Barriers and Bounds to Rationality: Essays on Economic Complexity and Dynamics in Interactive Systems

by Peter S. Albin

Edited with an Introduction by Duncan K. Foley

Contents

Chapter 1

Introduction

Duncan K. Foley

1.1 Dynamical systems

One of the most fruitful conceptions in scientific investigation has been the idea of representing the state of a system of interest (whether the system is physical, biological, or social) at a given time as a vector, $\mathbf{x}_t$ in a space X, the *state space* of the system. Interesting systems are those that change through time. Thus we are led to consider the idea that the evolution of a system is governed by certain laws that define a dynamical process. In a large class of situations it turns out that the state of the system in the current time period is the only information available about the influences on its state in the next time period. Thus whatever lawful regularities the system possesses can be summarized by the relation:

$$\mathbf{x}_{t+1} = F_{\mathbf{a}}(\mathbf{x}_t) \tag{1.1}$$

In this expression $F_{\mathbf{a}}$ is an operator on the state space and the vector $\mathbf{a}$ represents potentially changeable parameters of the system.

The operator $F_{\mathbf{a}}(.)$ may represent a deterministic or stochastic process. In the deterministic case the lawful regularities determine $\mathbf{x}_{t+1}$ uniquely given $\mathbf{a}$ and $\mathbf{x}_t$, while in the stochastic case some margin of uncertainty may remain about the exact value of $\mathbf{x}_{t+1}$. The systems studied in this book are deterministic. Since deterministic systems can be regarded as degenerate stochastic systems, the conclusions we draw about deterministic systems apply to at least some stochastic systems. Negative results and

counterexamples established for the deterministic case must apply a fortiori to stochastic systems as well.

The method of dynamical system analysis has been successfully applied to a huge range of phenomena in the physical and biological sciences, as well as to a wide range of social and economic problems. Thus a great deal of mathematical effort has gone into studying the behavior of dynamical systems like (1.1).

1.1.1 Linear dynamical systems

The possible range of behaviors of one class of dynamical systems, linear systems, is completely understood. In a linear dynamical system the operator $F_{\mathbf{a}}$ is a matrix, A, so that the system can be written:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t \tag{1.2}$$

The trajectories of this system can be decomposed into independent motions of three types (see Hirsch and Smale, 1974, for a complete exposition of this theory). One type of motion is geometric expansion or contraction along a particular ray. This motion is associated with a real, positive eigenvalue of the matrix A; the ray is the corresponding eigenvector. A second type is geometric expansion or contraction in which the system jumps from one side of the origin to the other on a particular line in alternate periods. This motion is associated with a real, negative eigenvalue of A; the associated eigenvector again determines the line to which the motion is confined. The third type is inward or outward spiraling in a particular plane. This motion is associated with a pair of complex eigenvalues of the matrix A; the corresponding real and imaginary components of the associated eigenvector define the plane in which it takes place. (The use here of the word "complex" to describe numbers having a real and an imaginary part is, of course, quite different from the use we will make of "complex" to describe properties of whole systems in what follows.)

The magnitude of the eigenvalues determines the stability of the corresponding motion. If an eigenvalue (or pair of complex eigenvalues) has a magnitude smaller than 1 the corresponding component of the motion will be stable, moving toward the origin. If an eigenvalue (or pair of complex eigenvalues) has a magnitude greater than 1, the corresponding component of the motion will be unstable, moving away from the origin indefinitely. Eigenvalues with magnitude just equal to 1 are neutral, neither stable nor unstable. The corresponding component of the motion is neutral as well: if the eigenvalue is $+1$, the system will remain indefinitely at any point on the line corresponding to its eigenvector; if the eigenvalue is -1, the system

will indefinitely oscillate in reflections around the origin; if a pair of complex eigenvalues has magnitude 1 (thus lying on the unit circle in complex coordinate space), the system, started at any point in the plane defined by the corresponding eigenvectors, will rotate indefinitely on a circle around the origin through the starting point.

The actual motion of a linear system starting from any initial point $\mathbf{x}_0$ can appear quite complex, but can always be decomposed into a combination of these simple independent motions.

1.1.1.1 Example: The linear oscillator

Consider the linear system defined by the equations of motion:

$$\begin{aligned} x_{t+1} &= \lambda \cos \omega x_t - \lambda \sin \omega y_t \\ y_{t+1} &= \lambda \sin \omega x_t + \lambda \cos \omega y_t \end{aligned} \tag{1.3}$$

Here $\lambda > 0$ and $\omega \neq 0$ are parameters that govern the equations of motion of the system. The matrix A in this case is

$$\begin{bmatrix} \lambda \cos \omega & \lambda \sin \omega \\ \lambda \sin \omega & \lambda \cos \omega \end{bmatrix}$$

The eigenvalues of this matrix are $\lambda(\cos \omega \pm i \sin \omega)$, which are stable if $\lambda < 1$, and unstable if $\lambda > 1$. The trajectories of this system from any arbitrary starting point except the origin are inward spirals if it is stable, outward spirals if it is unstable, and circles in the borderline case $\lambda = 1$. The sign of the parameter ω determines the direction of spiraling, and the magnitude of ω its speed.

1.1.1.2 Linearity and predictability

The possibility of decomposing the trajectories of linear dynamical systems into simple independent motions suggests that it might be possible to predict the motion of linear systems by observing their past trajectories. A clever predictor might use the past trajectories of this system to infer the values of the matrix A, and then could use standard mathematical techniques to extrapolate the future trajectory of the system. Sophisticated and robust techniques for estimating the structure of linear systems, even when the observations of the trajectories are contaminated by errors of observation, have been developed by econometricians, epidemiologists, physicists, and engineers who deal with time series.

Linear systems have the convenient property that their behavior in all regions of the state space is proportional to their behavior in a small neighborhood of the origin. Thus observation of a linear system in a small part of the state space is potentially sufficient to understand its behavior everywhere. This behavior contrasts sharply with those of systems that evolve in different ways at different scales.

Linear systems also have the convenient property that smooth changes in the parameters **a**, which in the case of linearity define the elements in the matrix A, lead to smooth changes in the behavior of the trajectories. The speed of expansion or contraction of a solution along a ray, and the location of the ray itself, may vary with the elements of A, but in a smooth and gradual way. Two real eigenvalues may coalesce into a pair of complex eigenvalues, their eigenvectors becoming the real and imaginary components of the complex eigenvectors, but this takes place so smoothly that the corresponding trajectories are close to each other for small changes in the parameters.

Linear systems may have many dimensions, and their trajectories may seem to be quite complicated, but in fact they are simple, because with enough information, even local information in one region of the state space, about the past trajectory of the system, it is possible to infer the future trajectory to any desired degree of accuracy.

1.1.1.3 Example: The linear oscillator continued

The linear oscillator we have studied exhibits the smooth dependence of linear systems on changes in parameters. It is convenient to rewrite this system in polar coordinates, (r, θ), where $r(x, y) = \sqrt{x^2 + y^2}$, and $\theta = \arctan(\frac{y}{x})$. In these coordinates the laws of motion become:

$$\begin{aligned} r_{t+1} &= \lambda r_t \\ \theta_{t+1} &= \theta_t + \omega \end{aligned} \tag{1.4}$$

In polar coordinates it is clear that the expansion or contraction of the system depends on the parameter λ, while its speed of rotation depends on the parameter ω. If we hold ω constant and increase λ from a value smaller than 1 through 1 to a value larger than 1, the trajectories change smoothly from inward spirals to circles to outward spirals as in Fig. 1.1. If we hold λ constant and vary ω, say from a negative value through 0 to a positive value, we see the spiraling on the trajectory slow down, stop, and then reverse direction, as in Fig. 1.2. When $\omega = 0$, the eigenvalues of the system are no longer a complex conjugate pair, but two real numbers, both equal to λ. In this situation the system decomposes into pure motions in

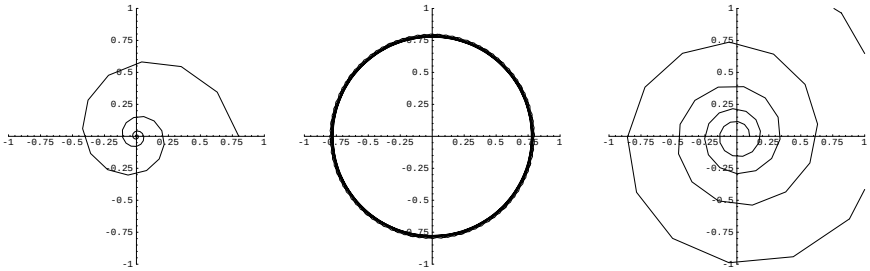

Figure 1.1: As λ moves from stable to unstable values, the trajectories of the linear oscillator change smoothly from inward spirals through a circle to outward spirals.

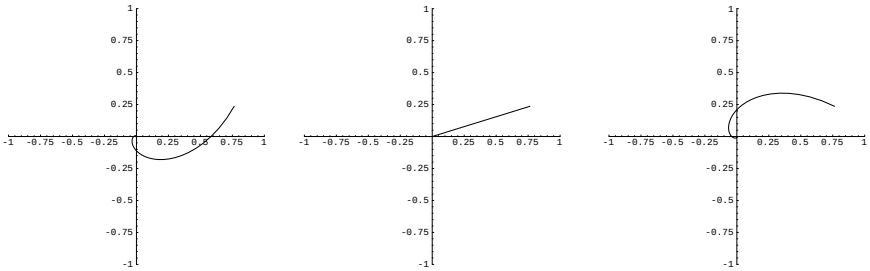

Figure 1.2: As ω moves from positive to negative values, the spiral trajectories of the linear oscillator slow down, and reverse direction. When $\omega = 0$, the system moves along a straight line toward or away from the origin.

the x- and y-directions, which are both unstable or stable, depending on the value of λ.

1.1.2 Nonlinear dynamical systems

The situation where the operator $F_{\mathbf{a}}(.)$ is not linear is quite different. Suppose, for convenience, that the origin of the state space is an equilibrium of the system, in the sense that the laws of motion require that if the system starts at the origin it will stay there:

$$0 = F_{\mathbf{a}}(0)$$

For points close to the origin the laws of motion can be approximated by expanding the operator $F_{\mathbf{a}}(.)$ as a Taylor's series:

$$\begin{aligned} x_{t+1} &= F_{\mathbf{ax}}(0)\mathbf{x}_t + \frac{1}{2}\mathbf{x}_t^T F_{\mathbf{axx}}(0)\mathbf{x}_t + \ldots + \\ &= (F_{\mathbf{ax}}(0) + \frac{1}{2}\mathbf{x}_t^T F_{\mathbf{axx}}(0))\mathbf{x}_t + \ldots + \end{aligned} \tag{1.5}$$

Here $F_{\mathbf{ax}}(0)$ represents the vector of first derivatives of $F_{\mathbf{a}}$ evaluated at the origin, and similarly for $F_{\mathbf{axx}}(0)$. This approximation suggests that we regard nonlinear systems as perturbations of linear systems, where the perturbing effects appear in the quadratic, cubic, and higher order terms of the power series. The second form emphasizes one important aspect of the nonlinearity of the system; the laws of motion of the system change as the system moves in the state space. In a nonlinear system it is no longer generally possible to make inferences about the global behavior of the system from local information.

1.1.2.1 Bifurcations and chaos

Nonlinear dynamical systems display a much greater range of behavior than linear systems. For example, if we gradually vary the parameters in a linear system with a pair of complex eigenvalues so that the eigenvalues move from being stable to being unstable, the corresponding motion changes from a stable spiral toward the origin, through neutral circling of the origin, to an unstable spiral away from the origin. In particular, the motion can remain bounded and cyclical only in the knife-edge case where the eigenvalues have magnitude exactly equal to 1.

In a nonlinear dynamical system matters are quite different. As a pair of complex eigenvalues moves across the unit circle (the *Hopf bifurcation*) the system becomes locally unstable, since the linear forces tend to push it in ever widening spirals. But as the system moves away from the origin, the nonlinear forces become more important. It is possible that the nonlinear forces will be stabilizing, so that the system will develop a stable limit cycle, generating bounded oscillations indefinitely for a range of parameter values and a range of initial conditions. This behavior is qualitatively different from the possible behavior of linear systems, which can produce indefinite bounded oscillations only in the fragile case where the parameters lead to a pair of neutral complex eigenvalues; even in this special case the cycle that results changes with each change of initial conditions.

The interplay of destabilizing linear forces and stabilizing nonlinear forces in a nonlinear system can produce even more complicated trajectories. The nonlinearities may couple the independent motions of the linear system. In this situation the system may be unable to achieve a limit cycle, because its motion disturbs a third variable, which in turn changes the

local laws of motion. The system may wander indefinitely in a confined part of the state space, so that it is globally stable, never exactly repeating its previous trajectory. The resulting patterns of motion are called *chaotic*, a terminology that does not do complete justice to the combination of organization and unpredictability that results.

