Chapter 26

Computational Modeling as a Philosophical Methodology

Patrick Grim

Since the sixties, computational modeling has become increasingly important in both the physical and the social sciences, particularly in physics, theoretical biology, sociology, and economics. Since the eighties, philosophers too have begun to apply computational modeling to questions in logic, epistemology, philosophy of science, philosophy of mind, philosophy of language, philosophy of biology, ethics, and social and political philosophy. This chapter analyzes a selection of interesting examples in some of these areas.

Computer Models in the Sciences and Philosophy: Benefits and Limitations

What qualifies as a computer model or a computer simulation has itself been subject to philosophical scrutiny, but without clear consensus. In a classic statement, T. Naylor (1966) defines computer simulation as

> a numerical technique for conducting experiments on a digital computer which involves certain types of mathematical and logical models that describe the behavior of . . . systems over extended periods of time. (p. 3)

On the other hand, Fritz Rohrlich (1991) and Paul Humphreys (1991), among others, emphasize the importance of tractability as a motivation for computer modeling. Humphreys builds that feature into his working definition:

> A computer simulation is any computer-implemented method for exploring the properties of mathematical models where analytic methods are unavailable. (1991: 501)

Most authors have not attempted strict definition, conceding that the notion of a "model" is vague and may even have several distinct senses (Fetzer 1999). Still, several important features of models are repeatedly emphasized in the literature: (1) models occupy a conceptual role somewhere between empirical data and traditional theory; (2) modeling represents a wide variety of techniques, rather than a single tool; and (3) model construction itself is part of the "art" of science (see especially Rohrlich 1991). There is also general agreement on reasons for welcoming computer modeling in particular: (1) increased mathematical tractability, particularly in understanding complex and dynamic processes over time; (2) a methodologically important vividness or graphic immediacy that is often characteristic of computer models; and (3) the possibility of computational "experiment." This last feature is clear to anyone who has worked with computer models, and is noted in almost every outline of computational modeling since Naylor (1966), quoted above.

A number of authors portray computer experimentation in general as a technological extension of an ancient tradition of thought experiment. It is this "experimental" aspect of computational modeling that has been seen as a particularly important addition to philosophical methodology. Kyburg (1998: 37) speaks of a "kind of philosophical laboratory or testing ground." Grim, Mar, and St. Denis (1998: 10) speak of "an important new *environment* for philosophical research," and Bynum and Moor (1998: 6) speak of computing as "a medium in which to model philosophical theories and positions":

> Computing and related concepts significantly enhance philosophy by providing a kind of intellectual clay that philosophers can mold and shape and study. Through computing, abstract ideas – which philosophers like to manipulate – can be instantiated and investigated. There is nothing wrong with good armchair reflection . . . But armchair reflection has its limitations. (1998: 2–3)

The exploration of abstract philosophical ideas by means of computer models offers a number of major benefits. One benefit is an astounding increase in manageable complexity. Although philosophers have long appealed to thought experiments, practical necessity has limited these to our individual human powers of calculation. As Bynum and Moor note, "armchair recursion doesn't recur very many times." With computer models, on the other hand, the computational ceiling is lifted on philosophical imagination. Complex interactions that previously could only be vaguely guessed at can now be calculated with ease, and consequences of such interactions can be revealed with a depth previously impossible. The development of complex systems over time could hardly have been envisaged at all before the computer, but has now become a topic of philoso-phical thought experiment in a wide range of areas.

Another benefit of computer modeling is that its methodological demands work as a counterforce against philosophical vices of imprecision, vagueness, and obscurity. The environment of computer modeling enforces "unflinchingly and without compromise, the central philosophical desideratum of clarity: one is forced to construct theory in the form of fully explicit models, so detailed and complete that they can be *programmed*" (Grim, Mar, & St. Denis 1998: 10). John Pollock has emphasized that one constraint imposed by computer modeling is simply that the theory at issue must actually work the way it is supposed to. "As mundane as this constraint may seem, I am convinced that most epistemological theories fail to satisfy it." The fact that computer modeling imposes demands of precision and detail "can have a very therapeutic effect on a profession that is overly fond of hand-waving" (Pollock 1998: 34).

Another benefit of a computational environment is the prospect of exploring possible variations on theory. With a computer model in place, variations on the model are generally easy. One can explore consequences of epistemological, biological, or social theories in slightly different environments or with slightly different parameters. The result is that the theory with which one begins may be replaced by a variation that appears more promising in action.

As has always been true of the interplay between technology and pure science, it is also possible for the application of philosophical ideas in computational models at the bottom to suggest new and intriguing philosophical ideas at the top. It was computer work in graphing the semantics of infinite-valued paradox, for example, that suggested the proof for a theorem on formal undefinability of chaos (Grim, Mar, & St. Denis 1998). Models developed in order to tackle old problems may open up new territory for philosophical exploration as well. A central question in Hobbes is how cooperation can emerge in a society of self-serving egoists. Game-theoretic attempts to answer that question have been developed and expanded in computer models (see Chapter 22, Game Theory). Those models have in turn raised further questions regarding evolution and ethics, rationality and justice, and the unpredictability of social behavior in even our simplest models (Axelrod 1984, Danielson 1992, Skyrms 1996, Grim, Mar, & St. Denis 1998).

Some benefits of computational modeling are evident even in aspects for which it is occasionally criticized. Formal models in general and philosophical computer models in particular are bound to be abstract. The high level of abstraction, however, can be seen not as a weakness but as an indication of potential power and promise. Distinct phenomena that appear in quite different contexts – biology and economics, for example, or logic and physics – may nonetheless have a similar structure or exhibit a similar dynamics. The search for patterns that hold across different disciplines has characterized new fields such as chaos theory and artificial life (see Chapter 15), and promises to be an area in which philosophical computer modeling could flourish as well. As Bedau (1998: 135) remarks,

> By abstracting away from the details of chaotic systems (such as ecologies, turbulent fluid flow, and economic markets), chaos science seeks fundamental properties that unify and explain a diverse range of chaotic systems. Similarly, by abstracting away from the details of life-like systems (such as ecologies, immune systems, and autonomously evolving social groups) and synthesizing these processes in artificial media, typically computers, the field of artificial life seeks to understand the essential processes shared by broad classes of life-like systems.

