

# Advancing the Art of Simulation in the Social Sciences

Robert Axelrod

School of Public Policy, University of Michigan, Ann Arbor, MI  
48109, USA

**Abstract.** Advancing the state of the art of simulation in the social sciences requires appreciating the unique value of simulation as a third way of doing science, in contrast to both induction and deduction. This essay offers advice for doing simulation research, focusing on the programming of a simulation model, analyzing the results and sharing the results with others. Replicating other people's simulations gets special emphasis, with examples of the procedures and difficulties involved in the process of replication. Finally, suggestions are offered for building of a community of social scientists who do simulation.

Published in Rosario Conte, Rainer Hegselmann and Pietro Terna (eds.), *Simulating Social Phenomena* (Berlin: Springer, 1997), pp. 21-40.

## 1. Simulation as a Young Field<sup>1</sup>

Simulation is a young and rapidly growing field in the social sciences.<sup>2</sup> As in most young fields, the promise is greater than the proven accomplishments. The purpose of this paper is to suggest what it will take for the field to become mature so that the potential contribution of simulation to the social sciences can be realized.

One indication of the youth of the field is the extent to which published work in simulation is very widely dispersed. Consider these observations from the Social Science Citation Index of 1995.

1. There were 107 articles with "simulation" in the title.<sup>3</sup> Clearly simulation is an important field. But these 107 articles were scattered among 74 different journals. Moreover, only five of the 74 journals had more than two of these articles. In fact, only one of these five, *Simulation and Gaming*, was primarily a social science journal.<sup>4</sup> Among the 69 journals with just one or two articles with "simulation" in the title, were journals from virtually all disciplines of the social sciences, including economics, political science, psychology, sociology, anthropology and education. Searching by a key word in the title is bound to locate only a fraction of the articles using simulation, but the dispersion of these articles does demonstrate one of the great strengths as well as one of the great weaknesses of this young field. The strength of simulation is applicability in virtually all of the social sciences. The weakness of simulation is that it has little identity as a field in its own right.

2. To take another example, consider the articles published by the eighteen members of the program committee for this international conference. In 1995 they published twelve articles that were indexed by the Social Science Citation Index. These twelve articles were in eleven different journals, and the only journal overlap was two articles published by the same person. Thus no two members published in the same journal. While this dispersion shows how diverse the program committee really is, it also reinforces the earlier observation that simulation in the social sciences has no natural home.

---

<sup>1</sup>I am pleased to acknowledge the help of Ted Belding, Michael Cohen, and Rick Riolo. For financial assistance, I thank Intel Corporation, the Advanced Project Research Agency through a grant to the Santa Fe Institute, and the University of Michigan LS&A College Enrichment Fund. Several paragraphs of this paper have been adapted from Axelrod (1997b), and are reprinted with permission of Princeton University Press.

<sup>2</sup>While simulation in the social sciences began over three decades ago (e.g., Cyert and March, 1963), only in the last ten years has the field begun to grow at a fast pace.

<sup>3</sup>This excludes articles on gaming and education, and the use of simulation as a strictly statistical technique.

<sup>4</sup>Three others were operations research journals, and the last was a journal of medical informatics.

3. As a final way of looking at the issue, consider citations to one of the classics of social science simulation, Thomas Schelling's *Micromotives and Macrobehavior* (1978). This book was cited 21 times in 1995, but these cites were dispersed among 19 journals. And neither of the journals with more than one citation were among the 74 journals that had "simulation" in the title of an article. Nor were either of these journals among the 11 journals where the program committee published.

In sum, works using social science simulation, works by social scientists interested in simulation, and works citing social science simulation are all very widely dispersed throughout the journals. There is not yet much concentration of articles in specialist journals, as there is in other interdisciplinary fields such as the theory of games or the study of China.

This essay is organized as follows. The next section discusses the variety of purposes that simulation can serve, giving special emphasis to the discovery of new principles and relationships. After this, advice is offered for how to do research with simulation. Topics include programming a simulation model, analyzing the results, and sharing the results with others. Next, the neglected topic of replication is considered, with detailed descriptions of two replication projects. The final section suggests how to advance the art of simulation by building a community of social scientists (and others) who use computer simulation in their research.

## **2. The Value of Simulation**

Let us begin with a definition of simulation. "Simulation means driving a model of a system with suitable inputs and observing the corresponding outputs." (Bratley, Fox & Schrage 1987, ix).

While this definition is useful, it does not suggest the diverse purposes to which simulation can be put. These purposes include: prediction, performance, training, entertainment, education, proof and discovery.

1. Prediction. Simulation is able to take complicated inputs, process them by taking hypothesized mechanisms into account, and then generate their consequences as predictions. For example, if the goal is to predict interest rates in the economy three months into the future, simulation can be the best available technique.

2. Performance. Simulation can also be used to perform certain tasks. This is typically the domain of artificial intelligence. Tasks to be performed include medical diagnosis, speech recognition, and function optimization. To the extent that the artificial intelligence techniques mimic the way humans deal with these same tasks, the artificial intelligence method can be thought of as simulation of human perception, decision making or social interaction. To the extent that the artificial intelligence techniques exploit the special strengths of digital computers, simulations of task environments can also help design new techniques.

3. Training. Many of the earliest and most successful simulation systems were designed to train people by providing a reasonably accurate and dynamic interactive representation of a given environment. Flight simulators for pilots is an important example of the use of simulation for training.