Chaotic systems are organized in that their asymptotic trajectories (or attractors) occupy only a part of the available state space, so that knowledge of the underlying dynamics of the system allows an observer to predict that they cannot assume certain configurations. (This type of knowledge would be of great practical value, for example, in speculating on the motions of financial asset prices. Even though an observer might not be able to predict the exact trajectory of prices very well, knowledge that certain related movements are impossible would allow for profitable informed speculation.) Furthermore, chaotic systems are locally predictable, because the laws of motion change only gradually over the state space. If the system returns to a position in state space close to an earlier trajectory, its trajectory will for a while follow the previous trajectory closely. Furthermore, chaotic systems produce statistically regular outcomes; every trajectory tends to spend the same proportion of time in different portions of the state space. Knowledge of these statistical properties is also of great practical value, since certain risks can be excluded from consideration.

1.1.2.2 Example: The perturbed linear oscillator

We can introduce a stabilizing nonlinear force into the linear oscillator easily by adding a quadratic term that tends to push the system back toward the origin when it gets far away from it. This is particularly transparent in polar coordinates:

$$\begin{aligned} r_{t+1} &= \lambda r_t - \lambda r_t^2 = \lambda r_t(1 - r_t) \\ \theta_{t+1} &= \theta_t + \omega \end{aligned} \tag{1.6}$$

This modified oscillator continues to behave like the linear oscillator when the linear part is stable, that is, when $\lambda < 1$, as Fig. 1.3 shows. The system continues to spiral inward to the origin. The reason for this is that in this case both the linear forces and the nonlinear forces are stabilizing: as the system approaches the origin the nonlinear forces, which are proportional to the square of the distance of the system from the origin, become negligible in comparison to the linear forces.

When we destabilize the linear system by increasing λ above 1, however, we see a dramatic difference between the behaviors of the linear and nonlinear systems. The linear system with $\lambda > 1$ always spirals indefinitely

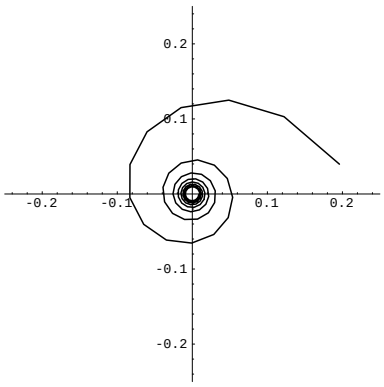

Figure 1.3: The perturbed linear oscillator behaves like its linear counterpart when $\lambda < 1$, spiraling inward toward the origin from any starting point.

outward from the origin, but the nonlinear system establishes a limited, stable oscillation at a particular distance from the origin that depends on λ, when λ is not too much larger than 1, as Fig. 1.4 shows. In this situation the nonlinear stabilizing forces are in conflict with the linear destabilizing forces on the system. When the system is close to the origin, the nonlinear forces are very weak compared with the linear forces, so the nonlinear system behaves like the linear system and spirals outward. But as the system moves farther from the origin, the stabilizing nonlinear forces become stronger, and eventually balance the destabilizing linear force, creating the observed limit cycle through a Hopf bifurcation.

If we continue to increase λ, the stable limit cycle itself begins to break up into more and more cycles as in Fig. 1.5, until finally the nonlinear oscillator reaches the chaotic state illustrated in Fig. 1.6. Trajectories of this system that start at very nearby points will diverge exponentially over time: small errors of measurement of initial conditions will lead to larger and larger errors of prediction.

Figure 1.7 illustrates a further nonlinearization of the oscillator. Whenever this system enters the small window of states, the value of λ is reduced to .5, the value of ω becomes 0, and the system moves on a straight line to the origin. We might think of the window as representing the solution to some difficult problem: the system casts around almost randomly seeking it, but once it has found it, finds its way to the ultimate equilibrium. The trajectories of this system diverge initially, but sooner or later all of them will converge on the origin. As we will see, this type of regime shift, in which the diversity of behavior of the system first expands and then contracts, is characteristic of highly complex systems.

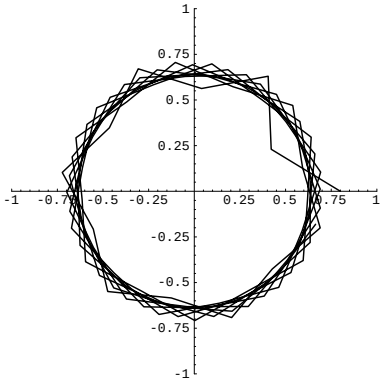

Figure 1.4: When λ is increased moderately above 1, the perturbed, nonlinear oscillator establishes a stable limit cycle at a distance from the origin depending on λ.

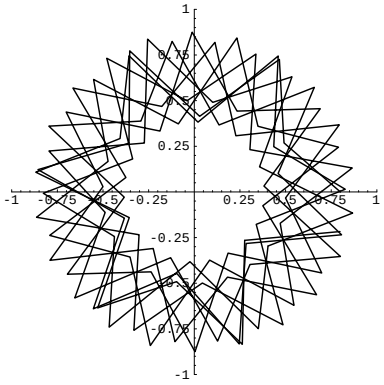

Figure 1.5: As λ increases above 1, the limit cycles of the perturbed oscillator break up into multiple cycles.

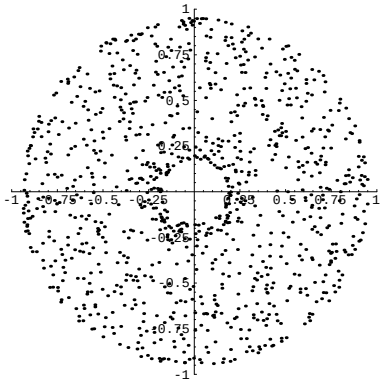

Figure 1.6: As λ approaches 4, the trajectories of the perturbed oscillator become chaotic. Here only the points of the trajectory are plotted.

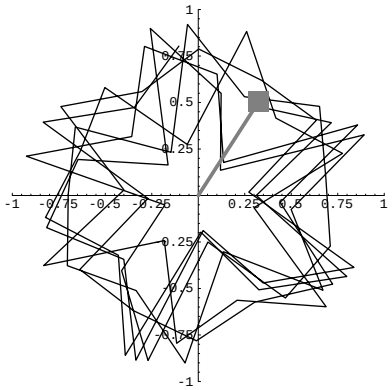

Figure 1.7: When the system enters the window (the gray square), its dynamics change and it moves on a straight line to the origin.

1.1.2.3 Dynamical systems from an informational viewpoint

A fruitful way to think of a stable dynamical system is as a signal processor. The input to the system is its initial condition, and its output is the attractor set representing the asymptotic trajectories. A dynamical system, for example, that maps each initial condition onto a finite number of point attractors can be seen as a pattern recognition device: the initial condition represents the data the system confronts, and the particular equilibrium to which it tends from the initial condition can be interpreted as the underlying pattern it associates with these data.

Chaotic systems, on the other hand, are unpredictable. It is relatively easy to predict the trajectories of a stable dynamical system with a simple attractor such as a point or a stable limit cycle. Wherever the system starts, its trajectory will approach the stable attractor, which itself has a relatively simple structure. Furthermore, trajectories that start from nearby initial conditions will actually become closer together as the system evolves. Thus small errors in measuring the initial condition of the system will become negligible factors in disturbing the accuracy of the predicted trajectory. Chaotic systems also tend to attractors, but the attractor is a geometrically complicated structure. One trajectory, starting from one set of initial conditions, may be very different in detail from the trajectory starting from a very slightly different set of initial conditions. In fact, in chaotic systems trajectories starting from slightly different initial conditions diverge progressively as the system evolves. Thus small errors in measuring initial conditions of a chaotic system are magnified over time, and lead to increasingly inaccurate predictions.

Chaotic systems are familiar aspects of human life, the weather being a paradigmatic case. Modern research suggests that chaos is the rule in real nonlinear dynamical systems, and that systems with simple attractors like points and limit cycles are relatively rare results of highly controlled interactions. Furthermore, the widespread existence of chaos suggests that many systems often viewed as being purely stochastic—that is, varying without any discernible order at all—may have the type of structure associated with chaotic dynamics.

1.1.2.4 Complexity and self-organization

Human experience also embraces the phenomena we associate with living organisms, which are, in principle, dynamical systems, moving on a trajectory through some large state space. Living organisms display behavior that is qualitatively different from chaotic systems like geological or meteorological interactions. First of all, living organisms are highly structured, and have powerful homeostatic mechanisms that stabilize important

aspects of their behavior. On the other hand, some living organisms, certainly human beings, produce state space trajectories that appear not to be even partially statistically predictable like chaotic systems. Human beings, for example, seek solutions to problems, which may involve an exploration of the relevant state space which does not have the repetitive features of chaotic motions. In solving a problem, a human being may pursue one approach, exhibiting one type of behavior, for a while, and then determine that this approach is a dead end, and suddenly (from the point of view of an external observer) shift to a qualitatively different type of behavior embodying an alternative approach to the problem. Furthermore, if the human is lucky enough to solve the problem, the problem-solving trajectory may come to an end. The terminal state is a kind of stable attractor for the problem-solving system, but is clearly not very well modeled by stable linear or nonlinear systems with simple attractors, because the trajectories leading to it may diverge before they eventually converge on the solution.

Thus there exist at least some nonlinear dynamical systems that are self-organizing and exhibit dynamic behavior qualitatively different both from simple stability and chaos. Although it is not easy to characterize these systems in general mathematical terms, they are called *complex, adaptive systems*. Their basis has to be nonlinear dynamics, since the motions of linear dynamics are not rich enough to support complex behavior. In complex systems the nonlinear dynamics can sustain self-organized and self-reproducing structures, and these structures in turn are capable of interactions that mimic the behavior of living organisms.

Complex adaptive systems are of great interest to human beings, since they include ourselves and other life forms, and in particular our processes of conscious thought. This book addresses the question of the degree to which the social interactions of human beings, especially their economic relationships, also constitute complex adaptive processes.

Human beings are very intelligent compared to systems with simple point attractors, though even these systems sometimes pose subtle problems to human understanding. We are intimately familiar with the behavior of massive objects falling to rest in the field of gravity, for example, though it is not so easy for us to throw a spear with force and accuracy at a distant target. Oscillatory systems pose a more complicated, but still conceptually soluble problem to human intelligence. The oscillatory system falls into a repetitive pattern that we can with long observation and attention recognize and learn to predict. But this process is not easy, either: many generations of human observation and thought were required to unravel the oscillatory motion of the planets, for example.

Chaotic systems pose an even higher barrier to human understanding, because their regularities are subtler, and require more observation and more careful analysis to discover. Our learning and knowledge of chaotic

systems take a structural, statistical form, in contrast to the precise predictions we can make about point-stable and oscillatory systems. We often regard chaotic systems as simply stochastic, through missing the subtle structures that determines their attractors. The deepest understanding we can reach of a chaotic system is a knowledge of its underlying laws of motion, even though this knowledge does not practically allow us to predict the evolution of the chaotic system in detail.

Complex adaptive systems pose novel problems for theories of human knowledge and learning that go beyond the realm of the chaotic. The problem is that these systems are inherently at the same (or perhaps a higher) level of complexity as our own consciousness, which provides the tools with which we learn and know the world. There does not seem to be any economical way to predict the problem-solving efforts of another human being, for example. We can try to solve the problem ourselves, thus reenacting the dynamical process which we are studying, but this effort may not (in many cases will probably not) replicate the problem solving of another person very closely. Even if we can solve the problem, that does not allow us to say very much about whether another person will solve it, or arrive at the same solution.

As we will see, the choice to regard human and social interactions as the aspects of complex systems poses deep questions for received models of human behavior, especially for the models of maximizing rationality that are the foundation of much existing social and economic theory.

1.1.3 Cellular automata as models of nonlinear dynamical systems

There is a bewildering variety of nonlinear dynamical systems, and it is unlikely that any very general laws apply to all of them. Our current understanding of these systems is largely the result of two research strategies. The first looks at the consequences of the simplest and most manageable nonlinearities and attempts to understand the resulting systems through rigorous mathematical analysis. The second seeks to identify very particular systems that are capable of a wide range of complex behavior and to study these examples through computer simulation to get clues as to the possible range of behavior of nonlinear systems in general.

The second strategy underlies much of the work reported in this book. One conceptually simple, but exceedingly rich model of nonlinear systems, the *cellular automaton*, serves as a laboratory for the exploration of the difficult general questions raised by nonlinearity and complexity. The cellular automaton is an array of cells with a particular geometry (arranged on a line segment, a circle, a segment of the 2-dimensional plane, or a torus,

for example). In each period of a simulation each cell is characterized by a certain *state*, chosen from a finite set of possibilities. This concept is motivated in part by quantum mechanical ideas, since in quantum theory the states of physical systems are confined to a finite or countable range of possibilities, and in part by certain common nonquantum physical systems, such as oscillators, that can be excited only at a certain range of frequencies, and saturating systems that move to a finite range of stable limiting configurations. The restriction of the states of the cells to a finite set is already a reflection of the underlying nonlinearity of the system being modeled.