One promise of philosophical computer modeling is highly abstract crossdisciplinary work of precisely this type.

Models in the physical and particularly the social sciences are also sometimes met with the objection that they are *mere* models: that the phenomena being modeled are complex in ways to which a simple model could not possibly do justice. The same is to be expected as an occasional response to philosophical computer modeling. In reply, it must simply be admitted that all models have major limitations built in. That is part of what makes them models. Models, like abstract laws, prove useful in both explanation and exploration precisely *because* they are simpler than the full phenomenon, and thus easier to handle and track. We need simple models because we need a simpler way of understanding complex phenomena, and because we need to separate what is important in what happens from the distracting but unimportant details.

One can then argue that neither the abstract level nor the simplicity of models requires further defense. With computational modeling as with any methodology, however, there are some real intellectual dangers. New methodologies always offer new ways of approaching particular *kinds* of questions. There is thus always a temptation to phrase questions in only those ways that the new method can handle, or to ask only those questions that the new method can easily address. We may end up considering only those versions of a theory that can be readily modeled, for example, or attending only to those types of theory that can be modeled at all. The only known cure for such a danger is to be aware of it. Although we do not want to ignore promising new techniques, we must be aware that they will inevitably come with limitations. For any promising new tool, there will always be further questions, equally deep and serious, for which it may *not* be the best approach.

Computational models carry a more specific danger as well. As models increase in sophistication in a particular tradition of model-building, they are inevitably built out of their simpler predecessors. Computational models often incorporate earlier passages of code wholesale. Thus, if early models in a tradition carry an inessential feature or an unexamined assumption, that feature or assumption is likely to remain, unquestioned and uncriticized, throughout later work as well. One of the earliest models to be applied to questions in economics, for example, was built using a quite particular balance of potential gains and losses: the gains and losses characteristic of the Prisoner's Dilemma, discussed further below. There is now extensive research in theoretical biology, evolutionary psychology, and philosophy using the descendants of that original model. But it is almost never asked whether the particular gains and losses built into the model reflect realistic assumptions for the specific applications at issue.

It must finally be admitted that individual models can certainly fail. Within both philosophy and the sciences a simple model may turn out to be too simple, or may be simple in the wrong ways. It may abstract not from accidental features but from fundamentally essential aspects of what is being modeled. The possibility always remains that a model captures too few aspects of the phenomenon, or the wrong ones. That models can go wrong in these ways is grounds for criticism of individual models, of course, but it constitutes no objection against a methodology of model-building in general. When a model falls short it quite generally suggests a better model.

The following sections emphasize several areas of current exploration in philosophical computer modeling. Evident in much of this work is a strong interdisciplinary or cross-disciplinary tendency. Modeling techniques developed primarily within physics have been applied within logic (Grim, Mar, & St. Denis 1998); techniques developed within computer science have been brought together with traditions in economics and sociology and applied to questions of ethics and social-political philosophy (Skyrms 1996, Danielson 1992, Grim, Mar, & St. Denis 1998). The comments of Clark Glymour and his collaborators with regard to "android epistemology" can be applied to philosophical computer modeling more generally:

> The force of this idea can be seen in the way in which it violates all kinds of traditional disciplinary boundaries in science, bringing together engineering and the life sciences, placing mathematical linguistics in the heart of electrical engineering and requiring moral philosophers to understand computation theory. University deans, forced to work within the old hierarchies, weep with frustration, and the work often has to be done in new "interdisciplinary" – and often undisciplined – research centers and institutes which escape the old categories. We live in interesting times. (Ford, Glymour, & Hayes 1995: xii)

The rest of this chapter is devoted to some key examples of philosophical modeling being done in these interesting times.

Logic

At a fundamental level, computers are logical machines: their basic operations can be outlined in terms of standard logical connectives. This immediately suggests the possibility of turning to computers as tools for extending work in traditional logic. One might expect intensive work on theorem-proving programs, for example, and indeed research of this type has an impressive history (see especially the bibliography on logic and automated theorem proving at <www.cs.cmu.edu/afs/cs/project/pal/bibs/Logictext.bib>). What is interesting, however, is that computers have also played a key role in the development of *non*traditional logical models.

Suppose we start from a set of premises P_1, and add a few more to create a larger set of premises P_2. In traditional logic, anything we could have deduced from premises P_1 will also be deducible from the inclusive set P_2: classical logics are *monotonic*. Much of everyday reasoning, however, seems to be nonmonotonic. If I am told that Tweety is a bird, I conclude that Tweety can fly. If given more information – that Tweety is a bird and a penguin, for example – I may withdraw that commitment. The need to handle "defeasible" or nonmonotonic reasoning of this kind quickly became evident in attempts at modeling patterns of reasoning in artificial intelligence, and the development of rival approaches is active and ongoing (Pollock 1998, Kyburg 1998, Gabbay, Hogger, & Robinson 1994).