4. Entertainment. From training, it is only a small step to entertainment. Flight simulations on personal computers are fun. So are simulations of completely imaginary worlds.

5. Education. From training and entertainment it is only another small step to the use of simulation for education. A good example, is the computer game SimCity. SimCity is an interactive simulation allowing the user to experiment with a hypothetical city by changing many variables, such as tax rates and zoning policy. For educational purposes, a simulation need not be rich enough to suggest a complete real or imaginary world. The main use of simulation in education is to allow the users to learn relationships and principles for themselves.

6. Proof. Simulation can be used to provide an existence proof. For example, Conway's Game of Life (Poundstone 1985) demonstrates that extremely complex behavior can result from very simple rules.

7. Discovery. As a scientific methodology, simulation's value lies principally in prediction, proof, and discovery. Using simulation for prediction can help validate or improve the model upon which the simulation is based. Prediction is the use which most people think of when they consider simulation as a scientific technique. But the use of simulation for the discovery of new relationships and principles is at least important as proof or prediction. In the social sciences, in particular, even highly complicated simulation models can rarely prove completely accurate. Physicists have accurate simulations of the motion of electrons and planets, but social scientists are not as successful in accurately simulating the movement of workers or armies. Nevertheless, social scientists have been quite successful in using simulation to discover important relationships and principles from very simple models. Indeed, as discussed below, the simpler the model, the easier it may be to discover and understand the subtle effects of its hypothesized mechanisms.

Schelling's (1974; 1978) simulation of residential tipping provides a good example of a simple model that provides an important insight into a general process. The model assumes that a family will move only if more than one third of its immediate neighbors are of a different type (e.g., race or ethnicity). The result is that very segregated neighborhoods form even though everyone is initially placed at random, and everyone is somewhat tolerant.

To appreciate the value of simulation as a research methodology, it pays to think of it as a new way of conducting scientific research. Simulation as a way of doing science can be contrasted with the two standard methods of induction and deduction. Induction is the discovery of patterns in empirical data.<sup>5</sup> For

---

<sup>5</sup>Induction as a search for patterns in data should not be confused with mathematical induction, which is a technique for proving theorems.

example, in the social sciences induction is widely used in the analysis of opinion surveys and the macro-economic data. Deduction, on the other hand, involves specifying a set of axioms and proving consequences that can be derived from those assumptions. The discovery of equilibrium results in game theory using rational choice axioms is a good example of deduction.

Simulation is a third way of doing science. Like deduction, it starts with a set of explicit assumptions. But unlike deduction, it does not prove theorems. Instead, a simulation generates data that can be analyzed inductively. Unlike typical induction, however, the simulated data comes from a rigorously specified set of rules rather than direct measurement of the real world. While induction can be used to find patterns in data, and deduction can be used to find consequences of assumptions, simulation modeling can be used as an aid intuition.

Simulation is a way of doing thought experiments. While the assumptions may be simple, the consequences may not be at all obvious. The large-scale effects of locally interacting agents are called "emergent properties" of the system. Emergent properties are often surprising because it can be hard to anticipate the full consequences of even simple forms of interaction.<sup>6</sup>

There are some models, however, in which emergent properties can be formally deduced. Good examples include the neo-classical economic models in which rational agents operating under powerful assumptions about the availability of information and the capability to optimize can achieve an efficient re-allocation of resources among themselves through costless trading. But when the agents use adaptive rather than optimizing strategies, deducing the consequences is often impossible; simulation becomes necessary.

Throughout the social sciences today, the dominant form of modeling is based upon the rational choice paradigm. Game theory, in particular, is typically based upon the assumption of rational choice. In my view, the reason for the dominance of the rational choice approach is not that scholars think it is realistic. Nor is game theory used solely because it offers good advice to a decision maker, since its unrealistic assumptions undermine much of its value as a basis for advice. The real advantage of the rational choice assumption is that it often allows deduction.

The main alternative to the assumption of rational choice is some form of adaptive behavior. The adaptation may be at the individual level through learning, or it may be at the population level through differential survival and reproduction of the more successful individuals. Either way, the consequences of adaptive processes are often very hard to deduce when there are many interacting agents following rules that have non-linear effects. Thus, simulation is often the only viable way to study populations of agents who are adaptive rather than fully rational. While people may try to be rational, they can rarely meet the requirement of information, or foresight that rational models impose (Simon,

---

<sup>6</sup>Some complexity theorists consider surprise to be part of the definition of emergence, but this raises the question of surprising to whom?

1955; March, 1978). One of the main advantages of simulation is that it allows the analysis of adaptive as well as rational agents.

An important type of simulation in the social sciences is "agent-based modeling." This type of simulation is characterized by the existence of many agents who interact with each other with little or no central direction. The emergent properties of an agent-based model are then the result of "bottom-up" processes, rather than "top-down" direction.

Although agent-based modeling employs simulation, it does not necessarily aim to provide an accurate representation of a particular empirical application. Instead, the goal of agent-based modeling is to enrich our understanding of fundamental processes that may appear in a variety of applications. This requires adhering to the KISS principle, which stands for the army slogan "keep it simple, stupid." The KISS principle is vital because of the character of the research community. Both the researcher and the audience have limited cognitive ability. When a surprising result occurs, it is very helpful to be confident that one can understand everything that went into the model. Simplicity is also helpful in giving other researchers a realistic chance of extending one's model in new directions. The point is that while the topic being investigated may be complicated, the assumptions underlying the agent-based model should be simple. The complexity of agent-based modeling should be in the simulated results, not in the assumptions of the model.