The cellular automaton advances through time with the cells synchronized to a common periodic clock external to the system. The state of each cell in the next period depends on the current state of the cells in the system as a whole. In most of the cellular automata that have been studied in detail, the evolution of any single cell depends only on the past states of the cells in a neighborhood containing it. Because of the physical motivation for much of the research in cellular automata, where the system being modeled consists of a large number of identical particles or subsystems, it is frequently assumed that all the cells obey exactly the same rule of evolution.

The simplest cellular automaton, for example, might consist of n cells arranged on a circle, each of which can be in one of two states (*on* or *off*). The neighborhood of each cell might consist of the k (perhaps 1) cells on either side of it. There are eight possible configurations of the three cells in a neighborhood, so the law of evolution of a two-state, radius-1 neighborhood circular cellular automaton can be described by a list of eight bits (0 or 1), interpreted as the next state of the cell when the current configuration of neighboring cells is described by each of the eight possibilities (000, 001, 010,...,111).

Figure 1.8 illustrates the evolution of a one-dimensional two-state cellular automaton in a format frequently used in this book. The cells are displayed horizontally, with the state of each cell in one period indicated by its shading (in a two-state cellular automaton typically white and black). The state of the system after a clock tick is plotted directly beneath the original state, so that the time evolution of the state of any cell can be traced vertically.

Cellular automata are obviously very simplified models, but they are capable of a remarkably wide and interesting range of behaviors. They can mimic, for example, all four of the types of nonlinear dynamic behavior we have discussed: stability toward a single state, stability toward a pattern of regular oscillation, nonrepetitive chaotic motions (within the limits imposed by the large, but finite number of states a finite cellular automaton can occupy), and complex evolutions with sharply changing regimes and

Figure 1.8: The evolution of a one-dimensional cellular automaton.

unpredictable stopping points.

Cellular automata have great research advantages, since they are easy and cheap to emulate on a computer, and exhibit such a wide range of possible nonlinear behavior. Certain cellular automata can emulate general-purpose computers (Turing machines), and are thus capable of producing output at the highest complexity level of formal language theory. They also have limitations as models of human social and economic interaction, which will be explored in some depth in this introduction and in the later chapters of the book. As a result of these limitations the logic of this research is largely that of the counterexample: if certain hypotheses about human behavior can be shown to be impossible or implausible in the environment of cellular automata, they will a fortiori be impossible or implausible in more general (and perhaps even more complex) nonlinear environments. But a demonstration that a certain strategy or behavior is plausible or functional in a particular cellular automaton may not generalize to other automata or to other nonlinear systems.

1.2 Dynamical systems in social and physical science

Social scientists have been tempted to apply the powerful methods of dynamical system analysis to social systems. In carrying out this project, however, one immediately runs into the fact that social systems comprise the interactions of conscious human beings, unlike physical systems in which the behavior of the elements arises purely from their physical properties. The later chapters of this book raise these issues sharply and concretely in a variety of research contexts, but it is worth reviewing them briefly here.

1.2.1 Local and global interaction

In most physical systems the forces carrying the interactions of particles or molecules decay rapidly with distance. It is natural in this circumstance to model the system as evolving through local interactions. Cellular automata reflect this presupposition by making the next state of any cell depend only on the current states of cells in it neighborhood.

Many models of human social interaction, on the other hand, assume that the agents in the society have at least some kind of global information that influences their behavior. In one standard economic model of markets, for example, every agent in the market is supposed to determine her net demand for the commodity in the next period on the basis of a market price established in the current period. But this market price in turn depends on the net demands of all the agents in the market, not just a limited subset that might be regarded as the informational neighbors of the agent.

The essay "Qualitative Effects of Monetary Policy in 'Rich' Dynamic Systems," chapter 4 in this volume, offers another instance of the same issue. Policy signals such as the short-term interest rates largely controlled by central bank intervention in markets are common information to all the firms in the economy. In deciding its investment policy the firm takes into account this global signal in addition to local information consisting of the current investment decisions of its neighbors. In addition to the local interactions modeled by classic cellular automata, there are at least weak global interactions mediated by markets and policy makers that have to be taken into account in economic models.

Systemwide interactions in cellular automata raise novel issues concerning the dynamics and evolution of complex systems. Economists believe, for example, that the relatively limited information contained in market prices is sufficient to stabilize and organize potentially chaotic market transactions. Classical studies of cellular automata give us a good sense of the complexity and instability inherent in even relatively simple patterns of local interaction in nonlinear dynamical systems, but we have very little feel for the impact of weak global interactions in these contexts.

1.2.2 Topology and geometry in physical and social models

Physical systems consist of particles located in a highly structured geometrical space (typically Euclidean space). In this situation there is no ambiguity about the distance between particles or the identity of the particles closest to a given particle at a given moment of time. This geometry carries over into cellular automata, though it is simplified in important

ways. Rather than considering a full three-dimensional space, it is computationally much easier to study the evolution of cellular automata in one- or two-dimensional spaces. The geometry of the spaces is further simplified by considering only a finite or countable set of lattice points as sites for cells. Rather than coping with the complexity of smoothly decaying interactions (like gravity or electrical fields), the cellular automata are assumed to respond only to the states of immediate neighbors within an arbitrarily chosen radius.

Notions of closeness have played a powerful and fertile role in social and economic theories as well, but are inherently much more ambiguous and rich. The simplest cases are those where the economic agents are deployed in a real physical space, for example, locational models in economics. The notion of closeness and neighborhood relevant to a spatial economic model is almost the same as in a physical model, since, for example, transportation costs are often proportional to geographical distance between agents.

But even this very favorable example of transportation costs as determining economic distance signals additional levels of mathematical difficulty. Transportation costs depend not only on geographical distance, but also on topographical features (mountains, rivers, and the like) and on transportation facilities (railroads, highways, canals, and so forth). In a city built on a grid plan the economically relevant distance between two points is not ordinary Euclidean distance, because travel on diagonals across the grid plan is not feasible. Furthermore, the relevant economic distance measured by the time or cost of movement of people or goods between two points may vary depending on traffic congestion and street capacity. The economic distance between a customer and supplier is really the cost of moving goods from one to another, which may also vary according to the good being shipped. Bulk sea freight may bring cities half the world away from each other economically closer for certain commodities than these cities are to their geographically much closer suburbs.

But spatial models are only one context in which the concept of social or economic distance plays a key role. In models of the business cycle demand is often assumed to diffuse through the economy from customer to supplier. The injection of a given amount of government spending in one sector of the economy creates a cascade of further spending as the immediate producers hire workers and buy supplies to meet the demand, creating demands for other sectors. In this context the relevant economic neighbors of an agent are determined not geographically but functionally in terms of their technological positions in the hierarchy of production. This example also indicates mathematical difficulties. To say that firm A is a neighbor of firm B because orders to firm A will create demand for firm B does not necessarily imply that firm B is reciprocally a neighbor of firm A. Demand for B may have no, or very weak, effects on the demand

seen by A. In Euclidean space the distance metric is symmetric, so that the distance from point A to B is the same as the distance from point B to A, but this may not be true for many relevant conceptions of economic distance. Economic distance is also evidently not necessarily additive: the distance from A to C is not necessarily the sum of the distances from A to B and from B to C, because A and C may have economic relations that are independent of B altogether.

The same considerations apply to models of the diffusion of technical change through economies. Technical change in one sector of the economy (say, the introduction of a new metallurgical process yielding a higher strength alloy) tend to ripple through the customers of that sector (the building industry can build higher buildings with fewer structural components). Bottlenecks in one sector are important incentives for technological changes in supplier sectors (power looms require stronger thread to reduce the costs of thread breakage). Cellular automata are natural simplified models of diffusion, but in the social context the problematic character of the geometry of the space in which the cells are located has always to be kept in mind.

1.2.3 Time and causality

Physical theories have progressed in part because of their decision to regard time and causality as flowing in only one direction. The past is allowed to influence the future but not vice versa. This presupposition is reflected in the cellular automata developed to model physical systems in the assumption that the state of the system in the next period depends only on its state in the current period. Human beings, on the other hand, act with an eye to the future consequences of their actions. Thus the future, at least through actors' conceptions of the future, plays a different role in social and physical interactions.

This difference in point of view in the physical and social sciences lies at the heart of some of the most vexing and controversial problems about modeling human action. One resolution is to distinguish between the actual future and agents' conceptions or expectations of the future. If we take agents' expectations as part of their current state, then we can rigorously regard their actions as determined only by the present and past, not by the future itself, even though their conceptions of the future are acknowledged as part of their motivation and calculation. This point of view has its own set of conceptual problems and tangles. Expectations, for example, are not directly observable, and as a result must be inferred from agents' actions, which the expectations are supposed to explain. It is tempting for economists to suppose that agents can use the same processes of rational calculation to form expectations that they are assumed to use in

allocating budgets and other resources, leading to the economic theory of *rational expectations*. One of the central questions addressed in this book is the plausibility of the assumption of fully rational expectations in environments evolving according to complex nonlinear dynamics. For example, is it reasonable to assume that agents can costlessly solve mathematical problems of any degree of complexity, or make measurements of the current state of their environment to any degree of accuracy? As we shall see, these issues take on a sharp clarity and poignancy in the models considered here.

1.2.4 Identity and diversity

Physical systems often consist of one or a small number of types of interacting particles. Many interesting phenomena can conveniently be analyzed in systems with only one type of particle, since the interactions of interest (exchange of momenta, resonance, and the like) can take place among identical particles. As a consequence the widely studied cellular automata models of complex system interaction assume that all cells are subject to identical laws of motion.

The situation in models of society and the economy is different, because diversity of interest is often central to social interactions. Take for example the simple standard textbook model of a market. A number of agents (which might be thought of as cells of a cellular automaton) respond to a publicly announced market price with a particular net (or, in economic terms, excess) demand for the commodity being exchanged in the market. A positive excess demand indicates that the agent wants to buy that amount of good at the announced price, while a negative excess demand indicates that the agent wants to sell. The problem to be solved by the market is to find a price at which the sum of the excess demands is zero, so that the market clears.

We could model this system as a cellular automaton in which the state of each cell represents the excess demand for the commodity. (It is traditional in economics to regard excess demands as continuous variables, but it would not do much violence to the conception to restrict the levels of excess demand to a finite, but perhaps large, set of possible values.) As we have already noted, the dependence of each agent's excess demand on an announced market price that reflects, say, the total excess demand in the previous period departs from the usual cellular automaton assumption that each cell's state depends only on the state of the neighboring cells in the previous period. But the traditional image of market clearing involves the division of the agents into buyers and sellers at the announced market price. If all the agents are identical, as in the cellular automaton model, they will all assume exactly the same state once a market price is announced. It

may still be possible to clear the market by finding a price at which every agent's excess demand is zero, but this is a somewhat restricted image of the activity of real-world markets, since if every agent's excess demand is zero there would be no actual transactions in the market.

Chapter 5 in this volume, "Decentralized, dispersed exchange without an auctioneer," deals with these modeling problems in some detail. In this essay the agents are identical in tastes (though not in endowments, and thus not in excess demand functions), but the assumption of a single market price announced to all of them is relaxed. Instead we assume that the agents trade with their neighbors, so that effectively each agent sees only a small part of the whole market. Because the agents are arrayed in a circle, the various submarkets corresponding to different neighborhoods are indirectly linked. Simulations show how such an indirect linkage can lead to some of the same results as the textbook model of a single market price, though not all.

The issue of diversity does not arise in all economic models. In fact, contemporary macroeconomic theory tends, for the sake of computability and conceptual simplicity, to make the assumption of a *representative agent*, that is, that the economic behavior arising from the interactions of many agents in the economy can be modeled as if it were the behavior of a single agent. One way this might happen is if all the agents were in fact identical (and there were no externalities, that is, nonmarket-mediated interactions between them). Here the cellular automaton assumption that all the cells are governed by the same law of motion would be consistent with the traditional economic model. These considerations, however, reveal the fact that the traditional economic models have built-in assumptions that simplify their trajectories. Take the textbook market model with identical agents as an example. Suppose that the agents start in a random configuration of excess demands. In the next period a price will be announced, and all of them will pass into the same state, responding to this market price. Then the market price will be adjusted to eliminate excess demand or supply, and all of the agents will move together from one state to another until they reach the equilibrium. This type of behavior clearly has low inherent complexity: there is no local interaction of the agents. In fact, the evolution of the system as a whole after the first period is simply an n-fold copy of the evolution of any individual agent.

To retain the possibilities of complex evolution in economic models we must either relax the economic assumption that all the agents respond to the same systemwide signals (such as market price) or relax the cellular automaton assumption that all the agents are identical.

1.3 Economic models of fully rational behavior

The essays in this volume critically examine some of the most fundamental assumptions in widely accepted economic models. For the sake of clarity in terminology, and as a service to readers who are not familiar with economic modeling practice, in this section we review the general form of economic models of rational behavior and consider the details of a representative sample of such models.

The central issues revealed by this survey concern the information-processing and computing capacities implicitly attributed to rational economic agents in various contexts. These issues become particularly acute, as we shall see, when the modeling context involves decisions over a large number of time periods and when the agent is uncertain about key aspects of the future environment.

