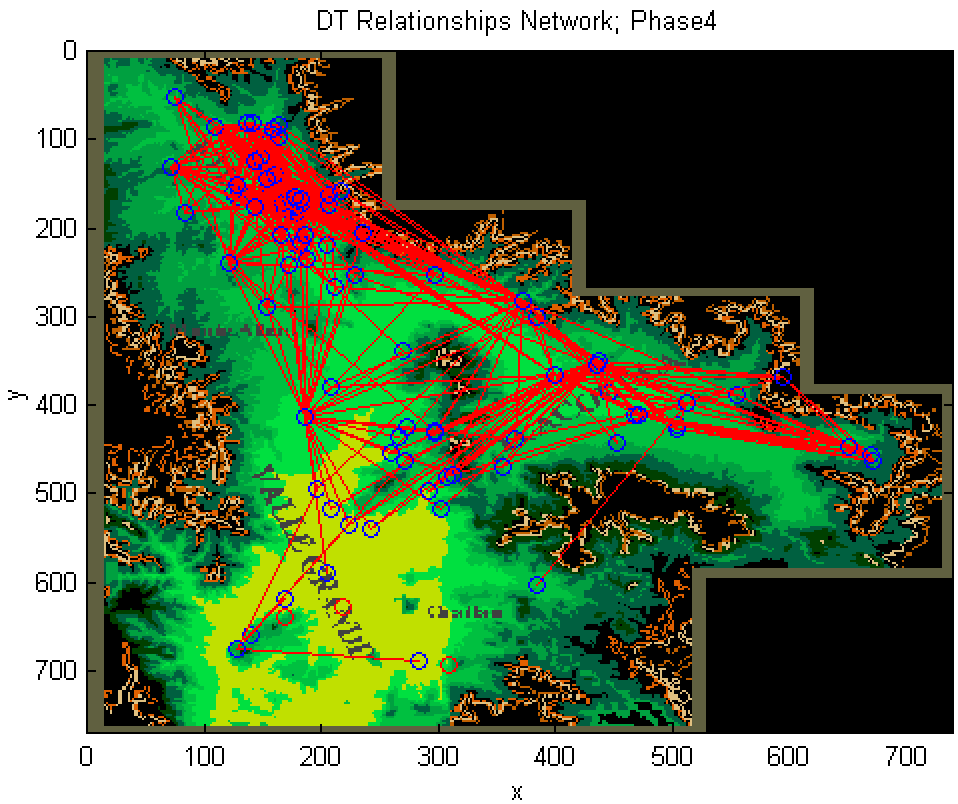

FIGURE 3 Alliance Network Formed among the Top 35 Largest Sites in the Valley in Terms of Population for the Rosario Phase Using the Generalized Warfare Rule Set

Figures 9 and 10 give the alliance networks produced by the two simulations for Monte Alban Ia (phase 5). As the warfare level increases in the valley, the differences between the two simulations tend to become much larger. The networks in Rosario are relatively similar; however, they are quite different here. In Figure 9, every big site seems to be associated with every other big site either directly or indirectly. There is little evidence for the emergence of a hierarchical organization. On the other hand, Figure 10 shows the emergence of a much more structured organization of big sites with the region of highest complexity associated with the Etla arm. The presence of highly productive agricultural land in that region produced excess resources used to generate alliance structures through warfare and foster the aggregation of resources through the use of these alliances. Less complex, but clearly complementary, networks are formed to the south in Valle Grande and to the east in Tlacolula.

Because the minimal warfare hypothesis clearly is unable to explain the aggregation of resources and agents necessary for the emergence of the state by Monte Alban Ia (phase 5), we focus only on the warfare model results for the last two phases. While it is unclear exactly when the state actually emerged in phase 5, it is known that the state has formed by Monte Alban Ic (phase 6). By Monte Alban II (phase 7), it had consolidated its grip on the entire valley. Figures 11 and 12 show the association between Monte Alban and the big sites in those

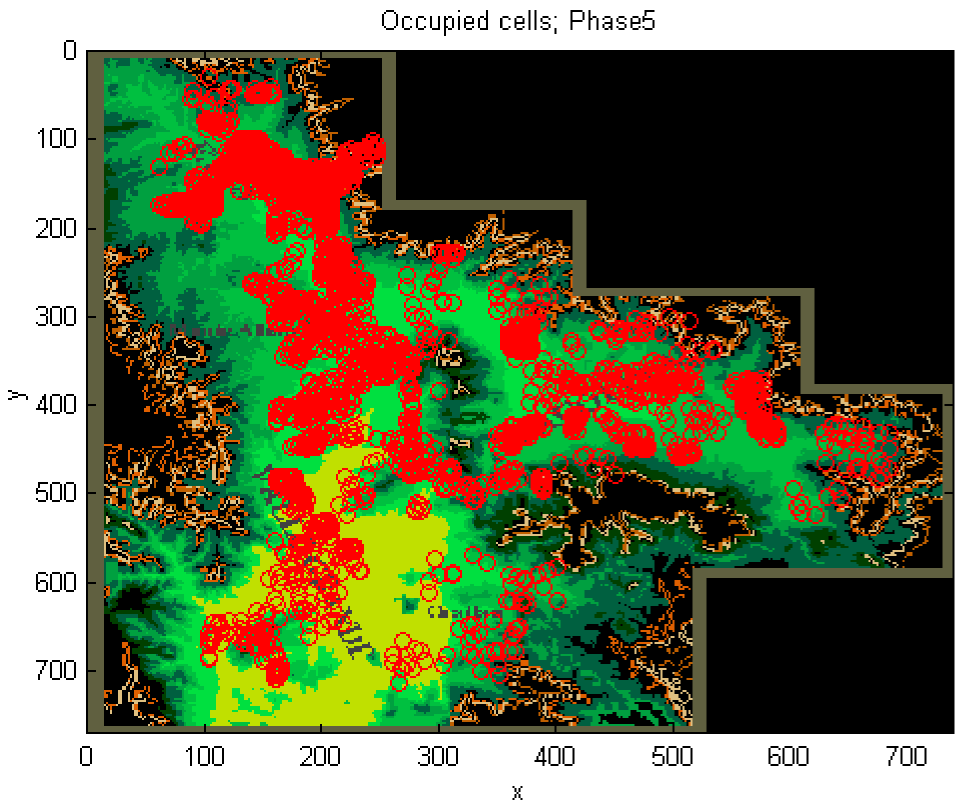

FIGURE 4 Distribution of All Occupied Sites in the Simulation for Monte Alban Ia (period 5) as Generated by the Minimal Warfare Model

two periods, respectively. In Monte Alban Ic, it now has links south into the Valle Grande arm in addition to the previous eastern connections. Sites known to have been at war with Monte Alban and conquered in this period are included in the graph as subjugates. Although the alliances now extend south for the first time, some of the nodes of relatively high complexity south of this point are still unconnected. Thus, not all of the valley has come under control of the state in terms of the alliance structure.

It is interesting when comparing Monte Albans' associations in Monte Alban Ic and Monte Alban I that no real differences in alliances between the two are evident, although the entire valley should have come under control of the state by then. This observation suggests that differential productivity and warfare, while important in explaining the properties of state emergence in the valley, are not the only factors involved. Other factors, such as craft production and trade, may have played a role in providing the additional resources needed by the state to secure these peripheral regions. In fact, it is clear that one of the practices used by the state was to place an embargo on trade with those who were in conflict with it. Monte Alban as a state became a center of craft production and, as a result, attracted additional resources and population. Likewise, external trade of its produce and craft would provide additional resources to aid in its securing of the valley. Further, peripheral sites would be more apt to be influenced by issues of trade with those outside of the valley. In a future simulation, we will place additional constraints on our agents related to craft production and trade at each site.

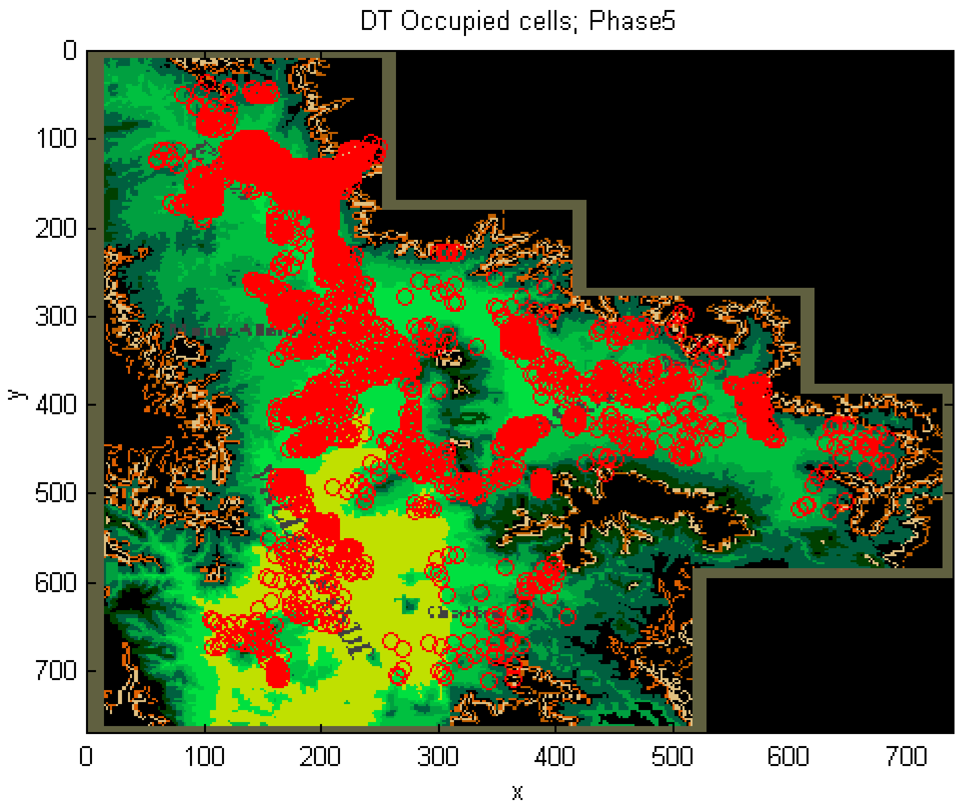

FIGURE 5 Distribution of All Occupied Sites in the Simulation for Monte Alban Ia (period 5) as Generated by the Generalized Warfare Model

CONCLUSION

In these experiments, we attempted to assess the impact of warfare on the process of state formation by looking at the impact on the generation of alliance networks that work as conduits for the aggregation of resources to be used in the formation of the state's complex social structures. The results of our experiments suggest several findings:

1. The model recapitulates many of the patterns associated with the emergent social complexity in the valley based solely on agricultural productivity and warfare concerns. Removing or reducing the influence of warfare significantly reduces the rate at which hierarchical structures emerge in the valley and the influence that Monte Alban has on the valley.

2. The results of the simulation allow us to monitor the extent of control that Monte Alban can exert over the valley in terms of the model. The results fit well with the archaeological data. However, some areas still do not come under the influence of Monte Alban in the model. These areas are often on the periphery of the valley and are engaged in trade with Monte Alban and the export of goods. We are adding rules for craft production and exchange, as well as external trade for agents. The fit of the model with these additional components will then be assessed.

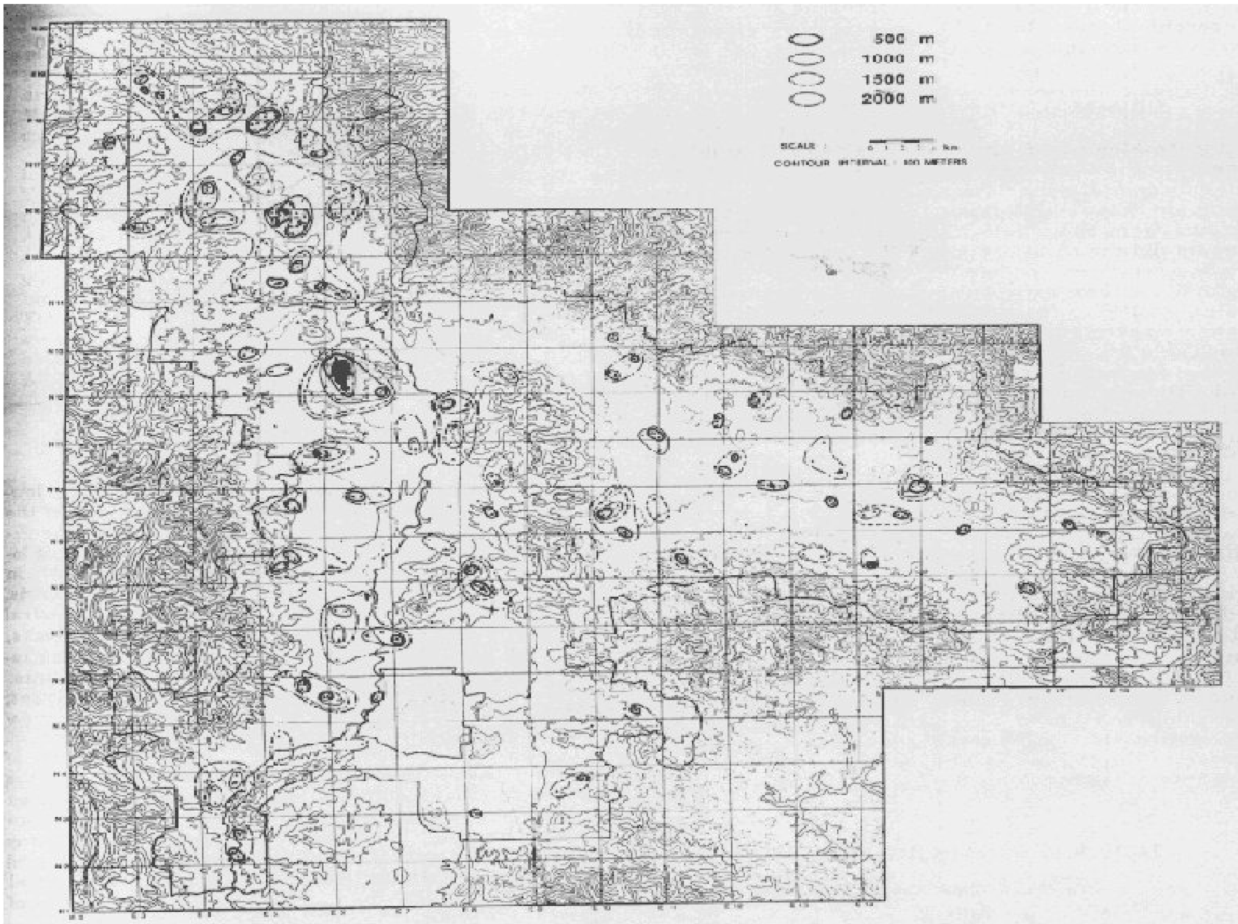

FIGURE 6 Actual Distribution of Site Aggregates for Monte Alban Ia (period 5) as Generated by the Site Settlement Survey

3. The simulation also suggests the presence of competing networks in the valley. Thus, the collapse of the state at Monte Alban in Monte Alban III might be profitably examined relative to weaknesses in its network structure that were exploited by its competitors. In other words, Monte Alban was envisioned to be succeeded by a number of smaller city states. The alliance networks generated here might be useful in suggesting the location and relationship between these emergent city states. In addition, it might be possible to identify the reasons and regions the larger network collapsed.

In summary, warfare clearly has an important role in the emergence of social complexity in the Valley of Oaxaca. Without its presence, agents have less opportunity to aggregate resources and population in a manner consistent with the needs of a complex social systems such as the state. Future work will investigate the impact of other factors such as trade and craft specialization.

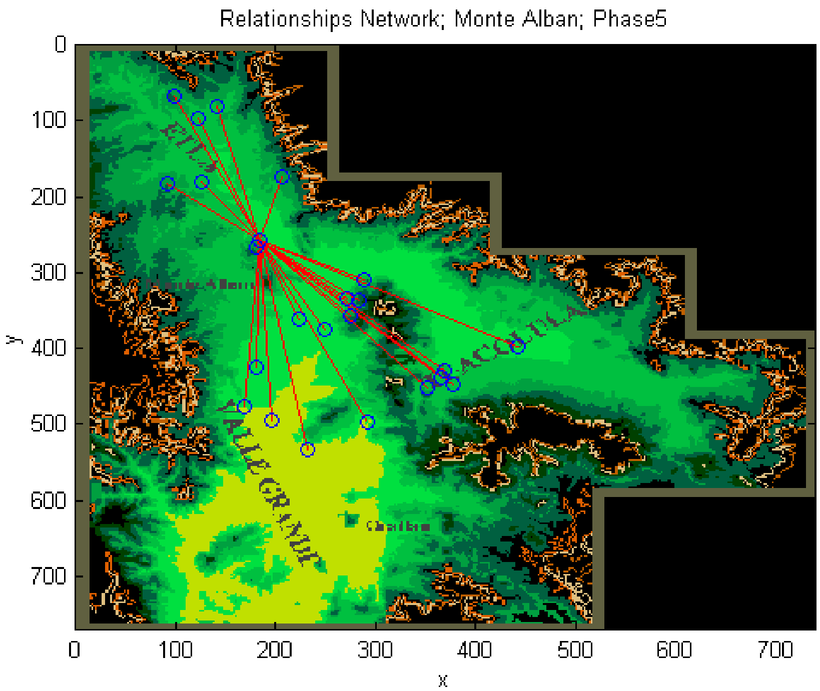

FIGURE 7 Alliances Associated with the Emergent Site of Monte Alban in Monte Alban Ia as Generated by the Minimal Warfare Model

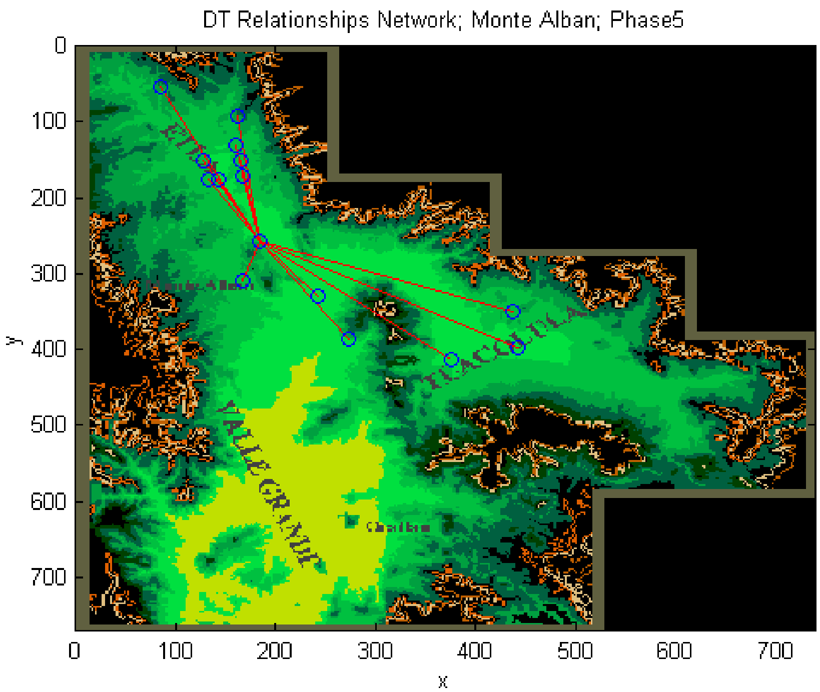

FIGURE 8 Alliances Associated with the Emergent Site of Monte Alban in Monte Alban Ia as Generated by the Generalized Warfare Model

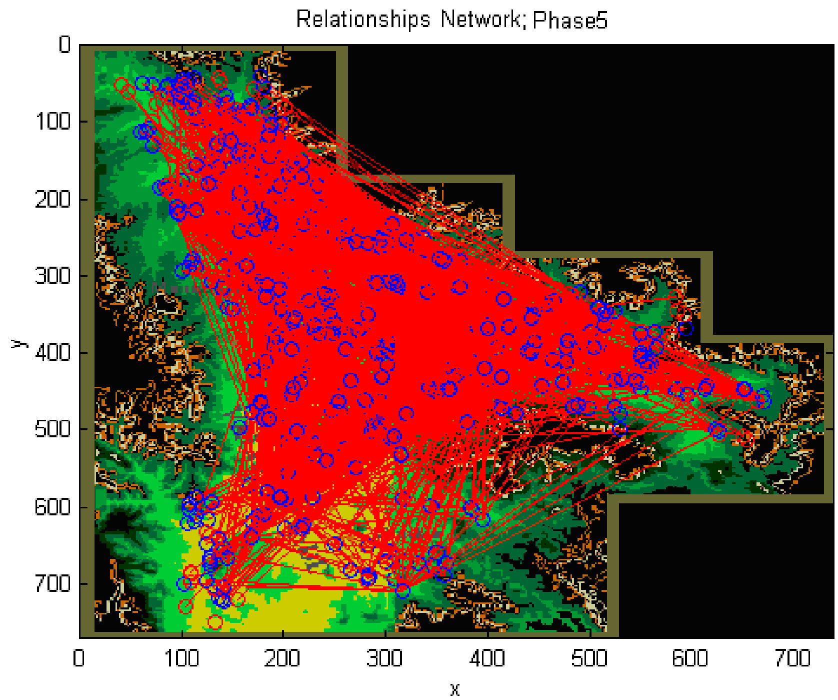

FIGURE 9 Alliance Network Generated for the Big Sites in the Valley in Phase Monte Alban Ia by the Minimal Warfare Simulation

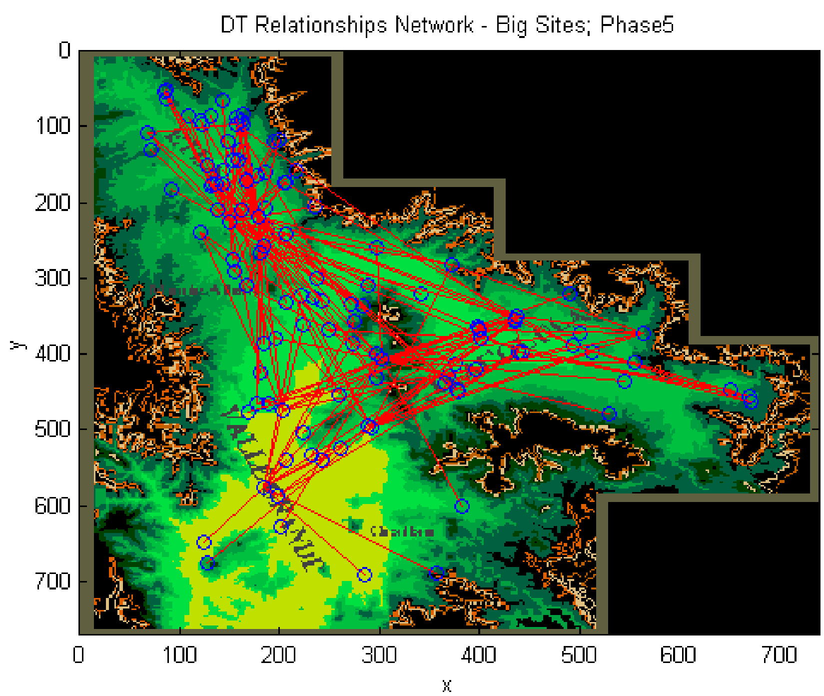

FIGURE 10 Alliance Network Generated for the Big Sites in the Valley in Phase Monte Alban Ia by the Generalized Warfare Simulation

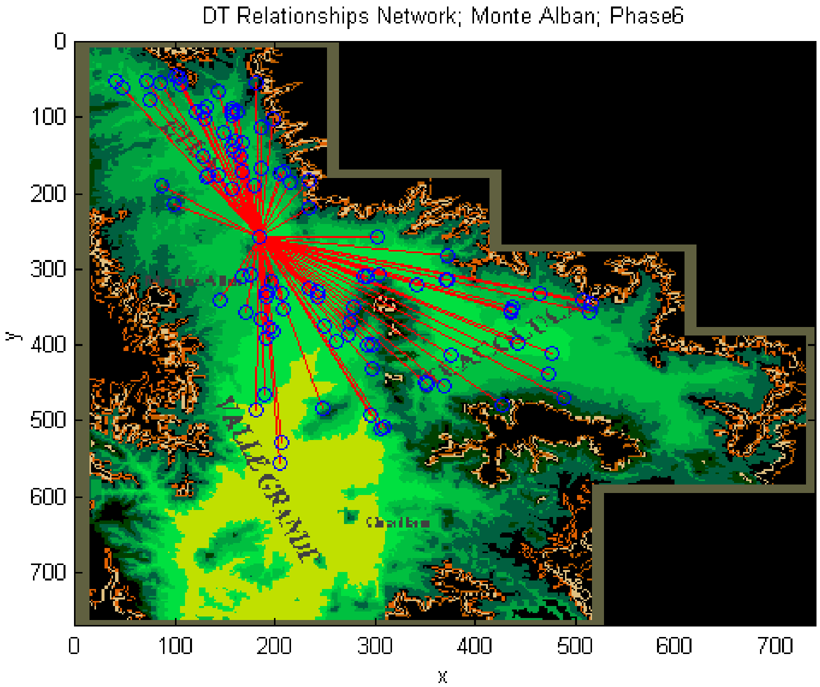

FIGURE 11 Associations between Monte Alban and Other Big Sites in the Valley in Monte Alban Ic (period 6) (Note the extension of alliances south of Monte Alban for the first time.)