As pointed out earlier, there are other uses of computer simulation in which the faithful reproduction of a particular setting is important. A simulation of the economy aimed at predicting interest rates three months into the future needs to be as accurate as possible. For this purpose the assumptions that go into the model may need to be quite complicated. Likewise, if a simulation is used to train the crew of a supertanker, or to develop tactics for a new fighter aircraft, accuracy is important and simplicity of the model is not. But if the goal is to deepen our understanding of some fundamental process, then simplicity of the assumptions is important and realistic representation of all the details of a particular setting is not.

### **3. Doing Simulation Research**

In order to advance the art of simulation in the social sciences, it is necessary to do more than consider the purpose of simulation. It is also necessary to be more self-conscious about the process of doing the research itself. To do so requires looking at three specific aspects of the research process which take place once the conceptual model is developed: the programming of the model, the analysis of the data, and the sharing of the results.

#### **3.1. Programming a Simulation Model**

The first question people usually ask about programming a simulation model is, "What language should I use?" My recommendation is to use one of the modern procedural languages, such as Pascal, C or C++. <sup>7</sup>

The programming of a simulation model should achieve three goals: validity, usability, and extendibility.

The goal of validity is for the program to correctly implement the model. This kind of validity is called "internal validity." Whether or not the model itself is an accurate representation of the real world is another kind of validity that is not considered here. Achieving internal validity is harder than it might seem. The problem is knowing whether an unexpected result is a reflection of a mistake in the programming, or a surprising consequence of the model itself. For example, in one of my own models, a result was so counterintuitive that careful analysis was required to confirm that this result was a consequence of the model, and not due to a bug in the program (Axelrod, 1997a). As is often the case, confirming that the model was correctly programmed was substantially more work than programming the model in the first place.

The goal of usability is to allow the researcher and those who follow to run the program, interpret its output, and understand how it works. Modeling typically generates a whole series of programs, each version differing from the others in a variety of ways. Versions can differ, for example, in which data is produced, which parameters are adjustable, and even the rules governing agent behavior. Keeping track of all this is not trivial, especially when one tries to compare new results with output of an earlier version of the program to determine exactly what might account for the differences.

The goal of extendibility is to allow a future user to adapt the program for new uses. For example, after writing a paper using the model, the researcher might want to respond to a question about what would happen if a new feature were added. In addition, another researcher might want someday want to modify the program to try out a new variant of the model. A program is much more likely to be extendible if it is written and documented with this goal in mind.

### **3.2. Analyzing the Results**

Simulation typically generates huge amounts of data. In fact one of the advantages of simulation is that if there is not enough data, one can always run the simulation again and get some more! Moreover, there are no messy problems of missing data or uncontrolled variables as there are in experimental or observational studies.

Despite the purity and clarity of simulation data, the analysis poses real challenges. Multiple runs of the same model can differ from each other due to

---

<sup>7</sup>For small projects, it may be easiest to program within a graphics or statistical package, or even a spreadsheet. For a discussion of alternative programming languages, see Axelrod (1997, 209f).

differences in initial conditions and stochastic events. A major challenge is that results are often path-dependent, meaning that history matters. To understand the results often means understanding the details of the history of a given run. There are at least three ways in which history can be described.

1. History can be told as "news," following a chronological order. For example, a simulation of international politics might describe the sequence of key events such as alliances and wars. This is the most straightforward type of story telling, but often offers little in explanatory power.

2. History can be told from the point of view of a single actor. For example, one could select just one of the actors, and do the equivalent of telling the story of the "Rise and Fall of the Roman Empire." This is often the easiest kind of history to understand, and can be very revealing about the ways in which the model's mechanisms have their effects over time.

3. History can also be told from a global point of view. For example, one would describe the distribution of wealth over time to analyze the extent of inequality among the agents. Although the global point of view is often the best for seeing large-scale patterns, the more detailed histories are often needed to determine the explanation for these large patterns.

While the description of data as history is important for discovering and explaining patterns in a particular simulation run, the analysis of simulations all too often stops there. Since virtually all social science simulations include some random elements in their initial conditions and in the operation of their mechanisms for change, the analysis of a single run can be misleading. In order to determine whether the conclusions from a given run are typical it is necessary to do several dozen simulation runs using identical parameters (using different random number seeds) to determine just which results are typical and which are unusual. While it may be sufficient to describe detailed history from a single run, it is also necessary to do statistical analysis of a whole set of runs to determine whether the inferences being drawn from the illustrative history are really well founded. The ability to do this is yet one more advantage of simulation: the researcher can rerun history to see whether particular patterns observed in a single history are idiosyncratic or typical.

Using simulation, one can do even more than compare multiple histories generated from identical parameters. One can also systematically study the affects of changing the parameters. For example, the agents can be given either equal or unequal initial endowments to see what difference this makes over time. Likewise, the differences in mechanisms can be studied by doing systematic comparisons of different versions of the model. For example, in one version agents might interact at random whereas in another version the agents might be selective in who they interact with. As in the simple change in parameters, the effects of changes in the mechanisms can be assessed by running controlled experiments with whole sets of simulation runs. Typically, the statistical method for studying the effects of these changes will be regression if the changes are quantitative, and analysis of variance if the changes are qualitative. As always in



statistical analysis, two questions need to be distinguished and addressed separately: are the differences statistically significant (meaning not likely to have been caused by chance), and are the differences substantively significant (meaning large enough in magnitude to be important).