The fact that computer models can offer vivid images of abstract phenomena is exploited for logical purposes in Grim, Mar, and St. Denis 1998. Here simple considerations from truth-table semantics are extended to construct "value solids," which portray in spatial terms the combinatory operation of connectives within particular systems (figure 26.1). Spatial representation of logical properties can make formal relations immediately apparent: in figure 26.1, for example, the duality of conjunction and disjunction is reflected in the fact that the value solid for "and" could be inverted and inserted into that for "or." Modeling of this sort has also produced some surprises, such as the persistent reappearance of the fractal Sierpinski gasket in a range of

's are logical
ı be outlined
ıectives. This
:y of turning
; work in tra-
ıtensive work
example, and
ın impressive
ıphy on logic
at <www.cs.
s/Logictext.
, is that com-
in the devel-
ıodels.
:mises P_1, and
:t of premises
'e could have
be deducible
al logics are
ing, however,
am told that
weety can fly.
Tweety is a
I may with-
'd to handle
oning of this
attempts at
ırtificial intel-
al approaches
998, Kyburg
on 1994).
an offer vivid
exploited for
nd St. Denis
from truth-
ıstruct "value
'ms the com-
'ithin particu-
epresentation
'mal relations
, for example,
lisjunction is
lue solid for
l into that for
lso produced
ent reappear-
in a range of

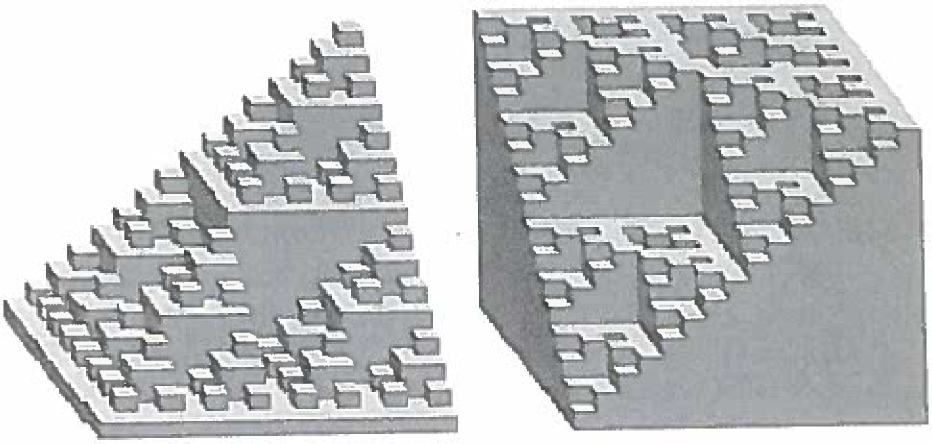


Figure 26.1: Value solids for AND (left) and OR (right)

value solids and the possibility of generating value solids by cellular-automata-like rules.

The fact that computer models can capture complex dynamics is used by Grim (1993) and Grim, Mar, and St. Denis (1998) in work on self-reference and paradox in infinite-valued logics. Informally presented, the Liar sentence ("This sentence is false") seems to produce an alternation between truth and falsity: "if it's true, since it says it's false, it must be false . . . but if it's false, and it says it's false, it must be true . . ." That dynamics is modeled in the first frame of figure 26.2. The authors consider relatives of the Liar within infinite-valued logics, such as the Chaotic Liar – "this sentence is as true as you think it is false" – which has a dynamics fully chaotic in the mathematical sense, illustrated for slight differences in initial estimated value in the second frame of figure 26.2.

Computer-modeling prospects for new approaches to logic are developed in a different way in John Barwise and Jon Etchemendy's *Hyperproof* (1994). In previous work, Barwise and Etchemendy had developed *Tarski's World* as a visual aid for teaching quantificational logic, with *Turing's World* as a similarly visual introduction to Turing machines. Sensitized from that experience to the power and ubiquity of visualization in reasoning, Barwise and Etchemendy's goal in *Hyperproof* is to expand logic beyond its current ties to sentential syntax to a formalization of information-processing that can exploit various forms of representation, diagrammatic as well as sentential (figure 26.3). Their purposes go far beyond the pedagogical. Traditional logic, as they portray it, has concentrated on only a "narrow slice" of a broader realm of valid information-extraction. "In the long run, logic must come to grips with how people use a multitude of representations in rigorous ways. This will force us to extend and enrich the traditional notions of syntax, semantics, logical consequence and proof . . . In the process, what seemed like a finished success story in philosophical and mathematical analysis will be refashioned in exciting new ways" (Barwise & Etchemendy 1998: 112).

Epistemology

One of the goals of epistemology is to understand how we come to know. It is related forms of engineering – epistemology "from the design stance," "epistemological engineering," or "android epistemology" – that have produced exciting research programs in philosophical computer modeling. It is clear that all of these will overlap with the tasks of artificial intelligence more generally (see Chapter 9, The Philosophy of AI and its Critique). Here as elsewhere in computer modeling, interdisciplinary collaboration is the rule rather than the exception.

John Pollock's OSCAR Project takes as its objective "the construction of a general theory of rationality and its computer-implementation

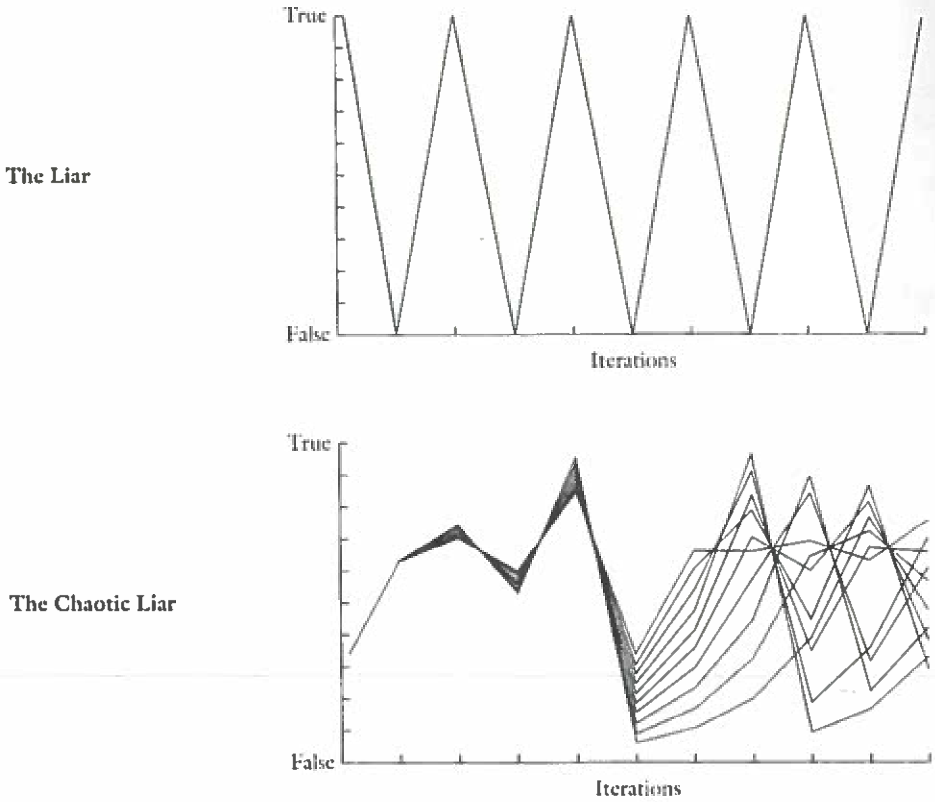
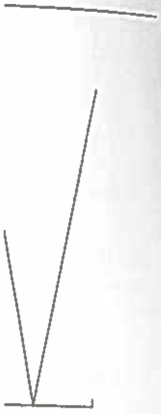