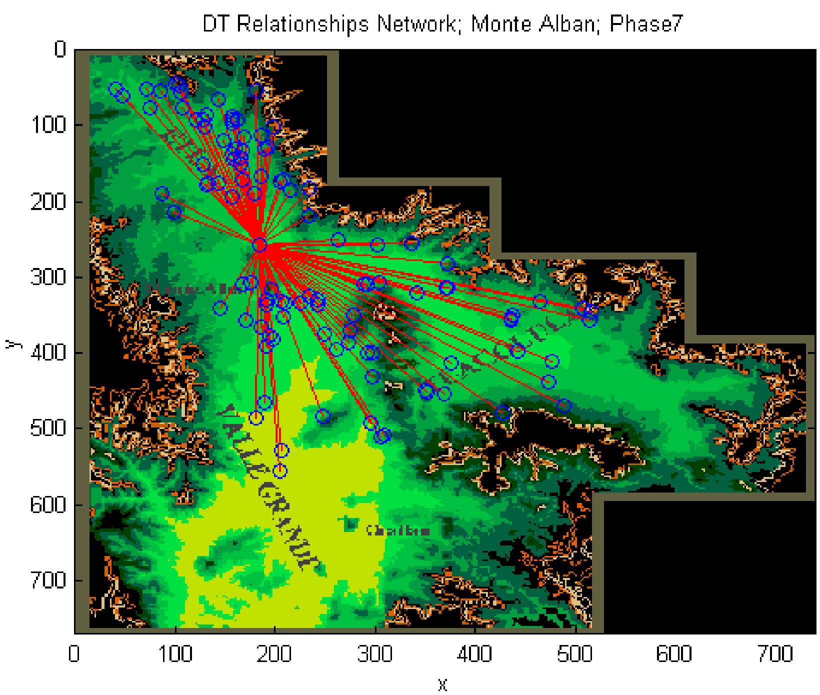

FIGURE 12 Associations between Monte Alban and Other Big Sites in the Valley in Monte Alban II (period 7)

BIBLIOGRAPHY

Blanton, R.E., 1989, *Monte Albán Settlement Patterns at the Ancient Zapotec Capital*, New York, NY: Academic Press.

Blanton, R.E., S. Kowalewski, G. Feinman, and J. Appel, 1982, *Monte Albán's Hinterland, Part I, the Prehispanic Settlement Patterns of the Central and Southern Parts of the Valley of Oaxaca, Mexico*, The Regents of the University of Michigan, The Museum of Anthropology.

Crowston, K., 1994, "Evolving Novel Organizational Forms," in *Computational Organization Theory*, K.M. Carley and M.J. Prietula, eds., Hillsdale, NJ: Lawrence Erlbaum Associates Publisher.

Fogel, D.B., 1995, *Evolutionary Computation — Toward a New Philosophy of Machine Learning*, IEEE Press.

Fox, M., M. Barbuceanu, M. Gruninger, and J. Lin, 1998, "An Organizational Ontology for Enterprise Modeling," in *Simulation Organizations Computational — Models of Institutions and Groups*, M.J. Prietula, K.M. Carley, and L. Grasser, eds., Menlo Park, CA/Cambridge, MA: AAAI Press/MIT Press.

Goldberg, D.E., 1989, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Boston, MA: Addison-Wesley Publishing Company, Inc.

Holland, J.H., 1975, *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI: University of Michigan Press.

Kowalewski, S.A., G.M. Feinman, L. Finsten, R.E. Blanton, and L.M. Nicholas, 1989, *Monte Albán's Hinterland, Part II, Prehispanic Settlement Patterns in Tlacolula, Etla, and Ocotlan, the Valley of Oaxaca, Mexico*, Vol. 1, The Regents of the University of Michigan, The Museum of Anthropology.

Marcus, J., and K.V. Flannery, 1996, *Zapotec Civilization — How Urban Societies Evolved in Mexico's Oaxaca Valley*, London, UK: Thames and Hudson Ltd.

Reynolds, R.G., 1984, "A Computational Model of Hierarchical Decision Systems," *Journal of Anthropological Archaeology*, No. 3, pp. 159–189.

Reynolds, R.G., 1994, "An Introduction to Cultural Algorithms," pp. 131–139 in *Proceedings of the Third Annual Conference on Evolutionary Programming*, A.V. Sebald and L.J. Fogel, eds., River Edge, NJ: World Scientific Publishing.

Reynolds, R.G., 1999, "The Impact of Raiding on Settlement Patterns in the Northern Valley of Oaxaca: An Approach Using Decision Trees," in *Dynamics in Human and Primate Societies*, T. Kohler and G. Gumerman, eds., Oxford University Press.

Reynolds, R.G., and H. Al-Shehri, 1998, "Data Mining of Large-Scale Spatio-Temporal Databases Using Cultural Algorithms and Decision Trees," in *Proceedings of 1998 IEEE World Congress on Computational Intelligence*, Anchorage, AK, May 4–9.

Reynolds, R.G., and A. Lazar, 2002, "Simulating the Evolution of the Archaic State," in *Proceedings of World Congress on Computational Intelligence*, Honolulu, HI, May 12–19.

Scacchi, W., 1998, "Modeling, Simulating, and Enacting Complex Organizational Process — A Life Cycle Approach," in *Simulation Organizations Computational — Models of Institutions and Groups*, M.J. Prietula, K.M. Carley, and L. Grasser, eds., Menlo Park, CA/Cambridge, MA: AAAI Press/MIT Press.

Utgoff, P.E., 1989, "Incremental Induction of Decision Trees," in *Machine Learning*, P. Langley, ed., Boston, MA: Kluwer Academic Publishers.

Weiss, S.M., and N. Indurkhya, 1998, *Predictive Data Mining A Practical Guide*, San Francisco, CA: Morgan Kaufmann Publishers, Inc.

A HYBRID MODEL OF DECISION MAKING IN CLOSED POLITICAL REGIMES

R. BHAVNANI, Department of Political Science, University of Illinois Urbana-Champaign*
D. BACKER, Department of Political Science, University of Michigan, Ann Arbor
R. RIOLO, Center for the Study of Complex Systems, University of Michigan, Ann Arbor

ABSTRACT

We present our preliminary efforts for an exploratory agent-based modeling project where the aim is to develop a tool for intelligence analysts to employ in studying decision-making processes in closed political regimes such as Iraq, North Korea, and Syria. Our hybrid of the landscape metaphor and the rule-based system approach captures the trade-offs that leaders of closed regimes face in attempting to balance power, policy preferences, and regime support — components of a utility fitness function — as well as differences in types of leaders that can result in departures from a strict notion of rationality and utility maximization. On the basis of the results of two simple experiments concerning succession, we observe that peaks in the utility landscape can arise in surprising places away from the ideal positions of both the leader and other elite because of compromises they are willing to make. Moreover, the interplay between leaders' basic tendency to maximize utility and their risk-sensitive heuristic rules leads to a high level of instability in a regime.

1 INTRODUCTION

We present our preliminary efforts to apply agent-based modeling methodologies to the study of decision-making processes in closed political regimes such as Iraq, North Korea, and Syria. Such an approach is plausible because these processes exhibit certain classic characteristics of a complex system. Chief among these characteristics is the difficulty of mapping outcomes to precipitating factors: leaders with ostensibly similar characteristics and interests act in ways that appear to be highly conditional on subtle variations in circumstances. At the same time, given the heterogeneity of actors and the political settings in which they operate, such a system becomes impractical to model statistically (with multivariate regression), mathematically (with differential equations), or using a game-theoretical approach. Moreover, because of the secretive and insular nature of these regimes, gathering detailed, reliable information on their members, interactions, and institutional structures presents a challenge for conventional analytical techniques. Thus, we conduct computer simulations by using an exploratory agent-based model (Bankes, 1993; Casti, 1997), using a hybrid of the landscape metaphor (Kollman, et al., 1992, 1998; Axelrod and Bennett, 1993) and the rule-based systems approach (Holland, et al., 1986).

Our model explores the trade-offs leaders face in altering the composition and policy stances of the regime to strengthen support from other elites, while at the same time keeping sight of their own preferences and power. The dimensions of the landscape correspond to an

* *Corresponding author address*: Ravi Bhavnani, Department of Political Science, University of Illinois, 361 Lincoln Hall, 702 South Wright Street, Urbana, IL 61801–3696; e-mail: rvib@uiuc.edu.

issue space, modulated by the distribution of power and interests among the agents. The basic goal of agents is to maximize their utility, as a function of the power they enjoy as well as the proximity of their preferences to those of the coalition to which they belong; leaders also care about support for their regime. Our objective is to explore how different levels of rationality or strategic play on the part of leaders affect the dynamics of regime survival.

Section 2 outlines an archetype of closed regimes, focusing on characteristics that are central to the structure and mechanics of decision-making processes. Section 3 discusses our efforts to develop a model of these processes, as well as our expectations regarding its evaluation and application. Section 4 describes the model. Sections 5 and 6 present and discuss some preliminary results from our first exploratory runs of the model, focusing on scenarios of leadership succession.

2 AN ARCHETYPE OF CLOSED REGIMES

Closed regimes share at least four characteristics that have a direct bearing on decision-making processes: concentrated power, an authoritarian command structure, high threat perception, and extreme secrecy. Individual regimes display these characteristics to varying degrees and can be distinguished by other characteristics, but we consider these four characteristics to be fundamental to our modeling exercise.

Power is concentrated in a closed regime, such that usually the regime consists of a small number of key decision makers (e.g., a paramount leader or the members of a ruling council or military junta). Whether one, several, or a group of individuals occupy positions of authority, they typically monopolize decisions on all matters, including political, military, economic, social, and cultural. This situation magnifies the importance of these key individuals, each of whom is likely to have a significant area of jurisdiction and considerable resources. Consequently, a suitable model of decision making in this context could logically contain relatively few agents. Yet, they would almost inevitably be specified in a manner that is more complex than a traditional agent-based model, both in terms of the number and the variety of traits as well as the actions they can take.

Related to the concentration of power among a small coterie of elites is the lack of institutional barriers to decision making. Often no consequential distinctions are evident among executive, legislative, and judicial authority: they are effectively part of a single command-and-control institutional structure, which has several implications. Most notably, one or a few powerful agents are in a position to quickly implement their preferred policy choices and to execute other significant decisions. In fact, formal decision-making processes might not be in place, or those that nominally exist might be trumped by informal norms and arrangements, if not the edicts of key actors. As a result, closed regimes are capable of taking swift action in many realms, not least political and military affairs, even though they may rank low on conventional measures of state capacity.

The inclination to consolidate authority is also inter-related with a heightened sense of insecurity among members of closed regimes. Such anxiety and uncertainty have several relevant by-products. First, cohesion among the governing elites tends to vary, thus invalidating assumptions of unitary action and undermining the prospect of stable, long-term, cooperative arrangements. In fact, punitive actions are probable within the ranks in the form of abrupt changes to the governing elites. In general, closed regimes are prone to aggression, either reactive or proactive, against challengers to the authority of the leader and other key figures. Consistent with this orientation, these regimes often display a well-ingrained political, religious, or ethno-cultural identity that emphasizes confrontation with and endorses the exclusion of nonadherents and other perceived outsiders. With this mindset, the members are prone to misperceptions, including overstating the extent of threats to authority. Responses, in turn, are highly subjective and can even appear irrational (e.g., retaliatory measures that immediately backfire).

In addition, leaders of closed regimes typically seek to insulate themselves from potential risks by restricting the flow of information. As observed earlier, the resulting secrecy and concealment complicate analysis. Unconventional means are often necessary to compile relevant details about leaders, events, and institutional arrangements; yet, our understanding of these aspects is inevitably incomplete or uncertain, if not speculative. From a modeling perspective, the information constraints also imply that agents, including those in positions of authority within the regime, are rarely equipped with the comprehensive information required to make fully rational decisions all the time. Rather, with at least some regularity, they (1) rely on proxies or guesses, (2) limit the number of options they consider, and/or (3) discount the future. We would expect, therefore, that they are subject to misperceptions and other miscalculations that reflect the noisiness of the information at their disposal, leading to mistakes in judgment.

3 MODEL DEVELOPMENT, USE, AND EVALUATION

3.1 Model Development

The process by which we approach our task merits further discussion, not least because of the unusual context. From a substantive perspective, we are dealing with an information-poor setting involving a small number of agents who are generally in a position to assert their will and also are prone to aggressive and seemingly irrational acts. Yet, the practical objective is to build a tool that helps intelligence analysts to fill in some of these gaps, to build and challenge their intuitions, and to attempt to make sense of various complex behaviors and processes.

At the outset, we concluded that far too many factors affect decision making in closed regimes for anyone to generate realistic (and defensible) predictions of the type, "a regime will go to war with 100% certainty," much less "it will invade its neighbor on Tuesday." Hence, we do not engage in a data-mining exercise to find a model that best fits a particular case or a set of cases, nor do we endeavor to construct a consolidated model. Rather, we develop an exploratory model to capture the key characteristics and behaviors of members of closed regimes (i.e., how elites respond to changes in the political, economic, and social environment, as well as each other's actions). Our objective is relatively modest: to afford additional insight into dynamics that *might* lead to a particular result or engender a certain phenomenon, with an emphasis on the *process* rather than on the outcome.

We build the model from the ground up, relying heavily on input from analysts as to what is important in terms of both foundational characteristics as well as problems and contexts that deserve in-depth study. We underscore the distinction between the direct use of such data in analysis compared with translating it into a more abstract form that shapes our conceptualization and implementation of a model. Our approach follows the latter route (i.e., harnessing the analysts' distinctive knowledge and expert interpretations to map between empirical data and our stylized representation of closed regimes, and vice versa).

We address this task in a number of ways. First, we gather pertinent details on each of the standard components of an agent-based model. At the broadest level, we ask a selection of intelligence analysts to identify (1) which factors are salient, and in what respect, to their evaluation of closed regimes; (2) which factors could reasonably be combined into general properties; (3) which factors could be considered exogenous, at least under particular circumstances or during certain time frames; and (4) which factors could be omitted from the model without much loss in precision. With respect to *agents*, we request information on the following:

- Number of key decision makers;
- Their positions/roles/status;
- Generic traits (e.g., beliefs, motivation, background, interpersonal decision style, degree of paranoia, proclivity to use violence) and unique characteristics (e.g., agent with a life-threatening illness);
- Constraints faced, particularly vis-à-vis obtaining information from other agents and the nonagent environment (e.g., specific agents might have more knowledge regarding military issues);
- Heuristics used to process information and make choices;
- Actions agents can take (e.g., commands, messages); and
- Agents' capacity to affect outcomes (e.g., control over the security apparatus, ability to influence economic policy).

In terms of the *connections among agents*, we focused on four factors:

- Which agents interact with whom and how often (e.g., formal vs. informal regime structures, fixed vs. dynamic structures, military junta vs. religious tribunals);
- The nature of decision making (e.g., degree of consensus, relative weight of electorate vs. selectorate);
- The information environment (e.g., sources and accuracy of information, degree of conflicting/reinforcing information, level of monitoring within and beyond member of inner circle); and
- The level of support along the chain of command (e.g., loyalty, compliance, defiance).

Notable *environmental factors* included stable aspects of the military, economic, political, and social realms, as well as shocks, triggers, and other short-run variables.

Second, with an eye toward capturing analysts' understandings of closed regimes, and to the extent possible formalizing where such structure does not yet exist, we request information on the sorts of mental models analysts currently employ in their work. In particular, we ask for descriptions of the following:

- *Informal models* — perspectives that employ intuitive, issue-dependent reasoning and therefore lack an explicit, consistent structure.

- *Semiformal models* — approaches that organize information in terms of at least some specific salient factors; analytical tools such as lists of actor attributes, classification schemes, and discrete if-then rules or other heuristics attributed to the actors.

- *Formal models* — explicit, well-defined paradigms, comprised of a standardized framework, established typologies and classification schemes, proven relationships among factors, and/or concrete theories that guide information processing and analytic reasoning.

Most of what the analysts discussed fell into the first two categories, especially the second (i.e., semiformal models). As a result, there were no well-defined models that we could implement directly, but rather a large number of discrete fragments, some better elaborated than others.

Third, we consolidate this knowledge, bridge the various regional and substantive perspectives of the analysts, and usher things in the direction of developing a generic model of decision-making processes in closed regimes. With this in mind, we ask the analysts to specify and rank model variants and mechanisms and to brainstorm about the problem areas they deal with most frequently, the regime environments in which they play out, and the mechanisms that are potentially at work.[2] We then look for any indications of overlap, if not consensus, as to what is critical.

Although the basic progression has been from specific to general, several caveats are in order. First, our working assumption has always been that this area of intelligence analysis exhibits too much variety, not to mention intricacy and uncertainty, for a single, comprehensive model to be a practical objective. A more plausible outcome, therefore, would be a modular set of models, each pertaining to a specific context (e.g., succession, repression, invasion) or class of phenomena (e.g., coalition politics). Certain underlying features and parameters could conceivably be shared by more than one model. These points of commonality, however, might be limited by the necessity of employing a number of different approaches to agent-based modeling, tailored to the relevant context.

Second, we are developing a tool to aid analysts, and logically it should be calibrated and attuned to their perspective. Analysts, for example, might have a series of related hypothetical scenarios that they wish to study (e.g., how the same regime responds to different economic, political, military, and cultural triggers). That is, the basic components of the model are essentially the same, but the analytical lens is directed at separate aspects of the environment. Or analysts might want focus on a specific outcome (e.g., regime collapse) and thus need to vary

[2] Since regular meetings with U.S. Department of Defense analysts were not feasible, we also intend to rely heavily on the domain experts hired for this project. Two of the domain experts are Middle East specialists: Ellen Lust-Okar (Political Science, Yale University) and Moshe Maoz (Political Science, Hebrew University of Jerusalem).

such components as composition and structure. One of our ultimate objectives, therefore, is to ensure that the end-user has the freedom to vary agent characteristics and decision rules, organizational structures, environmental conditions and dynamics, and perhaps even adaptive algorithm parameters, as well as to repeat actions under the same or different conditions and to test strategies that push the system to its extremes.

The importance of flexibility is not limited to user specification of starting conditions or interventions during a run. Ideally, analysts should have the capacity — in terms of both access to the source code and the knowledge required to make the appropriate changes — to easily modify virtually all the model mechanisms. The utility of this capacity is multifold. The models might be too simple or abstract to be of value for analysts, thus necessitating additional elaboration and sophistication. Moreover, developers cannot possibly include all the variants that different users want to explore. By the same token, we cannot anticipate what new questions would arise as users explore the dynamics produced by a particular version of the model. Constructing an intuitive framework and then providing the user with an "open box" should ensure that at least some analysts could devise their own distinctive and individually authentic versions of the model, given only modest programming skills.[3]

In sum, our goal in this project is to build a family of models, each attuned to the analysts' understanding of individuals and their behaviors in some closed regime setting (or set of related settings). Thus, we attempt to balance the need for (somewhat) general models, which can be applied to a range of related situations, with the need for a model that is not so abstract as to become divorced from analysts' views of the world.

3.2 Model Use and Evaluation

One important advantage of the agent-based modeling approach is that the model can be validated at multiple levels — at the level of individual agent behaviors and properties as well as in terms of emergent structure, relationships, and dynamics at the aggregate or system level. Given the complexity of the processes we are trying to model, however, validation of models in this area is especially difficult.

