AGENT-BASED MODELING: UNDERSTANDING OUR CREATIONS

Robert L. Axtell and Joshua L. Epstein

The defining feature of agent-based models is precisely that fundamental social structures emerge from the interaction of individual agents. The shorthand for this is that the collective structures grow "from the bottom up." As a result of research pioneered at SFI, there now exists an impressive—and rapidly growing—population of agent-based models, of coral reefs, ant colonies, bird flocks, forest fires, political empires, ecosystems, technological evolutions, and traffic patterns. In the case of our own Sugarscape model, an "artificial society," we have grown economies, epidemics, tribal formations, large-scale agent migrations, and other social phenomena characteristic of human societies.

COPING WITH COMPLEXITY

A major theme in all of this work is that a system of agents following very simple rules can exhibit very complex behavior. This is heartening. Indeed, were they incapable of generating complexity, agent-based models would be of little interest, certainly to social scientists. The problem, however, is this: *if we cannot understand these artificial complex systems any better than we understand the real ones then we haven't made progress.* How, then, do we cope with the complexity of our creations? It seems to us that there are two—intertwined—issues. First, given one of these agent-based models, we want to know, speaking bluntly: how good is it?

What empirical standards, in other words, does it meet? Second, what kinds of analysis and, in turn, what analytical tools, are necessary to address that question—to establish a given model's level of empirical performance? To begin our discussion of possible standards and necessary tools, it will help to have a familiar type of model in mind.

LEVELS OF PERFORMANCE LEVELS OF ANALYSIS

In agent-based computer models a heterogeneous population of individuals (agents) interact with each other, and perhaps with an artificial environment. An agent in these models is conveniently represented as an *object,* whose *instance variables* are the agent's internal states— like stocks of food or accumulated memories—and whose methods are the agent's behavioral rules, governing feeding, reproduction, and so on.

Imagine, then, an agent-based model of an ant colony. In it, individual ant "agents" are given rules of behavior which depend only on their internal states and on information from their immediate environment. That is, the ants move around according to local rules. Is the model a success? Well, what are success criteria and what tools are needed to test that they have been met?

A relatively weak criterion would be that the actions of *individual* ants be in *qualitative* agreement with the motions of real ants, e.g., that individuals ascend pheromone gradients in foraging for food. The requisite tool is correspondingly simple: a graphical depiction of ant movements overlaid on a density plot of pheromone concentration. To test that more demanding standards are being met requires other—generally more powerful—tools.

For example, perhaps one is interested in studying the *social* structure of ant foraging behavior. Now the model is successful only if the *population* of ants distributes itself realistically in the presence of multiple food sources. And a new analytical tool may be necessary to test that this criterion is met. In particular, when a colony of ants discovers two food sources it will first systematically exploit one before moving on to the other. So, we present our artificial ant colony with two food sources. If the simulated ants separate into two more or less equal groups of ants and simultaneously exhaust the food sources—which the modeler establishes through the use of dynamic histograms of the number of ants at the food sources and at the nest—then the model would be in disagreement with the behavior of actual ant colonies. At the previous level of performance, the model was successful. At this level, it is unsuccessful. But an additional tool, the histogram, was needed to establish that.

Of course, at this point, one might modify the rules of behavior for the individual ants, eventually getting the model to produce the desired social structure of ant foraging. To us it seems natural **to say** that a model replicating both the individual (pheromone-tracking) and group behaviors (sequential exploitation of food sources) is performing at a higher level than one meeting only the first (or second) criterion. Studying the performance of the model at ever-higher levels requires a cumulation of analysis routines.

For example, given that the ant agents have correctly reproduced the qualitative group behavior—foraging sites in sequence—one wonders whether quantitative agreement could also be reached. Clearly a model capable of producing a quantitatively-accurate simulation of ant social behavior at the macro-level would be superior to one which agreed only qualitatively. But to assess whether a given artificial ant colony produces macroscopic behavior which is in quantitative agreement with real-world data one must either build quantitative analytical routines into the simulation code or use canned statistical software **to** analyze off-line the data which the model has produced. The previous level of performance analysis required only plots of distributional information. At this new level, parameter estimation routines need to be available to the user, and it is preferable to have them on-board so that these computations can be accomplished in real-time.