Figure 26.2: Dynamics of the Liar and the Chaotic Liar over progressive iterations. The Chaotic Liar is shown for small differences in initial values.

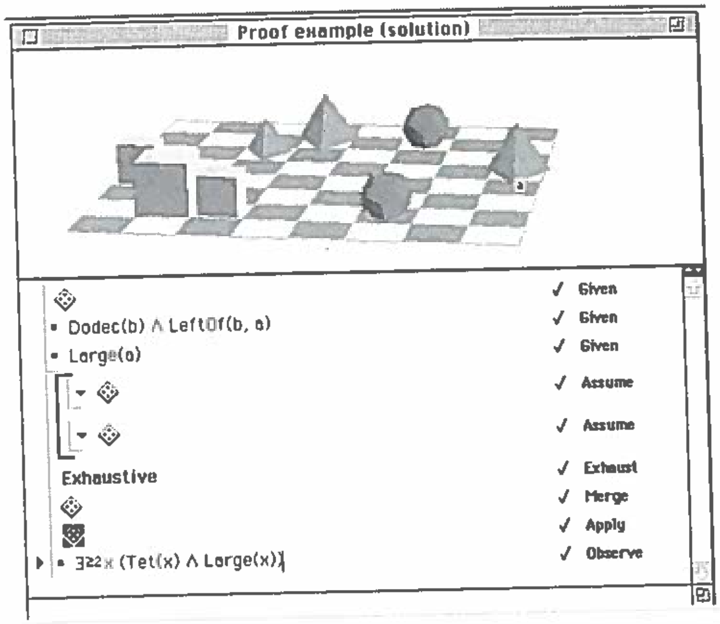
in an artificial rational agent" (Pollock 1998: 17). The general epistemic task is taken to be that of systematically updating a set of beliefs towards those which are warranted, and OSCAR is designed to exploit techniques of both defeasible and deductive logic toward that end. Pollock places particular emphasis on control of an epistemic system in terms of interests and goals. In epistemology, as elsewhere in philosophy, one of the benefits of computer modeling is the necessity of making assumptions explicit in a way that also exposes them to fruitful criticism. The fact that the program avoids Bayesian probability theory is taken as a point in favor of OSCAR's rationality by Pollock, but forms the basis for a number of criticisms in Kyburg (1998). The project is described in Pollock (1989) and (1995), and a current version of OSCAR is downloadable from <www.u.arizona.edu/~pollock/>.

Paul Thagard has used computational modeling in pursuing a wide range of issues in philosophy of science and epistemology more generally (see Chapter 23, COMPUTING IN THE PHILOSOPHY OF SCIENCE). ECHO was developed as a connectionist computational model of explanatory coherence, and Thagard has applied it to a number of examples from the history of science and in critique of other approaches (Thagard 1992, Eliasmith & Thagard 1997). He has also attempted to model major aspects of analogical thinking using the programs ARCS and ACME (Holyoak & Thagard, 1997). The deeply interdisciplinary character of much of his research is particularly clear in the work on induction (Holland, Holyoak, Nisbett, & Thagard 1987), in which Thagard works with two psychologists and a computer scientist in attempting to construct a computational model



aotic Liar is

omputational
;e of issues in
mology more
UTING IN THE
vas developed
al model of
'd has applied
the history of
r approaches
rd 1997). He
or aspects of
grams ARCS
, 1997). The
of much of
the work on
Nisbett, &
l works with
r scientist in
tional model



```
◈                              ✓ Given
▪ Dodec(b) ∧ LeftOf(b, a)      ✓ Given
▪ Large(a)                     ✓ Given
  ▾ ◈                          ✓ Assume
  ▾ ◈                          ✓ Assume
Exhaustive                     ✓ Exhaust
◈                              ✓ Merge
◈                              ✓ Apply
▸ ▪ ∃≥2x (Tet(x) ∧ Large(x))   ✓ Observe
```

Figure 26.3: A diagrammatic proof from Barwise and Etchemendy's *Hyperproof*

of scientific reasoning which avoids traditional philosophical problems.

The TETRAD project should also be mentioned as a rigorous computational attempt to treat epistemological issues regarding causal inference and probability with an eye to issues in experimental methodology (Spirtes, Glymour, & Scheines 2001). A very different computational approach explores the complex epistemological dynamics of competing information. Some of the information received by an epistemic agent is about the accuracy or reliability of information sources. In simple models, this can both produce a variety of epistemic chaos and suggest maps for the management of epistemic chaos (Grim, Mar, & St. Denis 1998).