### **3.3. Sharing the Results**

After cycling through several iterations of constructing the model, programming the simulation, and doing the data analysis, the final step in the research is sharing the results with others. As in most fields of research, the primary method of sharing research results is through publication, most often in refereed journals or chapter-length reports in edited collections. In the case of social science simulation, there are several limitations with relying on this mode of sharing information. The basic problem is that it is hard to present a social science simulation briefly. There are at least three reasons.

1. Simulation results are typically quite sensitive to the details of the model. Therefore, unless the model is described in great detail, the reader is unable to replicate or even fully understand what was done. Articles and chapters are often just not long enough to present the full details of the model. (The issue of replication will be addressed at greater length below.)

2. The analysis of the results often includes some narrative description of histories of one or more runs, and such narrative often takes a good deal of space. While statistical analysis can usually be described quite briefly in numbers, tables or figures, the presentation of how inferences were drawn from the study of particular histories usually can not be brief. This is mainly due to the amount of detail required to explain how the model's mechanisms played out in a particular historical context. In addition, the paucity of well known concepts and techniques for the presentation of historical data in context means that the writer can not communicate this kind of information very efficiently. Compare this lack of shared concepts with the mature field of hypothesis testing in statistics. The simple phrase " $p < .05$ " stands for the sentence, "The probability that this result (or a more extreme result) would have happened by chance is less than 5%." Perhaps over time, the community of social science modelers will develop a collection of standard concepts that can become common knowledge and then be communicated briefly, but this is not true yet.

3. Simulation results often address an interdisciplinary audience. When this is the case, the unspoken assumptions and shorthand terminology that provide shortcuts for every discipline may need to be explicated at length to explain the motivation and premises of the work to a wider audience.

4. Even if the audience is a single discipline, the computer simulations are still new enough in the social sciences that it may be necessary to explain very carefully both the power and the limitations of the methodology each time a simulation report is published.

Since it is difficult to provide a complete description of a simulation model in an article-length report, other forms of sharing information about a simulation have to be developed. Complete documentation would include the source code for running the model, a full description of the model, how to run the program, and the how to understand the output files. An established way of sharing this documentation is to mail hard copy or a disk to anyone who writes to the author asking for it. Another way is to place the material in an archive, such as the Interuniversity Consortium for Political and Social Research at the University of Michigan. This is already common practice for large empirical data sets such as public opinion surveys. Journal publishers could also maintain archives of material supporting their own articles. The archive then handles the distribution of materials, perhaps for a fee.

Two new methods of distribution are available: CD-ROM, and the Internet. Each has its own characteristics worth considering before making a selection.

A CD-ROM is suitable when the material is too extensive to distribute by traditional means or would be too time-consuming for a user to download from the Web. A good example would be animations of multiple simulation runs.<sup>8</sup> The primary disadvantage is the cost to the user of purchasing the CD-ROM, either as part of the price of a book or as a separate purchase from the publisher.

The second new method is to place the documentation on the Internet. Today, the World Wide Web provides the most convenient way to use the Internet. By using the Internet for documentation, the original article need only provide the address of the site where the material is kept. This method has many advantages.

1. Unlike paper printouts, the material is available in machine readable form.
2. Unlike methods that rely on the mail, using the Web makes the material immediately available from virtually anywhere in the world, with little or no effort required to answer each new request.
3. Material on the Web can be structured with hyperlinks to make clear the relationship between the parts.
4. Material on the Web can be easily cross-referenced from other Web sites. This is especially helpful since, as noted earlier, social science simulation articles are published in such a wide variety of journals. As specialized Web sites develop to keep track of social science simulations, they can become valuable tools for the student or researcher who wants to find out what is available.<sup>9</sup>
5. Material placed on the Web can be readily updated.

---

<sup>8</sup>A pioneering example will be the CD-ROM edition of Epstein and Axtell (1996) published by the Brookings Institution. The CD-ROM will operate on both Macintosh and Window platforms and contain the complete text as well as animations.

<sup>9</sup>An excellent example is the Web site maintained by Leigh Tesfatsion, Iowa State University. It specializes in agent-based computational economics, but also has pointers to simulation work in other fields. The address is <http://www.econ.iastate.edu/tesfatsi/abe.htm>.

A significant problem with placing documentation on the Web is how to guarantee it will still be there years later. Web sites tend to have high turnover. Yet a reader who comes across a simulation article ten years after publication should still be able to get access to the documentation. There are no well-established methods of guaranteeing that a particular Web server (e.g., at a university department) will maintain a given set of files for a decade or more. Computer personnel come and go, equipment is replaced, and budgetary priorities change. The researcher who places documentation on the Web needs to keep an eye on it for many years to be sure it did not get deleted. The researcher also needs to keep a private backup copy in case something does go wrong with the Web server being used.

The Internet offers more than just a means of documenting a simulation. It also offers the ability for a user to run a simulation program on his or her own computer. This can be done through a programming environment such as Java which allows the code that resides on the author's machine to be executed on the user's machine. A major advantage of this method of distributing a simulation program is that the same code can be run on virtually any type of computer. A good example is a simulation of a model of the spread of HIV infection. The description of the model, an article about its motivation, and a working version that can be run and even adapted by a distant user are all available on the Web.<sup>10</sup> One disadvantage of using Java is that it is slower in execution than a locally compiled program. Another disadvantage of using Java or a similar programming environment is that there is no guarantee that the standards will be stable enough to allow easy use in ten years.