With this in mind, analysts would map their data and knowledge about a regime into input parameters that describe the agent traits and decision rules, agent organization, environment, and adaptive processes. If, for instance, the regime was comprised of elites from the dominant ethnic group who shared military backgrounds and beliefs about an ethnically pure state, similar interpersonal and decision styles, and a proclivity for violence, then the agent traits would be set to mirror these real-world attributes. One of the challenges of creating models that would be useful to analysts is to make it easy for them to see the relationship between the relatively limited menu of model components and mechanisms, and the diverse set of factors that are believed to play important roles in the decision-making processes of closed regimes.

For that reason, our initial approach to evaluating the models involves establishing their face validity. In essence, the model should "behave" as experts would expect under a variety of

[3] This need for flexibility is a key reason why we use an object-oriented programming approach and build our model using an existing package of classes designed for constructing adaptable, powerful agent-based models (e.g., Repast).

simple conditions. We also intend to assess the validity of the model using standard, model-independent techniques, such as exploring the model's sensitivity to variations in parameter values. For example, we would want to establish that changes to a particular parameter yield the anticipated responses among agents and aggregate behavior. Not only should an association exist, but the sign, direction, and/or magnitude of this relationship should be consistent with analysts' expectations. This is not to suggest that all of the results generated by the model must match analysts' repository of prior knowledge. Rather, certain critical "facts" (e.g., *A* generally leads to *B*; *Y* never follows *X*; *Z* always increases over time) can be used to discern whether the model is consistent with reality. Many other aspects of the model may not be subject to definitive confirmation, yet they must be within the bounds of plausibility. In other words, analysts are not certain that something operates in a particular way, but it would make sense if it did. We must also allow for the possibility of counterintuitive results, despite nearly everything else being consistent with what analysts understand to be the case. Thus, our criteria for evaluating the model include whether it (1) corresponds in certain fundamental respects to the current understandings of intelligence analysts; (2) helps analysts to fill in some of their information gaps with plausible details and explanations; and (3) builds and challenges their intuitions, thereby aiding the process of understanding various complex behaviors and processes.

4 MODEL DESCRIPTION

In this section, we describe our first attempt at developing a model of decision-making processes in a closed regime. To reiterate, our efforts are informed and inspired by the needs of intelligence analysts for a simple, (somewhat) general, yet useful exploratory tool. At this stage, we use a hybrid landscape/rule-based approach for a variety of reasons.

First, the landscape metaphor nicely captures analysts' basic intuitions about closed regimes. In particular, the heterogeneity of key actors and policy options creates a complex set of alternative states. The task of the leader and other influential elites is to navigate this difficult terrain to their advantage. With the landscape approach, we can depict, both conceptually and visually, the tension between the leader's desire for power, the wish to set policy to mirror his/her own preferences, and concern for support (so as not to be overthrown). Similarly, we can represent other agents' trade-offs between acquiring power and achieving policy goals.

Second, by including a set of rule-based heuristics, we can attribute simple characteristics to the leader and explore the resulting variation in behavior. While we could capture the diversity in leadership styles by using more complicated utility functions, analysts deemed the approach of introducing heuristics to capture exceptions from "rational" (i.e., utility-maximizing) behavior to be more intuitive because individuals in these positions often operate with first-order norms or principles (e.g., survival).

5 AGENTS

The model has three sets of agents: the leader, individuals who hold formal positions in the regime hierarchy (Tiers 1 and 2), and other elites (Tier 3). The tiers are generally shorthand for the amount of influence agents enjoy in the regime, including their status and capabilities.[4] At

[4] That being said, Tier 3 agents can enjoy more power than those who hold formal positions in the regime hierarchy.

this stage in development, the tiers do not entail anything as far as relationships among agents nor are limitations imposed in terms of movement both within and across these different levels. Such considerations are reserved for future extensions.

The agents have several traits. First are preferences [0,10] and salience weights [0,1] on each of a set of M policy issues. (For the experiments reported in this paper, $M = 2$.) For simplicity at this early stage of model development, the user specifies agents' preferences and saliencies at the start of a run to reflect the "state of the regime" the user wishes to explore.[5] Second is the power, if any, agents enjoy from a formal position in the regime hierarchy (*r*_power). Third is the agent's "independent" or extra-regime power (*i*_power). All power is issue-specific to allow for the prospect that agents might be more capable of influencing particular policy decisions (e.g., military generals would be strong in their own domain but weak in others). Unlike preferences, power does not remain constant but can vary as the result of the leader's actions as well as exogenous shocks. Initial values for each power value (*r*_power and *i*_power) on each of the M dimensions are drawn from [0,1].

5.1 Agents' Utility Functions

The basic goal of all agents is to maximize their utility. The leader's utility function includes three components:

- *D*, distance between preferences and policies of the regime (weighted by the salience of the issue);
- *P*, total power (both regime-based and independent); and
- *S*, total support for the regime (the aggregation of the salience- and power-weighted policy distances of all agents, including the leader).[6]

Each of these components is weighted [0,1] as part of a linear utility function for the leader L:

$$U\_L\ (D, P, S) = \alpha_D * D + \alpha_P * P + \alpha_S * S\ .$$

[5] Our current implementation of this model allows the user to drag agents to new preference positions and to use Repast probes to change other agent characteristics during a run (e.g., to explore how particular changes would alter the utility or support landscapes). Future versions of the model may introduce endogenous variation in issue preferences and weights.

[6] In detail, total support for a regime policy position R is the sum of the support from all agents, including the leader. Support is calculated independently over each policy dimension, so we assume $M = 1$ dimensions. The support s for an agent A with policy preference I, salience W, and total power (*i*_power + *r*_power) P is $s = \mathrm{diff}(I, R) * I * W * P$, where diff($I$, R) is a function of the difference between the agents' preferred position I and the regime policy R (on the dimension in question). That is, diff(I, R) indicates how much the agent "likes" the policy (independent of the agent's salience and power). For $R = I$, diff(I, R) is 1.0, providing for maximum support, and for R maximally distant from I, diff(I, R) = 0, providing no support. We use a diff(I, R) function that has two parameters that allow adjustment from a simple linear function to sigmoid-shaped functions with various thresholds and degrees of steepness. In particular, for $d = |R - I|$, $\mathrm{diff}(d) = (1 + h^s) * (1 - d)^s / (1 - d)^s + h^s)$. For the experiments reported here, $h = 2.0$ and $s = 3.0$, which, gives a steep drop, with diff(2.0) ~ 0.5.

In the experiments reported here, all of the alpha coefficients are 0.33 (e.g., the leader cares equally about policy distance, power, and support for the regime). All other agents differ from the leader in that they do not take account of support for the regime. As such, their utility function is the weighted sum of just two components — policy distance and total power. In our experiments, they weight the components equally.

5.2 The Landscapes

Our initial focus is on the trade-offs faced by the leader of a closed regime in making personnel and policy decisions. We help the model user to understand these trade-offs by means of two $M + 1$ dimensional landscapes, one for leader utility and the other for regime support, where M is the number of policy/issue dimensions. Again, for simplicity, we consider models where $M = 2$.[7]

Thus, the x and y dimensions of the landscape correspond to two issues, each with a range of policy options [0,10]. While the specific nature of these two issues is irrelevant, it is logical that they would be at least somewhat unrelated, so that it is possible for an individual agent to have more or less extreme preferences on a given dimension, irrespective of what that agent feels about the other issue.[8] Where agents array themselves in this policy space, as well as where the leader locates the regime's position, determines the support that the regime enjoys from elites and, in turn, affects the leader's utility. On the support landscape, therefore, the z values at policy x, y indicates the support the regime would have if the regime position was set to x, y (all else unchanged). Similarly, the z values on the leader's utility landscape indicate the utility the leader would have when regime policy is set to the corresponding x, y positions.

The agents in the model generally do not know the shape of these landscapes. To know the shape, they would need to know the preferences, saliencies, power, and utility weights for all the other agents in the model. While such knowledge might be viewed as an upper bound on the "rationality" an agent could exhibit, for the domains of interest in our study, the agents have much less knowledge and are not able to do all the requisite information-gathering calculations necessary to search such landscapes.

5.3 Agents' Actions

In the experiments described here, we focus entirely on the leader's actions. All nonleader agents are essentially without volition. They are not permitted to refuse an offer from the leader to join the regime hierarchy, much less to challenge a demotion or firing. Nor do they exercise any direct control over the position of the regime, although obviously their preferences factor into the leader's decision-making style. We readily acknowledge that these actions occur in practice, and we intend to endow agents with such capabilities in future iterations of the model. For now, however, the model can be thought of as a simplification, applicable only to those settings where the leader basically enjoys absolute authority, subject only to the presence of other elites who

[7] The model implementation supports an arbitrary number of dimensions, although for landscape displays, the user can only display two at a time.

[8] One dimension, for instance, might reflect an agent's preferences with respect to religious authority.

exert some influence via the policy stances they hold, the independent power they possess, and the formal positions they occupy within the regime.

Specifically, the leader can take three types of actions. The first is to shift the policy of the regime. The second is to reallocate regime power, either by removing an agent from a formal post (“fire”) and bringing in another agent as a replacement (“hire”), or by demoting an agent and consolidating power in the agent’s own hands. The third is to do nothing on a given step. The key question is, “How does the leader decide which of those actions (limited to one) to take at each step?” In our experiments, we explore two factors that characterize the leader’s decision-making style. One is the leader’s rationality (i.e., how often and how effectively does the leader take actions that increase utility). The other is the leader’s risk type (i.e., does the leader make decisions based on simple heuristics [if-then rules] that reflect a propensity to either risk-averse or risk-taking decisions, without being guided by explicitly considering the effect those decisions would have on the leader’s utility). The risk-averse (RA0) and risk-taking (RT0) heuristics are described below. Clearly, many agent characteristics and cognitive processes could influence these decision-making styles. The model described here uses just a few mechanisms and parameters to capture at least some of the important aspects.

In short, at each time step, the leader does the following: (1) checks to see if conditions are such that RA0 or RT0 (for risk-averse and risk-taking leaders, respectively) should be executed, and if so, takes the action specified by the appropriate rule; and (2) if a heuristic is not activated, tries to make a rational decision (i.e., one that increases utility). In particular, a highly rational leader tries to find an alternative regime policy or distribution of *r*_power that increases utility. However, a less rational leader sometimes makes random changes in policy or power distribution.[9] Risk type also affects how the leader looks for higher-utility options, as described below.

- *RA0.* Risk-averse leaders are disposed to perceive that other powerful agents in the regime pose a threat to their authority. As a result, they demote (or replace) agents whose power exceeds a certain threshold.[10] In our experiments, if agents have total power equal to or greater than 75% of their total power, the risk-averse leader tries to reduce the agents’ *r*_power to the point where their total power is equal to or less than 50% of the leader’s power. As a result of this “demotion,” the leader transfers this *r*_power to himself.

- *RT0.* Risk-taking leaders, by contrast, are prone to shift regime policy to satisfy their personal preferences, regardless of whether such a move costs them as far as support for the regime. In particular, RT0 is activated when the regime’s position deviates too far from the leader’s ideal. In our experiments, the difference threshold is 5.0 (i.e., 50% of the range of possible preference values). When RT0 is activated, the leader moves the regime policy to reduce the difference between his/her preferences and the regime policy by 75%.

9 The reasons for such random moves can include a lack of information, mistaken calculation or judgment, and errors in policy implementation.

10 All of the thresholds, rates, and other values that are used in the leader’s decision processes are parameterized in the model, so they could be subject to exploration in future experiments.

- *Highly rational leaders.* Highly rational leaders find alternative regime policies or distributions of *r*_power that increase their utility. A fully rational leader (in the *homo economicus* sense) would consider all possible alternative actions and then pick the one that would maximize utility. We consider leaders who have bounded rationality; that is, they can only consider N alternatives at a time. As such, varying N could be one way to adjust the rationality of the agents. In our experiments, the leaders consider only one alternative per step. If that alternative would result in higher utility, the highly rational leader would make that change.[11] Thus, this kind of search for a higher-utility alternative state of the regime can be thought of as a kind of (weak) hill-climbing algorithm.

Each time the leader tries to find a better option, the agent has an equal probability of considering either an alternative policy or the effect of reallocating *r*_power from one agent (i.e., fire that agent) to another agent who has no *r*_power (i.e., hire that agent). When looking for an alternative policy to evaluate, the leader picks one candidate by adding Gaussian noise (mean = 0; StdDev = policySD) to the current regime policy (on each dimension). The policySD parameter controls the range of policy alternatives considered. Thus, in the experiments reported here, risk-averse leaders have a low policySD (0.05); that is, they only consider alternative policies that are close to the status quo. On the other hand, risk-taking leaders use a higher policySD (2.0), meaning that they might consider policy alternatives that are relatively far from the status quo.

When looking for alternative distributions of *r*_power, the leader randomly picks one agent with *r*_power for possible firing and one agent without *r*_power for possible hiring. The leader then calculates the utility that would result from that transfer of *r*_power and makes the change if utility is increased. The leader's risk type affects this process only for risk-averse agents. In particular, a risk-averse leader accepts hiring an agent only if that agent's total power is less than 75% of the leader's total power.

- *Low-rationality leaders.* As mentioned above, leaders can adjust their rationality (i.e., the ability to find higher utility alternative regime configurations) in many ways. We took a very simple approach: each time leaders try to increase their utility by the search process described above, there is a probability that they would instead make a random change (i.e., accept a policy or hire/fire alternative without considering the consequences for their utility). In particular, for the experiments described below, low-rationality leaders make random moves 80% of the time, whereas 20% of the time, they consider utility before accepting a change to the regime policy or *r*_power distribution.

5.4 Exogenous Shocks

We introduce random exogenous shocks to include the effects of windfalls and losses on each agent's independent power. Every time step, each agent (including the leader) has a

[11] We assume the leader can calculate the expected utility of an alternative. Adding errors of various kinds to this calculation would be another way of adjusting the rationality of agents.

probability (probExogShockRate) that the independent power (*i*_power) on one randomly selected policy dimension will increase or decrease. The amount of change is obtained by making a random draw from a zero mean, normal distribution, with standard deviation of probExogShockSD. In our experiments, probExogShockRate = 0.1 and probExogShockSD = 0.2. If a shock causes *i*_power to go below 0, *i*_power would be set to 0. Because *i*_power is bounded by 0, this exogenous shock process induces a bounded random walk on power; as a result, an agent's power slowly increases over time. Because our runs are relatively short, this long-run change in total power is not important. However, in future versions, we intend to normalize power to ensure that total power does not change (and thereby distort the balance of factors in the utility calculations).

6 DESCRIPTION OF EXPERIMENTS

Our objectives in conducting experiments at an early stage in model development are to explore the basic dynamics of the model, to identify measures and tools that will be useful for analysis (e.g., landscape displays, ability to move agents by dragging), to conduct some early "face validation" exercises to determine if the model runs plausibly, and to examine the effects of various parameters. To accomplish these tasks, we conduct two simple experiments to explore the dynamics of leadership succession in a closed regime.

Both experiments include multiple scenarios. In each scenario, the starting conditions are the same: the incumbent, who is risk averse and highly rational, enjoys strong support from Tier 1 agents (those who hold high-level formal posts in the regime and have low levels of independent power), and weak support from both Tier 2 agents (those who hold low-level posts in the regime and have high levels of independent power) and Tier 3 agents (elites who hold no formal posts in the regime but have high levels of independent power). In Experiment 1, we alter the type of the leader, whereas in Experiment 2, we leave the leader's type fixed but alter the relative balance of power between the leader and the Tier 1 inner circle compared with the other (Tier 2 and 3) agents.

Figure 1 shows the initial state of the model (i.e., Step 1) for all three scenarios of Experiment 1. The display labeled "Power" shows all 10 agents, arrayed in the two-dimensional space based on their respective policy preferences. The size and shape of the symbol used to represent an agent conveys the total power (i.e., regime-related plus independent) enjoyed in each policy dimension: the larger the symbol, the higher the power; ovals imply more power on one dimension, corresponding to the longer axis, whereas circles imply equal power on both dimensions. The leader is identified by red, those with formal positions in the regime are black, and all others are blue. In this experiment, the leader is located in the lower right corner; the red lines show "connections," which are purely illustrative at this stage, to the two Tier 1 agents. The regime's position, indicated by a green square in the upper right corner, is also included for reference.[12]

[12] We also generate a "Preferences" display, not presented here due to space limitations, which shows all 10 agents arrayed in the two-dimensional space based on their respective policy preferences. The only difference is that the size and shape of the symbol used to represent an agent conveys the salience weights that he/she accords to each policy dimension: the larger the symbol, the higher the weights; ovals imply a greater weight on one dimension relative to the other, whereas circles imply equal weights on both dimensions. The regime's position, indicated by a square, is also included for reference.

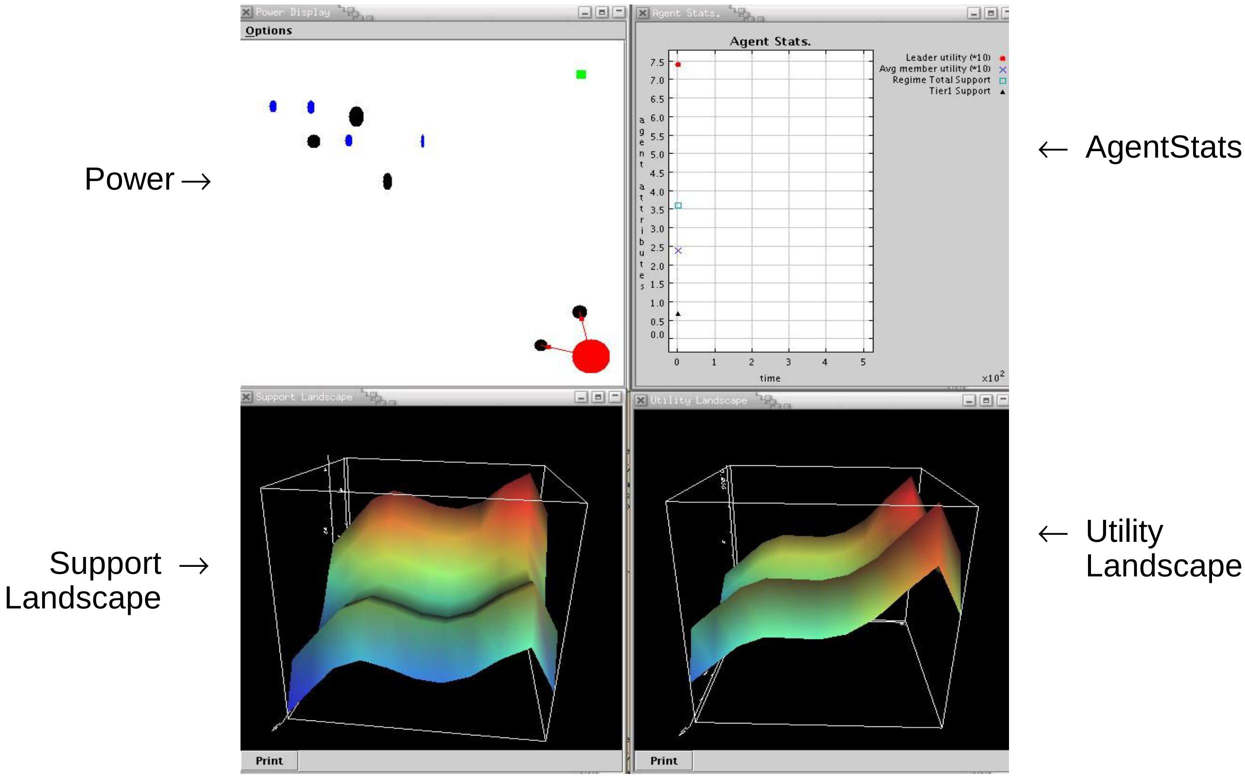

FIGURE 1 Experiment 1: Screenshot at Step 1

The display labeled "Support Landscape" depicts the support all agents would afford (the *z* dimension) for a regime position at any one of the potential locations in the policy space (the *x* and *y* dimensions) given their salience weights and preferences, as well as the power they currently enjoy in each policy dimension. It is worth noting that both local and global maxima are possible, reflecting the particular distribution of agents' preferences, the nonsymmetric nature of their salience weights, and the nonlinear shape of each individual's support function.

The display labeled "Utility Landscape" depicts the leader's utility if he/she was to shift the regime's position around in policy space (the *x* and *y* dimensions—given preferences, current power, and support). Of significance here are two maxima with very similar utilities. Both entail extreme positions on one dimension (i.e., to the far right of the landscape). The leader is effectively indifferent, however, between setting the regime's position on the other dimension to be equivalent to the leader's (and close to that of the Tier 1 agents) or accepting a compromise on that dimension that satisfies the Tier 2 and 3 agents. This circumstance can arise when there is a divergence between what the leader and the leader's close allies, on the one hand, and the other elites, on the other, consider to be the most salient issue(s). This heterogeneity of interests likewise explains why the utility landscape is not symmetric.

The display labeled "AgentStats" tracks four measures at each time step: the leader's utility, average utility of the agents, total support for the regime, and support from the two Tier 1 agents. For simplicity, we report on aggregate support and leader utility.

One objective of the simple experiments conducted is to explore the trade-offs a leader faces in attempting to balance two separate considerations:

- Maximizing utility, especially vis-à-vis obtaining more power and bringing regime policy in line with the leader's policy preferences; and

- Bolstering support, a component of the leader's utility function.

In addition, support constitutes an important measure of how likely the regime is to survive. Of course, many other measures are also of interest (e.g., the average utility of the elites represented by agents in the model, which could serve as an alternative measure of potential disenchantment).

7 RESULTS

Experiment 1 assesses the impact of different types of successors.[13] In Scenario A, the successor is far more prone to taking risks than his predecessor. That said, both individuals share a disposition to make calculated moves — and therefore display high levels of rationality — as well as the same policy preferences. In Scenario B, we introduce a successor who again is risk taking, but far less likely than his predecessor to make rational choices. In Scenario C, the successor shares the predecessor's aversion to risk, but is far more prone to making choices without regard for whether they are rational. These scenarios, which are summarized in Table 1, permit us to explore the effects of the two moving parts in the model: risk and rationality.