Finally, even if the ant colony, *qua colony*, is working right, it may be desirable to directly compare numerical data on the behavior of individual agents with that produced by an agent-based simulation. That is, one may wish to assess the performance of a model quantitatively not only in terms of the *macro-structures* produced, but also by comparing the simulated *micro-structure* with data on real-world micro-structures. For example, imagine that there exists data on the initial positions of individual ants and their seed foraging behaviors over the period of an hour. If a model could accurately come up with the number of seeds gathered by the individual ants over the course of the hour then it would be in quantitative agreement at this microscopic level. In order to establish this a modeler would have to build into the source code the equivalent of a database, so that the state variables of individual ants could be inspected at the user's discretion. Now, it may well be that the data necessary to carry out such a study have simply never been collected. But, as often happens in science, theoretical work can precede and guide empirical work.

We can summarize these levels of agent-based model performance and analysis as follows:

Level 0: The model is a *caricature* of reality, as established through the use of simple graphical devices (e.g., allowing visualization of agent motion);

Level 1: The model is in *qualitative agreement with empirical macro- structures*, as established by plotting, say, distributional properties of the agent population;

Level 2: The model produces *quantitative agreement with empirical macro-structures*, as established through on-board statistical estimation routines; and finally,

Level 3: The model exhibits *quantitative agreement with empirical micro-structures*, as determined from cross-sectional and longitudinal analysis of the agent population.

These levels are progressive in the sense that satisfactory performance of a model at level N implies that it is also satisfactory at level N-i. This addresses a problem with agent-based models that many of us have worried about: *cumulativeness*. That is, once one has built an agent-based model of some phenomenon—say ant foraging—has one learned anything which can be brought to bear on other models of the same phenomenon?

Software Problems and Solutions Unique to Agent-Based Models

The answer to this question is complicated by the fact that many different sets of agent local rules might produce the same kind of global output. That is, the mapping from micro-rules to macro-structures may be many-to-one. In such circumstances, one may encounter quite different models, all having similar behavior—all producing caricatures of ant foraging, say. When this happens one may be skeptical that anything cumulative has been achieved, although some intuition about the multiplicity of this mapping has probably been produced. However, if some of the models perform at a higher level than others then one has learned something cumulative. For instance, say that two competing ant foraging models have the agents following pheromone gradients, but one model supplements the agents with simple memory. If both models produce interesting caricatures of overall colony behavior in the presence of multiple food sources but only the model with agent memory gets the distribution of agents around the sources qualitatively right, then one reasonably feels that something has been learned about the importance of memory in such models.

LET A THOUSAND ARTIFICIAL FLOWERS BLOOM

To-date it is perhaps fair to say that most work with agent-based models has taken place at the lower levels, mostly due to the novelty of the modeling approach. However, particular models often display elements of several levels. For example, in our own Sugarscape model we have devoted a significant fraction of the overall development effort to creating specialized analytical routines for the purpose of assessing the performance of our model at Levels 1 and 2. We have found it useful to not only plot the distribution of wealth and note that it is highly skewed, as in real societies (Level 1 agreement), but to also compute Lorenz curves and Giiii coefficients in search of quantitative (Level 2) agreement with empirical data. Analysis routines comprise somewhere between 1/3 to 1/2 of the entire code, a ratio that has stayed more or less constant as additional behavioral modes have been added to the model. Most recently, in thinking about Level 3 performance— looking at individual agents — we have concluded that nothing short of a full real-time database engine, capable of performing cross-sectional and longitudinal analysis of the agent population, will suffice. The necessity of tying the simulation system to such a database will certainly drive the amount of code dedicated to analysis to well beyond 50% of the total.

Attempts to build models meeting the “caricature” standard could powerfully advance certain fields. For example, since Thucydides, the study of international relations has centered on the macrophenomena of arms races, alliances, and wars. It would be healthy for political scientists to try to “grow” these phenomena. Those purporting to know *why* the international system looks as it does might attempt to specify the *rules* they think the *agents* (states) are executing, put them on a computer, and see if those agents and rules in fact generate a world that looks more or less recognizable. It is at this lowest, “caricature,” level that agent-based models are usefully thought of as software laboratories, in which alternative rule systems can be quickly and qualitatively studied.