Despite the need to assure the durability of one's own Web site, placing documentation and perhaps even executable programs on the Internet has so many advantages that it is likely to become an important means of providing material needed to supplement the publication of simulation research.<sup>11</sup>

#### **4. Replication of Simulations**

Three important stages of the research process for doing simulation in the social sciences have been considered so far: namely the programming, analyzing and sharing computer simulations. All three of these aspects are done for virtually all published simulation models. There is, however, another stage of the research process that is virtually never done, but which needs to be considered. This is replication. The sad fact is that new simulations are produced all the time, but rarely does any one stop to replicate the results of any one else's simulation model.

---

<sup>10</sup>The site is <http://www.nytimes.com/library/cyber/week/1009aids.html>.

<sup>11</sup>Documentation and source code for many of my own agent-based models are on the Web at <http://pscs.physics.lsa.umich.edu/Software/ComplexCoop.html>.

Replication is one of the hallmarks of cumulative science. It is needed to confirm whether the claimed results of a given simulation are reliable in the sense that they can be reproduced by someone starting from scratch. Without this confirmation, it is possible that some published results are simply mistaken due to programming errors, misrepresentation of what was actually simulated, or errors in analyzing or reporting the results. Replication can also be useful for testing the robustness of inferences from models. Finally, replication is needed to determine if one model can subsume another, in the sense that Einstein's treatment of gravity subsumes Newton's.

Because replication is rarely done, it may be helpful to describe the procedures and lessons from two replication projects that I have been involved with. The first reimplemented one of my own models in a different simulation environment. The second sought to replicate a set of eight diverse models using a common simulation system.

The first replication project grew out of a challenge posed by Michael Cohen: could a simulation model written for one purpose be aligned or "docked" with a general purpose simulation system written for a different purpose. The two of us chose my own cultural change model (Axelrod, 1997a) as the target model for replication. For the general purpose simulation system we chose the Sugarscape system developed by Joshua Epstein and Rob Axtell (Epstein and Axtell, 1996). We invited Epstein and Axtell to modify their simulation system to replicate the results of my model. Along the way the four of us discovered a number of interesting lessons, including the following (Axtell, Axelrod, Epstein and Cohen, 1996):

1. Replication is not necessarily as hard as it seemed in advance. In fact under favorable conditions of a simple target model and similar architectures of the two systems, we were able to achieve docking with a reasonable amount of effort. To design the replication experiment, modify the Sugarscape system, run the program, analyze the data, debug the process, and perform the statistical analysis took about 60 hours of work.<sup>12</sup>

2. There are three levels of replication that can and should be distinguished. We defined these levels as follows.

- a. The most demanding standard is "numerical identity", in which the results are reproduced exactly. Since simulation models typically use stochastic elements, numerical equivalence can only be achieved if the same random number generator and seeds are used.

- b. For most purposes, "distributional equivalence" is sufficient. Distributional equivalence is achieved when the distributions of results cannot be distinguished statistically. For example, the two simulations might produce two sets of actors whose wealth after a certain amount of time the Pareto distribution

---

<sup>12</sup>We were able to identify only two cases in which a previous social science simulation was reprogrammed in a new language, and neither of these compared different models nor systematically analyzed the replication process itself. See Axtell et al. (1996).

with similar means and standard deviations. If the differences in means and standard deviations could easily have happened solely by chance, then the models are distributionally equivalent.

c. The weakest standard is "relational equivalence" in which two models have the same internal relationship among their results. For example, both models might show a particular variable as a quadratic function of time, or that some measure on a population decreases monotonically with population size. Since important simulation results are often qualitative rather than quantitative, relational equivalence is sometimes a sufficient standard of replication.

3. In testing for distributional equivalence, an interesting question arises concerning the null hypothesis to use. The usual logic formulates the problem as rejection of a null hypothesis of distributional identity. The problem with this approach is that it creates an incentive for investigators to test equivalence with small sample sizes. The smaller the sample, the higher the threshold for rejecting the null hypothesis, and therefore the greater the chance of establishing equivalence by failing to find a significant difference. One way to deal with this problem is to specify in advance the magnitude of the difference that will be considered meaningful, and then use sample sizes large enough to reliably detect this amount of difference if it exists. (For more details see Axtell et al. 1996).

4. Even seemingly minor differences in two models can prevent the attainment of distributional equivalence. In the model of cultural change that we studied, the agents were activated at random. When this model was implemented in Sugarscape, the agents were sampled without replacement, meaning that each agent was activated once before any agent was activated a second time. Unfortunately, in the original implementation of the model (Axelrod, 1997a), the agents were sampled *with* replacement. This seemingly minor difference in the two versions of the model made a noticeable difference in some very long simulation runs. Had the model not been replicated, the effect of the sampling decision would not have been appreciated.

This systematic replication study demonstrates that replication is a feasible, although rarely performed, part of the process of advancing computer simulation in the social sciences. The lessons suggest that further replication would be worthwhile. The concepts and methods developed for this particular study suggest how further replications could be performed. The observation that seemingly small differences mattered suggests that it would pay to find out whether this experience was typical or not. In particular it would pay to replicate a diverse set of simulation models to see what types of problems arise.