1.3.1 The rational choice program

The cornerstone of received economic theory is the idea that human agents behave rationally. Rationality is supposed to underly the predictability of human behavior, and thus to establish it as a candidate for systematic scientific investigation.

The concept of rationality itself has no content without further elaboration and specification. For the economist rationality has come to mean behavior that can be viewed as maximizing some consistent mathematical function of behavioral and environmental variables. The rational agent can be seen as pursuing a definite and unambiguous goal, and her actions are predictable in this light. The most general form of the rational model, then, is that the agent's actions $\mathbf{x}$ maximize her objective function $u_{\mathbf{a}}(\mathbf{x})$, where $\mathbf{a}$ represents parameters describing the environment. Thus we predict that the agent's behavior can be described by a lawful relation $\mathbf{x}(\mathbf{a}) = \arg\max_{\mathbf{x}} u_{\mathbf{a}}(\mathbf{x})$. That is, the agent chooses the action $\mathbf{x}$ that maximizes her objective function $u_{\mathbf{a}}(\mathbf{x})$ given the environment $\mathbf{a}$. In particular, the economist will try to infer the objective function u from the agent's observed behavior, and then predict that her actions in the face of a change in the environment will follow the law $\mathbf{x}(\mathbf{a})$.

There are some well-known problems with this program of explanation and prediction. In the first place, as suggested earlier, the hypothesis of rationality puts no observational restrictions on an agent's actions. We can always rationalize behavior by positing an appropriate objective function. For example, one might think it irrational for an agent to bet simultaneously for and against a certain event at odds that guarantee her a loss. But

if her social position, for example, requires her participation in betting, she may in this way be regarded as rationally meeting social norms at the lowest cost. Second, if we allow the agent to change her mind over time, that is, for the objective function u to change from one time to another, no practical observations can disconfirm the hypothesis of rationality. What appears to be irrational or inconsistent behavior may simply reflect a change in the function being maximized between observations. Third, the program of rationality seems to run afoul of Occam's Razor, in that the utility function that is assumed to mediate between the environment and behavior is inherently unobservable, so that it is not clear what explanatory advantage its presence in the theory confers.

Furthermore, it is not obvious that rational behavior is more predictable than irrational behavior. If we take self-destructive mental illness, for example, as a paradigm of irrational behavior, it may be possible to establish highly confirmable and replicable patterns of behavior associated with particular diseases. On the other hand, the behavior of a presumably rational stock speculator, making complex calculations based on a constant flow of new and subtle information, may be very difficult to predict lawfully.

Economists have long been aware of these problems with the concept of rationality. Economic theory proceeds on a series of more or less tacit assumptions that address them. For example, when economists speak of objective functions, or utility functions, or preferences, there is an implicit assumption that they remain invariant over the relevant period of analysis. Despite the difficulty of actually observing utility functions, economic models typically assume that they are known fully to the economist, or take a particular functional form that makes them amenable to mathematical manipulation. Economic models implicitly assume that the legitimate arguments of the utility function are confined to a certain range of relevant factors, such as the physical consumption of the individual, or her money income, and in most (though not all) cases rule out social status, ritual perfection, aesthetic taste, and the like as arguments in the utility function.

Much of existing economic and social theory consists of the application of this rational choice paradigm to the explanation of human behavior and social interaction. This effort constitutes a rational choice research program, which seeks explanations for all social and economic phenomena through identifying the preferences of the agents involved and uncovering the rationality inherent in their actions.

The essays in this book, however, pursue another fundamental set of questions raised in the work of Herbert Simon (1978, 1984) which challenge the rational choice program. In assuming that the individual behaves so as to maximize a utility function in a given environment, the economist also assumes that the agent can costlessly process the information describing the

environment and costlessly compute the optimal policy for coping with it. In the past fifty years, however, the emergence of computers has, paradoxically, both drastically lowered the costs of computation and information processing and made us acutely aware of these costs and of the limitations of computability. The investigations reported in this book suggest that it is impossible to formulate the hypothesis of rationality consistently so as to take systematic account of the costs of computation.

It is tempting to think that the problem of computation and information processing costs could be resolved by moving the rational decision to a higher level, constructing a new rational choice problem in which computation is one strategy. This idea, in other words, tries to formulate a metaproblem, in which one component of the agent's behavior is the decision to undertake computation to improve her estimates of the consequences of various fundamental actions. This approach, however, cannot solve the problem, as Sidney Winter (1975) and John Conlisk (1988) have pointed out, because the resulting metachoice problem is just as difficult to solve computationally as the original problem. As we will see in more detail later, one characteristic of computationally complex problems is that there is in principle no way to predict the cost of solving them short of undertaking a computation that does in effect solve the problem. Computation costs might lead an agent to be content with an approximately maximizing, thus only approximately rational solution to her maximization problem: Herbert Simon has emphasized this type of behavior as *bounded rationality*. Computational complexity, on the other hand, raises issues not of costs but of logical feasibility, and thus appears to constitute a *barrier to rationality*.

1.3.2 Individual decision models—intertemporal optimization

The best way to understand how expectation formation in economic models leads inevitably to the consideration of nonlinear dynamic systems and the related problems of complexity is to look at some typical models in detail.

The simplest type of rational decision model considers a single agent acting in effective isolation. Traditional economic theory views this situation as a paradigmatic case of modeling and explanation. The traditional economics research program seeks to generalize the features of this type of model to more complex social interactions that involve several agents.

1.3.3 The finite-horizon Ramsey problem

Consider the situation of an agent with a lifetime of T periods, who has no interest in leaving a bequest, and who begins life with a fund of wealth

W_0, which can be invested at a constant real interest rate r, and faces the choice in each period of how much of her wealth to consume. Thus $W_1 = (1+r)(W_0 - C_0)$, and similarly for later periods, $W_{t+1} = (1+r)(W_t - C_t)$. For mathematical convenience, let us assume that the agent's objective function is the sum of the natural logarithms of her consumption over her lifetime, each period's utility being discounted at the constant factor $\beta < 1$. This objective function has a natural interpretation. In each period the agent experiences a utility proportional to $\log(C_t)$. The shape of the logarithm function reflects a declining marginal utility of consumption, that is, the agent puts lower values on additional units of consumption the higher her consumption already is in each period. Utility in the future counts less for the agent than utility in the present, and this temporal discounting is represented mathematically by multiplying the utility of consumption in period t by β^t, where, since $\beta < 1$, the discount factor declines geometrically with time.

This model can be summarized as a mathematical programming problem:

$$\begin{aligned} &\text{choose} && (C_0, C_1, \ldots, C_T) && \text{so as to} \\ &\text{maximize} && \log(C_0) + \beta \log(C_1) + \ldots + \beta^T \log(C_T) \\ &\text{subject to} && W_{t+1} = (1+r)(W_t - C_t) && \text{for} \quad t = 0, 1, \ldots, T \\ & && W_{T+1} \geq 0 \\ & && W_0 \quad \text{given} \end{aligned} \tag{1.7}$$

The constraint $W_{T+1} \geq 0$ is required to make the problem well determined: otherwise the agent could always increase her utility by consuming more and going deeper into debt. Since the logarithm function has the limit $-\infty$ as $C \to 0$, it is clear that the agent will be better off consuming some finite amount in each period than choosing not to consume at all. Furthermore, the agent has no reason to choose a policy with $W_{T+1} > 0$ under the assumption that she has a finite known lifetime. She will be better off consuming her whole wealth in period T, so that $W_{T+1} = 0$ on the optimally chosen path.

There are a variety of equivalent mathematical approaches to the solution of this problem. One approach effectively transforms the problem of finding the optimal consumption policy into the problem of finding the trajectory of a particular dynamical system, thus revealing the close links between the problem of rational optimization and dynamical systems.

To see how this link is established, let us use the *Lagrangian method* to solve the programming problem. We introduce a series of Lagrange

multipliers, which economists often call *shadow prices*, $\beta^t P_t, t = 0, 1, \ldots, T$, corresponding to the $T + 1$ constraints. The Lagrangian function is then:

$$\mathcal{L}(\{C_t, W_{t+1}, P_t\}_{t=0}^T) = \sum_{t=0}^{T} \beta^t (\log(C_t) - P_t(W_{t+1} - (1 + r)(W_t - C_t)))$$

The solution to the programming problem corresponds to a saddle point of the Lagrangian function: at the optimum choice it must not be possible to reduce the value of Lagrangian by changing the $\{P_t\}$, nor to increase it by changing $\{C_t, W_{t+1}\}$. The first condition ensures that the constraints are met, since if $W_{t+1} > (1 + r)(W_t - C_t)$, the coefficient of P_t would be positive, and it would be possible to reduce the value of the Lagrangian by increasing P_t. The second condition ensures that the objective function is maximized.

The first-order conditions for the problem, which ensure local optimality on the path by testing whether shifting consumption from one period to the next will increase the value of the Lagrangian are:

$$\beta^{-t} \frac{\partial \mathcal{L}}{\partial C_t} = \frac{1}{C_t} - (1 + r)P_t = 0$$

$$\beta^{-t} \frac{\partial \mathcal{L}}{\partial W_{t+1}} = -P_t + \beta(1 + r)P_{t+1} = 0$$

$$\beta^{-t} \frac{\partial \mathcal{L}}{\partial P_t} = -W_{t+1} + (1 + r)(W_t - C_t) = 0$$

The first equation tells us that $(1 + r)P_t$ is the marginal utility of consumption in period t, and allows us to eliminate $C_t = 1/((1 + r)P_t)$. Then we are left with a dynamical system in the variables (W, P):

$$\begin{aligned} P_t &= \beta(1 + r)P_{t+1} \\ W_{t+1} &= (1 + r)(W_t - \frac{1}{(1 + r)P_t}) \end{aligned}$$

The optimal policy must be one of the trajectories of this dynamical system. We have two boundary conditions, W_0 given, and $W_{T+1} = 0$, which choose out a particular trajectory which is optimal. To look at this in a slightly different way, we could consider an arbitrary choice for P_T, which would, from the first dynamic equation, determine the whole path of P_t. Plugging this solution into the second dynamical equation, we would determine the path of W_t. If we start with a very high P_T, the whole path of P_t will be high, consumption will be low, and $W_{T+1} > 0$, which

is not optimal. If we choose a very low P_T, consumption will be high and $W_{T+1} < 0$, which is infeasible. The optimal path corresponds to the choice of P_T which just makes $W_{T+1} = 0$.

This relatively simple problem reveals the basic structure of intertemporal optimization. Local optimality considerations establish a dynamical system, whose trajectories are the candidates for the rationally optimizing decision. Some of these trajectories are infeasible because they violate economically important constraints. There are typically many feasible trajectories, each of which leads to a different level of the objective function over the whole time path; the agent must then choose the one that yields the highest value for the objective function, by choosing a trajectory that respects the boundary conditions on the problem.

The problems of complexity and computation costs arise in rational intertemporal decision making when the dynamic systems that result become chaotic or complex. In these cases the rational agent faces the formidable problem of working out the full trajectory corresponding to a whole range of initial choices; only when she has accomplished this task can she determine which one is in fact optimal.

1.3.4 Market models

Many economic models represent the economy as a group of agents independently optimizing, with market prices acting so as to coordinate their potentially inconsistent decisions. The very influential Walrasian model of market equilibrium, for example, posits an auctioneer who cries out a system of market prices. Each agent then maximizes her utility subject to her budget constraint at these prices, and announces her excess demands to the auctioneer. In general the agents' plans will not be consistent: their excess demands will not sum to zero. In the Walrasian model, the auctioneer then adjusts the market price vector so as to try to eliminate the market excess demand. Equilibrium results when the auctioneer has succeeded in finding a vector of prices at which the sum of the agents' excess demands is zero.

Walrasian models in their full generality are quite difficult to solve, because the excess demands as functions of the announced market price vector can be arbitrary nonlinear functions. In order to get some insight into the structure of Walrasian equilibrium, economists often look at simplified economies in which it is easier to analyze the excess demand vectors mathematically. One popular simplification is to assume that all the agents in the economy are exactly alike in that they have the same preferences, technology, and endowment. This assumption might be viewed as a first approximation to an economy in which the agents differ only to a small degree. It has the virtue of greatly simplifying the analysis of equilibrium, but

the disadvantage that in the resulting equilibrium no trading takes place. Since all the agents are identical, they will always report the same excess demand vector in response to the same market prices, so that the only way to achieve equilibrium is to announce prices at which each individual agent's excess demand vector is zero.

One way to compute the equilibrium price vector in identical-agent models is to maximize a representative agent's utility over the technology and endowment constraints, using the technique of Lagrange multipliers. The resulting production and consumption plans will be the equilibrium production and consumption plans, and the vector of Lagrange multipliers is equal to the equilibrium relative price vector.

Because identical-agent models are simpler than the general Walrasian model, any degree of complexity that can arise in identical-agent equilibrium models must a fortiori be possible in the general Walrasian model. The converse conclusion, of course, does not hold: it may be possible to demonstrate simple structure in identical-agent models that does not hold in general. An example is the observation that equilibrium in identical-agent economies always implies no actual trading among the agents, a conclusion which does not hold in general.