However, it is also true that in many scientific fields there are probably **few** problems remaining for which significant progress can be made with simulations which are not developed past the level of “caricature.” There is simply too much research which has already come; too much is already known. Yet at the next higher level, Level 1, there may be many problems which could profit from agent-based modeling. For example, an artificial ecology model in which food webs emerge having the same qualitative structure (connectivity, hierarchy) as real food webs could help ecologists to understand, say, how such structures evolve in response to environmental fluctuations. In some fields there may be very few problems left for which Level 1 models would provide important new information. Economics is a highly quantitative discipline and any agent-based artificial economy must be equipped with tools for quantitatively analyzing its performance at Level 2 or above.

Appropriate objectives will, of course, vary with the state of a field, available hardware and software, modeling resources, and so on.

AGENT-BASED MODELS

Now, when we speak of "understanding our creations," we must include our computer programs themselves: "Is my program doing what I want it to?" Although one can rarely have more than a probabilistic answer to this crucial question, there are both important pitfalls and powerful diagnostic tools unique to agent-based simulation, which we need to appreciate if we are to have confidence in our results.

Software "bugs" can have special characteristics in agent-based models. We often focus on emergent macro-structures in our models. However, certain pathologies in the agent population due to "bugs" may not be revealed by the macro-structures. For example, imagine that some agents have their internal data over-written occasionally due to an array indexing error in some (seemingly) unrelated procedure. An agent whose wealth is actually 100 might have her wealth field over-written to 0. Since 0 is not an impossible value it goes unnoticed in building up, say, a wealth histogram. To make matters worse, imagine that memory management by the operating system is moving the agents around in memory so that the agent being modified by the "bug" is different each time the "buggy" procedure is called. Such software problems are difficult to discover due precisely to the highly distributed nature of agent-based models. Indeed, the "robustness" of macro-structures to perturbations in individual agent performance— "graceful degradation," to borrow a phrase from neural networks—is often a property of agent-based models and exacerbates the problem of detecting "bugs." By contrast, it is also possible that agent-based models can display sensitive dependence on agent initial conditions. For example, in our Sugarscape model we have studied one set of runs in which particular agents were

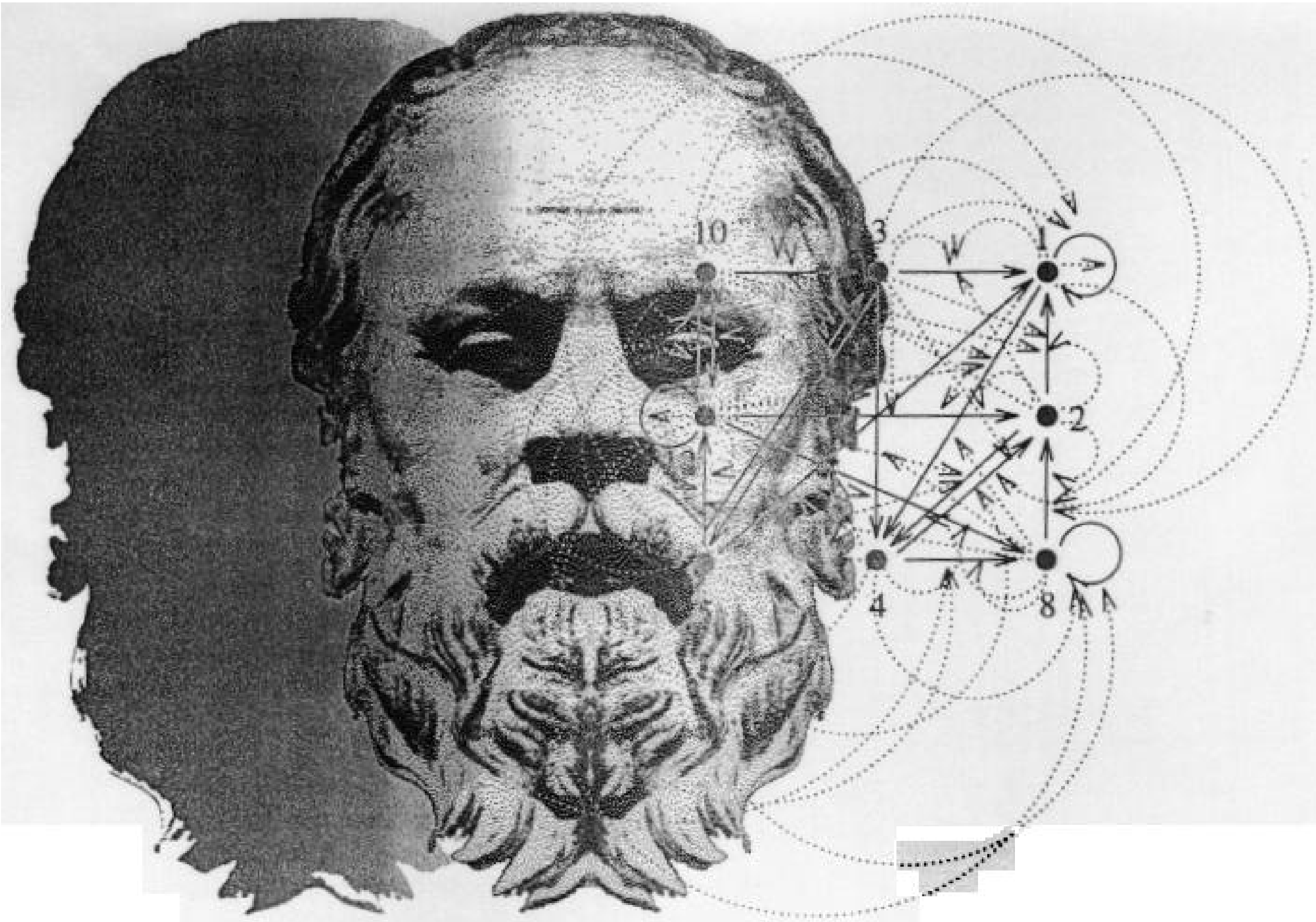
removed at the start of each run, yielding very different societal evolutions. Some agents matter a lot—they are critical or keystone agents—while others are not so important. When a model is known to have such sensitive dependence then one should be particularly wary of the possible existence of "bugs," since these could produce very large changes in the output of the model. Interestingly, the agent-based modeling approach offers novel ways to systematically address questions of software validity.