TABLE 1 Three Succession Scenarios

	Leader Characteristics	
Experiment 1	Risk	Rationality
Scenario A	Risk taking	High
Scenario B	Risk taking	Low
Scenario C	Risk averse	Low

Results from Experiment 1 support our basic intuitions concerning leadership types and levels of rationality. Our results, presented in Table 2, are based on averages drawn from 30 runs of the model. For each measure (e.g., leader utility, Steps 90–99), we report four numbers. To obtain these numbers, we first average leader utility over the specified 10 steps to obtain a mean m_i and standard deviation s_i over those 10 steps. In the set of four numbers for leader utility (Steps 90–99), the first number (0.7442) is the mean of the m_i's from the 30 runs; the second number (0.0025) is the standard deviation over those m_i's; the third number (0.0071) is the mean of the s_i's from the 30 runs; and the fourth number (0.0051) is the standard deviation over those

[13] At this stage, whether or not the incumbent dies, is killed, or is forcibly removed from office has no bearing on the nature of succession.

TABLE 2 Experiment 1 Results

Scenario		Leader Utility Mean	Leader Utility In-Run StdDev	Regime Support Mean	Regime Support In-Run StdDev
A					
Steps 90–99	Risk averse, rational	0.7442 (0.0025)	0.0071 (0.0051)	4.0932 (0.4357)	0.0611 (0.0444)
Steps 100–109	Risk taking, rational	0.7695 (0.0061)	0.0094 (0.0060)	3.7623 (0.4999)	0.1920 (0.1385
Steps 190–199	Risk taking, rational	0.8081 (0.0169)	0.0065 (0.0054)	4.0392 (0.5969)	0.0784 (0.1201)
B					
Steps 90–99	Risk averse, rational	0.7432 (0.0026)	0.0053 (0.0036)	4.0468 (0.4680)	0.0571 (0.0376)
Steps 100–109	Risk taking, irrational	0.7639 (0.0142)	0.0156 (0.0063	3.4713 (0.3967)	0.2769 (0.1543)
Steps 190–199	Risk taking, irrational	0.7627 (0.0206)	0.0171 (0.0065)	3.3656 (0.5129)	0.2736 (0.1605)
C					
Steps 90–99	Risk averse, rational	0.7435 (0.0032)	0.0084 (0.0059)	3.8048 (0.4211)	0.0791 (0.0344)
Steps 100–109	Risk averse, irrational	0.7427 (0.0028)	0.0070 (0.0049)	3.7495 (0.3531)	0.0761 (0.0366)
Steps 190–199	Risk averse, irrational	0.7427 (0.0017)	0.0048 (0.0025)	3.8269 (0.6307)	0.0568 (0.0258)

s_i's. The first number indicates the "average" value of the measure over the specified steps; the second number indicates how much that average varies from run to run; the third number indicates how variable the measure is within the sampled steps of 1 run (the in run standard deviation); and the fourth number indicates how much that variability differs from run to run.

Thus, in Scenario A, where a risk-averse and highly rational leader is succeeded by an equally rational, risk-taking leader, the successor's utility increases slightly immediately after the leadership change (Steps 100–109) and more substantially as time unfolds (Steps 190–199). By contrast, regime support drops after the change, though it later recovers, albeit not to the level enjoyed by the risk-averse predecessor. To explain these results, we look at the underlying dynamics more closely. The risk-averse leader remains stuck on a smaller hill in the upper right portion of the utility landscape, given that the leader only shifts policy minimally. Figure 2 displays the support and utility landscapes for this run at Step 99, just prior to succession. At Step 100, as depicted in Figure 3, the risk-taking successor assumes power and shifts policy (using the RT0 heuristic) sufficiently to increase utility, hence, the upward tick in the AgentStats display. Support for the regime falls, however, because policy now veers away from the preferences of Tier 2 and 3 agents. Moreover, this shift effectively induces more variation in support for the regime after the change. As the rational, risk-taking leader restructures the regime by hiring and firing agents to make up for the loss in support, the contours of the support landscape change dramatically to reflect the increased influence of new regime members with different preferences, salience weights, and independent power than the agents they replaced in formal posts, which effectively leads to more variation in support for the regime after the change (see Figure 4).

In Scenario B, where the successor is more prone to taking risks and less rational than the predecessor, the successor's utility increases after the change (as does the variance in utility), whereas support for the regime decreases (with a corresponding increase in variance). The

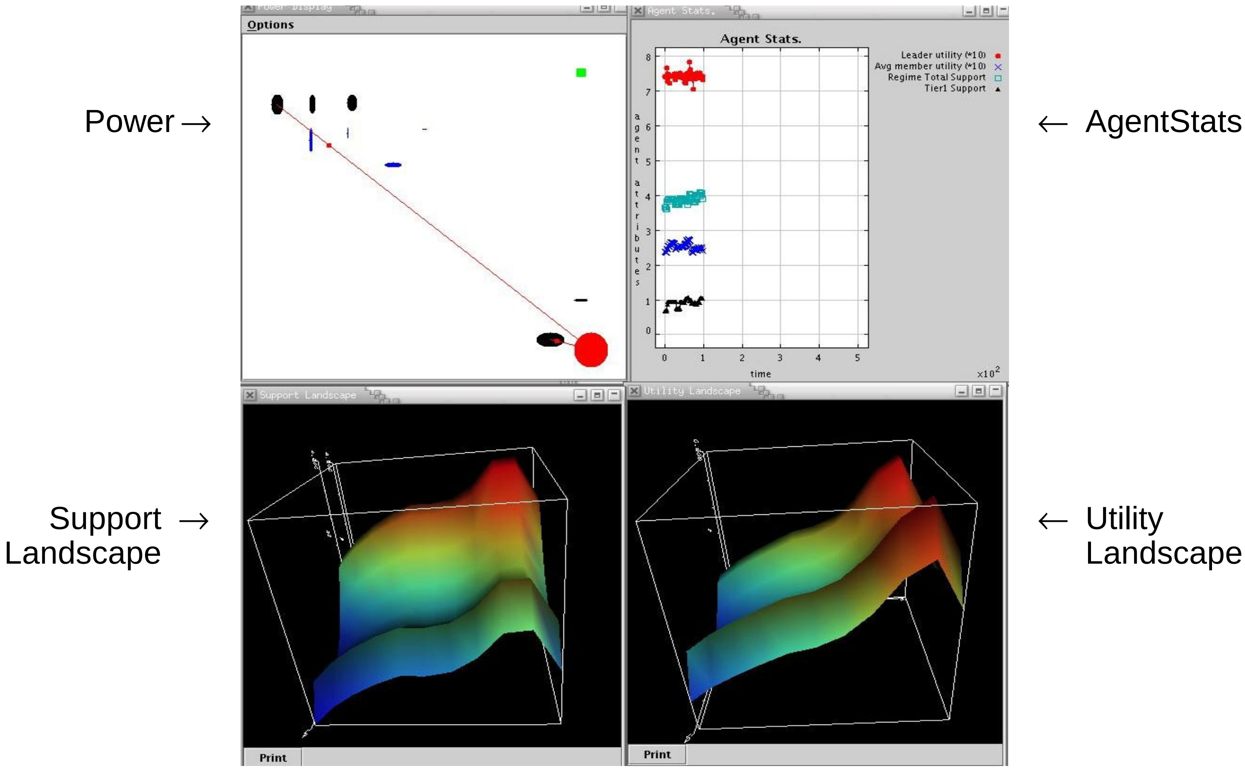

FIGURE 2 Experiment 1, Scenario A: Presuccession Screenshot (Step 99)

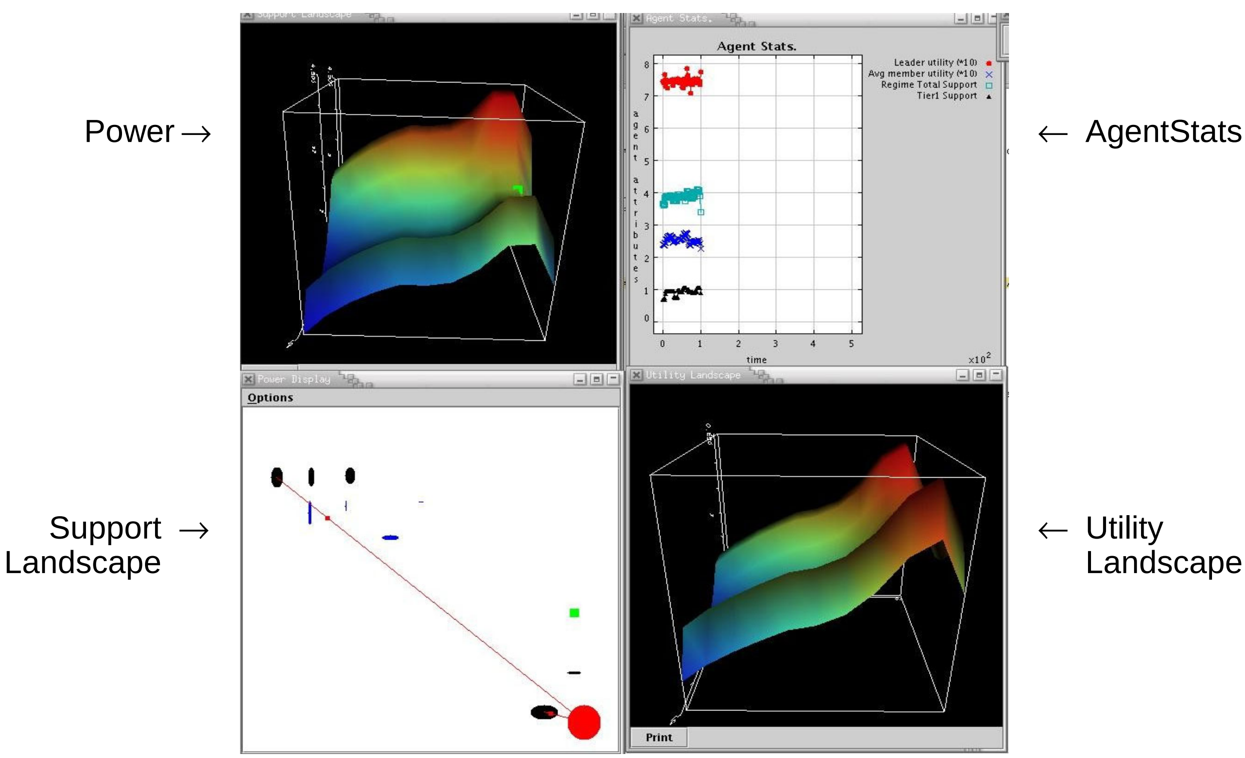

FIGURE 3 Experiment 1, Scenario A: Postsuccession Screenshot (Step 100)

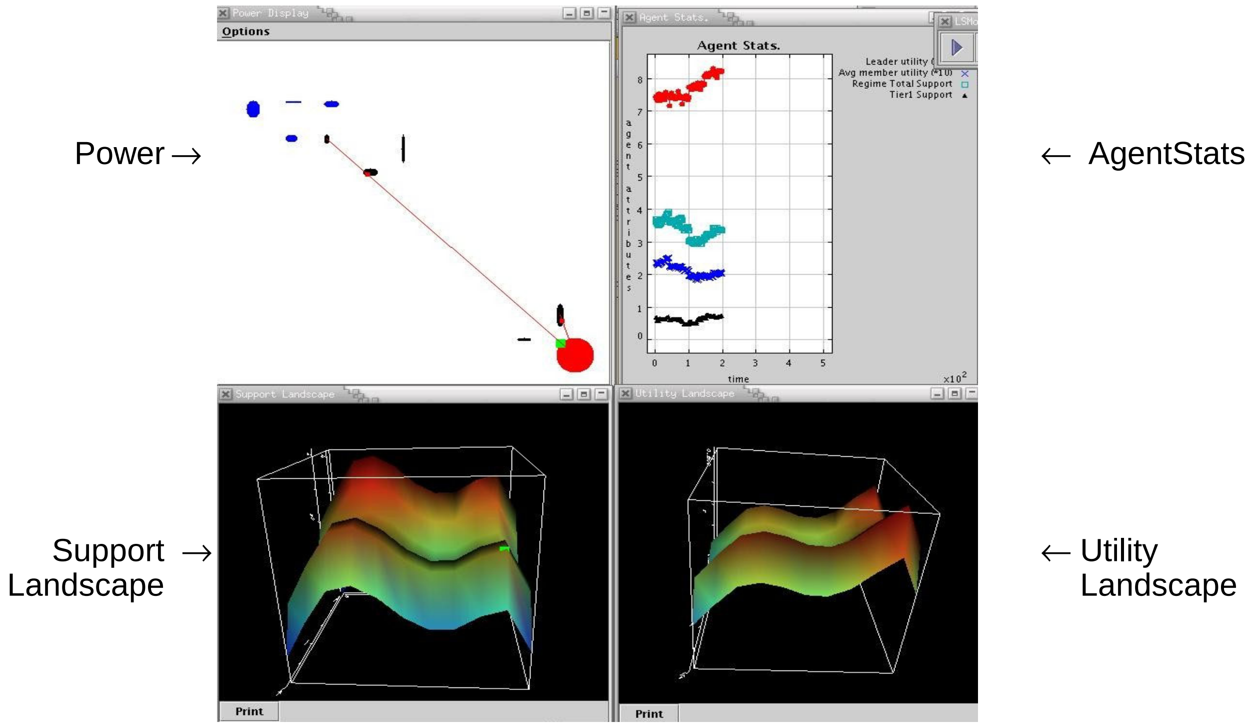

FIGURE 4 Experiment 1, Scenario A: Post-Succession Screenshot (Step 200)

successor's propensity to shift policy to mirror both his preferences and low rationality, as expressed through random changes in policy and allocation of *r*_power, accounts for the increased variation in utility and support after the succession.

Finally, in Scenario C, the risk propensity of the leader and successor is held constant; we simply examine the effect of a change in rationality. The leader's utility remains stable after the change as does regime support. The fact that the successor does not consider dramatic changes in the composition and position of the regime offsets the vulnerability in making decisions that are not to the successor's benefit.

In Experiment 2, we vary the ratio of power in the "leader's corner" (i.e., the cluster in the lower right formed by the leader and the two Tier 1 agents) to power in the upper left corner (Tier 2 and 3 agents). This permits us to further explore the tension between the leader's policy preferences and the location of the support peaks. As in Scenario A of Experiment 1, the incumbent is risk averse and highly rational and is succeeded by an equally rational, risk-taking leader. The results from Experiment 2 are summarized in Table 3. Note that a ratio of 1.00 indicates that the leader and Tier 1 agents have a combined total power equal to Tier 2 and 3 agents in the upper left corner of the landscape. As this ratio decreases, it reflects the increase in total power of Tier 2 and 3 agents.

When the ratio is 1.0, support initially peaks in the leader's corner; however, a local maximum exists in the upper right corner where policy starts. At Step 100, the jump in policy taken by the risk-taking successor (as a result of RT0) to the lower right corner therefore

TABLE 3 Experiment 2 Results

Scenario[a]		Regime Support Mean	Regime Support In-Run StdDev
Ratio of LR/UL = 1.00			
Steps 90–99	Risk averse, rational	2.5725 (0.4376)	0.0660 (0.0437)
Steps 100–109	Risk taking, rational	2.7452 (0.4243)	0.1704 (0.1347)
Steps 190–199	Risk taking, rational	3.5724 (0.6805)	0.1058 (0.1530)
Ratio of LR/UL = 0.75			
Steps 90–99	Risk averse, rational	2.7661 (0.4179)	0.0566 (0.0340)
Steps 100–109	Risk taking, rational	2.8522 (0.3271)	0.1636 (0.1434)
Steps 190–199	Risk taking, rational	3.5865 (0.6902)	0.1134 (0.1271)
Ratio of LR/UL = 0.50			
Steps 90–99	Risk averse, rational	3.4217 (0.5392)	0.0665 (0.0294)
Steps 100–109	Risk taking, rational	3.3440 (0.4810)	0.3154 (0.2603)
Steps 190–199	Risk taking, rational	4.1976 (0.7027)	0.1994 (0.1918)
Ratio of LR/UL = 0.33			
Steps 90–99	Risk averse, rational	4.0798 (0.5385)	0.0564 (0.0240)
Steps 100–109	Risk taking, rational	4.0051 (0.5317)	0.5190 (0.3241)
Steps 190–199	Risk taking, rational	4.4978 (0.6937)	0.2851 (0.3774)

[a] LR = lower right; UL = upper left.

improves support slightly just after the change, and more at Steps 199–200. As the initial ratio of power falls (i.e., Tier 2 and 3 agents have more power), the trend is toward a smaller increase or even a decrease in support just after the successor changes policy, and a smaller increase in support for the regime at the end of the runs. Also, a dramatic increase occurs in the instability of support (the in-run standard deviation) for lower in/out power ratio cases.[14]

Figure 5 presents a screenshot from one set of runs at an in/out power ratio of 0.50.[15] Support peaks in the upper left corner where Tier 3 agents are located at (and before) Step 99. Once the risk-taking successor assumes power (Step 100), the propensity (via heuristic RT0) to shift policy to reflect his preferences to the lower right corner comes into conflict with rational (utility-hill climbing) policy shifts. Thus, right after the succession, the successor's RT0 is activated, which moves policy to the lower right corner, near the leader's position. Utility-increasing policy shifts then moves policy back to one or the other of the peaks in the remaining three corners of the landscape (see Figure 6). Eventually, policy moves far enough away from the successor's preferences to reactivate RT0, moving policy back to the lower right. These

14 The leader's total power is fixed at 3.5, and Tier 1's total power is fixed at 1.5. For a ratio of 1.0, the total power of Tier 2 and 3 agents is 5.0; for a ratio of 0.33, the total power of Tier 2 and 3 agents is 15. The ratio in Experiment 1 is about 0.75.

15 The in/out power ratios reflect the initial state of the model. As the leader begins to reallocate power, and as exogenous changes in power occur, these ratios change.

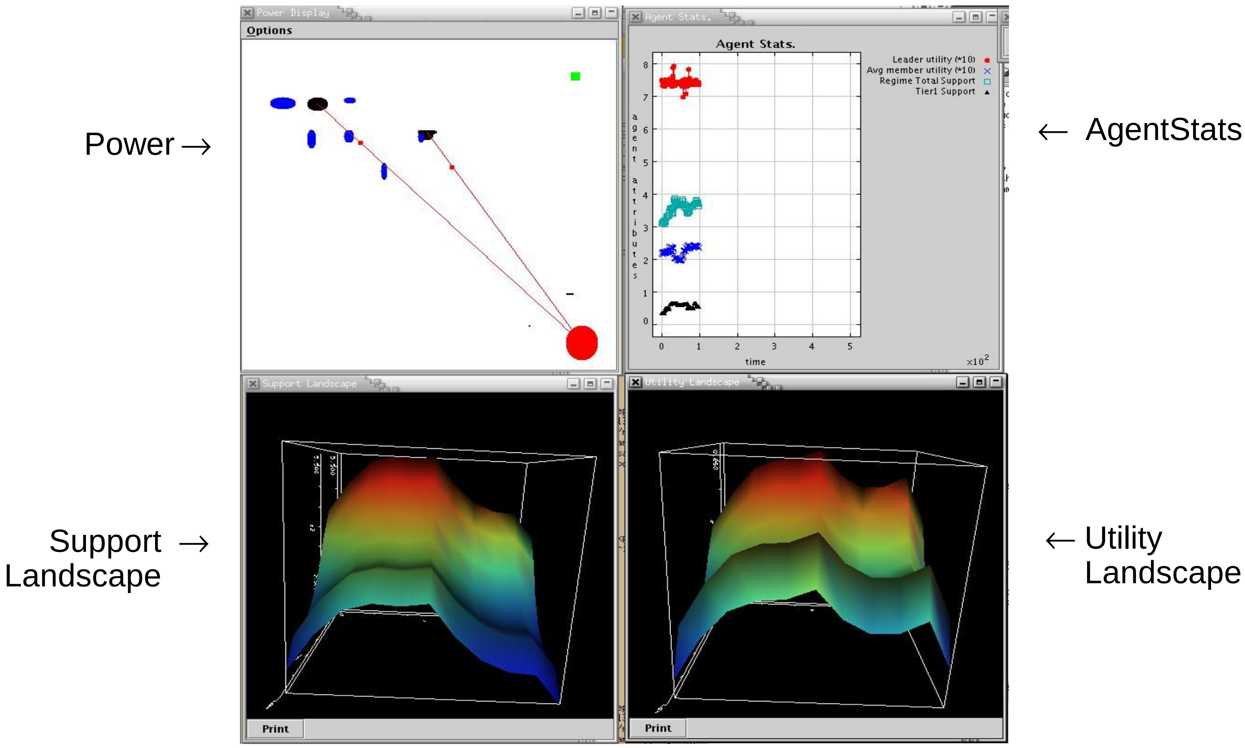

FIGURE 5 Experiment 2: Presuccession Screenshot (Step 99)

conflicting policy moves continue for the rest of the run, generating high in-run support variance (0.3154 for in/out power ratio = 0.5 at Steps 100–109). However, by the end of the run (Steps 190–199), the successor also reallocates power to mitigate the loss in support from shifting policy to reflect his preferences and to a drop in instability of policy and support.

To summarize, the leader's behavior seems plausible in our initial set of experiments, given the composition of the regime and the decision rules and behaviors the leaders have in this model. Observation of a number of runs reveals a variety of histories and outcomes, showing the kind of perpetual novelty that is a hallmark of complex adaptive systems (Holland, 1995; Axelrod and Cohen, 2000). In this model, that novelty is driven by the exogenous shocks and the resulting actions of the leader (i.e., moving the regime policy and reallocating *r*_power). Clearly, much more remains to be done to understand the dynamics of this model under different conditions. For instance, the rule for the risk-averse leader (RA0) was seldom used in our runs and thus had little effect on these results. Perhaps other parameter settings (e.g., a lower threshold for triggering RA0) or other initial allocations of power or exogenous shock dynamics would result in runs in which RA0 was more of a factor.