1.3.4.1 The corn-steel model and the stable manifold

To get a concrete sense of the logic of intertemporal equilibrium in market models, consider an economy (see Burgstaller, 1994) with two goods, corn, K, and steel, S, each of which is the sole input into the production of the other. Thus to produce 1 unit of corn in the next period requires a_{SK} units of steel this period; to produce 1 unit of steel next period requires a_{KS} units of corn this period (perhaps to feed workers who produce the steel). The stocks of both corn and steel depreciate at the same constant rate δ each period. The typical agent of the economy consumes only corn in each period, C_t, but has no way to produce it except by producing steel first, so that she has a motive for producing both goods. The typical agent starts with an initial endowment of corn, K_0, and steel, S_0. She has to decide how much of the corn to consume and how much to devote to the production of steel for the next period. We assume as in the previous example that the agent's utility in each period is equal to the logarithm of the amount of corn produced, and that she discounts utility with a factor $\beta < 1$. Thus the typical agent's utility maximization problem can be written:

$$\text{choose} \quad \{C_t, S_{t+1}, K_{t+1}\}_{t=0}^{\infty} \geq 0$$

$$\text{so as to maximize} \quad \sum_{t=0}^{\infty} \beta^t \log(C_t)$$

$$\begin{aligned} \text{subject to} \quad & a_{SK}(K_{t+1} - (1-\delta)K_t) \leq S_t \qquad t = 0, 1, \ldots \\ & a_{KS}(S_{t+1} - (1-\delta)S_t) \leq (K_t - C_t) \qquad t = 0, 1, \ldots \\ & K_0, S_0 \quad \text{given} \end{aligned} \tag{1.8}$$

A convenient and economically insightful way to solve this problem is to use the Lagrangian method. Let $\beta^t P_{St}$ and $\beta^t P_{Kt}$, $t = 0, 1, \ldots$, be the Lagrange multipliers (shadow prices) associated with each of the two constraints in each period. Then the Lagrangian function is:

$$\begin{aligned} & \mathcal{L}(\{C_t, S_{t+1}, K_{t+1}, P_{St}, P_{Kt}\}_{t=0}^{\infty}) = \\ & \sum_{t=0}^{\infty} \beta^t \log(C_t) \\ & - \sum_{t=0}^{\infty} \beta^t P_{St}(a_{SK}(K_{t+1} - (1-\delta)K_t) - S_t) \\ & - \sum_{t=0}^{\infty} \beta^t P_{Kt}(a_{KS}(S_{t+1} - (1-\delta)S_t) - (K_t - C_t)) \end{aligned}$$

The optimal policy again corresponds to a saddle point of the Lagrangian, a sequence $\{C_t, S_{t+1}, K_{t+1}, P_{St}, P_{Kt}\}_{t=0}^{\infty}$ at which no marginal change in $\{C_t, S_{t+1}, K_{t+1}\}_{t=0}^{\infty}$ can raise the value of the Lagrangian, and no marginal change in $\{P_{St}, P_{Kt}\}_{t=0}^{\infty}$ can lower the value of the Lagrangian. The second condition assures us that the constraints must be satisfied, and the first that the utility function is at a maximum, since if it were not, it would be possible to increase the value of Lagrangian by adjusting the consumption and production plans.

These first-order conditions can be written:

$$\beta^{-t} \frac{\partial \mathcal{L}}{\partial C_t} = \frac{1}{C_t} - P_{Kt} = 0, \qquad t = 0, 1, \ldots$$

$$\beta^{-t} \frac{\partial \mathcal{L}}{\partial K_{t+1}} = \beta P_{Kt+1} + a_{SK}(1-\delta)\beta P_{St+1} - a_{SK} P_{St} = 0, \qquad t = 0, 1, \ldots$$

$$\beta^{-t} \frac{\partial \mathcal{L}}{\partial S_{t+1}} = \beta P_{St+1} + a_{KS}(1-\delta)\beta P_{Kt+1} - a_{KS} P_{Kt} = 0, \qquad t = 0, 1, \ldots$$

$$\beta^{-t} \frac{\partial \mathcal{L}}{\partial P_{St}} = a_{SK}(K_{t+1} - (1-\delta)K_t) - S_t = 0, \qquad t = 0, 1, \ldots$$

$$\beta^{-t} \frac{\partial \mathcal{L}}{\partial P_{Kt}} = a_{KS}(S_{t+1} - (1-\delta)S_t) - (K_t - C_t) = 0, \qquad t = 0, 1, \ldots$$

From the first set of equations, we see that P_{Kt} measures the marginal utility of consumption of corn in period t, which must always be a strictly positive magnitude, given the assumption of logarithmic utility. If we substitute $\frac{1}{P_{Kt}}$ for C_t in the last equation, we can eliminate C_t altogether, and we get a nonlinear dynamical system in the variables $\{P_K, P_S, K, S\}$:

$$P_{Kt} = \beta((1-\delta)P_{Kt+1} + \frac{1}{a_{KS}}P_{St+1})$$

$$P_{St} = \beta(\frac{1}{a_{SK}}P_{Kt+1} + (1-\delta)P_{St+1})$$

$$K_{t+1} = (1-\delta)K_t + \frac{1}{a_{SK}}S_t$$

$$S_{t+1} = (1-\delta)S_t + \frac{1}{a_{KS}}(K_t - \frac{1}{P_{Kt}})$$

$$K_0, S_0 \quad \text{given}$$

As in the simpler optimal consumption model with a single asset, the optimal consumption/production path for the agent must be one of the trajectories of this dynamical system. But the system has four variables, $\{P_K, P_S, K, S\}$, and the terms of the problem impose only two initial conditions, K_0 and S_0. Any arbitrary choice of P_{K0} and P_{S0} will lead to a trajectory that satisfies the local conditions for optimality. The agent, in order to solve the problem, must then evaluate her utility on each of these candidate trajectories in order to find the optimal one.

In this example the problem is somewhat simplified because the dynamical subsystem involving the shadow prices is independent of the quantities, and is also linear. Analysis of the eigenvalues and eigenvectors of this subsystem reveals that the eigenvector corresponding to the smaller of its two roots in magnitude is positive, while the eigenvector corresponding to the larger of its two roots in magnitude has prices of opposite sign. Any solution that activates the larger root will eventually lead to one of the prices becoming negative, which signals a trajectory that cannot be economically optimal. Thus the only trajectories that are candidates for the optimal policy are those that set the prices of steel and corn on the eigenvector corresponding to the smaller of the two roots, that is, those on the *stable manifold* of the dynamical system. This requirement adds one boundary condition to the problem.

This still leaves the question of the absolute magnitude of the shadow price of corn. As in the optimal consumption model, if we choose the initial shadow price of corn very low, consumption will be high, and the stock of corn will eventually become negative, which is economically infeasible. If

we choose the initial shadow price of corn very high, consumption will rise so slowly that the agent will accumulate stocks of corn and steel that will never provide her with increased consumption because of their enormous depreciation costs. The optimal policy corresponds to a choice of the initial level of P_{K0} that maintains feasibility and avoids overaccumulation of the stocks. Once again, the choice of the optimal policy requires the agent to look at the complete trajectories of a dynamical system in order to weed out infeasible or suboptimal policies.

Once we have the optimal solution, with its shadow prices, we also have the market equilibrium for this model. If the market prices are announced proportional to the shadow prices of the optimal policy, the utility maximizing agent will be led to the optimal policy, which will also clear the market since it is feasible. The important mathematical point to be gleaned from this example is that the establishment of a Walrasian equilibrium requires (or is equivalent to) the determination of the trajectories of a dynamical system. In the cases we have looked at, the dynamical systems that arise are fairly simple and exhibit only a small range of the possible spectrum of dynamical repertoire of nonlinear systems. In these cases it is possible to convince oneself that the discovery of the market equilibrium is computationally feasible for highly motivated and clever agents, who might use methods of trial-and-error extrapolation to determine the consistent current prices of assets.

1.3.4.2 The potential complexity of equilibrium paths in investment models

The preceding examples show the close connection between rational choice in an intertemporal framework and the behavior of dynamical systems. From a mathematical point of view the determination of an optimal intertemporal consumption/production policy is equivalent to the solution of a boundary value problem for a dynamical system. The solution of these problems is relatively easy in the examples we have studied, because the trajectories of the dynamical system are relatively simple.

But even highly simplified intertemporal choice models can give rise to complex trajectories. M. Boldrin and L. Montrucchio (1986), for example, have shown that an arbitrary dynamical system can be interpreted as arising from an intertemporal model of optimal consumption with several capital goods (or equivalently, the identical-agent market equilibrium for the same situation) without violating any of the common assumptions of economic models, such as diminishing returns. In particular this means that the trajectories rational agents have to compute may have any degree of complexity that dynamical systems can exhibit: they may be chaotic, have sensitive dependence on initial conditions, and so forth.

Jess Benhabib and Richard Day (1981), and a substantial literature their work has spawned, have shown that chaotic behavior can arise in a wide range of economically relevant intertemporal models. These mathematical facts raise serious questions for the methodological plausibility of the Walrasian equilibrium concept that underlies these models. If Walrasian equilibrium can be established only through the interaction of agents who must carry out computations we know to be impossible or impractical, what warrant do we have for regarding Walrasian equilibrium as a relevant model of real market interactions?

1.3.5 Game theory models

Competitive market equilibrium theory manages to approach the problem of analyzing the interactions of many rational agents through the assumption that the number of agents is very large, so large, in fact, that each agent can ignore the effects of her actions on any other particular agent. Under these circumstances each agent can be viewed as maximizing against a market price system on which her behavior has a negligible impact. The potential complexity of interagent interactions is finessed by the assumption that all these interactions are mediated by market prices. We have seen that taking computational costs into account calls into question the viability of long-accepted formal solutions to the problem of competitive interaction.

But there are many important spheres of social and economic life where the assumption of perfect competition is a violent distortion. In the first place, perfect competition itself is an ideal approximation to the complex behavior of economic systems with large but finite numbers of agents. In order to achieve a completely satisfactory theory of competitive equilibrium it would be necessary to demonstrate exactly how this approximation works, that is, how the behavior of a large system of interacting agents actually converges to the perfectly competitive limit as the number of agents increases without bound. In order to carry out this program, we need to have a theory of what actually happens when competition is less than perfect.

In fact, it is a common observation of life in market-directed societies that competition always breaks down at least locally. When one comes to make a transaction, no matter how competitive the market, one has to deal with one or a small number of other agents, and there is always some latitude in price, quality, or other dimensions of the transaction for bargaining. These noncompetitive spheres may be negligible from the point of view of the whole economy, or market, and the economic theorist may reasonably choose to abstract from them for the sake of a simple and tractable the-

ory. But a complete theory should at least embrace and account for these pervasive features of economic reality.

Finally, there are many extremely important economic interactions that simply do not involve a large number of competitive agents. Many important markets are oligopolies, in which a few very large sellers interact with an acute awareness of the impact of their behavior on each other. In many cases these interactions are of critical importance for economic development. A large investment in the construction of a transportation facility may be justified if complementary large directly productive investments will be forthcoming to generate a sufficient level of traffic. The investment in direct production, in turn, may be profitable only if the transportation facilities are available to bring the product to market. In this situation the productive investor and the transportation investor are inherently involved in a strategic interaction, in which each has a strong incentive to outguess the other, or, at least, to anticipate the other's behavior.

The formal analysis of situations of strategic interaction leads to the general theory of games. A game is formalized by specifying the number of players, their strategic options, the information conditions that link their actions, and the payoffs or utilities that characterize their evaluation of the outcomes of any combination of strategic choices. The research program of game theory hopes to find general characterizations of the behavior of rational agents interacting strategically. This program is a kind of generalization of the theory of rational agents interacting on competitive markets to the analysis of strategic interactions.

Here again the chapters in this volume show the critical importance of the issues of complexity and computational costs. As we will explain, even games of relatively simple structure rapidly give rise to computational problems of extremely high complexity, in which the assumptions of rationality are impossible to maintain. Once again, the logical structure of the argument is that of counterexample. If serious computational issues arise in relatively simple contexts, they must, a fortiori, be present in more complicated contexts.

1.3.5.1 Cournot-Nash equilibrium

There is a vast literature on various general solution concepts for abstract games (see, for example, Binmore and Dasgupta, 1986). The most popular starting point for most analysis of noncooperative games is John Nash's proposed equilibrium, generalizing an idea of Auguste Cournot. The Cournot-Nash equilibrium singles out strategy choices for the players of a game that have the property that no player can improve upon her outcome unilaterally, by changing her strategy, assuming that the strategies of the other players will remain unchanged.