Michael Cohen, Rick Riolo and I took up this challenge. We selected a set of eight core models to replicate. We selected these models using six criteria: (1) their simplicity (for ease of implementation, explanation and understanding), (2) their relevance to the social sciences, (3) their diversity across disciplines and types of models, (4) their reasonably short run times, (5) their established heuristic value and (6) their accessibility through published accounts. Most of the eight models meet at least five of these six criteria. To be sure we included some

models that we could completely understand, we selected one model from each of the three of us. The core models were:

1. Conway's Game of Life from 1970 (see Poundstone 1985),
2. Cohen, March and Olson's Garbage Can Model of Organizations (1972),
3. Schelling's Residential Tipping Model (1974;1978),
4. Axelrod's Evolution of Prisoner's Dilemma Strategies using the Genetic Algorithm (1987),
5. March's Organizational Code Model (1991),
6. Alvin and Foley's Decentralized Market (1992),
7. Kauffman, Macready and Dickenson's NK Patch Model (1994, See also Kauffman 1995.)
8. Riolo's Prisoner's Dilemma Tag Model (1997).

Cohen, Riolo and I implemented each of these models in the Swarm simulation system developed at Santa Fe Institute under the direction of Chris Langton.<sup>13</sup> In each case, we identified the key results from the original simulations, and determined what comparisons would be needed to test for equivalence. After a good deal more work than we had expected would be necessary, we were able to attain relational equivalence on all eight models. In most cases, the results were so close that we probably attained distributional equivalence as well, although we did not perform the statistical tests to confirm this.

We hoped to find some building blocks that were shared by several of these models that could provide the basis for a set of useful simulation techniques. Instead, we found little overlap. On the other hand, Riolo and Ted Belding developed a useful tool for running batch jobs of a simulation program to execute experimental designs.<sup>14</sup>

The most important discovery we made in replicating these eight models is just how many things can go wrong. Murphy's Law seemed to be operating at full strength: if anything can go wrong it will. Listing the problems we discovered and overcame may help others avoid them in the future. Or if they can not be avoided, at least they might be found more easily having been clearly identified at least once before.

The list below does not include the errors that we made in reimplementing the models, since the discovery and elimination of our own errors are just part of the normal process of debugging programs before they are regarded as complete and

---

<sup>13</sup>Ted Belding did the replications for the models of Schelling, and Alvin and Foley. For details on the Swarm system, see the Santa Fe Institute Web site at [www.santafe.edu](http://www.santafe.edu).

<sup>14</sup>This tool, called Drone, automatically runs batch jobs of a simulation program in Unix. It sweeps over arbitrary sets of parameters, as well as multiple runs for each parameter set, with a separate random seed for each run. The runs may be executed either on a single computer or over the Internet on a set of remote hosts. See <http://pscs.physics.lsa.umich.edu/Software/Drone/index.html>.

ready for publication.<sup>15</sup> Instead, the list below includes the problems we found in the published accounts or the programs that they describe. It should be noted that while these problems made it more difficult for us to replicate the original results, in no case did they make a major difference in the conclusions of the published accounts.

The first category of problems was ambiguity in the published descriptions. Ambiguities occurred in the description of the model, and in the presentation of the numerical results. Ambiguities in the description of the model included the order in which the agents should be updated, and what to do when there was a tie. Ambiguities in the description of the model included the meaning of a variable in a figure, and the divisor used in a table. Some of these ambiguities in the published descriptions were resolved by seeing which of two plausible interpretations reproduced the original data. This is a dangerous practice, of course, especially if multiple ambiguities give rise to many combinations of possibilities. When the original source code was available (as it was for five of the models), we could resolve ambiguities directly.

The second category of replication problems was gaps in the published descriptions. In two cases, published data was not complete enough to provide a rigorous test of whether distributional equivalence was achieved or not. In one of these cases, the author was able to provide additional data. The other gap in a published description occurred when a variable in the program could take on values of +1, 0 or -1, but was described in a way that made it appear to have only two possible values.

The third category of replication problems was situations in which the published description was clear, but wrong. One example was a case where the criteria for terminating a run of the model was not the same in the text as it was in the runs of the model for which data were reported. In another case, the description in the main text of an article was inconsistent with the appendix of the same article. Finally, there was a case in which the description in the text was a clear, but an inaccurate description of the model embodied in the source code.

The fourth and final category of replication problems were difficulties with the source code itself. In one case, the only source code available was from a printout so old that some of the characters were smudged beyond recognition. The last case was probably the most interesting and subtle of all. After a good deal of effort we tracked down a difference between the original program and our reimplementation to the difference in the way two computers represented numbers. While both computers represented floating point numbers with considerable precision, they could differ in whether or not two numbers were exactly the same. For example, is  $9/3$  exactly equal to  $2 + 1$ ? In one implementation of the model it was, but in another implementation it was not. In

---

<sup>15</sup>A great deal of effort was sometimes required to determine whether a given discrepancy was due to our error or to a problem in the original work.

models with nonlinear effects and path dependence, a small difference can have a cascade of substantive effects.

## **5. Conclusion: Building Community**

This paper has discussed how to advance the state of the art of simulation in the social sciences. It described the unique value of simulation as a third way of doing science, in contrast to both induction and deduction. It then offered advice for doing simulation research, focusing on the programming of a simulation model, analyzing the results and sharing the results with others. It then discussed the importance of replicating other people's simulations, and provided examples of the procedures and difficulties involved in the process of replication.