8 DISCUSSION

We have presented an initial look at an exploratory modeling project, where the aim is to develop a tool for intelligence analysts by applying methodologies for representing nonlinear

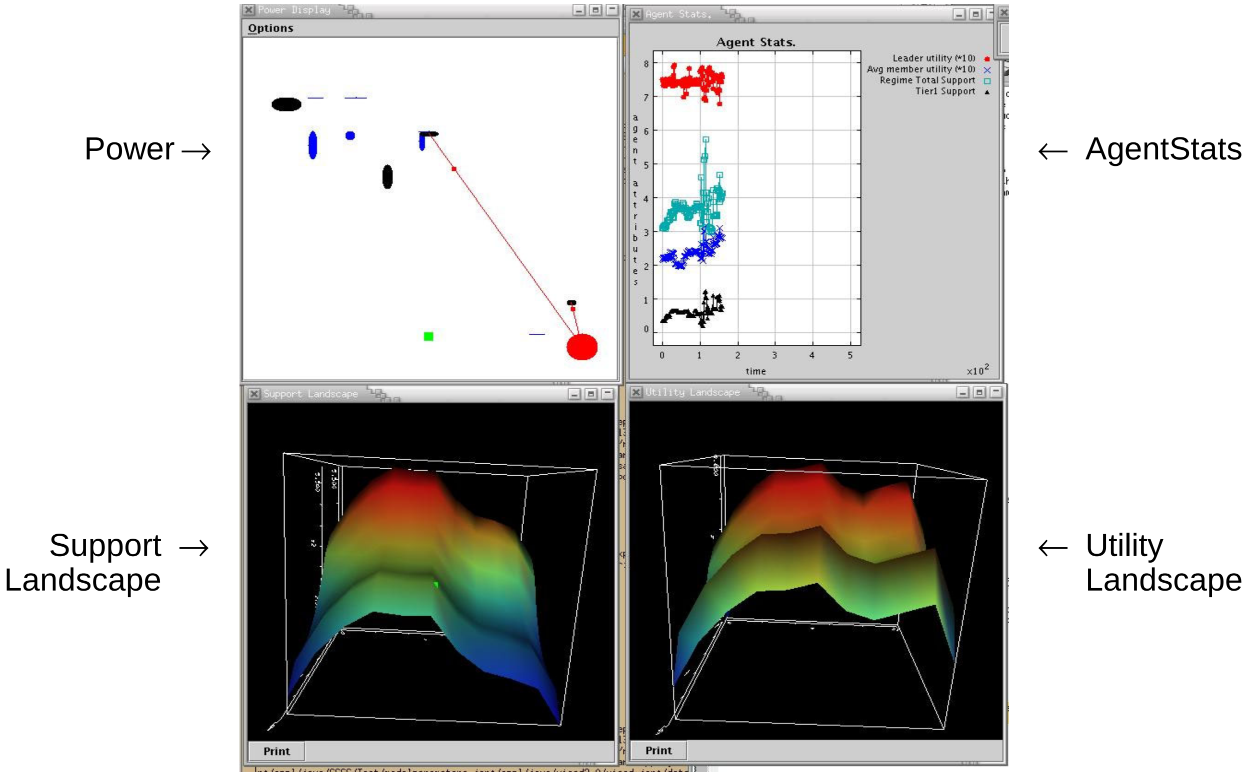

FIGURE 6 Experiment 2: Postsuccession Screenshot (Step 157)

complex adaptive systems to the study of decision-making processes in closed political regimes. We opt for a hybrid of the landscape metaphor and the rule-based system approach to capture the trade-offs that leaders face in attempting to balance power, policy preferences, and regime support — the three components of a utility fitness function — as well as differences in leaders' types, which can result in departures from strict rationality and utility maximization.

Our goals at this early stage of model development are modest: to construct a basic framework and to begin to assess its validity and relevance by means of simple experiments that seek to represent real-world phenomena of interest to intelligence analysts. The experiments reported in this paper focus on the dynamics of succession. In particular, we devise a set of simulations that looks at how changes in the leader's type, rationality, and level of strategic play via this transition of authority affect the leader's utility and the support for the regime from other elites. Our efforts are clearly not exhaustive; rather, they are merely an initial set of experiments that employ one specific version of the model. Nevertheless, even this foray yielded some interesting dynamics. Of note, we observe that peaks in the utility landscape can arise in surprising places (i.e., away from the ideal positions of both the leader and other elites) as the result of compromises they are willing to make. Moreover, the interplay between the leader's basic tendencies to maximize utility and the leader's sensitive heuristic rules also led to much more instability in the regime. The leader would sometimes change the regime's policy to make it close to his preferences, and other times change policy and reallocate power among other agents to raise utility by garnering support for the regime.

In the immediate future, we plan to continue the process of exploration and validation of this and related agent-based models of closed regimes. For the current model, we need to test the dynamics using extreme parameter settings, different arrangements of agents' preferences, different distributions of power among agents, different weights between issues and among the components of utility, and different levels of rationality, both in the context of succession and in other applied settings. Our efforts to date have been aided in no small measure by the feedback received from analysts and domain experts who observed demonstrations of earlier versions of this model. Not only did these meetings stimulate thought and provide us with an opportunity to discuss which factors are important, which are not, and how to include important factors and mechanisms in simple models, but the response also indicated that the model matched the way (some) analysts view (some) regimes, and that our notion of leadership behavior and priorities seemed plausible.

Ultimately, we intend to modify the current model to reflect the observations and ideas we have received (and continue to receive) from analysts and domain experts. Some factors and mechanisms that are not in the current model, but which were identified as crucial for understanding decision making in closed regimes, include formal and informal organizational structure, as well as making agents' power a function of the relationships they have with other agents. In addition, we expect to represent political groupings within the regime itself (i.e., factions and coalitions) as well as among those who are outside the regime, including opposition elements, to reflect influences of competing forces and alternatives to the regime.

9 ACKNOWLEDGMENTS

This paper documents work in progress on the project, "Simulating Closed Regimes Using Agent-based Models," which is funded by the U.S. Department of Defense (DoD) and administered under the auspices of the Center for the Study of Complex Systems (CSCS) at the University of Michigan, Ann Arbor. The views expressed in this paper are those of the authors alone and do not reflect the opinions of the DoD or CSCS.

10 REFERENCES

Axelrod, R., and S. Bennet, 1993, "A Landscape Theory of Aggregation," *British Journal of Political Science* **23**:211–233 (April).

Axelrod, R., and M. Cohen, 2000, *Harnessing Complexity*, New York, NY: Free Press.

Bankes, S., 1993, "Exploratory Modeling for Policy Analysis," *Operations Research* **41**:(3)435–449.

Casti, J., 1997, *Would-be Worlds: How Simulation Is Changing the Frontiers of Science*, New York, NY: John Wiley and Sons.

Holland, J., K. Holyoak, R. Nisbett, and P. Thagard, 1986, *Induction: Processes of Inference, Learning, and Discovery*, Cambridge, MA: The MIT Press.

Holland, J., 1995, *Hidden Order*, Reading, MA:Addison-Wesley.

Kollman, K., J. Miller, and S. Page, 1992, "Adaptive Parties and Spatial Elections," *American Political Science Review* **86**:929–937.

Kollman, K., J. Miller, and S. Page, 1998, "Political Parties and Electoral Elections," *British Journal of Political Science* **28**:139–158.

DISCUSSION: MACROEVOLUTION

JOHN PADGETT, The University of Chicago, Moderator

John Padgett: By way of introducing this discussion, I would like to say that I find all of these papers of more than casual interest because, as some of you know, I study concrete historical events and am frequently in the position of trying to relate the historian side of myself to my model-building side. This tension between model building on the one hand and empirical studies on the other is something that is central to all of these papers.

I would like to focus my remarks more on methodology than on theory. As I reflected on these papers, it struck me that even though I agree thoroughly with the theoretical optimism of this conference and the value of agent-based modeling, criteria for the success of these exercises remain somewhat less than clear in my mind. What defines the difference between a successful and an unsuccessful agent-based model? All of these things are interesting and of value, but how do we evaluate them? The criteria are multiple or a little murky, I think, so I wanted to use this as an opportunity to reflect on this issue. Now, please direct your questions to the speakers.

Claudio Cioffi-Revilla: Claudio Cioffi, George Mason University. I have questions for Bob [Reynolds] and John [Padgett]. My question for John is, would you consider cosmology and climatology scientific disciplines or historical disciplines, given their subject matters?

The question for Bob is very different and is specific to the simulation that he presented. First, I want to say that I am happy to see the very impressive progress you have made on this project in a very short time, particularly in terms of the domains and technologies that you have brought together in a unique way for the first time. One thing I am unclear about, though, and it is to me the central question here, is what are the explicit theoretical and archaeological — and I mean "dirt" archaeological — criteria for the identification of a state, as opposed to, say, a chiefdom? Of course, this is from Wright's and Johnson's and Flannery and company's literature. But, friendly as I am to this enterprise, if you take off the contours of the Oaxaca Valley so that I cannot tell where in the world we are, and if you do not tell me the names of the phases, which I know are identified with chiefdom and state stages, and you just show me the process with graphics, I cannot tell that a state formed, because I cannot recognize in the simulation results the diagnostics of a state in such a way that I could map them onto the theoretical and archeological indicators. Maybe I missed it, but I just did not see it.

Robert Reynolds: That is a very interesting question. Lars-Erik [Cederman] talked this morning about the four levels of sociological investigation, and your question hits right in there. In fact, I have been rethinking what I am doing in terms of his presentation, so I will answer your question in those terms. Basically, I have worked at the network level, and what you see are these networks of flows forming out; they reflect the differential distribution of resources in the valley. When I think of a state, I think of a political organization, which is often hierarchical, with specialists and decision makers. None of that is visible at this level. What we have are indicators that a state may have formed. The next level is to move down and actually simulate the formation of these political structures.

At this point, I wanted to see what the underlying network of flows might be to support a state relative to the differences in productivity. One of the things that I could point to as a success is the ability to see the relationship that Monte Albán, which is literally at the center of the state, has to other sites in the valley. If, in fact, it is completely connected to other sites, it has control in some sense over the flows of resources between them. The other thing is that Monte Albán actually became a center of complexity; it was not something that we programmed in explicitly. Of course, the distribution of resources in other potential sites made that emerge. So right now we are looking at the network level, but we want to move down to the political level.

Tim Kohler: Tim Kohler, Washington State. I would like to follow up on Claudio's remark, because I had a similar reaction in that one of the archaeological signatures we might expect for a state-level organization is a series of flows that go from local centers to regional centers, and then from regional centers to the capital of the system. In your case, all of the flows were between each site and a primary center. I was wondering why. What is it about your framework that keeps us from seeing the secondary and primary centers?

Reynolds: It often has to do with the statistics. Actually, the sites I showed you were just the big sites. In fact, we were talking about 500 or 600 sites at any given time. We were just plotting the relationships of the big sites. If you plot those out, you will see a hierarchical structuring of site size that relates to a distribution of second- and third-tier sites that are assumed to be associated with the state. I did not present that. We are still putting those graphs together, but that is an excellent point.

As John [Padgett] was saying, you produce a lot of data, and one of the things you do is sort through it and try to visualize it in some way. One of the tough things involves trying to integrate the GIS with the simulation and back again, while trying to sort through the complexity of getting the 35 big sites. If we put all the sites in, it would be "pizza" again. It is quite a challenge to find some structural patterns that emerge and reemerge over and over and then try to understand and represent them.

One of the implications of your question, which I think is a challenge, is trying to generalize this model. Can I, for example, take this model and move it to another site somewhere in the world, plug it in, and see if it is going to work? To what extent can this model generalize? Can I take some other GIS and plug it in? That is another issue. Actually, think I could, if I had the GIS expertise. I do not know if it is going to work, but the processes I have right now are just warfare, alliance building, and so on. A lot of other historical examples of state formation deal with those principles in different environments, and most of this is modular, so I could take out one productivity module and put in another one. So I think there is potential there. A lot of it has to do with the methodology of agent-based modeling in exploiting this portability from one place to another.

Kohler: One of John's points was that those of us working at the interface between modeling and history might want to consider a systematic search for counter-factuality as a good approach, and I entirely agree with that. But I would like to point out, in passing, my interpretation of Steve Goodhall's first paper, if we can remember at this point what that was about. That gave us a counter-factual demonstration in a rather indirect way, if I understood it properly. I have not read the paper.

We know, for example, that in explaining altruism in human societies, the mechanism of group selection is gaining great ground, and the idea underlying that is that groups having a high proportion of altruists will tend to survive better. If you have enough isolation of groups with some subsequent mixing at appropriate stages, you can get high levels of cooperation and get that fixed in populations.

What is interesting, then, is that Steve's results tend to show that altruism does not work. So the counter-factuality is that we know that in human societies altruism, in fact, operates at extremely high levels. Therefore, I think that we could say that what Steve has is not a model for any kind of recognizably human society, but perhaps a model for a chimp society or maybe some kind of a proto-hominid society in which, for example, there is very little food sharing or other altruistic behaviors of any sort, except some limited amount between mothers and offspring, for example.

So there is some counter-factuality there, and it would be extremely interesting to see what would happen with that. I share your concern about why it is specifically that altruism fails in these societies, and I would like to see the mechanism for that worked out very precisely. I had, in fact, the same two candidate explanations for that as you did, and I think that should be expanded upon. Maybe Steve would like to respond to that.

Steven Goodhall: Steve Goodhall, Compuware. I absolutely agree. Your point about this not being a human society model is absolutely correct. It is a pre-human, and maybe a very far pre-human, model. It shows things that we know do not happen as you ascend to higher levels of social organization, so as we build on this model, do the structures that we put in place to make this model operate at a higher level of consciousness, if you will, a higher level of decision making, start to exhibit a benefit for altruism? That is, I think, one major point for going forward.

I also have the same thoughts on why the collapse is taking place, and it is certainly on my list of things to look at next. I am inclined to suspect in this context that it is the fact that everybody runs out at the same time and cannot survive, but then there is the question of why do they do that, as opposed to making better decisions? If you look at the dynamics of these runs, there are a lot of unexploited resources in these models. We are not getting close to the carrying capacity of the space, and I think that is another thing that we can watch improve as we put better processes in place.

Kathleen Carley: Kathleen Carley, Carnegie Mellon University. In the last couple of days of this conference, a lot of people have been talking about representing networks, and this relates to something I mentioned to Mike [North]. In talking about visualizing networks, these two papers and the paper this morning on electric power represent the potential advantages that considering networks provide in understanding things. The graphs that I produce are in fact networks, and within the model, flows are taking place. If the flows do not take place, the network will fall apart. So I was very fascinated to hear that we had something very much in common, although we are separated by historically long periods of time.

I think, though, that in the toolkit session, people were talking about graphical objects and visualizing graphs, and we might do that. I think all three of these papers basically show the advantage of networks and the power that a network focus has on presentation. But it also challenges the people who are responsible for the toolkits to support that visualization. The question is, how do we do it? But I think I will ask you that question, because you did not get the chance to comment this morning.

Michael North: Mike North from Argonne National Laboratory. I would like to say a couple of things. The first is that there is not a lot of altruism in our model either, and in some ways, you could say that the electric markets, at least in California, were archaic states, so I guess there is probably a lot more commonality than we realize.

There is a lot of interest right now in developing network libraries. Repast is an example, but other toolkits are doing this as well. I think that representing a network is probably not the hardest part: it is the visualization that you talked about. In fact, we have some things we could show you later. In particular, Chick [Macal] put together some very interesting network visualizations that he has developed that are not strictly part of the current modeling toolkits; we export the data and bring them into these visualization tools. We have placed a big emphasis on that in our own research programs — developing better ways of looking at these networks, and hopefully rolling that into the toolkits at some point as well.

I think that probably the most important thing here is that a large amount of research has already been done in this area, and it is not just in agent-based modeling. We are talking to people in the visualization community to try to learn what they know and bring that in, which is part of a larger point I was going to make later in closing the panel, relating to a reproachment with existing fields. I will just foreshadow it here. As agent-based modeling matures, it is not only going to need to find its place among other fields, but learn from them. We need to go from things like folk psychology or folk sociology that have been used in the past and go to real sociology and real psychology. In the same way, we have to stop folk visualization and start using some of the real tools that are available.

Invited Speaker:

Scott E. Page

THE INTERPLAY OF DIFFERENCES

S.E. PAGE, The University of Michigan, Ann Arbor*

ABSTRACT

Agent-based modeling would be less challenging and of little practical interest without the ability to include diversity in models of physical, biological, human, institutional, and artificial systems. Including diversity entails constructing and analyzing models involving agents that exist in *different* spatial and social locations, that interpret the world in various ways, and that are capable of taking a range of actions. This paper examines the role of diversity in agent-based modeling and begins with a review of such classic models as Conway's Game of Life, Wolfram's Rules 30 and 110 for two-dimensional cellular automata, Schelling's segregation model, Watt's Small World framework, the Page and Hong diversity model, Arthur's Bar Problem, and an incredibly important model called the Prisoners' Dilemma. A new interpretation of existing research is offered. The discussion at the end of the paper encourages other scholars to explore, research, and play with models consisting of diverse agents.

1 INTRODUCTION

The term "interplay of differences" refers to the substantial and various impacts of the inclusion or enabling of diversity in physical, biological, human, institutional, and artificial systems. Without diversity not only would there be very little physics and no biology, economics, politics, or sociology, there also would be little use for agent-based models. It is relatively easy to construct and analyze a model of identical agents, that is, agents that exist in the same spatial and social locations, that interpret the world in the same ways, and that take the same action. Such a model, however, would not be very interesting or enlightening. It is more challenging — and more interesting — to move agents outside their environment and place them in social, economic, and family networks; endow them with distinct interpretations of the world and different intellectual toolboxes; assign them to roles and responsibilities within institutional and organizational structures; and enable them to define and evolve labels or types that simplify their interaction with others.

In this paper, we present a brief, but broad, discussion of diversity and a new interpretation of existing research. Our discussion is intended to encourage other scientists to explore, research, and play with models of diverse agents more than it is intended to be a coherent and complete summary from on high. Agent-based modeling research is itself bottom-up.

This paper is organized around four themes whose titles are borrowed from literature. Each of these sections emphasizes a particular theme but also explores other topics ranging from architecture, to *E. coli*, to decks of playing cards.

---

* *Corresponding author address*: S.E. Page, Center for the Study of Complex Systems, Departments of Political Science and Economics, The University of Michigan, Ann Arbor, MI 48106; e-mail: spage@fiore.physics.lsa.umich.edu.

2 DIVERSITY IN THE CLASSICS

Our examination of the role of diversity in agent-based models begins with a review of some of the classic models in complex systems. These “classic” models include Conway’s Game of Life; Wolfram’s Rules 30 and 110 for two-dimensional cellular automata (CA); Schelling’s segregation model; Watt’s Small World framework; the Page/Hong diversity model; Arthur’s Bar Problem; and a novel little game called the Prisoners’ Dilemma, which is now recognized as an incredibly important model.

2.1 Game of Life

Most members of the agent-based modeling community have played with Conway’s Game of Life. This model permits users to experience how simple rules enable elaborate patterns from cycles, to gliders, to glider guns.[1] Figure 1 shows the “figure eight” pattern from the Game of Life. This configuration, which resembles a figure eight at a 45-degree angle, creates a cycle of length 8 according to the Game of Life rules for a two-dimensional CA.

The Game of Life is designed with limited diversity on a few dimensions; only one type of agent is included. The agent can choose from two actions. All agents follow the same behavioral rule; they do not tremble or mutate. The spatial diversity is over a two-dimensional grid with a checkerboard topology transformed into a torus. Although the diversity in the model’s assumptions is limited, the Game of Life exhibits incredible diversity at the level of emergent structures. Furthermore, the Game of Life has proven to be a universal computer, so in a sense, this simple model can do anything. The fact that a little diversity can go a long way is a recurrent theme in this paper.

			X	X	X
			X	X	X
			X	X	X
X	X	X			
X	X	X			
X	X	X			

FIGURE 1 The Game of Life

2.2 A New Kind of Science: CA Rules 30 and 110

Although the Game of Life is captivating and evocative, it is at best a definitive example and at worst an anomaly; it probably lies somewhere in between — in the land of supportive anecdotes. A more compelling argument runs through the many pages of Steven Wolfram’s *A New Kind of Science*. Much of Wolfram’s analysis focuses on one-dimensional CA using a nearest-neighbor topology. Its spatial structure is less diverse than that of the Game of Life and

[1] In the Game of Life, each cell on a checkerboard has eight neighbors. A cell can be alive or dead. A live cell remains alive if two or three of its eight neighbors are alive. A dead cell only comes to life if exactly three of its neighbors are alive. Time proceeds in discrete steps with synchronous updating.

certainly far less interesting than what we encounter in the real world, irrespective of scale. If we were to look from quarks, to genes, to molecules, to neurons, to people, to computers, to firms, to nations, and to planets, it would be difficult to find many interaction topologies as simple as the one-dimensional CAs that Wolfram, for the most part, studies.

Furthermore, Wolfram allows only two types of actions, and, as in the Game of Life, all of the agents follow the same behavioral rule and interpret the world in an identical way. They do not adapt, and they do not tremble. In fact, Wolfram does not appear to think of the cells in his CAs as agents but as automata, which is what they are. Thus, many of Wolfram's models — like the Game of Life — should be seen as belonging to a class of models that assume fixed spatial diversity and moderate action diversity. In the space of the possible, this class occupies a small but important location.

Depending on the source, *A New Kind of Science* can be interpreted in several ways — from revolutionary to provocative to repetitive. In this paper, we not to pass judgment on all of those keystrokes; rather, we highlight what readers have obviously gleaned: it is possible to do some incredible things with two-state, one-dimensional CA. We can create randomness — what could be more diverse than that — and we can compute anything provided that we start in the right place.

On the basis of the figures presented in Wolfram's book (lacking permission none are reprinted here), we might conclude that these CA rules can produce anything. In one sense, that is true temporally: if Rule 30 truly does generate a random sequence, then eventually it produces any finite sequence of 0's and 1's along a single-cell location over time.[2] In another sense (i.e., spatially), it is not true. A long time ago Arthur Burks and others described "Garden of Eden" states. These automata states cannot be attained temporally. That is, if we look at all of the cells in a particular time step (a horizontal strip), there will be some configurations that we never see for a given rule.

We reiterate this distinction because it is an important one: Rule 30 can give any vertical slice, but only a subset of the possible horizontal slices. Those that Rule 30 cannot give belong to the Garden of Eden. Those Garden of Eden states must be the initial states of the system if they are to be visited, so those visits are either brief or, for equilibrium Garden of Eden states, eternal. Therefore, at least for two-dimensional CAs, saying that a little diversity goes a long way is eerily accurate because, while it goes quite far, it does not go all of the way of being able to generate all states.