This idea has considerable intuitive appeal, since at least in a repetition of the same game it is hard to see why players would be content to continue with strategies when alternatives offering a higher payoff are available. But it also has serious methodological difficulties. To begin with, in general the set of Cournot-Nash equilibria of a game is large, and in many cases includes outcomes that are uninteresting or implausible. Thus the Cournot-Nash equilibrium appears to be incomplete as a theory: we require further considerations to reduce the multiplicity of equilibria. But even when the problem of multiplicity of equilibrium has somehow been dealt with, there is some question as to why we should expect agents to choose Cournot-Nash equilibrium strategies. It is plausible that if an agent knew for sure that her fellow players would play their equilibrium strategies she will play hers, since it is by definition the best response she can make. But why should she have any confidence that her fellow players will behave in the way predicted by the equilibrium? The motivation for the Cournot-Nash concept of equilibrium lies in considerations of symmetry that have more potency as a mathematical postulate than credibility as a theory of human behavior. Once we acknowledge the possibility that an agent might doubt that her opponents will follow the predicted strategies, we must also admit the possibility that the opponents might doubt her commitment to it as well, and the whole structure of the equilibrium concept unravels.

But, as the chapters in this volume show, there are serious computational issues raised by the Cournot-Nash concept of equilibrium as well. Even if players have implicit faith that some version of the Cournot-Nash theory will correctly describe the outcome of their interactions, and even if the theory can be made to single out a unique outcome, how confident can we be that the agents can carry out the computations necessary to characterize the equilibrium?

1.3.5.2 The one-shot Prisoners' Dilemma

A great deal of the work relating game theory to larger social theory issues has centered on a particular game situation, the Prisoners' Dilemma.. This game poses particularly sharply the classic social dilemma of opportunism and the paradoxes of the pursuit of self-interest. In its simplest form the Prisoners' Dilemma involves two players, each of whom has two strategies, often called *cooperate* and *defect*. If both agents cooperate, they share a reward. But if one cooperates and the other defects, the defector gets a larger reward, while the cooperator suffers a penalty. If both defect, they share a smaller penalty. Thus regardless of what the other player does, it is in the direct individual interest of each player to defect: defection is a dominant strategy. The only Cournot-Nash equilibrium of the Prisoners' Dilemma is mutual defection. But this is disturbing because both players would

be better off if they avoided the temptation of defection and cooperated. In economic terms the "invisible hand" that supposedly guides individual competitors pursuing their own self-interest to a socially desirable outcome misguides them in the Prisoners' Dilemma.

The Prisoners' Dilemma can be seen as an abstract model of a huge range of human interactions. Hobbes's state of nature can be viewed as a Prisoners' Dilemma, and Hobbes's proposed solution in the form of a coercive external force a major rationalization of institutions of power. Thoughtful economists from Adam Smith on have recognized that the benefits of the invisible hand depend on the existence of a stable institutional structure (such as property rights) supported by political power which can prevent competition from destroying itself. Wherever the development of property rights lags behind the emergence of economically important interests, as in the area of environmental pollution, a social space emerges where the opportunism modeled by the Prisoners' Dilemma has a free run.

Hobbes's Leviathan may be the explanation for the absence of private armies in modern industrial societies, but there are lots of other spheres of life where cooperative behavior survives without the support of the organized violence of the state. People manage to function in families, neighborhoods, clubs, bureaucracies, and firms where cooperative behavior is essential and the direct police power of the state weak or absent. There is a need for a theory of what Robert Axelrod (1984) has called the "evolution of cooperation," the spontaneous maintenance of cooperative behavior in the face of the strong temptations modeled by the Prisoners' Dilemma.

1.3.5.3 Repeated games

The argument that agents caught in a Prisoners' Dilemma will fall into the mutual defection equilibrium is strongest when we imagine them to be playing the game just once. If agents know they are going to interact many times, it may be easier for them to establish a pattern of mutually advantageous cooperation. This consideration leads to the model of a *repeated game*, which is a major focus of the essays in this volume. The agents in a repeated game know that they will encounter each other many times in situations structured the same way. Each agent can condition her strategic choices on the past behavior of the other player.

The structure of repeated games is inherently much more complex than the structure of the corresponding one-shot games. If each player has just two strategies available in each one-shot game (like cooperation and defection in the Prisoners' Dilemma) the record of a player's behavior in a game repeated n times consists of a vector of dimension n, each component recording the player's strategy choice on one round. There are 2^n such pairs of vectors. In principle an agent's strategy in the $n+1$st round

of the repeated game could be conditioned on this full information, giving rise to a game with 2^{n+1} strategies for each player on the nth round. As n increases, the number of strategies becomes very large indeed. A full strategy for the entire repeated game, would be a plan telling what the player will do on the first round, then what she will do on the second move depending on what the opponent did on the first round (two possibilities), then what she will do on the third round depending on what the opponent did in the first two rounds (four possibilities) and so on. Even if the player knows the game will be repeated a finite number of times, and were to be fortunate enough to know the exact strategy chosen by her opponent, the examination of her own enormous strategy set in a search for a best response, the core of the Cournot-Nash concept, is a daunting prospect.

Once again we see the issues of computational feasibility and costs arising naturally and centrally in the development of a theory of completely rational action. The difficulties of computation strongly suggest truncating the strategic choice problem in some plausible way. The full strategy problem, for example, requires the player of the game to remember the behavior of her opponent on every previous round; in a game repeated N times, she would need N memory cells to devote to this task. Since memory (either human or machine) is expensive, we might reasonably restrict her strategy choices to those that could be implemented with a smaller number of memory cells. This is a practical and plausible, but not strictly rational strategy. Strictly rational logic would require the agent to consider whether she would be better off allocating $1, 2, \ldots, N$ memory cells to this game. Of course, to decide how many memory cells the game is worth, the agent would have to solve an even much more complex problem, and spend even more computing resources to do it. In the essays in this book the cost of computation is often represented, crudely to be sure, by limitations on the computing power of the agent.

Suppose that the agent can devote only 1 memory cell to a repeated game, for example. Then she is restricted in her strategy choice to those strategies that are sensitive only to whatever statistic she chooses to record. If she remembers only the last choice of her opponent, she has reduced the size of her strategy space from $\sum_{n=1}^{N} 2^n$ to 4 (since she has two strategy choices on the $n+1$st round of the game for each of the two possible plays of her opponent on the nth round). In this much reduced strategy space it may be easier to evaluate strategies.

1.3.5.4 The repeated Prisoners' Dilemma

The specific case of the Prisoners' Dilemma has been the subject of considerable theoretical attention. A moment's thought shows that even in a repeated Prisoners' Dilemma of known finite length N, defection is still a

dominant strategy. On the last round, it is clear that defection dominates, since both players know that this is the last episode of interaction. But if each knows the opponent will defect on the last round, there is no point to cooperating on the next-to-last round, and so forth back to the very first round. To get around this analytical hurdle, it is common to assume that the players face a positive probability of the game continuing on every round, so that there is at least the possibility that the prospect of future cooperation can influence their behavior at all times.

The full analysis of the repeated Prisoners' Dilemma on rational grounds remains a major research area. There may be a very large set of Cournot-Nash equilibria for the metagame, but we do not know very much about this set. Mutual defection is always a Cournot-Nash equilibrium, since if you know your opponent will defect there is no point in your cooperating. One equilibrium that seems to be capable of sustaining cooperation is a trigger strategy: the player cooperates on the first round, but shifts to permanent defection if her opponent ever defects. If the game will continue a long time with high probability, the prospective losses from facing permanent defection can outweigh the one-shot gain from defection against a cooperator, making the trigger strategy a best responses to itself, and therefore a Cournot-Nash equilibrium. The trigger strategy requires only one bit of information to implement, a record of the opponent's worst behavior in the game up until the present. Thus it has the advantage of computational cost-effectiveness. Everyone has probably encountered people who adopt trigger strategies in life, but they are notoriously difficult people to get along with. The trigger strategy is too unforgiving of errors in perception or execution, for example, to be a very good method for sustaining cooperation in real human interactions.

Considerable attention has been given to another low-computation cost strategy, tit-for-tat. An agent playing tit-for-tat cooperates on the first round of play, and thereafter plays whatever strategy the opponent played on the last round. In effect, tit-for-tat punishes defection by defecting just once, as long as the opponent returns to a policy of cooperation. Tit-for-tat defends itself better against unconditional defection than unconditional cooperation does, since it shifts promptly to defection on encountering a defecting opponent. It also requires only a single memory cell to implement, a record of the opponent's play on the last round. But it is not a best response to itself, and therefore not a Cournot-Nash equilibrium.

1.3.5.5 Repeated local interaction multiperson Prisoners' Dilemma on the torus

The complexity of the Prisoners' Dilemma even in the two-person repeated game model is high, but plausible and interesting extensions of the two-

person game lead rapidly to even more complex situations. Peter Albin in Chapter 6 below proposes such a generalization, which reveals close links between game theory and cellular automata: the local interaction multi-person Prisoners' Dilemma played on a lattice. In this model each player occupies a node of a lattice in two dimensions. The upper and lower boundaries of the lattice are identified, as are the right and left boundaries, so that in effect the players find themselves on the surface of a torus. The payoffs to an agent depend on the number of her neighbors (including herself) who cooperate in the game. The payoff structure generalizes the Prisoners' Dilemma in that defection is a dominant strategy (no matter what your neighbors do, your one-shot payoff is higher if you defect than if you cooperate), but the payoffs to uniform cooperation are higher than to uniform defection. As in the original Prisoners' Dilemma, private incentives are at sharp variance with social goals.

This is a simple and plausible abstract representation of social interactions and the collective action problem. In real societies neighborhood effects (or small group effects) are extremely important to the welfare of individuals. The externalities modeled by the Prisoners' Dilemma correspond to a wide range of real social problems. The cleanliness of the street depends on the actions of the relatively small number of users; the psychological health of an office depends on the resistance of fellow workers to rumormongering and scapegoating; the effectiveness of education depends on students' reluctance to cheat. But an agent's neighbors in one interaction are involved in turn with other agents with whom the first agent has no direct contact. Just as we are exposed to disease by the behavior of people we do not know through people we do know, our social environment can be influenced by the behavior of strangers who interact with our friends.

Despite its simplicity and appeal as an abstract model of social interaction, the complexity of Albin's local interaction multiperson Prisoners' Dilemma considered from the point of view of rational choice theory (say, Cournot-Nash equilibrium) is staggering. Strictly speaking, the relevant state of the repeated game at any moment from the point of view of one of the agents depends not just on her own and her neighbors' past behavior, but on her neighbors' neighbors' past behavior as well, in fact, on the behavior of all the agents indirectly linked through the neighborhood structure. Since each agent's strategic choices can in principle depend on her neighbors' past behavior, which in turn may be conditioned on the neighbors' neighbors' past behavior, all of the agents in the whole society are indirectly linked. A full Cournot-Nash equilibrium has to consider the best response of each agent to the strategies, not just of her neighbors, but of all the agents in the society. The problems of computational cost and feasibility obtrude immediately and unavoidably.

As Albin shows in his essay, there are actually two levels of computa-

tional complexity problems in this kind of situation. On the one hand it may in principle be feasible to set up algorithms to predict the detailed evolution of the system in response to a particular action of one agent (say, a change in her strategy choice), but very expensive because of the chaotic nature of the ramifications. The cost of the computation constitutes a bound to rationality, and suggests limiting the strategic plans of agents to a subset of all those that are in principle relevant. For example, we might limit the agent to strategies that are conditioned on the past behavior of her neighbors (since she may have difficulty even observing the behavior of her neighbors' neighbors and further degrees of separation). Even here, as the discussion of the two-person Prisoners' Dilemma shows, the complexity of the strategy space is quite high, and may exceed reasonable estimates of the computing power available to an agent. These considerations lead us to study boundedly rational agents who pursue their self-interest within limits imposed by computational costs.

But Albin's analysis also reveals a further remarkable problem implicit in the multiperson Prisoners' Dilemma. Albin shows by construction that there are bounded rationality strategies (one, e.g., which is mathematically equivalent to John Conway's cellular automaton Life) which will lead to a level of complexity in the social interactions represented that is equivalent to that of a general-purpose computer or Turing machine, or to an unrestricted formal language. In this context the computational issues that arise when an agent tries to work out the consequences of a change in her strategy are not just those of cost, but include issues of feasibility. If the society as a whole has the complexity level of a general-purpose computer, it will be impossible for any other general-purpose computer to work out its evolution except by direct simulation. To carry out the program of rational explanation of behavior in this context would require positing that each individual agent in society had some way of simulating the potential evolution of a system of interlinked Turing machines. At this point the rational explanation program runs into deep paradoxes of self-reference. Albin explores these paradoxes in Chapters 2 and 6 of this volume.

1.4 Definitions and measures of complexity

We can see that issues of complexity arise inexorably in social and economic theory once we contemplate seriously the possibility of nonlinearity in the underlying lawful behavior of agents. The development of economic theory in particular has managed to finesse these issues by focusing attention on idealizations like perfect competition, which abstract from the interactions of individual agents, and on situations in which the structure of equilibria is relatively simple, like the saddle-point stable trajectories of intertempo-

ral accumulation models. Some such maneuver was probably functional to the progress of social science, especially during periods in which the mathematical and philosophical understanding of complexity issues was relatively undeveloped. But if complex interactions indeed play a fundamental role in the evolution of social reality, this abstraction may have imposed a high price on the relevance of accepted economic and social theory.