One final theme needs to be addressed, namely the building of a community of social scientists who do simulation. This paper began with the observation that simulation studies are published in very widely dispersed outlets. This is an indication that social science simulators have not yet built strong institutional links across traditional disciplinary boundaries, even though the work itself is often interdisciplinary in content and methodology. Certainly, the very existence of conferences like this one demonstrates that a community of simulators can and should be formed, and that the early steps are underway. The question now is what it would take to promote the growth and success of social science simulation. My answer comes in three parts: progress in methodology, progress in standardization, and progress in institution building.

This paper has already discussed suggestions for progress in methodology. The next step is to begin to establish the internal structure and boundaries of the field. In particular, converging on commonly accepted terminology would be very helpful. A host of terms is now used to describe our field. Examples are artificial society, complex system, agent-based model, multi-agent model, individual-based model, bottom-up model, and adaptive system. Having commonly accepted distinctions between these terms could certainly help specify and communicate what simulation is about.

Hand-in-hand with developing the terminology, a shared sense of the internal structure and boundaries of the field is needed. For example, simulation in the social sciences might continue to develop primarily within the separate disciplines of economics, political science, sociology and so forth. There are powerful forces supporting disciplinary research, including the established patterns of professional education, hiring, publication, and promotion. Nevertheless, if simulation is to realize its full potential there must be substantial interaction across the traditional disciplines.

Progress requires the development of an interdisciplinary community of social scientists who do simulation. Progress also requires the development of an even broader community of researchers from all fields who are interested in the simulation of any kind of system with many agents. Certainly, ecology and



evolutionary biology have a great deal to offer for the study of decentralized adaptive systems. Likewise, computer science has recently started to pay a great deal of attention to how large systems of more or less independent artificial agents can work with each other in vast networks. And mathematics has developed some very powerful tools for the analysis of dynamic systems. Even the playful field of artificial life offers many insights into the vast potential of complex adaptive systems. Conversely, social scientists have a great deal to offer evolutionary biologists, computer scientists and others because of our experience in the analysis of social systems with large numbers of interacting agents.

There are a variety of institutional arrangements that will facilitate the development of these two communities of simulators. These arrangements include journals devoted to simulation,<sup>16</sup> professional organizations, conference series,<sup>17</sup> funding programs, university courses,<sup>18</sup> review articles, central Web sites,<sup>19</sup> email discussion groups,<sup>20</sup> textbooks, and shared standards of research practice. Early examples of these institutional arrangements already exist. To realize the full potential of computer simulation will require the development of these institutional arrangements for community building. Who should be better able to build new institutions than the researchers who use simulation to study real and potential societies?

## Appendix: Eight Models Used For Replication

Here is a brief description of the eight models selected by Michael Cohen, Robert Axelrod, and Rick Riolo for replication. For a fuller description of the models and their results, see the cited material. For more information about the replications see our Web site at <http://pscs.physics.lsa.umich.edu/Software/CAR-replications.html>.

1. Conway's Game of Life, 1970 (See Poundstone, 1985).

---

<sup>16</sup>Examples of journals that have been favorable to simulation research include the *Journal of Economic Behavior and Organization*, and the *Journal of Computational and Mathematical Organization Theory*.

<sup>17</sup>An example is the series of workshops on Computational and Mathematical Organization Theory. See <http://www.cba.ufl.edu/testsite/fsoa/center/cmot/history.htm>.

<sup>18</sup>The Santa Fe Institute already has two summer training programs on complexity, both with an emphasis on simulation. One program is for economists, and one program is for all fields. The University of Michigan Program for the Study of Complexity has a certificate program in Complexity open to all fields of graduate study.

<sup>19</sup>Very useful web sites are [www.santafe.edu](http://www.santafe.edu), [www.econ.iastate.edu/tesfatsi/abe.htm](http://www.econ.iastate.edu/tesfatsi/abe.htm), and [pscs.physics.lsa.umich.edu/pscs-new.html](http://pscs.physics.lsa.umich.edu/pscs-new.html).

<sup>20</sup>One such discussion group for social science simulation is organized by Nigel Gilbert <[Nigel.Gilbert@soc.surrey.ac.uk](mailto:Nigel.Gilbert@soc.surrey.ac.uk)>.

Comment: Although this is not a social science model, it is one of the earliest and most influential simulations of artificial life.

Metric (i.e., interaction neighborhood): 2 dimensional cellular automata.

Rules: An agent stays alive if 2 or 3 neighbors are alive, otherwise it dies. New agent is born if exactly 3 neighbors are alive.

Sample result: Complex dynamic patterns arise from very simple rules applied to simple initial conditions such as a glider or a R pentomino.

## 2. Cohen, March and Olsen's Garbage Can (1972)

Comment: This is one of the most widely cited social science simulations.

Metric: organizational relations

Rules: An organization is viewed as collections of choices looking for problems, issues and feelings looking for decision situations in which they might be aired, solutions looking for issues to which there might be an answer, and decision makers looking for work.

Sample results: The timing of issues, and the organizational structure both matter for outcomes.

## 3. Schelling's Tipping Model (1974, 1978)

Comment: This is an early and well known simulation of an artificial society.

Metric: 2 dimensions, 8 neighbors

Rule: A discontented agent moves to nearest empty location where it would be content. An agent is content if more than one-third of its neighbors are of the same color.

Sample result: Segregated neighborhoods form even though everyone is somewhat tolerant.

## 4. Axelrod's Evolution of Prisoner's Dilemma Strategies (1987)

Comment: This study is widely cited in the genetic algorithms literature.