Philosophy of Mind, Philosophy of Language, and Philosophy of Biology

It could be argued that the entire field of AI qualifies as computational modeling in philosophy of mind. Traditional debates regarding innatism and empiricism, the character and limits of the human mind, and even freedom and consciousness are now debated with illustrations drawn from competing computational architectures. Work by John Searle, Daniel Dennett, Jerry Fodor, and David Chalmers features prominently in the philosophical debate, if not so prominently in the details of computational modeling. Research by the Churchlands is noteworthy for framing contemporary debates in terms of current resources in neurophysiology and computer science, with an emphasis on neural networks as models for both the power and the peculiar inscrutability of the workings of the human mind (P. S. Churchland & Sejnowski 1993, P. M. Churchland 1995). A variety of attempts to approach issues in philosophy of mind using the tools of dynamical systems theory are represented in Robert Port and Timothy van Gelder (1997).

Some attempts have recently been made to apply tools of computational modeling to issues in the philosophy of language. Structural mapping approaches to analogy and ambiguity appear in Holyoak and Thagard 1997. They form the core of a program designed to generate and interpret metaphors in Steinhart and Kittay 1994 and

Steinhart 1995. Computer modeling can also be expected to play an important role in the inevitable conflict between Chomskian models of linguistic representation and language learning and alternative connectionist proposals (McClelland, St. John, & Taraban 1992). A game-theoretic attempt to understand linguistic convention initiated in Lewis's *Convention* (1969) is further developed with the tools of replicator dynamics in Skyrms's *Evolution of the Social Contract* (1996). In this same tradition, simple computational environments which show emergence of coordinated signaling behavior are offered as models for a theory of meaning as use in Grim, Kokalis, Tafti, & Kilb 2000 and 2002, and Grim, St. Denis, & Kokalis 2003.

Much as AI can be seen as a computational version of philosophy of mind, Artificial Life (ALife) can be seen as a computational version of philosophy of biology (see Chapter 15). Mark Bedau uses ALife to illustrate both his theory of life as supple adaptation (1996) and his consideration of emergent phenomena in biology (1997). In time, it seems inevitable that tools from AI and ALife will be brought together to answer questions in philosophy of mind and philosophy of biology. Some simple mathematical models in this direction are offered in Peter Godfrey-Smith 1996.

Ethics, Social and Political Philosophy

Computational modeling and ethics might seem an unlikely combination, but these are linked in an intriguing interdisciplinary history. Game theory (see Chapter 22) was developed by von Neumann and Morgenstern (1944) as an attempt at a mathematical theory applicable to economics and political strategy. The Prisoner's Dilemma, a two-person game that seems to capture a basic tension between individual and collective advantage, quickly became a paradigm for work in aspects of economics, theoretical sociology, and eventually theoretical biology. Played in terms of "cooperations" and "defections" on each side, the Prisoner's Dilemma has been referred to as the *e. coli* of social psychology.

In 1980, political scientist Robert Axelrod invited experts in game theory from various fields to submit programs for a Computer Prisoner's Dilemma Tournament (Axelrod 1984). Submitted strategies played 200 games against all other strategies, themselves, and a strategy that chose responses at random. The winner of that tournament was a strategy called "Tit for Tat" (TFT). Cooperate with TFT and it will cooperate with you on the next round. Defect against TFT and it will defect against you. The fact that such a cooperative strategy triumphed in the first tournament was a surprise. Its continued success in later tournaments, where its reputation clearly made it the strategy to beat, was a further surprise. Axelrod and Hamilton (1981) replaced the tournament competition with an "ecological model," which employs the replicator dynamics of theoretical biology: more successful strategies "breed" to occupy a larger percentage of the population. Here too TFT triumphs. The affinity of these results with the Hobbesian question of how social cooperation can grow in a community of self-serving egoists is striking, and the contemporary models inevitably drew the attention of philosophers. Could the triumph of something that looked like altruism in this formal model be telling us something about the dynamics or nature of social cooperation and ethics? Brian Skyrms has pursued this cluster of philosophical questions using the tools of computer modeling, adding also techniques drawn from dynamical systems or chaos theory. "Using these tools of evolutionary dynamics, we can now study aspects of the social contract from a fresh perspective" (1996: p. x). In *Evolution and the Social Contract*, Skyrms shows that an evolutionary model using replicator dynamics goes beyond rational decision theory in producing particular principles of fair division and a "law of mutual aid." The potentially chaotic dynamics of a bargaining game is illustrated in figure 26.4.

Grim, Mar, & St. Denis (1998) emphasize spatialized models of emerging cooperation. Figure 26.5, for example, shows a spatialized conquest by TFT in a field of 8 simple strategies. Nowak and Sigmund (1992) showed that a greater level of cooperation ("generous TFT," which forgives defection against it in 1/3 of all cases) triumphs in versions of Axelrod and

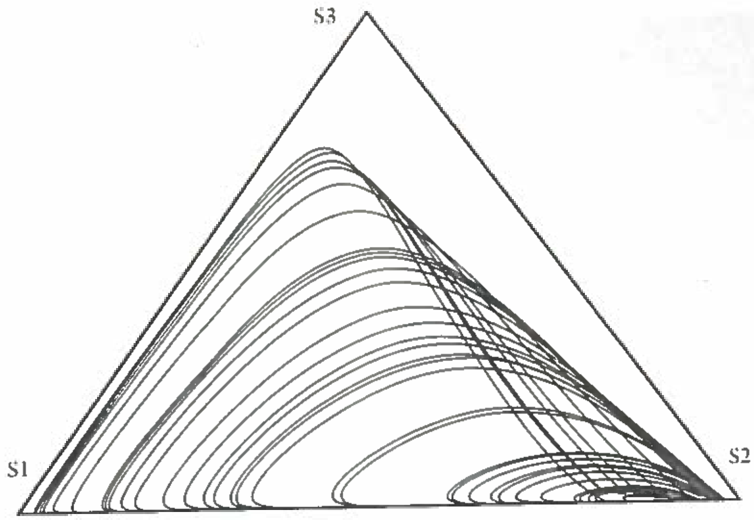


Figure 26.4: A chaotic attractor in game-theoretical dynamics (from Skyrms 1997)

Hamilton's model that are arguably more realistic in incorporating stochastic or probabilistic imperfections. Grim, Mar and St. Denis show that spatialization of such models results in an even higher level of cooperation. They also consider a version of the spatialized model with "fuzzy" values of intermediate cooperation and defection, show a formal undecidability result for the Spatialized Prisoner's Dilemma, and take some first steps toward applying the model to questions of racial discrimination.