The emergence of computers and of a mathematical theory of language has transformed our understanding of these issues at an abstract level. Albin's work in this volume shows how central these concerns are to the further development of social theory. One of the key achievements of computational and linguistic theory has been to propose explicit definitions of system complexity, and to provide a qualitative classification and quantitative measures of complexity. A brief review of these developments can help put the essays in this volume in context.

The intuitive concept of complexity that lies behind most measures is that simple systems are easy to describe with a small amount of information, while the description of a complex system requires a large amount of information. This idea can be unpacked in several different contexts, which turn out to be closely related. We can, for example, measure the complexity of a system by the size of the smallest computer program capable of describing it. More complex systems require larger programs. Or we could measure the complexity of the system by the computer resources required to represent it, in time and memory. A more complex system requires more time and memory to describe than a simple one. Or we could measure the complexity of a system by the richness of the language required to describe it. A simple system can be described by a minimally structured language, while a complex one requires a language rich in possibilities.

1.4.1 Computational complexity

Mathematicians have always been interested in solving problems, that is, developing algorithms that allow a computing system (up until 1940 usually a human equipped with pencil and paper) to transform one description of a situation (the problem) into another description in a more functional and usable form (the solution). Mathematics has also from its earliest beginnings recognized a relation between the computational tools available and the range of problems that can be solved. Greek geometry, for example, stipulates constructions with straight-edge and compass as a computational limit: the trisection of an angle is not possible with these tools, though it is possible with the addition of more powerful computational devices. Those mathematicians who pursued the practical solution of real problems, such as the computation of artillery trajectory tables, or the compilation of ephemeridae describing the motion of heavenly bodies, became acutely

aware of the wide gap between the discovery of methods that in principle are capable of solving certain problems, and the implementation of these methods in practical situations.

Our understanding of the relation between systems and the sophistication of the machines needed to describe or represent them has been most completely developed for *formal languages* (see Hopcroft and Ullman, 1979, and Révész, 1983, for thorough surveys of these results). The next two sections summarize these accounts. This theory establishes the existence of parallel four-level hierarchies of complexity for machines and languages. A similar four-part hierarchy appears in the study of nonlinear dynamical systems, such as cellular automata as well, and plays a central role in Albin's theory of the complexity of social and economic systems.

1.4.2 Linguistic complexity

Noam Chomsky (1959, 1963) defines a formal language as a set of rules for producing strings from a finite alphabet of symbols. These rules establish certain strings as words (or sentences) in the language and define procedures for producing new words by substitution in existing words. The complexity of a language in Chomsky's hierarchy corresponds to the restrictions imposed on the procedures for producing new words.

A brief sketch of this theory and its main results can help to clarify the mutual relationship of linguistic, computational, and nonlinear dynamical system complexities that informs Albin's work.

A formal language is defined by a *grammar*, which is defined by a finite *alphabet* T of terminal symbols, a finite set V of *intermediary variables*, a finite set P of *productions*, rules for substituting new strings of symbols and variables in existing strings, and a *distinguished variable* S, which serves to start the productions of the language. The productions take the form $P \rightarrow Q$, where P is a string composed of one or more variables together with zero or more terminals, and Q is a string composed of any combination of variables and terminals. The terminal symbols are analogous to the words of a natural language (nouns, verbs, and so forth); the variables are analogous to grammatical forms (sentences, clauses, parts of speech); and the productions are analogous to the grammatical rules of a language. The productions of a natural language like English might take such forms as: <sentence> → <noun phrase> <verb phrase>, and <noun> → "car," indicating that the variable <sentence> can be replaced by a noun phrase followed by a verb phrase, and that the variable <noun> can be replaced by (among many other things) the terminal "car" to produce grammatical (but not necessarily meaningful) English expressions.

The potential complexity of the language generated by a grammar depends on what restrictions, if any, are placed on the production rules.

The languages produced by grammars satisfying the preceding definition with no further restrictions on the production rules are called *unrestricted languages*. The unrestricted languages include the most complex formal languages (and all the simpler ones as well).

The first step in establishing Chomsky's hierarchy of complexity of languages is to consider the subset of grammars for which all the productions $P \rightarrow Q$ have the length of Q at least as long as the length of P. The languages generated by such grammars are called *context-sensitive languages*, because it is possible to prove that all their productions have the form $P_1PP_2 \rightarrow P_1QP_2$, where Q is a nonempty string. This kind of production rule allows the substitution of Q for P in the context $P_1 \ldots P_2$, but not necessarily in other contexts.

Note the monotonicity inherent in derivations arising from context-sensitive languages: each application of a production to a word must result in a word no shorter than the one we started with. This need not be true for an unrestricted grammar: in general the application of productions might first lengthen and then shorten the words produced. This monotonicity greatly reduces the complexity of the languages generated. For example, if we are trying to "parse" a word, that is, decide whether or not it could have been generated from a given grammar and by what sequence of applications of production rules, we know that a given word from a context-sensitive language can have as its antecedents only expressions of the same length or less, a finite set. Thus we can be sure of reaching a definite answer in parsing context-sensitive languages, since we have only a finite number of possible paths by which the word could have been generated. But if an unrestricted grammar is not context-sensitive, a word of a given length could be generated by intermediary expressions of any length whatsoever, thus raising the specter that any parsing procedure might have to run for any length of time before reaching a conclusion. It might even be impossible for us to design a parsing procedure guaranteed to return an answer for any word of such a noncontext-sensitive language in finite time.

The next subset in Chomsky's hierarchy consists of languages generated by grammars whose productions take the form $P \rightarrow Q$, where P is a string of variables (and Q a string composed of terminals and variables), without contextual restrictions, called the *context-free languages*. The not immediately obvious fact is that the context-free languages that do not contain the empty string are a strict subset of the context-sensitive languages, so that they share the monotonicity property of the context-sensitive languages, and have further complexity-reducing structure as well.

The final subset in the Chomsky hierarchy consists of languages generated by grammars with production rules of the form $P \rightarrow T$ or $P \rightarrow TQ$, where P and Q are variables and T is a string of terminal symbols, called the *regular languages*. The regular languages are a strict subset of the

context-free languages, and exhibit a further reduction in complexity.

The Chomsky hierarchy of complexities of formal languages is a key to the characterization and analysis of the complexity of social and economic systems because of its close relation to the complexity of computational devices, on the one hand, and to the complexity of nonlinear dynamical systems (e.g., cellular automata), on the other. As we have seen, we are led to represent social and economic systems by nonlinear dynamic models, and forced to consider the agents in social and economic interactions as operating with meaningful costs of computation. Thus, as Albin's work shows, we must come to grips with the complexity hierarchy of social and economic systems through the same methods and tools that have been developed to analyze linguistic and computational complexity.

1.4.3 Machine complexity

A hierarchy of abstract models of computing devices parallels the Chomsky hierarchy of formal languages. The simplest of these models, the finite automaton, has a finite number of internal states, and can read input symbols from a tape. The automaton makes a move by reading a symbol from the tape and entering a new state which depends on the symbol read and its current state. The automaton is defined by stipulating its starting state, its allowable moves, and a particular subset of states which signal that the automaton has accepted the input. A pocket calculator is a finite automaton. It reads the alphabet of symbols present on its keyboard, and processes them. If it encounters a sequence of symbols that are uninterpretable, it enters an error state. In all states except the error state the calculator has accepted the input stream. It is plausible to regard the finite automaton from the economic point of view as the cheapest form of computational capacity available to a decision maker.

The languages accepted by finite automata are precisely the regular languages that occupy the lowest rank in the Chomsky hierarchy. Conversely, a finite automaton can be constructed to accept any regular language as well, so that the correspondence between finite automata and regular languages is exact.

A consequence of this exact correspondence is that it is impossible to construct a finite automaton to recognize a context-free language which is not a regular language. We can, however, supplement the computational capacity of a finite automaton by adding an unbounded pushdown stack on which the automaton can store and retrieve symbols on a first-in, last-out basis. The resulting pushdown automata can recognize all context-free languages (including, of course, regular languages). Conversely, given any context-free language, it is possible to construct a pushdown automaton that will recognize it, so that the correspondence between pushdown

automata and context-free formal languages is exact. The pushdown automaton still represents a relatively primitive form of computing machine, since its access to the information stored on the pushdown stack is highly restricted: only the last stored symbol is available to influence the course of the computation at any stage; the automaton can access symbols deeper in the stack only by discarding those stored later above. A decision maker equipped with a pushdown automaton, however, is in a position to recognize and respond to a much larger and more complex set of patterns in the world than one equipped only with the computational capacity of a finite automaton.

A logical next step in extending the computational capacity of an automaton is to add a second pushdown stack, thus creating a two-pushdown automaton. A two-pushdown automaton has access at any stage to all of its stored information, since it can shift the symbols from one stack to the other without losing any. The second pushdown stack provides the automaton with the ability to recognize languages of any complexity: a two-pushdown automaton can be constructed to recognize any formal language and, conversely, so that the two-pushdown automata correspond exactly to the Chomskian unrestricted formal languages.

The two-pushdown automaton is equivalent in its computational power to the abstract Turing machine, the generally accepted abstract model of a general-purpose computer capable of implementing any conceivable systematic problem-solving procedure. The Turing machine is an automaton which operates by reading and writing symbols on a single one-sided infinite tape, with the possibility of moving forward and backward on the tape depending on its state and the input stream. The class of Turing machines also is equivalent to the class of unrestricted languages in complexity.

As we might expect, there is an intermediate level of computing power between the pushdown automata and the two-pushdown automata corresponding to the context-sensitive languages. This class of automata includes two-pushdown automata with bounded storage (where the bound on storage may depend on the length of the input word), called *linear bounded automata*. Intuitively, an automaton reading a word from a formal language is inverting the production rules that govern the language, thereby replacing parts of the word with the expressions that might have produced it. The monotonicity of context-sensitive language production rules assures that an automaton applying their inverse can never lengthen the word, so that the space necessary for the work is bounded by the size of the input word itself. Since the context-free languages without the empty string are a strict subset of the context-sensitive languages, a linear bounded automaton can also recognize all of the context-free languages.

From an economic point of view, the bounded storage of the linear bounded automaton represents a qualitatively different computation cost

from the unbounded two-pushdown automaton. A decision maker using a linear bounded automaton to analyze a problem can confidently budget the cost of the computational resources required, while a decision maker using a general two-pushdown automaton can never be sure how large the computational costs of any given analysis might grow to be. This difference in computational costs represents an economic trade-off in computational capacity as well, since the costly general two-pushdown automaton can carry out all the analyses possible for the linear bounded automaton and many others as well.

1.4.4 Decidability, computational complexity, and rationality

A fundamental question about any problem is whether there exists a computational procedure for solving it that will eventually come up with an answer. This question can be rigorously formulated as whether or not a Turing machine (or equivalently, a two-pushdown automaton) exists which, presented with the problem as input, will halt in finite time with the answer. Problems of this type are *decidable*. One of the central mathematical discoveries of the century is that undecidable problems exist (an idea which can be expressed in a variety of forms, including formal language and automaton theory).

A second practical question about any decidable problem is how much resources (e.g., time and machine memory) its solution will require. A decidable problem may still become intractable if it requires computational resources that grow rapidly with the size of the problem itself.

A decision maker trying to carry out the rational choice paradigm in a complex environment faces the fundamental problem of the computational cost of predicting the consequences of her various actions so as to choose the best action in the situation. The theory of computability and computational complexity suggests that there are two inherent limitations to the rational choice paradigm. One limitation stems from the possibility that the agent's problem is in fact undecidable, so that no computational procedure exists which for all inputs will give her the needed answer in a finite time. A second limitation is posed by computational complexity in that even if her problem is decidable, the computational cost of solving it may in many situations be so large as to overwhelm any possible gains from the optimal choice of action.

1.4.5 Dynamical systems and computational complexity

A rational decision maker confronting a dynamical system needs a machine that can evaluate the consequences of her actions to carry out the program of rationality. The analysis of the computation power required of the machine is thus closely connected with the cost the rational agent must sustain.

Albin approaches this problem by associating the economic cost of computation to the hierarchy of computing complexity: finite automata, pushdown automata, linear bounded automata, and two-pushdown automata (Turing machines). These classes of computation can also be linked to the character of dynamical systems.

A dynamical system with a unique globally attracting equilibrium state from this point of view corresponds to a finite automata and to regular languages. A finite automaton needs no memory to record any of the data it has read, and is therefore very cheap. A rational agent confronting a single-point attractor dynamical system needs, therefore, only very primitive computational resources.

Dynamical systems with a periodic attractor correspond to pushdown automata and to context-free languages. The memory needed here, and thus the cost of the device, is proportional to the length of the repetitive patterns the dynamical system will generate. It is notable that most human beings can implement programs representing only very short periodic attractors without becoming confused and losing track of the problem they are working on. Most human beings can work out the dynamics of a two- or three-state system (e.g., what happens if their spouse is angry at them or is friendly with them), but very few can carry on chains of recursion of even ten periods, not to speak of the hundreds or thousands that easily arise in complex financial transactions or complex diplomatic or military confrontations. Even periodic systems, if they involve a large number of degrees of freedom, can challenge the computational resources built into our brains by evolution.