The two classic models discussed so far include limited diversity of actions with a dash of fixed spatial diversity tossed in so that we can look at pretty pictures. We have ignored diversity in behavioral rules and representations, which could be either hard wired or allowed to emerge, as well as diversity in spatial representation and social connectivity. These types of diversity are discussed below.

---

[2] In other words, if we look at a vertical line of cells, we might see anything.

2.3 Schelling's Segregation Model

In later introductions to *David Copperfield*, Charles Dickens wrote that he had a favorite child and that child's name was David Copperfield. If there is a competitor to the Game of Life as the "best of the simple," as the model in moments of ambition we most would like to call our own, it must be Schelling's model of tipping. Scholars ranging from Cederman to Axtell to Gaylord adore this model for its elegance and intuitions. Schelling's model has two types of agents, called here Type 1 and Type 2. The agents are placed randomly on a two-dimensional grid. These agents can move if they are unhappy with their current location. In some instances, these agents follow a single rule: move if more than X% of your neighbors are of the opposite type.[3] Schelling finds that the system tips. In other words, a few agents move and continue to move until eventually the system segregates into regions of Type 1 and Type 2 agents even when the tolerance threshold is set at 40% so that agents only relocate if more than 40% of their neighbors are different.[4]

The clarity and power of Schelling's message resonated across the social sciences: we cannot infer agent-level racism despite the segregation seen in American cities as detailed in Denton and Massey's *American Apartheid* or in high school or college cafeterias. Moreover, we must always be careful not to commit the error of inferring focal or obvious microlevel forces from macrolevel patterns, the so-called ecological fallacy. That the transition from micro to macro can be paradoxical with so few moving parts is astounding. This is the real beauty of Schelling's model.

We can learn another lesson from Schelling's model: other agents can use identifiable differences as food for both thought and action. When agents are allowed to evolve strategies and actions, they often use anything to gain the advantage. Types are frequently something at their disposal. This phenomenon can also be seen in other agent-based models. A good example of this phenomenon can be found in Riolo, Axelrod, and Cohen (2001), whose agents use identifiable tags to create epochs of cooperation in the Prisoners' Dilemma.

2.4 Watts' Small Worlds Framework

In the Small Worlds framework of Duncan Watts and others, agents belong to close knit groups (small worlds), and they also have an occasional random connection (e.g., the man your sister's roommate dated who moved to Seattle and who you saw at the airport once). One way to appreciate the beauty of the diversity in this model is to think of it as the offspring of two other models and to compare it to its rather boring parents. One parent has all random connections, whereas the other has all local connections. If we consider a collection of agents with random connections, we can prove all sorts of things about degrees of separation and spanning trees, but none is very surprising or interesting. Alternately, if we consider a collection of agents with local connections, whether in cliques or on a lattice, we again find little beauty in the answers to such questions as: "Are all agents connected? Yes. What is the separation between two agents? It depends on how far apart they are on the checkerboard."

[3] Schelling allows the two populations to have different thresholds for moving. From our own experiments, this diversity of thresholds creates significantly more tipping.

[4] The "tipping" is not as pronounced as one might expect from Schelling's account of the model. In fairness, however, his technology did not allow him to calculate the extent of tipping as easily as we can now.

By combining the random connections with the local connections, the Small Worlds model generates degrees of separation between agents that are remarkably close to what arise in the random model, and the model accomplishes this feat with relatively few random connections. The Small Worlds model has many other useful properties of which most of us are familiar. The point is that these properties only rely on a little diversity.

2.5 Brian Arthur's Bar Problem

The Bar Problem, like many of the models that we now explore, can be found in Schelling's *Micromotives and Macrobehavior*. Brian Arthur's implementation of the Bar Problem includes a collection of agents, each of whom must decide whether to go to the bar. All of the agents have identical payoff structures. They want exactly 60 agents to be at the bar. The agents evolve rules to decide whether to go to the bar based on the time series of past attendance. These rules are simple. Arthur restricts the set of possible functional forms that they can take. Yet, the agents collectively can evolve strategies, which lead to approximately the correct number of agents going to the bar each time. If all agents use the same rule, then either all or none of the agents go each week. Instead, they evolve a diversity of rules that works. When several agents' rules tell them not to go, other agents' rules kick in and tell them to go. It is a form of stability through diversity, a point we revisit later.

If we compare the Bar Problem to the Santa Fe Artificial Stock Market model, which by comparison is baroque, unwieldy, and difficult to analyze, we begin to see the reason for limited diversity. Too much diversity on too many dimensions gives us a mangle that is hard to sort out. A little diversity appropriately placed, however, can generate just enough complexity to prick our interest.

2.6 Hong and Page's Diversity Model

In the Hong and Page model on diverse agents solving problems, agents are endowed with diverse encodings of problems — called *perspectives* — and diverse analytical tools to solve those problems — called *heuristics*. We can think of heuristics as playing cards, and each agent has up to 13 cards. Simple combinatoric arguments show that an agent could have an enormous number of collections — as many as the number of possible bridge hands. Finally, we have a model that does not have a little diversity; rather, it has diversity in abundance.

Diverse collections of agents perform incredibly well. In fact, under a set of fairly unrestrictive assumptions, we find (and prove) that a relatively small group of randomly chosen diverse but intelligent agents will outperform a group of the best agents on a hard problem. So, for example, in a model with five possible heuristics, each agent is endowed with five. The 10 best individual agents performed worse than 10 randomly chosen agents because the best agents use similar approaches in solving the problem, so two heads are not much better than one. Also, all of the agents are smart, so the random agents have good heuristics as well.

The first lesson from the Hong and Page model is that small collections of simple, diverse agents can be collectively extremely bright. Even though the model can include substantial diversity and still work, all of the diversity is not needed to make the point that diversity trumps ability. This statement is an extension of an idea that has been around since Adam Smith and expanded upon by Hayek. Recall the old lines about how markets aggregate individual knowledge. The railroad worker knows about railroads; the farmer knows about wheat; and therefore the economy knows about everything. But while the Smith-Hayek argument about markets as super agents is really a comment on the power parallel processing along dimensions, the Hong and Page result is on the power of simultaneous diverse processing that overlaps dimensions.

The second lesson from the Hong and Page model is that to include lots of diversity, which is done in this model by allowing agents to have their own heuristics and perspective, as well as to have all agents interact in a soup, agents cannot influence one another. In the Hong and Page model, the agents work on a solution to a problem, so they all interact, but with the problem solution not with one another.

2.7 The Prisoners' Dilemma

The Prisoners' Dilemma is by far the most famous, popular, analyzed, and modeled game. Agent-based Prisoners' Dilemma models have shown that agents can learn to evolve cooperative strategies like "Tit for Tat." However, they can only learn this skill if one of two conditions is satisfied:

- Either the space of strategies must be small: *limited strategy diversity* or

- Agents must be connected by a topology and not in a "soup": *limited number of interacting players*.

Beginning with a soup of many players who can evolve many strategies means that it will be necessary to wait a long time for cooperation to occur. In fact, cooperation will not be able to gain a foothold, so the strategy must be sophisticated *and* general. It is not that cooperation is impossible in Prisoners' Dilemma; it just takes time.

The point is that the early models of agents by John Miller, Robert Marks, Bob Axelrod, Nowak and May, and others that generated cooperation in the Prisoners' Dilemma did so by only admitting limited diversity. Agents did not have big strategy spaces, and they were also often small in number or placed on a lattice.

If a little diversity goes a long way, what can be said about a lot of diversity or even a modicum of diversity? Do they take us even further? Surprisingly, the answer is often no. Agent-based modelers extol the virtues of stark models, from Axelrod with his KISS principle (Keep It Simple Stupid) to Wolfram with his empirical evidence that three-state, five-neighbor CA rules; they are not as interesting on average as two-state, three-neighbor CA rules. Even the Epstein and Axtell Sugarscape model, which by comparison appears Guadian with its sugar and spice, epochs of cooperation (everything nice?), trade, and disease, is relatively simple at its core.

In the end, we have a paradox. The most compelling agent-based models of complex adaptive systems do not have many moving parts — what many would consider a defining characteristic of something complex. Diversity is necessary for complexity in that complexity models require diversity to get them started. This initial diversity can be in types, in behavior, in actions, or in space. In the most compelling models, only a little diversity has been added along one or two of these dimensions.

3 CHILDREN'S LITERATURE

The second theme of this paper concerns the diversity or variation that arises from mutation and the process by which those mutation rates are chosen or evolved. A single work of children's literature is used to frame this discussion: *Goldilocks and the Three Bears*. In brief, in the story, Goldilocks happens upon the bears' cabin and finds three chairs, three bowls of porridge, and three beds. When Goldilocks tastes the porridge, she discovers that one bowl is too hot; one is too cold; and one, the baby bear's bowl, is "just right." The metaphor of finding the temperature that is just right applies in agent-based modeling in multiple contexts. When people talk about the exploration/exploitation trade-off, the edge of chaos, or the evolution of evolvability, they are really talking about finding the baby bear's porridge, though being good academics, they use much more jargon.

In most biological, ecological, and social science models, mutation is the primary cause of diversity. Crossover in biological contexts and partial imitation in human ones often bootstraps this diversity to create even more. It is important to note that crossover alone may not guarantee full exploration of the space depending on the initial population. All of us have learned the same lessons when it comes to mutation rates: too much mutation leads to a system that boils and bubbles and never settles. Too little mutation leads to a system that freezes. Somewhere in the "in between," the system comes alive, generating patterns and complexity.[5] We focus next on the discovery and properties of that in-between region.

3.1 Exploration and Exploitation

We first examine the trade-off between exploration and exploitation. Scholars of agent-based modeling and complex adaptive systems have devoted substantial time and energy developing models that explicate the trade-off between exploration and exploitation: too much exploration and the system eternally boils; too much exploitation and the system stabilizes prematurely. If a model is built to compute something or to find an efficient outcome, in the former, the system never stabilizes and therefore fails to exploit successful structures. In the latter, the system settles on a bad solution, possibly the best in the initial population. In contrast, if the model was intended to represent some physical, biological, or social phenomenon, and too much exploration has been done, we have only random agents bumping into one another. If we have too much exploitation, however, the system is not interesting. Therefore, if we imagine an exploration/exploitation slide bar, we want to position it in "the interesting in between," that is, the region that separates stasis and randomness.

To see the exploration/exploitation trade-off in action, consider the following problem. Each of two agents must choose whether to graze on some subset of 30 fields. An agent gets a payoff of one from being the only grazer on a particular field, a payoff of negative two from being one of two grazers on a field, and a payoff of zero if the agent does not graze on the field. In this multiple location commons problem, there are 2^{30} socially efficient equilibria — one for each partition of the fields between the two agents. Table 1 provides data on average payoff in period 100 from a series of runs in which the mutation rate varies from 0.0001 to 0.20 from

[5] Oscar Wilde once quipped, "Moderation is a fatal thing. Nothing succeeds like excess." The opposite is true in agent-based models. Excess is a fatal thing. Nothing succeeds like moderation.

a model using a genetic algorithm with a population of size 30 and a simple tournament selection operator. Performance equals the percentage of the maximal possible payoff achieved.

TABLE 1 The Exploration/ Exploitation Trade-off

Mutation Rate	Performance
0.0000	23.33
0.0005	66.67
0.0010	83.33
0.0015	100.00
…	…
0.0200	100.00
0.0205	100.00
0.0210	100.00
0.0215	96.67
0.0220	96.67

This simple example shows that with too much exploitation, we cannot find the optimum and with too much exploration, we cannot exploit successful structures. This example also shows an enormous range for acceptable mutation rates. So while there may be an optimal mutation rate (i.e., the one that on average would most quickly enable agents to locate and maintain the optimum), we need not expend much energy to find it. In this case, the slide bar has a huge margin of error, which often occurs in agent-based models. Mutation rates can vary substantially without much effect on outcomes.

3.2 The Edge of Chaos

The slide bar metaphor also applies in a loose way to the concept of the “edge of chaos” put forth by Langton and others on the basis of some simple models. The edge of chaos as initially formulated was intended to capture the idea that a small increase in action interdependency can shift a system from being “complex” to being chaotic. Life and even the Game of Life were thought to exist on this precipice at the edge of chaos. As Crutchfield first explained and John Miller and I have elaborated in a simpler context, the edge of chaos may not exist, at least in the case of one-dimensional CAs. Instead as Miller and I describe, CA rules may be classified as structured (all 0’s and all 1’s are stable), unstructured (a CA of all 0’s gets mapped to all 1’s and vice versa), and partially structured (both all 0’s and all 1’s go to the same state). Chaotic rules belong to the unstructured class, and complex rules belong to the structured class. In other words, they are not near each other in CA rule space, so there is no edge. The edge appeared to exist because the CA rules were projected onto a one-dimensional measure, the number of 1’s in the CA rule, making the complex and random CAs neighbors on that projection even though they are not close neighbors in the space of CA rules.

All theory aside, there is a looser sense in which there is an edge of chaos. If we plot the efficacy of a system or of a search algorithm as a function of the mutation rate (see Table 1), we sometimes see gradual increases up to the optimal mutation level followed by a sharper drop-off in performance. This extreme decline in performance — from a little push of our exploration/exploitation slide bar toward more exploration — suggests “an edge.” The “of chaos” part can be attributed to marketing. Systems that haphazardly move between states need not be chaotic in the formal sense.

The above example did not appear to have much of an edge. However, if we increase the number of agent types to 10 but keep the environment essentially the same, the environment becomes more complex and therefore creates more of an edge in payoffs from an increase in mutation. With 10 types, assume that an agent gets a payoff of 9 if the agent is the sole grazer but loses 10 for each other agent that grazes. This scenario imposes a much stiffer penalty for stepping on what has evolved to be another agent’s turf. This change in payoffs alone causes the drop-off to appear more severe than in the previous example, but as we see in Table 2, the reduction is even larger after taking the payoff difference into account.

TABLE 2 The Exploration/ Exploitation Trade-off Revisited

Mutation Rate	Performance
0.0175	100.00
0.0180	100.00
0.0185	90.37
0.0190	100.00
0.0195	75.93

Considering both the 2-agent and the 10-agent versions of our simple multiple grazing model, it seems fair that the mutation slide bar can create an edge, if necessary.

Kauffman offers an alternative formulation of the “edge of chaos” and life existing at its edge based on the NK model. Kauffman finds that as K (i.e., the number of connections between agents) is increased, performance increases and then decreases. A later discussion looks at the similarities between increasing mutation rates and increasing the number of types of agent interactions. At this time, it is sufficient to recognize that Kauffman’s edge is similar in spirit but different in construction from Langton’s and that both suffer from some logical flaws.

3.3 The Evolution of Evolvability

The previous discussion suggests that the best mutation rates balance exploration and exploitation to avoid falling off the edge into the “boiling porridge” region. The adage “moderation in all things” appears to apply even to diversity. But we know more than that. The payoff as a function of rate of mutation is single peaked in the mutation rate. We also know that single-peaked functions can be solved easily by either gradient ascent methods or evolutionary search. If we combine these two ideas, we bump into an irresistible concept: "the evolution of evolvability."

We first encountered this phrase in a lecture by Mark Bedau. In a fascinating (but for our purposes perhaps overly complex) model, Bedau and Packard allow the mutation rates to evolve. While the ideas of evolving search parameters did not originate with Bedau and Packard, they show profound insight by recognizing that by evolving the mutation rate into the "baby bear" region, the system is evolving not only good system-level performance but also has the ability to evolve. Moreover, in light of the fact that performance appears single peaked in mutation rate with a substantial central plateau, even if the system's ability to evolve the mutation rate was crude, evolution would succeed. It would lead to evolvability. Even better, if the system exhibits an edge of chaos and if it also begins with a mutation rate that is too high, then any change that leads to lower mutation rates would be of high marginal value, and we would expect it to be even easier to evolve evolvability.

Can we infer from these few models that scholars have uncovered a partial explanation for the existence of successful evolutionary systems? We think so. At a minimum, we have not found evidence to the contrary, that is, that getting the mutation rate in the evolvability region would be hard. Biological constraints on how mutation rates might change notwithstanding, getting the mutation rate correct appears substantially easier than evolving the Krebs cycle, the human eye, or a pocket watch.

To quickly summarize the second theme, we see that diversity drives system performance. Without it the system has no ideas or attributes to explore, but with too much diversity the system boils. Agents cannot maintain and exploit successful building blocks. In some cases, ramping up the diversity in the population appears to lead to gradual increases in performance followed by a precipitous fall. This collapse can be thought of as an edge, as a cliff on a rugged landscape, or, if to push the metaphor, as the edge of chaos, where chaos is loosely interpreted as a system that never settles down or exploits successes. Finally, since payoffs appear single peaked in the mutation rate, the evolution of mutation rates that enable evolution to occur, the so-called evolution of evolvability, appears a plausible conjecture.

4 POETRY

We begin this section with a reminder of two of the most overquoted lines of poetry in the English language. They relate to two themes from complex systems, namely path dependence and lever points. These poems resonate, and we examine what agent-based modeling and complex systems say about that. The first line of poetry comes from Robert Frost: "...two roads diverged in a yellow wood… ." This phrase has obvious uses in advertising colleges and the like. The idea is that if you take the right road now, you will be happier. The second line of poetry comes from Shakespeare, and it puts a negative spin on the same theme: "There is a tide in the affairs of men which taken at the flood leads on to fortune. Omitted all the voyages of his days are bound in shallows and in miseries." Both poets are talking about how decisions in our lives affect our futures, that history matters. Frost is talking about the path (to him the entire path matters), whereas Shakespeare is talking about a lever point, a critical juncture.

These lines profoundly affect people because they highlight and romanticize those rare instances, those times when we make choices that influence our lives. Most days not like that; the courses of our lives are much more robust. Lots of butterflies flap their wings without changing the next word we type. It is this unromantic robustness that we focus on. Is it related to diversity in any way?

From an ecological perspective, there seems to be one obvious positive relationship between diversity and robustness. Diversity as used here means diversity of species and not so much diversity within a species. If we reduce the number of species in a particular ecosystem, we might destabilize it, and the ecosystem may not be able to stabilize itself without the loss of many other species. The impact would be especially severe if we eliminated what is known as a keystone species. But this line of argument is ridiculous because if we introduced a new species, we might also destabilize the system. Therefore, more diversity would also appear to make the system less robust.

Rather, we ask, are more diverse ecosystems better able to respond to fluctuations in their environment or the extinctions of some species? This question is difficult to answer. To see why, we describe two extremely similar models. In the first, diversity implies substantially more robustness, whereas in the second, the opposite is true.

4.1 The Route Selection Model

Imagine a world in which there are 20 time periods per cycle and 20 possible actions. Suppose that each agent must take each action once in each cycle to survive and that each action takes exactly one period. We can then write the behavior of an agent during a cycle as an element of the permutation group on 20 objects. To complete the model, suppose that the payoff from an action during a particular cycle is decreasing in the number of others taking that action in that period. Think of agents who have to visit a set of stores. They would prefer to go to less crowded stores. Therefore, everyone wants to be doing the opposite of what everyone else is doing.

Given a number of agents, there can be an enormous number of efficient collections of routes. Some are rather simple. For example, with five actions, denoted by A, B, C, D, and E, if equal numbers of the agents choose the sequences, ABCDE, BCDEA, CDEBA, DEABC, and EABCD, then the collection of sequences is efficient. Other efficient collections are also much more diverse. Consider the collection in which exactly one agent chooses each member of the permutation group on five elements. Since an equal number of these sequences have each action in each location, the collection is efficient.

We now compare the robustness of these two collections or sequences. Suppose that one action (action E) and one time period are eradicated. The five sequences in the population are then ABCD, BCDA, CDAB, DABC, and ABCD. The first and fifth sequences are identical. Forty percent of the agents could benefit by changing their sequences. Given the simplicity of the model, the agents can stabilize into a new configuration, but it takes time.

In contrast, when agents each choose a unique strategy from the permutation group on five elements, losing one action and one time period has much less of an effect. In fact, the collection of sequences remains efficient because the permutation group on five elements is transformed into five copies of the permutation group on four elements once one action has been dropped. This example shows a strong connection between diversity and robustness.

In the second example, each agent chooses a subset of N fields on which to graze sheep as in an earlier example. But what we want to do is change the payoffs so that each agent wants to graze with each other agent on exactly *k* fields. If there are only two types of payoffs and a field is dropped, there are only two possibilities: the agents met there or they did not. If they did not meet, then their actions remain efficient. If they did meet on that field, then the agents only need to find one new field on which to meet, a task that should be accomplished with a few timely mutations.

If we increase the diversity so that each of 10 agents must meet each other type on exactly *k* fields, we see more interesting phenomena. First, suppose that the 10 agents have settled on a pattern of grazing that is not diverse: *all 10 agents choose the same k fields*. The dynamics that would arise if we wipe out one field are not identical to what we would see with only 2 agents. If they did not graze on the field that is wiped out, then they are still optimally arranged. If they did graze on the field, then they have to coordinate on a single new field or evolve some pattern in which each meets another once. These processes would take some time. Second, suppose that the agents have diverse grazing patterns. Reducing the number of fields is certain to wipe out several matches between pairs of agents. Agents then need to find new fields upon which to match one another, but coordinating on a new set of fields is not easy unless an excess of fields exists. Assuming that an excess of fields does not exist, we are faced with two possibilities. In the first case, they cannot find a field, so efficiency falls and the system is less robust. In the second case, they find another field, but in doing so they meet with other agents more often than they would like, which could cause further relocations. The two possibilities are lower performance or a rather elaborate sequence of action changes. Therefore, we cannot help but think that the second system is less robust than the first.

These two examples suggest that the relationship between diversity and robustness is too subtle to merely say that increasing diversity increases robustness or that increasing diversity decreases robustness. We have seen two similar models: one case where diversity increases robustness and another case where it decreases. To gain a better understanding of what is going on, we need to turn to science fiction.

5 SCIENCE FICTION

The third theme in this paper is the most provocative, the least advanced, and the most speculative. It is science fiction for the simple reason that we have not advanced our argument to the point where it could be called scientific fact. This section should be seen as an attempt to make some sense of the first three themes. We argue that the number of types, the number of interactions, the mutation rate, and the spatial arrangement of agents contribute to the viability of a complex adaptive system similarly and that the complexity models that we most appreciate and understand all satisfy what we call the *limited interplay*. We want to distinguish the word interplay from the word in that you and I could interact but you might not affect me, as is the case in the Hong and Page diversity model, which supports all sorts of diversity. If we interplay, then what you do affects our fitness or payoff.