Peter Danielson characterizes his work in this tradition as "artificial morality," intended to combine game theory and artificial intelligence in the development of "instrumental contractarianism" as an ethical theory. Building on the work of David Gauthier's *Morals by Agreement* (1986) and constructing a range of PROLOG models, Danielson's attempt is to show at least that it can be rational to be moral. Danielson seems more willing than other researchers in the area, however, to use modeling as an argument for something more: for a version of ethical naturalism in which morality simply *is* that strategy that proves successful. In some forms of such a view, such as Michael Ruse's Darwinian naturalism, the conclusion is that morality is something other than what we have thought it to be: "Morality is no more than a collective illusion fobbed off on us by our genes for reproductive ends" (Ruse 1991: 506). The use of computational models to demonstrate naturalistic conclusions of this sort is contested in Grim, Mar, and St. Denis 1998. They offer as a counterexample the success of certain discriminatory strategies, which play TFT with others of their own color but always defect against outsiders. Strategic success cannot simply be identified with morality, they argue, since discriminatory strategies are clearly successful in such models, but it is clear that analogous racial discrimination is morally wrong. How social strategies may develop or propagate is one question; whether they should be judged as genuinely ethical is another.

Powerful new tools are now available for further research in this general tradition. TIERRA and the later AVIDA are ALife platforms that may be customized to pursue questions in both philosophy of biology and social and political philosophy. TIERRA is available by anonymous ftp from <ftp://alife.santafe.edu>. A good introduction to AVIDA, packaged with the software, is Adami, *Introduction to Artificial Life* (1998). SWARM is a powerful general platform for agent-based modeling, developed by members of the Santa Fe Institute as a way of offering researchers in various fields a powerful "lingua franca" for computational experimentation. The program can be downloaded at <www.swarm.org>. A good introduction to this platform is Luna and

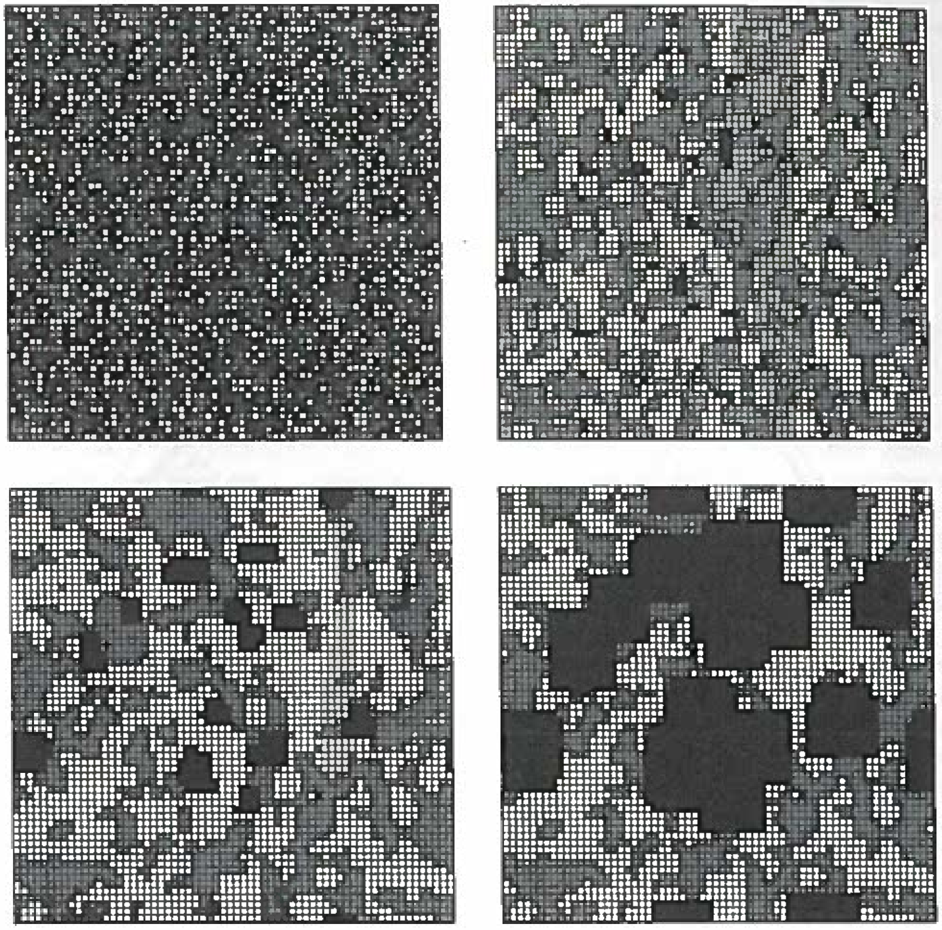


Figure 26.5: Progressive conquest by Tit for Tat, shown in black, in an array of 8 simple strategies. TFT eventually conquers the entire field.

Stefansson (2000), which offers a SWARM instantiation for the Spatialized Prisoner's Dilemma as one of its opening examples.

Conclusion

Computational modeling offers not a single systematized method but an enormous toolbox of potential models and techniques. A number of basic modeling tools – the tools of dynamical systems, neural nets, and cellular automata, for example – have found application across the physical and social sciences. The result has been active model-borrowing between different research programs and a flourishing interdisciplinary awareness. The application of computational modeling to philosophical questions has just begun, and first returns are promising in a number of areas.