DATA GATHERING AGENTS

By mimicking the way human societies gather data about themselves—namely, by sending specialized data-gathering agents out to observe regular (non-data-gathering) agents—it is possible to uncover software bugs. Such agents can be part of the regular agent population, executing the same rules the regular agents execute, but having their rule systems supplemented by data-gathering activities. They track-down bugs by finding and FIRS1 reporting anomalous events—agents with unusual or rapidly-changing internal states, "forbidden" behavioral modes, catatonia, and so on. These agents cannot guarantee software validity but, if they are given enough clock cycles, they can go a long way toward increasing a programmer's confidence that his or her program is operating correctly. Incidentally, the use of data-gathering agents may allow us to systematically study how the presence of observers distorts that which is under observation, an enduring problem for field anthropologists and other social scientists!

DESIRABILITY OF ANALYTICALLY TRACTABLE SPECIAL CASES

Modelers in the agent-based tradition are often most interested in the transient, non-equilibrium, non-stationary behavior of their creations. However, it is clearly very desirable, given the complexity of



agent-based software, to be able to construct special cases of the model which are analytically tractable. For example, imagine an artificial economy model in which agents are in all respects neoclassical, except that preferences change due to cultural transmission. And suppose prices are observed to vary erratically. To study whether this nonequilibrium price behavior is caused by the changing preferences one could merely “turn off” the cultural transmission behavioral mode, yielding a model with fixed preferences and—dare we say it?—equilibrium behavior, for which there are known analytical results in the literature. It may also be possible to develop new analytical results. Ultimately, as has occurred in other sciences, interesting behavior may inspire the development of entirely new formal methods for analyzing agent-based models. After all, for confidence that *we* “understand our creations,” there’s nothing like outright proofs.

CONCLUSION

In summary, there are many ways to go about understanding our complex agent-based creations. Our own Sugarscape work certainly reflects the high variance of the field as a whole:
different areas of the research are at different levels. We have some caricatures, some qualitative agreements with macroscopic observables, some quantitative agreements, a few glimmers of hope that certain of the rules our agents follow might actually be close to real human behaviors, and a body of formal mathematical work applicable to some special cases. But, we have tried to move beyond mere simulation and have begun to develop tools for examining the specimens we grow. For, as Socrates would surely have said, “The unexamined Life is not worth living.”