We begin by developing a connection between rates of mutation and the number of agents with whom an agent interacts. The connection is straightforward: if an agent interacts with 10 other agents, if any one of those agents mutates its action, then the agent under consideration may also have to change its action. Thus, the effective mutation rate could be 10 times as high as when the agents interact with only one agent. First, we provide an example where this is exactly the case. Second, we show how spatial arrangements in effect reduce the number of types, thereby allowing models to stay complex for two reasons:

- The systems might boil.
- Spatial segregation enables local clusters of agents of the same type to survive.

This leads to our discussion of the subtle and confusing relationship between diversity and robustness.

5.1 Mutation and Type Equivalence

We call the first functional form the *single grazer commons problem* (SGCP). The SGCP has N fields. Each agent can choose some subset of those fields to graze. In this way, the actions of the agents can be written as binary strings of length N. The payoff from being the only agent on a particular field equals 1, and the cost of having 2 agents on the field equals some number larger than 1. Given these payoffs, an optimal configuration of agents has exactly 1 agent grazing on each field.

It is straightforward to show that for any N and for any M, a configuration is an equilibrium in action space if and only if it is the optimal configuration. If a configuration is not optimal, at least one of two conditions must be met:

1. Two agents are grazing on the same field.
2. There exists a field upon which no agents are grazing.

If Condition 1 holds, one of the agent types that is grazing on the same field as another can increase its utility by no longer grazing on that field. If Condition 2 holds, any agent type can increase its utility by grazing on that field.

This first model nearly satisfies the *reducible other agents property*. This means that the strategies of all of the other agents can be aggregated to form a strategy that could be that of a single agent. If we change the payoff structure so that the payoff equals 0 if agents do not graze, 1 if they graze and no one else grazes, and minus the number of agents if they graze and some other agent or agents also graze, then the reducible other agents property would be strictly satisfied. If the reducible other agents property is satisfied, increasing the number of agents is equivalent to a linear increase in the mutation rate. Suppose that the system has stabilized. At this point, the interplay in SGCP for a single agent is approximately the same whether 1 agent has a mutation rate of 0.01 or 10 agents have a mutation rate of 0.001. Out of equilibrium, a similar though less precise correspondence occurs. If system performance is good, an agent does not have much interplay. There are not many other actions by agent types that influence the agent. Therefore, in this model we see that it is possible to create an equivalence between the number of types and the mutation rate. It is also clear that increasing the mutation rate increases interplay. If an agent has settled on a particular strategy or action, it is as if the other agents do not interplay with that agent. If the agent keeps experimenting, however, the other agents do interplay with the agent.

5.2 Interplay Not Number of Types Determines Complexity

The second functional form is that by virtue of the payoff structure SGCP has low interplay regardless of the number of types. If the interplay hypothesis is correct, the number of types should not matter much for time of convergence. Table 3 shows time to an efficient equilibrium as a function of the number of types that support that intuition. This and other results from this section use a genetic algorithm with a population of 30 to represent each agent and employ a standard tournament selection operator.

TABLE 3 Time to Efficiency in the SGCP

Agent Types	Number of Periods
2	23.52
3	30.88
4	32.98
5	40.18
6	45.64
7	47.10
8	53.46
9	56.48
10	62.64

These data show that the time increases only slightly because initially 0's and 1's are equally likely to be in our initial populations of strings. When 10 types are present, the strings in an efficient configuration have many fewer 1's.

The second model is the same as discussed in the previous section, where each agent wants to meet each other agent on exactly k out of N. In this game, there are N choose k equilibria in which all of the agents choose the same k fields. There are also more diverse equilibria as well as inefficient equilibria. Suppose that $N = 4$ and $G = 2$. Consider the following configuration:

Type 1: 1110
Type 2: 1101
Type 3: 1011
Type 4: 0111
Type 5: 0111

It is straightforward to show that the Type 1, 2, and 3 agents match each of the other types on exactly two fields. Agents 4 and 5 match each other on three fields. Since Types 4 and 5 have the same action, it suffices to show that Type 4 cannot improve by changing its action. A simple calculation shows that if the Type 4 agent chooses not to graze in any field, to graze in one field, or to graze in two fields, the agent is worse off. Similarly, if the Type 4 agent grazes in four fields, the agent is also worse off. Therefore, the Type 4 agent must graze in exactly three fields. It turns out that any action that grazes in exactly three fields gives the same payoff.

Therefore, once the population settles into the configuration such that each of the five agents grazes on exactly three fields, it never leaves that configuration.

It should be clear from this example that this model has much more interplay than the previous model and that the interplay increases in the number of agent types (see Table 4). What we see in these data is that time to efficiency increases dramatically in the number of types because the interplay increases as the number of types increases.

TABLE 4 Time to Efficiency

Agent Types	No. of Periods
2	2.08
3	3.60
4	6.16
5	10.02
6	18.46
7	57.90
8	272.32
9	867.96
10	1650.52

5.3 Space and Interplay

Many of our models of complex systems that generate interesting patterns take place over networks and not in soups. In this section, we describe some real-world experiments that corroborate agent-based theory and tie the model back to the concept of interplay.

In a recent paper in *Nature*, Kerr, et al. (2002) conduct experiments on a real-life rock-paper-scissors game. Cells in *E. coli*, a form of bacteria, contain genes that encode a toxin. They also contain genes that encode a protein that makes the cell immune and a gene that encodes a protein that causes the toxin to be released. This allows us to classify the types of expressed cells as resistant to the toxin (R), sensitive to the toxin (S), or toxic (T). These cell types interact according to a rule not unlike the children's game, rock-paper-scissors. T cells replace S cells because S cells are sensitive to the toxin. S cells have a growth rate advantage over R cells, so over time S cells replace R cells, and to complete the cycle R cells replace T cells because they also have a growth advantage.

If these three types of cells are placed on a two-dimensional lattice and interact locally, diversity is maintained. S cells grow faster than R cells, and R grow faster than T, but as the S become prominent locally, the T cells begin to grow very fast. When the locality assumption is relaxed the toxic cells (T's) wipe out all of the sensitive cells (S's) and then the resistant cells (R's), which grow faster than the T cells take over the population.

Kerr, et al. ran experiments in three environments: flask, static plate, and mixed plate. In flask, all three strains of *E. coli* were well mixed. In static plate, interactions are mostly local, and in mixed plate, there is some mixing and some locality. They found that diversity was maintained in the static plate environment but not in the other two. Local interactions and dispersal would seem to be necessary to maintain diversity.

They make the very nice point that if producing the toxin costs money, we could see these nontransitive relationships all of the time. In fact, think of the Prisoners' Dilemma with three types of players: All *C*, *TFT*, and All *D*. The All *D* players defeat the All *C* players; the *TFT* players defeat the All *D* players; and the All *C* players grow faster than the All *D* players. Others have run agent-based models with similar strategies and found that if the agents are in a soup, one strategy wins out, but if the agents are on a relatively sparse network or on a grid, diversity is preserved.

We would like you to contemplate the obvious point that spatial segregation decreases interplay. It is difficult to construct a model with 10 types of agents growing in a soup that remain viable, but it is relatively easy to do so on a grid because the types isolate themselves and do not interplay with the other types.

5.4 Interactions and Interplay

In Kauffman's NK and NKC models, the N term does not have much of an effect on performance. What matters are the K and C terms. In the NK model, if we think of each site on Kauffman's strings as an agent, then K is a great proxy for interplay. The problem that we have with the NK model is that the interplay is fixed and random. The agents have a limited ability to lower the amount of interplay. There are no micro foundations that explain why one site interplays with another in a system in which agents can spatially position themselves and choose actions that generate certain levels of interplay. This is why Kauffman's "evolution" to the edge of chaos argument is tenuous. He dials up K, but each time he wipes out all of the existing connections and creates new ones. Not only would this not happen in a system, it also misses a golden opportunity (i.e., to establish firmer foundations for the evolution of evolvability). Packard and Bedau show how this can be accomplished through an evolvable mutation rate. But as this science fiction discussion hints, it may be interplay that matters, and evolvability may evolve through type reduction, through spatial segregation, or through increasing or decreasing the number of interactions and the potential interplay.

6 ACKNOWLEDGMENTS

I would like to thank Troy Tassier for reading earlier drafts of this work and Troy and Cosma Shalizi for their input on this project. This research is supported by a grant from the James S. McDonnell Foundation.

7 REFERENCES

Axelrod, R., *The Complexitiy of Cooperation, Agent-based Models of Competition and Collaboration* (Princeton Studies in Complexity), Princeton, NJ: Princeton University Press.

Hayek, F.A., 1976, *The Sensory Order*, Chicago, IL: The University of Chicago Press.

Kauffman, S., 1995, *At Home in the Universe: The Search for Laws of Self-organization and Complexity?*, New York, NY: Oxford University Press.

Kerr, B., M.A. Riley, M.W. Feldman, and B.J.M. Bohannan, 2002, "Local Disperal Promotes Biodiversity in a Real-life Game of Rock-paper-scissors," *Nature* **428**:171–174, July.

May, R.M., 1999, *Stability and Complexity in Model Ecosystems* (Princeton Landmarks in Biology), Princeton, NJ: Princeton University Press.

Riolo, R., R. Axelrod, and M.D. Cohen, 2001, "Evolution of Cooperation without Reciprocity," *Nature*, **414**:441–443, Nov.

Schelling, T.C., 1978, *Micromotives and Macrobehavior*, New York, NY: W.W. Norton & Company.

Watts, D.J., 1999, *Small Worlds*, Princeton, NJ: Princeton University Press.

Wolfram, S., 2002, *A New Kind of Science*, Champaign, IL: Wolfram Media, Inc.

DISCUSSION: THE INTERPLAY OF DIFFERENCES

C.M. MACAL, Argonne National Laboratory, Moderator

Charles Macal: Many of us who have used concepts from complexity sciences to model social systems have, from time to time, wondered about some seemingly miscellaneous things, such as how Conway's Game of Life relates to Schelling's segregation model, or how route selection is just another form of poetry, or possibly how small-world networks are apparently different from Kauffman's rugged landscapes. Or perhaps you have pondered how Wolfram's Rule 30 could be similar to a new kind of science fiction.

Well, Scott Page has been thinking deeply about these and other matters, and he is here to share some findings from his journeys with us in a paper titled, "The Interplay of Differences." Scott Page is a member of the Center for the Study of Complex Systems in the Department of Political Science and Economics at the University of Michigan. He has been working in the areas of complex systems, economics, and political science his entire career. He is a member of the Santa Fe Institute (SFI) Economics Program, where he founded and leads the annual SFI economic summer boot camp.

[Presentation by Page]

Macal: Thank you very much, Scott, for those wide-ranging and thought-provoking ideas. I think that we should consider a few questions, comments, or observations at this time.

Unidentified Speaker: I was reminded of the movie *The Graduate* and of the word "plastics," and I was thinking "entropy" — "entropy interplay." Are we thinking about trying to take agent-based science up to the level of science?

Scott Page: My impression is that work is proceeding on two parallel scientific tracks. One involves people who are trying to do foundational mathematics on complex systems and link interplay and entropy or something like that. The other is an effort going on by more mortal mathematicians like myself who are asking whether we can build up a science of complex systems that looks not unlike the science of economics or the science of physics that is taught to undergraduates. Can we construct a set of simple models with some simple rules that are correct in themselves and that can be correct to the 95 to 99% level when the models are expanded globally? So I think there are parallel tracks.

You are absolutely correct. The question is, can we make this; can we come up with core concepts, core models? I think that we are coalescing a set of classic models, and now the question has become a set of classic tools.

Macal: Thank you. Does anyone else have questions ? … Actually, I have a question based on your last answer. You suggested building core models, or perhaps core examples, but is it the notion that the kind of factual theorems, which are the foundation of, say, physics and some of the other sciences, are not necessarily the goal of complexity science researchers? Is there a little wiggle room?

Page: Yes, I think that there certainly are people who would like to have fundamental theorems, but that the method itself suggests that we want some wiggle room because the set of things we can prove is not all the things that we are interested in. So I think that one of the things that — one of the philosophies — we can prove, and then we can explore just beyond that. I think that there is a sense in which we are allowing ourselves wiggle room because people are going to accept …. I think there are certain professions that are worse at this than others, and economics is a prime example of one in which, if you cannot prove it, it is not interesting, regardless of how interesting it is. I think as a group of people working in …. Even if we cannot quite prove it, if it is interesting, it is interesting. Right?

And there is a sense … where somebody proved something, like seven claims, and it was for a specific functional form. So I did simulations with three other equally important functions and the theorem was not true for those functions.

There is a sense in which, in complex systems, it is easy to make theorems — anything to get good about smaller things — and it is hard to do it in general things. And so ... a more statistical view of things.

Macal: So perhaps we could prove theorems at the 99% level.

Page: Yes, or you could do things like, say, Jim Crutchfield — a classic. Jim Crutchfield did a very nice paper where he decided there are more zeroes and more ones. He gets something that is right about 97% of the time. He cannot prove it is the best rule, but he can prove how the rule works, and then he gets his emergent substructures and these particles and how the particles act. So you group all the parts of it, and there is a sense in which in the paper there are results, but the big theorem is missing. But even though the big theorem is missing, all the subtheorems are there. And it may be because there is no big theorem.

Macal: Okay, one more question.

Michael North: Mike North, Argonne National Laboratory. Following up on what you just said, there are sciences that have different standards for wiggle room. There are sciences that have a different standard of proof or a different view of proof, like biology, for instance, where they obviously are not proving things mathematically, but if you are successfully classifying things, successfully demonstrating these things exist, that is sufficient. So in that sense, the fact that this is computational makes it natural to compare it to math. But if we were to compare it to other areas, say biology or something like that, we might have a more reasonable standard for what we are trying to do.

Page: Right. There is a key question of what counts as a proof. I have an anecdote to support you in this. Sir Duncan Watts did a small study out of Columbia in the Political Science Department. He was explaining this small stuff, how it explains six degrees … system. But these political scientists were saying that … comparative status … explanation…. It is not comparative status; it is not science.

There was a sense in which how we were trained is the way you — how you change — the function of those, and you are right. In different fields, say in physics, the physicists like to be used as an explanation.

So I think it is also the case that we are bound by scientific methods that have grown up around saying, it is 95% sufficient presenting statistical data. It is an artifact, right? Now we have computers where, if something is 95.1, how is the behavior … 94.9? You have to figure out where your data line….

Closing Panel

DISCUSSION: CLOSING PANEL

C.M. MACAL, Argonne National Laboratory, Moderator
M. NORTH, Argonne National Laboratory
F. ROJAS, The University of Chicago
S. GAILMARD, The University of Chicago

Charles Macal: A great deal of information has been presented during this workshop, and now I'd like to introduce the members of the closing panel: Mike North from Argonne National Laboratory, Fabio Rojas from the University of Chicago, and Sean Gailmard from the University of Chicago. They will make comments as they see fit, and then we can open up for further discussion.

Michael North: I found this workshop to be very interesting, and certainly a diverse collection of people attended. I'd like to bring up some points that seemed to be common themes throughout most of the talks, but then I'd like to specifically reference someone who, in my opinion, did the best job of bringing these ideas out. In keeping with the idea of leading to a very positive, happy ending, I have chosen seven, which is a lucky number, points to address. I hope this [format] will work out well.

The first point has to do with multilevel or nested modeling, and the idea that, when we start to develop and model these systems, we need to move more toward nesting or embedding agents within agents or levels of agents. I think David Sallach actually introduced that idea. I thought that was an interesting example of something that we need to think more about as a community. People have been doing this, but more needs to be done in terms of not only nesting within a given model, but also comparing between models and taking ideas, for instance, from individual or small group models and combining them with these much larger institutional models to see what things we can find. So I think that nesting and embedding are important.

The second point is the idea of rapprochement with other fields. An important part of the maturation of agent-based modeling and computational social science in general is to realize that there are many other fields with many things to say. Lars-Erik [Cederman] did a very good job with that, and Kathleen [Carley] actually took me to task, I believe, after my talk, mentioning the idea that social network analysis has been around for a long time, and I agree that it has in various forms. Visualization of networks has also been around for a long time. There's much to be learned from these fields.

Lars-Erik also mentioned issues with social science, social theory. David Sallach mentioned the idea of folk sociology in forming many of the models and how we need to graduate from that folk sociology into a more robust, theoretically grounded version of sociology, because much is already known. Our future is a rapprochement with the other fields that do exist, moving out of that naïve folk state into a more mature relationship with the other subspecialties and sciences. I think that's very important.

The third point I'd like to make concerns the whole idea of balancing and reconciling these notions that Kathleen Carley talked about of transparency and veridicality. I think that's very critical because even here we saw a tension between simplicity — I'll say simplicity — and complexity in models to choose a slightly less accurate but more widely used version of the statement. However, in the discussions, we heard [comments from people] who liked the simple models, and they challenged those who had more complex models. Why can't this be simpler? Also, the people with the more complex models often challenged the people with the simpler models. Why can't you add more realism or more detail?

This idea is something we're going to need to work out because it creates, I think, an important tension, and a useful tension, to find that right balance. However, it's something that we need to give more thought to in the future, lest we have a schism where we break up into complex and simple agent-based modeling in the long term, because I think they should coexist. They exist on a continuum, and we need to find the right balances for different types of models.

The fourth point, which Kathleen Carley mentioned, is the idea of education. I'm going to try to extend that point slightly by saying that education covers at least four different constituencies. The first, and probably nearest to my heart, are the model developers. We need to educate more people on agent-based modeling and to increase the level of skill of those who are already creating these models.

I also think, though, that model builders are very important, and these people are different than the developers. The developers are writing code, but the model builders are thinking about what the models should do. The developers are deciding what should be here and what should not be there. They're the people who are the genuine social scientists thinking about the social problems, and I think that we need to do more to educate them on what happens at each stage of the development, what these tools can do, what they cannot do, and how they can be effectively used.

The model operators make up another constituency. It's very common that the people who physically run the models are not the people who originally designed them, and this raises a host of issues in terms of glass box versus black box problems, making sure that the operators understand what it is that they are operating, particularly if the models are deployed and given to people who did not originally design them, which is starting to happen in some areas.

The final group is what I'll call model consumers. These are the people who look at the results of models; they are the policymakers. Other people are making decisions based on the results that are coming out of these models, and the [policymakers] make up a unique constituency with special educational requirements. Kathleen [Carley] and others talked about things that we can do to make these models trustworthy and credible for these people, and that's very important. We need to provide much more help to people who probably will never rise to the level of actually building a model, nor should they have to, but yet they still need to make informed choices on these very complex problems.

The fifth point is the idea of tools, and this area is, the way I see it, starting some things and stopping others. In particular, I mean starting things such as more flexible topologies, networks, and combinations of grids and networks. At the same time, I also mean stopping other things in the sense that stopping just the grid over and over again. I don't mean to say that there's anything wrong with grid models, but the idea is that we need to rise above that level of simplicity, I guess we would call it, and to try to develop, not necessarily more realism, but simply a wider range of possible things to be considered. That's a point that I've tried to emphasize, and Scott Page also repeatedly mentioned the idea of grids, grids, grids. Well, grids are fine, but it's important to test a wider range of spaces.

The next point, my sixth point, is support for diversity. Again, this is a point that Scott Page made extremely well, and that is that one of the things I liked most about this community and also the other agent-based modeling communities. It is the idea that we're actually willing to listen to someone else's model, listen to someone talking about something outside of our very narrowly focused subspecialties, the idea that many, if not all of us, feel that we can actually learn something from some other discipline.

Okay, this is actually a shocking revelation in many parts of the world, and it's outright heresy in others. I think that it is something very good about our community and something that we need to work hard to encourage and continue, lest we eventually vulcanize as well and fail to learn from other disciplines.

My final point, and this is John Padgett's comment, has to do with verification and validation. I would venture to guess that there are a reasonable number of people in this room who aren't sure what the definitions of verification and validation are, and that is not meant as an insult. I may be one of them, hopefully not. I actually looked up those words in the dictionary before I gave this speech.

But the basic idea is that we — and this includes myself — need to do more to verify and validate these models: verify, check the model against its design, validate, making sure that that design, in fact, has something to do with the real world in the way we claim it does. That seems to be an important part of our future as these models become more important, as they start to make larger comments on science and larger comments on the world. As John [Padgett] said, we need to make sure that those comments are actually correct. Again, that applies to me as much as anyone else.

[My intention is not to end on a] negative note, but every person should have one slide that explains what they did to check these models, at least one, and that's just a good start. I'm saying you should fill it up with stuff that you did to check these models. It's very commonly done in other areas. Some people did it; some people didn't. That's something we can do in the future. We can make sure that we not only do the verification, but that we also tell other people that we did it. One thing that's often missed [is the verification]. I've heard it said a couple times [in the discussions that followed the presentations]. How did you check this model? I didn't see you checking it at all. Then they'll explain all the things they did offline during the questioning. It's even part of the education, but we need to present those things up front so people know what we did.

In general, though, I was very excited by the things I saw, and I have to say that I learned a lot during this conference myself as a person. I saw many things I hadn't seen before, and I think I was very heartened by that fact, and the fact that at least half the audience stayed to listen to this closing panel well after the scheduled time is also a compliment. I'd like to thank you for being here.

Macal: It's just one minute after the scheduled time. Thank you, Mike, for those comments. Fabio [Rojas], would you care to make some comments, please?

Fabio Rojas: Yes. I'd also like to thank you for waiting to hear me. My comments are more of a personal nature rather than a technical nature. My perspective is that I'm finishing up my Ph.D. in sociology, and I'm starting to get into agent-based modeling. My other perspective is that one of my main research interests is in the evolution of intellectual communities. In fact, my dissertation's on the evolution of various academic disciplines and intellectual groups. So I see this workshop as almost an ethnography, rather than just a transferring of information from people to me. I'll go through a couple of observations quickly.

The first observation [comes from a question I asked myself]. How do things catch on in intellectual communities? How are they imported from one intellectual community into another? A couple of sociologists have written about it, and they argue that you need a good catch phrase. So we have a couple of catch phrases like "edge of chaos" or "Rule 30" or whatnot. In sociology, we have our own catch phrases like "strength of weak ties" or "embeddedness." And so in order to really understand how agent-based modeling can contribute to established academic communities such as sociologists, it's important to figure out where you can get the real key or core lessons. You could think of it like this: if you don't understand anything else, if you're scared of math, and you really don't even know how to use a mouse and a keyboard, you should know this about agent-based models.