It must be recognized that novel techniques carry some intellectual risks. Powerful new methods inevitably draw attention to those questions, or those forms of questions, for which the methods hold out the most promise. In philosophy as elsewhere it must be remembered that the new techniques should take their place among a range of traditional tools for approaching a range of perennial questions. It is also important to beware of fixing on a few simple models too early. Continuing development of new variations is important in order to avoid narrow assumptions and to facilitate wide exploration.

Progress in philosophical computer modeling is already proceeding so swiftly that any overview is bound to be partial and incomplete. In this chapter, the attempt has been to emphasize the general promise of such an approach by highlighting a few examples of intriguing and noteworthy current work.

Acknowledgments

I am grateful to Jason Alexander, Mark Bedau, Terrell Bynum, John Etchemendy, James Fetzer, Henry Kyburg, Paul Humphreys, James Moor, Brian Skyrms, and Eric Steinhart for helpful input, and to members of the Group for Logic & Formal Semantics and Luciano Floridi for careful comments on earlier drafts.

References

Adami, C. 1998. *Introduction to Artificial Life.* New York: Springer-Verlag. [A good introduction to the AVIDA program for artificial life. Advanced undergraduate and above.]

Axelrod, R. 1984. *The Evolution of Cooperation.* New York: Basic Books. [A major source for game-theoretic modeling regarding competition and cooperation, and a good introduction for any interested reader.]

—— and Hamilton, W. 1981. "The evolution of cooperation." *Science* 211: 1390–6. [Application of replicator dynamics to earlier game-theoretic work regarding competition and cooperation. Technical but clear.]

Barwise, J. and Etchemendy, J. 1994. *Hyperproof.* Stanford, CA: CSLI and Cambridge University Press. [A system of computer-enhanced logic. Advanced undergraduate and above.]

—— and ——. 1998. "Computers, visualization, and the nature of reasoning." In T. Bynum and J. Moor, eds., *The Digital Phoenix.* Oxford: Blackwell, pp. 93–116. [Working reflections on computer prospects for new approaches to logic. Undergraduate and above.]

Bedau, M. 1996. "The nature of life." In M. Boden, ed., *The Philosophy of Artificial Life.* Oxford: Oxford University Press, pp. 331–57. [Bedau's theory of life as supple adaptation illustrated in terms of artificial life. Suitable for any interested reader.]

——. 1997. "Weak emergence." In J. Tomberlin, ed., *Philosophical Perspectives: Mind, Causation, and World* (vol. 11). Oxford: Blackwell, pp. 375–99. [One of the best philosophical treatments of the notion of emergent phenomena. Requires some philosophical background.]

——. 1998. "Philosophical content and method of artificial life." In Bynum & Moor 1998: 37–47. [A good general discussion of artificial life as computational philosophy of biology. Undergraduate and above.]

——. 1999. "Can unrealistic computer models illuminate theoretical biology?" In A. Wu, ed., *Proceedings of the 1999 Genetic and Evolutionary Computation Conference Workshop Program,* pp. 20–3. Available on-line at <www.reed.edu/~mab/papers.htm>. [An argument for the virtues of abstract modeling. Suitable for any interested reader.]

Burkholder, L. 1992. *Philosophy and the Computer.* Oxford: Westview Press. [A valuable anthology of computer-assisted philosophical work earlier than Bynum & Moor 1998. Different contributions at different levels of difficulty.]

Bynum, T. and Moor, J., eds. 1998. *The Digital Phoenix: How Computers are Changing Philosophy.* Oxford: Blackwell. [An important anthology of work in computer modeling, from which several other references are drawn. A good introduction for anyone.]

Churchland, P. M. 1995. *The Engine of Reason, the Seat of the Soul.* Cambridge, MA: MIT Press. [A philosophically informed review of current work in neural nets, including also neurophysiology. Suitable for any interested reader.]

Churchland, P. S. and Sejnowski, T. 1993. *The Computational Brain.* Cambridge, MA: MIT Press. [A philosophically informed review of current work in neurophysiology, including also work on neural nets. Occasionally technical but clearly written.]

Danielson, P. 1992. *Artificial Morality: Virtuous Robots for Virtual Games.* New York: Routledge. [Uses formal game-theoretic modeling to argue for a form of ethical naturalism. Undergraduate and above.]

Eliasmith, C. and Thagard, P. 1997. "Waves, particles, and explanatory coherence." *British Journal for the Philosophy of Science* 48: 1–19. [An application of ECHO to the history of the wave-particle debate in physics. Some background in physics required.]

Fetzer, J. 1999. "The role of models in computer science." *The Monist* 82: 20–36. [A critical examination of various approaches to models. Advanced undergraduate and above.]

Ford, K., Glymour, C., and Hayes, P. 1995. *Android Epistemology.* Menlo Park, CA: AAAI Press/MIT Press. [An anthology of philosophical reflections on epistemological modeling. Undergraduate and above.]

Gabbay, D., Hogger, C., and Robinson, J. 1994. *Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 3: Nonmonotonic Reasoning and Uncertain Reasoning.* Oxford: Clarendon Press. [An overview of formal work in defeasible and nonmonotonic logics. Very technical, graduate and above.]

Gauthier, D. 1986. *Morals by Agreement.* Oxford: Oxford University Press. [A contractarian model important for Danielson's "artificial morality" models. Undergraduate and above.]

Godfrey-Smith, P. 1996. *Complexity and the Function of Mind in Nature.* Cambridge: Cambridge University Press. [Mathematical models relevant to both philosophy of mind and philosophy of biology. Advanced undergraduate and above.]

Grim, P. 1993. "Self-reference and chaos in fuzzy logic." *IEEE Transactions on Fuzzy Systems* 1: 237–53. [Dynamical models for the behavior of self-referential sentences in fuzzy logic. Sometimes technical, graduate and above.]

——, Kokalis, T., Tafti, A., and Kilb, N. 2000. "Evolution of communication in perfect and imperfect worlds." *World Futures* 56: 179–97. [A spatialized model in which simple patterns of communication emerge as behavioral coordination within a community. Advanced undergraduate and above.]