Metric: everyone meets everyone

Rule: A population of agents play the iterated Prisoner's Dilemma with each other, using deterministic strategies based upon the three previous outcomes. (There are  $2^{70}$  such strategies.) A genetic algorithm is used to evolve a population of co-adapting agents.

Sample result: From a random start, most populations of agents first evolve to be uncooperative, and then evolve further to cooperate based upon reciprocity.

## 5. March's Organizational Code (1991)

Comment: An good example of learning in an organizational setting.

Metric: 2 level hierarchy

Rules: Mutual learning occurs between members of an organization and the organizational code. The organizational code learns from the members who are

good at predicting the environment, while all members learn from the organizational code.

Sample result: There is a trade-off between exploration and exploitation. For example, there can be premature convergence of the organizational code and all the agents on incorrect beliefs

#### 6. Alvin and Foley's Decentralized Market (1992)

Comment: A good example of simulation used to study the robustness of markets.

Metric: 1 dimensional ring

Rules: Exchange is initiated by agents who broadcast costly messages indicating their interest in trade. Trade is accomplished by bilateral bargaining between pairs of agents. Agents use information from previous attempts at local trade to calculate their search strategies.

Sample result: Limited rationality with decentralized advertising and trade can do quite well, giving a substantial improvement in the allocation of resources and average welfare.

#### 7. Kauffman, Macready and Dickenson's NK Patch Model (1995. See also Kauffman 1995)

Comment: A very abstract model with an interesting result.

Metric: 2 dimensions

Rules: Each agent's energy depends on state of several agents, forming a rugged NK landscape. The entire 120x120 lattice is partitioned into rectangular patches. For each patch all possible single spin flips within the patch are examined, and one is randomly chosen which leads to lower energy within the patch.

Sample result: Ignoring some of the constraints (effects on agents beyond the current patch) increases the overall energy temporarily, but is an effective way to avoid being trapped on poor local optima.

#### 8. Riolo's Prisoner's Dilemma Tag Model (1997)

Comment: A realization of John Holland's theme about the value of arbitrary tags on agents.

Metric: soup (anyone can meet anyone)

Rules: Pairs of agents meet at random. If both agree, they play a 4 move Prisoner's Dilemma. An agent is more likely to agree to play with someone with a similar "color" (tag). Strategies use 2 parameters: probability of C after C, and probability of C after D. Evolutionary algorithm determines next generation's population.

Sample result: Tags provide a way for reciprocating agents to attain high interaction rates, but then their success is undermined by "mimics" with the same tag. Although the meaning and success of a particular tag is temporary, tags help sustain cooperation in the long run.

## References

- Alvin, P., & Foley, D. (1992). Decentralized, dispersed exchange without an auctioneer. *Journal of economic behavior and organization*, 18, 27-51.
- Axtell, R., Axelrod, R., Epstein, J. & Cohen, M. D. (1996). Aligning simulation models: a case study and results. *Computational and mathematical organization theory*, 1, 123-141.
- Axelrod, R. (1987). The evolution of strategies in the iterated Prisoner's Dilemma. In *Genetic algorithms and simulated annealing*, Lawrence Davis (ed.). London: Pitman; Los Altos, CA: Morgan Kaufman, 32-41.
- \_\_\_\_\_, (1997a). The dissemination of culture: a model with local convergence and global polarization. *Journal of conflict resolution*, 41, 203-26. Reprinted in Axelrod (1997b).
- \_\_\_\_\_, (1997b). *The complexity of cooperation: agent-based models of competition and collaboration*. Princeton, NJ: Princeton University Press.
- Bratley, P., Fox, B. & Schrage, L. (1987). *A Guide to Simulation*. Second Edition. New York: Springer-Verlag.
- Cohen, M. D., March, J. G., & Olsen, J. (1972). A garbage can theory of organizational choice. *Administrative science quarterly*, 17, 1-25.
- Cyert, R. and March, J. G. (1963). *A behavioral theory of the firm*. Englewood Cliffs, N. J., Prentice-Hall, 1963.
- Epstein, J. & Axtell, R. (1996). *Growing artificial societies: social science from the bottom up*. Washington, DC: Brookings and Cambridge, MA: MIT Press.
- Kauffman, S., Macready, W. G., & Dickinson, E. (1994). Divide to coordinate: coevolutionary problem solving. Santa Fe Institute Working Paper, 94-06-031.
- Kauffman, S., (1995). *At home in the universe*. Oxford and New York: Oxford University Press. See especially 252-64.
- March, J. G., (1978). Bounded rationality, ambiguity and the engineering of choice. *Bell journal of economics*, 9, 587-608.
- \_\_\_\_\_, (1991). Exploration and exploitation in organizational learning, *Organizational science*, 2, 71-87.
- Poundstone, W. (1985). *The recursive universe*. Chicago, IL: Contemporary Books.
- Riolo, R. (1997). The effects of tag-mediated selection of partners in evolving populations playing the iterated Prisoner's Dilemma. Santa Fe Institute Working Paper, 97-02-016.
- Schelling, T. (1974). On the ecology of micromotives. In *The corporate society*, Robert Morris (ed.). 19-64 (See especially 43-54).
- \_\_\_\_\_, (1978). *Micromotives and macrobehavior*. New York: W. W. Norton. (See especially 137-55.)
- Simon, H. A., (1955). A behavioral model of rational choice. *Quarterly journal of economics*, 69, 99-118.