Once I came to that conclusion, I started thinking about what sociologists actually know about agent-based models. At the University of Chicago, there aren't too many courses that deal with this, but some does filter into the literature in other courses on other substantial topics. One model that comes up frequently was mentioned many times during this conference: the garbage can model. Prisoners' Dilemma also came up very frequently in all sorts of classes that noncomputationally oriented people like to take. The Schelling model is extremely popular amongst sociologists, because they're very interested in residential segregation. The common thought among these three models is that they're all simple. This idea gets back to the division between simple and complex models, but if anything else, simple is great advertising.

For example, it would be very difficult for me to say ... let me tell you about bio-war. If I went back and saw my colleagues on Monday, I don't think I could explain bio-war very well. But I could explain some of the other models that I've seen. And that's not to "dis" bio-war or any other complex models, but just introduce an issue that sometimes advertising is important. Packaging is very important, and sometimes it's good to ask, what's the one thing that our intellectual community — and hopefully now that I'm starting to work with people developing my own models I can say that our intellectual community — what can we come up with that's our one calling card to another intellectual community we'd like to reach?

A great example of that, which some people in this room may not like, was the economic imperialism experienced by political science in the 1960s, 1970s, and 1980s. Economists came up with a very simple, straightforward, easy-to-explain theory of human behavior that was very flexible. That was their calling card: utility maximization. Any problem — marginal cost analysis, utility maximization — here's my card. I think a lot gets lost in that transmission, but it's important to remember that sometimes successful spread of ideas, whether you like the economic approach, political behavior's a different issue, but it is successful to some degree, and did impose some order —intellectual order — in political science.

My second observation is that I feel that some kinds of research areas need to be developed. I'm not saying that no one's researching anything, but in the sense that we need to have "core fields." The example that comes to mind in the history of science is David Hilbert, who, at the International Congress of Mathematicians in 1900, said, “Here are 20 problems that mathematicians can work on. You don’t have to work on them, but they seem to hit on core issues.” That [comment] really helped to push a number of fields in very positive directions in trying to solve these problems. I think it would be very interesting if the agent-based modeling community could come up with a David Hilbert and say that it’s important to have a flexible modeling framework that can be adapted to many different situations, but let’s try to work out — and Scott Page brought mentioned this frequently — some fundamental principles, or let’s choose some basic problems that we’re really going to try to solve that will help connect our theoretical foundations to our applications.

Another observation is that we should accept a division of labor. For example, from my perspective, it seems that academics are more attracted to simple models than are policymakers. Physicists probably like a simple equation or two, whereas engineers are willing to live with messy simulations, and I don’t think that’s a bad situation. I think that a lot of the argument of one versus the other could be saved for just admitting there’s a division of labor; that’s good. Academia’s often about basic science and basic research, and so maybe simple models can help you there. But actually predicting something in a country, or a bridge collapse, or a flood, or some complex event may require an abandonment of that. It’s okay to have a division of labor.

I’m going to end on a negative note, but I’m trying to make it a positive note. A sociologist colleague of mine has written a book called *Chaos of Disciplines*, and he makes the argument that interdisciplinary academic groups are doomed to failure because people need to become successful, and the way they become successful is by appealing to core audiences and academic groups, and they have to reproduce to some degree what is happening in existing academic communities.

So one way around that, something I learned from my own research in studying the history of computer science, was that early on one way that computer scientists were able to say that they were a real academic group with their own science was they had good, positive backing from the federal government. Don’t get me wrong; I’m not saying the ‘feds’ should sponsor everything. What I mean is that, at various universities, the federal government gave some money for a mainframe, and then there was a research group around that mainframe. After a number of years, the mainframe developed into its own, the research group around the mainframe developed into its own really autonomous group with its own research problems, and that eventually established computer science as something more than just applied engineering, or library science. I don’t know if people know that, but computer science used to be part of librarianship programs, because computers were used to store and sort information.

So having concrete projects — institutionally backed concrete projects — can really help an academic, an interdisciplinary academic group, get its own identity and tell the whole world that we have our own things, and we have something useful for you to use.

On that note, I’ll end [my comments] and pass it on to Sean [Gailmard].

Sean Gailmard: Thank you. My name is Sean Gailmard, and I’m an Assistant Professor in Public Policy at the University of Chicago. My perspective comes from the work that I do on applications of game theory, especially in political science. I am also a survivor of the Santa Fe Institute Work Camp and had a lovely time there [laughter].

Coming from that perspective and having a strong interest in agent-based modeling, it's interesting to see some similarity and some overlap in an agent-based modeling conference and a formal theory or applied game theory section of panels at a political science conference. The reason that these people get together is because they share a technique-based or approach-based overlap of interests in dealing with an incredibly wide-ranging set of problems. On a formal theory panel, you find people who are interested in bureaucracy and legislative structure, comparative politics, electoral behavior, international relations, and you name it; it spans the entire range of the discipline. What they share is a belief that simple analytical models can reveal important insights about all those different types of things.

That's exactly the same kind of thing that brings everybody together at an agent-based modeling conference, only the diversity of interest is really magnified. It also throws in an interesting wrinkle that you don't have on a formal theory panel at a political science conference. [That wrinkle] is in some cases a strong disagreement about the appropriate role of anything approaching optimizing behavior or taking that even as a useful benchmark in model building. There's common agreement in formal theory panels that, sure, you do, and there's — and even departures from it, for example, agent-based work presented on those kinds of panels — can be presented in a way that's useful in that framework. And at an agent-based modeling conference, that's a little tougher to find.

So while I'm the last person that should prognosticate or pontificate about where agent-based modeling is heading, one possible issue that it might have to deal with — if it chooses to see itself as a coherent community — is how those types of differences play out. I think it relates a lot to some of the stuff that Scott [Page] brought up in his talk on diversity. It's bringing a diversity of approaches into this subdiscipline or set of techniques, and that brings with it benefits and costs. That may be one important point — an important bridge — that needs to be crossed as this set of techniques develops and remains coherent.

I also am struck by the number of times at a conference like this people talk about the importance of agent-based modeling in some sense turning a corner or maturing or developing. It's an interesting notion of self-awareness, of some good stuff happening in this field, and interesting things: first of all fun and second, for the applied social sciences, useful techniques that can used, that agent-based modeling brings to the table. And so, one constantly gets the sense of asking, how is this set of techniques or this set of tools going to turn the corner?

Unlike Fabio, I have no systematic interest, and I'm certainly not writing anything like a dissertation on it, but I do again analogize to what is more readily available for me, and that is the experience of game theory in economics and in political science. Game theory reached a point in economics where no self-respecting economist can get by, can get through graduate school or we hope life, without knowing the core models in game theory. And that *should* also be true in political science, and for many political scientists, and we hope more and more every day that it is true.

But it's a success story in terms of something that turned a corner and at one point *wasn't* past that corner. It started out with some fanfare and then was pretty boring for a good long while, and it wasn't until ... so what caused it to turn that corner? When did it become something that was really important, critical to have in everybody's toolkit? I'm not a true historian of economic theory, but it does seem that it was the development of information economics — models of incomplete information and information economics — that really allowed that corner to be turned, because it was at that point that a set of questions could be answered with a new modeling framework in a compact way that people were always interested in and that couldn't be answered as well as before.

For example, George Akerlof's paper on the market for lemons ["The Market for Lemons: Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* **84**:488–500, August 1970] is, by his own admission, clunky, and when cast as a game of incomplete information, is actually very simple. The same is true of Mike Spence's model on job market signaling ["Job Market Signaling," *Quarterly Journal of Economics* **87**(3):355–379, August 1973].

So game theory became important because it provided a set of techniques that allowed people to answer questions that they had always been interested in. It helped them formulate and formalize exactly what those questions were and how to answer them. It seems to me that agent-based modeling offers many of the same kinds of promises. Adaptive behavior, combined with complex network interaction and spatial structures, these kinds of things aren't well handled with simple equation-based models that people from my own perspective are most comfortable and familiar with.

That's why Schelling's model was quickly a big hit. It revealed something useful that you couldn't immediately see with other types of models. [Ken] Kollman, [John] Miller, and [Scott] Page did work on policy choice in a multi-dimensional space by politicians. Formal political scientists had wrestled with and struggled with these issues for a long time and had not been able to make a great deal of headway with them because of the lack of preference-induced equilibrium conditions. A new set of techniques allowed very simple statements to be made unequivocally about what kinds of things we should expect in those situations. Those sorts of discoveries about questions that people in these disciplines have always been interested in are what allows something, it seems to me, in my own off-the-cuff view of it, to turn the corner.

Those are things that I think agent-based modeling has. So how do you bring more of them out? If I could do that, I would be doing it according to some recipe. But I will just end by saying, how does that happen? As far as recipes go, the best thing we can do — I'll just echo what Michael and Fabio said — is to adhere to good modeling discipline, making things nice and simple, so that simple, but important, lessons can be gleaned and understood right away without too much modeling complexity. It's very important to build that into these models, and I think it will help to get to the point where, at agent-based modeling conferences, we're not wondering how we can turn the corner and how we can mature, but perhaps we're even reflecting on the golden days when that maturation and corner-turning actually took place.

Macal: Thank you for those comments and thoughts. Do we have any questions, or would anybody like to make any comments or statements? Queries perhaps? Concluding remarks?

Scott Page: Recently, I looked back through two books — and this echoes the things that have already been mentioned — *Thinking Strategically,* by Dixit and Nalebuff, and *Harnessing Complexity,* by Axelrod and Cohen. The first is a book on how to use game theory in management, and the subtitle is something like, "How to Win in Politics, Business, and Everyday Life." *Harnessing Complexity* is a book of ideas and complex systems that I'd apply to those same areas, and the subtitle is something along the lines of, "Organizational Adaptation in a Complex World."

One of the things that struck me in looking at those two books is that one reason game theory's resonated so well in the classroom and also in the business world is that it's simple: here's how you win and here's the optimal thing to do. People just love hearing that. You want to go to college and have somebody tell you, here's how to win, and here's how to do the optimal thing. There's this incredible honesty, in comparison, in the Axelrod-Cohen book, where they say that it's a path; you're doing the best you can; things will go up; things will go down. Here are some general rules to live by. But there's no sense in which you win. There's no sense in which the game tree suddenly stops and there's a big parade and it's over. There's a sense that every day it's a new set of decisions, and those decisions feed back upon themselves. I thought it was interesting.

Then at this conference [referring to another conference], there was Erik Beinhocher from McKenzie, and he said that …. McKenzie, as you all know, got serious egg on its face by having a big thing out on what are effective companies, and the big effective company was Enron in their story. And that didn't quite work. So they commissioned a huge study looking at thousands of firms over 63 years in 15 industries, and they found in fact that there are no successful companies. There are *none*. Companies that have been around a long time actually underperformed relative to other companies, so that the whole idea of a successful company is a myth. But yet, there was one point where they must have followed Dixit-Nalebuff and won, and then they sort of stopped winning or something.

I think that there's a deeper sense in that it's — Kathleen [Carley] is probably a better person to talk about this than me — a very complex thing. Firms have standard operating procedures; they have ways of seeing the world; they have routines; they have heuristics; they do things in a complex space; and at times they win and at times they lose. And it strikes me that it's complex. I think the challenge for us, and this plays off what all three of the panelists are saying, is that if you look out at the world, it doesn't jump out at you to say "equilibrium." It does jump out at you to say "complex." How do we translate that fundamental core reality into a set of tag lines, models, or rules for the road? Axelrod and Cohen tried to make rules for the road. It's just that they're so wishy-washy in a way that they don't sell as well as they probably could.

Nigel Gilbert: Just a quick comment that, of course, one ought really to apply that same idea to the notion that one can plan the development of a specialty like agent-based modeling, which, of course, the sciences are just as complex as organizations. And so, we can dream about the way in which agent-based modeling can take over the world, or at least take over the social sciences, but whether it will or not and how we might do it and whether there's a recipe or not, are all, I think, complex issues. Perhaps we oughtn't spend too much time looking at our navels in that respect. Get on with it, and hope for the best.

David Sallach: Yes, I'd like to respond to that by saying that I think another way of framing the question is how will it sweep all before it. That may not be the relevant question, but I think there is a relevant question, and that is you have a fairly novel, fairly innovative set of techniques with some new insights and so forth, and it has to be in dialogue with the traditional disciplines. The dialogue with the traditional disciplines has not always been tremendously rich or tremendously sought after, and there are reasons for that on both sides.

The traditional disciplines sometimes see the models as being too simple to even address the questions that they're primarily interested in. Much of the time, modelers see the traditional disciplines, especially sociology, political science, and so forth, as being incredibly messy, just incredibly messy. I think that there *is* a way of furthering that dialogue that has to do with complexity and complexity theory. Certainly, it means being able to say clearly to ourselves, to people that see the value of models and so forth, that many practitioners of the traditional disciplines are deeply immersed in the complexity. They may not have formal models of it, but they are immersed in the complexity, and that's why these dialogues have to go on between the two camps. I think that while I'm sure Nigel [Gilbert] is right, that it's impossible to predict how to make it such that this particular epistemology will prevail and that we could waste a lot of time on that. The place where I think we would *not* waste time is in finding and fostering dialogues with the traditional disciplines.

Gailmard: Actually, I think that's interesting about the complexity versus simplicity angle. In economics, one complaint that I think is sometimes fairly leveled against weak agent-based modeling is that it's *way* too complex, and people revel in simulation details and what happened in Simulation 73, and how it's different from what happened in Simulation 84. As far as the economics goes, no one cares. As social science goes, it's not important. Part of that is probably a selection effect for the kinds of people that got interested in agent-based modeling.

What I think it actually means, though, is there are different dimensions of simplicity and complexity. Agent-based modeling may be too simple on one end, too complex on the other…. We're right back to it. It's about finding the right balance. Again, I guess the best thing to do, or one thing to do, as in every other modeling enterprise, is look at the best ones and try to do the best you can to do.

That's really hard and doesn't give you any guidance to go on, but when you're done with the modeling exercise, when you look at your own and compare it to that of Schelling, you realize that it's definitely not as good. But you might even get some traction in being able to say that it's not even a *simple* story about something; it's incredibly complicated, and people would just be bored with that besides me.

Claudio Cioffi-Revilla: I just remembered something I hadn't thought about in a few years, and it may be pertinent to some of the ideas that are being discussed by way of conclusion to these proceedings. It's something that I call Klein's thesis. It refers to something I read in a book by the late mathematician Morris Klein. It's about mathematics and the understanding of the world. Klein makes a very important point in this delightful book. I highly recommend it. It's a lot of fun. It has history and philosophy of mathematics.

His main thesis is that, in the physical and life sciences, the reason why people study mathematics and learn these tools is not to create new mathematics or even to be good mathematicians, because mathematicians are in charge of that. The reason that a physicist picks up semester after semester of mathematics is to understand something about the real world of physics. In other words, how apples fall and how galaxies expand, and how they make spirals as opposed to something else, and so on. So in my own physics background as an undergraduate, this is a really, really instrumental view of mathematics, which is, of course, offensive to a mathematician. That's a separate discourse between those two communities.

But be that the case, the thesis is that physical scientists pick up mathematics and use mathematics not for their own sake, but because these are special tools that highlight aspects of the real world — insights about the way the real world works — that are otherwise inaccessible through any other means. They are not accessible through statistical observation or through participant observation or through any other way of witnessing or observing a phenomenon. The book deals with many examples in physics that we understand about the way the physical world works only and exclusively through the medium of mathematics. So that's Klein's thesis. He didn't call it that, but I think that's a succinct way of referring to it.

Fifteen or maybe 16 years ago, Paul Johnson from the University of Kansas invited a number of people — actually, he sent out a call for papers to put together a special issue of a journal that has changed names since then. It's one of these Pergamon journals in applied math. I think it's called *Mathematical and Computer Modeling.* I wrote a paper [for that journal] that applied Klein's thesis to a very specific problem in political science and that was the study of war. I asked, what do we know today in 1985 that we didn't know before, absent the application of mathematical methods, from stochastic processes, differential equations, mark-off processes, game theory, whatever, it doesn't matter, but through the medium of such things?

The reason I mention this is because it would be nice at some point — maybe it's too soon — but at some point it would be nice to have a clear demonstration of aspects of the real world of social life and social systems and processes that we understand today through the medium of computational modeling and related approaches that would otherwise be inaccessible because there's no way to see and understand these things in a scientific sense other than through these tools.

I think that's a high bar, but it's an important test, and a very effective one. Maybe this is a long-term task, but I think the day may come sooner rather than later when we might be able to draw up an inventory of a dozen important things that we're able to understand about the way society works through the medium of computational modeling.

Macal: Thank you, Claudio. The second law of thermal dynamics seems to dictate that our Agent 2002 conference comes to an end, in spite of our best intentions to extend it. We've certainly had a long week. Many of us started on Monday with the toolkit developers' meeting, followed by two days of Repast training, followed by the developer conference on Thursday, followed by the presentations on Friday and Saturday. In any case, I think that we've advanced our collective thinking about the area, based on the fine presentations and general discussions.

We plan to publish the workshop proceedings. I want to thank all the speakers who prepared papers in advance. On behalf of David [Sallach] and the university, as well as Argonne, we'd like to thank everyone for their attendance and attention and general energy. We hope to see you again next year.

List of Attendees

List of Attendees

Bulent Acma
Anadolu University

T.K. Ahn
Indiana University

Konstantinos Alexandridis
Michigan State University

Henry Allen
Wheaton College

Kelcy Allwein
Defense Intelligence Agency

David Backer
University of Michigan

Jonathan Bendor
Stanford University

Matthew Berland
Northwestern University

Ravi Bhavnani
University of Illinois Urbana-Champaign

Dariusz Blachowicz
Argonne National Laboratory

Marilyn Brandt
Rosentiel School of Marine/
Atmospheric Science,
University of Miami

Lisa Brouwers
Stockholm University/KTH

Daniel Brown
University of Michigan

Doug Bryan
Independent Consultant

Derek Bunn
London Business School

Roger Burkhart
Deere Company

Kathleen Carley
Carnegie Mellon University

Laura Carlson
Indiana University

Lars-Erik Cederman
Harvard University

Jen-Chien Jack Chang
Pittsburgh Supercomputing Center

David Chavalarias
Ecole Polytechnique

John Christiansen
Argonne National Laboratory

Claudio Cioffi-Revilla
George Mason University

Richard Cirillo
Argonne National Laboratory

Nick Collier
University of Chicago

Noshir Contractor
University of Illinois Urbana-Champaign

Jean Czerlinski
DRW Trading

Margot Damaser
Hines VA Hospital

Daniel Diermeier
Northwestern University

Elenna Dugundji
University of Amsterdam

Mary Ebeling
University of Surrey

Sean Gailmard
University of Chicago

Les Gasser
University of Illinois Urbana-Champaign

Nigel Gilbert
University of Surrey

Steve Goodhall
Compuware Corporation

Laszlo Gulyas
Harvard University

Nobuyuki Hanaki
Columbia University

Stephen Haptonstahl
Northern Illinois University

Michael Heaney
University of Chicago

Adam Hirsch
Woods Hole Research Center

Tom Howe
University of Chicago

Phil Huxtable
U.S. Department of Defense

Marco Janssen
Indiana University

Hyeran Jo
University of Michigan

Julie Jupp
Sydney University

G. Jiyun Kim
University of Michigan

Jeongseob Kim
Argonne National Laboratory

Matthew Koehler
Center for Army Analysis

Tim Kohler
Washington State University

Thomas Lairson
Rollins College

Doug Lauen
University of Chicago

Alina Lazar
Wayne State University

Debbie Wei Lin
University of California, San Francisco

Yung Liu
Argonne National Laboratory

Victor Lofgreen
University of Chicago

Charles Macal
Argonne National Laboratory

Karen May
Araiza Ownership Ventures

John McManus
NCORE/RSMAS/University of Miami

Jennifer Mills
NSWC Dahlgren

Tun Myint
Indiana University

Robert Najlis
Indiana University

Daniel Nepstad
Woods Hole Research Center

Michael North
Argonne National Laboratory

Serhan Ogur
Illinois Commerce Commission

Fernando Oliveira
London Business School

Elinor Ostrom
Indiana University

David O'Sullivan
The Pennsylvania University

John Padgett
University of Chicago

Scott Page
University of Michigan

N. Paramesh
University of New South Wales

Miles Parker
BiosGroup, Inc.

Dawn Parker
Indiana University

Alexander Peterhansl
Columbia University

Randy Picker
University of Chicago

Shamus Ragen
University of Chicago

William Rand
University of Michigan

Mary Rasmussen
University of Illinois at Chicago

Lisa Reyes
University of Chicago

Robert Reynolds
University of Michigan

Seung-Kyu Rhee
Korea Advanced Institute of Science and Technology

Rick Riolo
University of Michigan

Sergio Rivero
Federal University of Rodonia

Russell Robbins
Rensselaer Polytechnic Institute

Duncan Robertson
University of Oxford

Fabio Rojas
University of Chicago

David Sallach
University of Chicago

Desmond Saunders-Newton
Defense Advanced Research Projects Agency

R. Keith Sawyer
University of Washington

Carl Simon
University of Michigan

Luiz Solorzano
Woods Hole Research Center

Ricardo Sosa
Sydney University

Fred Stafford
University of Chicago

Wesley Stevens
Stevens Associates

Ron Sun
University of Missouri

Craig Swietlik
Argonne National Laboratory

Pamela Sydelko
Argonne National Laboratory

Eric Tatara
Illinois Institute of Technology

Mehmet Tetiker
Illinois Institute of Technology

Parkash Thimmapuram
Argonne National Laboratory

William Thomas II
NSWC Dahlgren

Michael Ting
University of North Carolina

Seth Tisue
Northwestern University

Cenk Undey
Illinois Institute of Technology

Josh Unterman
Northwestern University

Shalini Venkataraman
University of Illinois at Chicago

Harko Verhagen
Stockholm University/KTH

Maria Luisa Villanueva
University of Miami

Jesse Voss
University of Wisconsin

Alexander Wendt
University of Chicago

Robert Whallon
University of Michigan

Uri Wilensky
Northwestern University

Benjamin Wu
Sandia National Laboratories

Aletta Yniguez
Rosentiel School of Marine/Atmospheric
Science, University of Miami

Abigail York
Indiana University