——, ——, ——, and ——. 2002. "Evolution of communication with a spatialized genetic algorithm." *Evolution of Communication* 3: 105–34. [A genetic algorithm model showing emergence of simple patterns of communication in order to capture food and avoid predation. Advanced undergraduate and above.]

——, Mar, G., and St. Denis, P. 1998. *The Philosophical Computer: Exploratory Essays in Philosophical Computer Modeling.* Cambridge, MA: MIT Press. [A rich sampler of philosophical computer modeling by a single research group, including a CD-ROM with animations and all source code. Undergraduate and above.]

——, St. Denis, P. and Kokalis, T. 2003. "Learning to communicate: the emergence of signaling in spatialized arrays of neural nets." Group for Logic & Formal Semantics, Dept. of Philosophy, SUNY Stony Brook, Research Report no. 01-01, forthcoming in *Adaptive Behavior.* [A neural net model showing emergence of simple patterns of communication in order to capture food and avoid predation. Advanced undergraduate level and above.]

Holland, J., Holyoak, K., Nisbett, R., and Thagard, P. 1987. *Induction: Processes of Inference, Learning, and Discovery.* Cambridge, MA: MIT Press. [A better philosophy of science through computer modeling. Graduate and above.]

Holyoak, K. and Thagard, P. 1997. "The analogical mind." *American Psychologist* 52: 35–44. [An attempt to model reasoning by analogy. Advanced undergraduate and above.]

Humphreys, P. 1991. "Computer simulations." In A. Fine, M. Forbes, and L. Wessels, eds., *PSA 1990*, vol. 2: 497–506. East Lansing, MI: Philosophy of Science Association. [A philosophical examination of computer models in the sciences. Occasionally technical, advanced undergraduate and above.]

Kyburg, H. 1998. "Epistemology and computing." In Bynum & Moor 1998: 37–47. [Includes criticism of Pollock's OSCAR project. Suitable for any interested reader.]

Luna, F. and Stefansson, B., eds. 2000. *Economic Simulations in Swarm: Agent-based Modeling and Object Oriented Programming.* Boston: Kluwer Academic. [A good introduction to a new platform for social, and biological experimentation developed by the Santa Fe Institute and downloadable at <www.swarm.org>. Different elements accessible at advanced undergraduate and graduate level.]

McClelland, J. L., St. John, M., and Taraban, R. 1992. "Sentence comprehension: a parallel distributed processing approach." In L. Burkholder, ed., *Philosophy and the Computer.* Oxford: Westview, pp. 34–56. [An opening salvo in the growing controversy between Chomskian linguistics and connectionist models. Suitable for any interested reader.]

Naylor, T. 1966. *Computer Simulation Techniques.* New York: Wiley. [An early text, now of mostly historical interest. Advanced undergraduate.]

Nowak, M. and Sigmund, K. 1992. "Tit for tat in heterogeneous populations." *Nature* 355: 250–52. [Introducing stochastic imperfection into game-theoretic models of cooperation. The result: greater generosity. Advanced undergraduate and above.

Pollock, J. 1989. *How to Build a Person: A Prolegomenon.* Cambridge, MA: MIT Press. [Problems faced and lessons drawn from the OSCAR project. Undergraduate and above.]

——. 1995. *Cognitive Carpentry.* Cambridge, MA: MIT Press. [Problems faced and lessons drawn from the OSCAR project. Undergraduate and above.]

——. 1998. "Procedural epistemology." In Bynum & Moor 1998: 17–36. [An outline of the OSCAR project. Undergraduate and above.]

Port, R. and van Gelder, T., eds. 1997. *Mind as Motion: Explorations in the Dynamics ff Cognition.* Cambridge, MA: MIT Press. [An anthology of work modeling psychological processes in terms of dynamical systems theory. Different contributions at different levels.]

Rohrlich, P. 1991. "Computer simulations in the physical sciences." In A. Fine, M. Forbes, and L. Wessels, eds., *PSA 1990,* vol. 2: 507–18. East Lansing, MI: Philosophy of Science Association. [A philosophical examination of computer models in the sciences. Graduate and above.]

Ruse, M. 1991. "The significance of evolution." In P. Singer, ed., *A Companion to Ethics.* Oxford: Blackwell, pp. 500–510. [A particularly strong version of ethical naturalism. Suitable for any interested reader.]

Skyrms, B. 1996. *Evolution of the Social Contract.* Cambridge: Cambridge University Press. [A seminal source of formal game-theoretic modeling relevant to questions in ethics and social and political philosophy. Undergraduate and above.]

——. 1997. "Chaos and the explanatory significance of equilibrium: strange attractors in evolutionary game dynamics." In C. Bicchieri, R. Jeffrey, and B. Skyrms, eds., *The Dynamics of Norms.* Cambridge: Cambridge University Press. [Game-theoretical dynamics with an emphasis on chaos. Sometimes technical, graduate and above.]

Spirtes, P., Glymour, C., and Scheines, R. 2001. *Causation, Prediction, and Search,* 2nd ed. Cambridge, MA: MIT Press. [A sophisticated attempt to model causation and prediction well enough to improve research design. Technical, graduate and above.]

Steinhart, E. 1995. "NETMET: a program for generating and interpreting metaphors." *Computers and Humanities* 28: 383–92. [An attempt at programming semantic "fields." Undergraduate and above.]

—— and Kittay, E. 1994 "Generating metaphors from networks." In J. Hintikka, ed., *Approaches to Metaphor.* Dordrecht: Kluwer Academic, pp. 41–94. [An attempt at programming semantic "fields." Undergraduate and above.]

Thagard, P. 1992. *Conceptual Revolutions.* Princeton: Princeton University Press. [The ECHO program applied to examples from the history of science. Undergraduate and above.]

Von Neumann, J. and Morgenstern, O. 1944. *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press. [The seminal work on game theory. Graduate and above.]