Context-free languages can exhibit a long-distance correlation in data strings: an opening parenthesis requires the eventual appearance of its matching closing parenthesis, but an indeterminately large amount of material (including many more parenthesis pairs) may intervene. This type of long-distance correlation has not received much attention in economic models, though in principle it may play an important role. For example, the settling of debts has this recursive character. Agent B may borrow from agent A, opening a parenthesis that will be closed by the eventual repayment of the debt. But agent B may then turn around and lend to C,

opening a second parenthesis. C's repayment of B is the precondition for the resolution of the original loan by B's paying A.

Dynamical systems with a chaotic attractor correspond to linear bounded automata and to context-sensitive languages. Small deviations in a trajectory at one time ramify indefinitely over time and space, and are never resolved by a closing parenthesis. This spreading out of chaotic trajectories corresponds to the monotonicity property of production rules for context-sensitive languages. While problems in this complexity class are typically decidable in principle, they are often computationally intractable because of rapid increases in computational cost associated with increases in the complexity of the problem, the time horizon of interest, or the accuracy of solutions required.

As Albin points out in his essay on the metalogic of economics (Chapter 2) if we regard human beings as having a complexity at the least equivalent to that of Turing machines, then economic and social systems composed of interacting human beings are in principle more complex than the context-sensitive languages that correspond to linear bounded automata, corresponding at least to the complexity of unrestricted formal languages. The distinguishing characteristic of these systems is the nonmonotonicity of the derivations of their states. If we choose some measure of complexity (such as the length of a word generated, or the number of transactions we observe in an economic market) to characterize the state of the system, we cannot rule out the possibility that it evolved from an even more complex state. Human beings, for example, may adopt apparently simple modes of behavior on the basis of extremely complex and involuted reasoning that considers much more subtle possibilities.

The representation of dynamical systems as a part of the rational choice program thus raises two important questions involving computation. The first is the more important from a practical perspective: how much will the effective computation of the behavior of the system cost? This question, as we have argued, poses a bound to the degree of rationality achievable in any given situation. The second is the more important from a theoretical point of view: is it possible in principle to compute the functions posited by the rational choice program? This question, when answered in the negative, poses an absolute barrier to the rational choice program.

The essays in this volume address these two questions in a variety of economic modeling contexts.

1.5 Complexity in cellular automata

Cellular automata are, as we have seen, a simplified representation of general nonlinear dynamical systems. Because they are defined by very simple

rules, it is possible to investigate a substantial subset of the possible cellular automata in some detail by simulation, even within the limits posed by the power of contemporary computers.

Stephen Wolfram's studies of cellular automata are the foundation of the economic and social models in this book.

Wolfram proceeds by classifying cellular automata in terms of the dimension of the lattice on which they exist, the number of states, k, each site can occupy in each step, the radius defining the limits of one-step influence of sites, r, and the rule of evolution governing the evolution of the automaton. One-dimensional cellular automata are the cheapest to simulate and the results of the simulation can be visualized in two-dimensional plots, with the sites of the automaton arrayed horizontally and time represented vertically. Wolfram and others have also looked intensively at two-dimensional cellular automata. As the dimension of the lattice increases, the number of rules possible increases enormously, so that exploration by simulation becomes an increasingly difficult program to carry out.

In general a cellular automaton rule can allow the next-step state of each site to depend in an arbitrary way on the states of the cells in its neighborhood. For example, in a one-dimensional cellular automaton with $k = 2$ and $r = 1$ the next state of each site depends on three bits representing the current state of itself and its left- and right-hand neighbors. The state of this neighborhood can take on eight configurations. In an arbitrary rule, each of these configurations is mapped onto a 1 or a 0 representing the next state of the center cell, so there are 256 possible cellular automaton rules. Wolfram cuts down the number of possibilities by restricting his attention to subsets of the possible rules that reflect natural physical assumptions. In physical systems, for example, there is no reason to think that influences on one side of a site are any different from influences on another side, so that it is natural to concentrate attention on *symmetrical* rules. In many physical systems the influences of neighbors are at least approximately additive, so that it makes sense to look at *totalistic* rules, in which the next state of the cell depends only on the total number of neighbors (including itself) in each state, or to *outer totalistic* rules, in which the next state of the cell depends separately on the total number of neighbors (not including itself) in each state and the state of the cell itself. In many physical systems the 0 state represents a minimum energy state of a cell, so that it is plausible to require that when the neighborhood is all in the 0 state, the next state of the cell is 0 (*legal* rules).

These restrictions are plausible physically, there is no strong indication that the behavior of the totalistic, legal, or symmetric class of cellular automata is unrepresentative of the behavior of the general case, and the restrictions greatly increase the efficiency of the simulation exploration program by restricting the parameter space it must examine. But it is worth

noting that each of these physically plausible restrictions on cellular automaton rules is less compelling in a social than a physical context. If we think, as Albin does in some of the essays in this volume, of the cells as representing firms, then the concept of "neighboring firm" might be thought of as "suppliers" and "customers." In this interpretation the next state of firm might be thought of as depending on the current state of its suppliers and customers, since they determine the availability of inputs and the demand for output from the firm. If we were to array the firms on a line with suppliers to the left and customers to the right, the assumption of symmetry in the cellular automaton representation would not necessarily hold, since the impact of the state of its supplier on the future state of the firm could be quite different from the impact of the state of its customer.

Similarly, totalistic rules make considerable sense in some social and economic contexts, but are less attractive in others. In the exchange model of Chapter 5, for example, an agent's decision to undertake costly advertising to enter the market depends on her estimate of the exact offer prices of her neighbors, information which cannot easily be summarized by a single statistic derived by addition. Even the restriction to "legal" rules, in which the state of all 0s in a neighborhood determines a 0 for the next state of the cell, might not be attractive in some economic models. If the cell represents a possible production activity, and the neighborhood similar production activities, 0s in neighboring cells might represent a competitive opportunity, and lead to spontaneous generation of activity in a particular location.

Since the rules governing the evolution of cellular automata are so simple, one might hope that it would be possible to derive simple and general algorithms to predict the evolution of cellular automata from initial conditions. Wolfram's study is motivated largely by exploring the limits of this program. His method is to look through simulation at the evolution of a large number of cellular automata from either simple seeds (single nonzero sites) or random seeds, to see what patterns emerge over time. His striking findings are the basis for a number of the essays in this volume.

1.5.1 Complexity types

Wolfram finds that one-dimensional cellular automata can be classified into four basic complexity levels. (He conjectures that these same four levels also suffice for higher dimensional cellular automata.) The first three levels of complexity are analogous to various types of attractors in nonlinear dynamical systems.

Type 1 cellular automata evolve to uniform states from arbitrary initial conditions. Alteration of the initial conditions leads to changes that die out

over time as the system seeks out the uniform state. They are analogous to dynamical systems that have a unique stable equilibrium.

Type 2 cellular automata evolve to periodic patterns of states from arbitrary initial conditions. Small change in initial conditions may change the phase of the periodic limit, but not its basic pattern. They are analogous to dynamical systems that have periodic attractors.

It is not difficult to predict the evolution of type 1 and type 2 systems: once we have seen the pattern that evolves for a few initial conditions, we pretty well know what is going to happen in any further experiment we do on the system. Small errors in the measurement of the initial conditions lead to errors in prediction which decrease over time.

Type 3 cellular automata evolve to nonrepeating, complicated patterns from arbitrary initial conditions. A small change in initial conditions changes features of the evolutionary pattern on a larger and larger scale with time. They are analogous to dynamical systems with chaotic behavior and chaotic attractors. While it is difficult to predict the evolution of type 3 cellular automata in detail, it is possible to find statistical regularities in the patterns that evolve, such as the average number of cells in each state, and entropy measures of the diversity of patterns developed. In some cases it is possible to identify types of patterns that can never evolve in the automaton, despite the large and constantly changing patterns that do evolve. Type 3 cellular automata, like the context-sensitive languages, exhibit a monotonically increasing range of changes emanating from a local perturbation of initial conditions.

Type 4 cellular automata produce propagating irregular structures from arbitrary initial conditions. A change in initial conditions can lead to changes that propagate coherently for large distances. The structures that evolve from similar initial conditions may be quite different. In type 3 cellular automata the effects of changes in initial conditions spread out constantly in space over time, but in type 4 systems the effects may spread and then contract. As a result there is no way to tell in a type 4 system how large the intermediate structures leading to a particular configuration might be: a relatively simple final configuration of the system may be the result of the distillation of immense and very complex intermediate structures. The type 4 cellular automata are thus analogous to formal languages produced by unrestricted grammars.

From the dynamical systems point of view, type 4 cellular automata lie on the boundary between types 2 and 3. A disturbance in the initial conditions of a type 2 cellular automaton monotonically subsides into the periodic attractor, and a disturbance in a type 3 system monotonically explodes through the whole space. A disturbance in a type 4 system neither implodes nor explodes, but propagates irregularly through the system.

Type 4 cellular automata are thus models of "complexity on the edge of chaos," systems just on the boundary between subcritical stability and supercritical instability.

1.5.2 Computability, predictability, and complexity in cellular automata

Type 3 and type 4 cellular automata represent the boundaries and barriers to rationality that are the theme of the essays in this book.

Type 3 cellular automata produce the propagating, irregular patterns characteristic of chaotic dynamical systems. A small change in the initial conditions of the system leads to ever-widening changes in its eventual configuration. Thus the computational cost of predicting the effect of a change in a type 3 system rises with the length of the horizon and with the accuracy of prediction required. An agent attempting to carry out the program of rational action in a type 3 environment faces these costs. For even fairly simple representations of real social and economic interactions, the costs of implementing a fully rational strategy become unreasonably large. In such a situation the agent has to accept limited computation as a bound on her rationality. She may, like the agents in some of the papers below, deploy limited computational capacity to improve her strategy by adopting one of a number of boundedly rational strategies, but she cannot reasonably be supposed to carry out the full computations presupposed by the hypothesis of full rationality.

Type 4 cellular automata pose even more fundamental paradoxes for the conception of rational action. Type 4 cellular automata are conjectured to be capable of general computation, that is, of simulating a Turing machine. Systems of this level of complexity pose difficult problems for a decisionmaker. A given small change in initial conditions (e.g., representing a change in the behavior of the rational agent) may lead to a series of ramifying consequences which simplify over time into the emergence of a particular state (perhaps a state very much desired or disliked by the agent). But there is no way for the agent to find out whether or not this is true except to simulate the whole system in all its complexity. In undertaking this computation, the agent will run into frustrating barriers to her carrying out the program of rational choice. There is no way for her to tell whether the simulation in question will arrive at the desired answer in a finite time, or indeed any answer in finite time. She may let her computer run for a long time; at any moment she has no way of knowing whether it is just a few steps away from resolving the question or is caught in an enormous loop from which she can never garner any useful information.

One of the main themes of the essays in this volume is how easy it is for

type 4 environments to appear in relatively simple and standard models of human social and economic interaction.

1.6 Modeling complex social and economic interactions

1.6.1 Self-referencing individual agents

In "The metalogic of economic predictions, calculations and propositions" (Chapter 2), Albin addresses the paradoxes of computability in the context of rational economic decision making at the most abstract and general level.

The central point is that economic agents are at least as complex as Turing machines, and the economy (or any subsystem of the economy, such as a firm or an industry or a market) is a system made up of agents with this complexity level. An agent (or indeed, an economist) trying to predict the behavior of such a system in relevant dimensions faces a problem exactly equivalent to the problem of calculating the value of an uncomputable function. The difficulty arises because there is no shortcut available to simulate the behavior of the complex agents who make up the system. The consequences of a change in initial conditions (say, the behavior of an agent herself) can be worked out only by predicting the complex reactions of the other agents in the system, and there is no way to do this except by a complete simulation of these other complex systems. In this chapter Albin establishes the necessary logical mapping from a general social system to the theory of computable functions, and shows that the undecidability propositions of computation theory translated into the economic sphere imply the impossibility in general of computing the economic outcome of a change in initial conditions.

One might hope that this difficulty is essentially a problem of approximation, and that the agent might, with finite computational resources, reach an acceptable approximation to the true consequences of the change in initial conditions in finite time. But Albin shows that the same logic that applies to the computation of the consequences of an action themselves applies to the problem of determining how good an approximation any particular feasible computational method will produce.

For example, economists have often based policy prescriptions on their evaluation of the social welfare implications of changes in taxes or subsidies. But in the general case, the computation of the allocation that will result from a given change is undecidable; the economist could carry out her analysis only by simulating the economic system in detail, a task she could not accomplish with finite resources.

For another example, the manager of a firm considering the consequences of a pricing decision, would in principle have to simulate the full reactions of all the other agents in the economy to a contemplated price increase, taking into account the reactions of her customers to the change, and the rippling effects of their changed actions on other related sectors of the economy.

It is important to appreciate the fundamental nature of this result, as well as some of its limitations and qualifications. What is at issue here is not the quantitative efficiency of computers or information collection systems. The problems of undecidability in computational theory arise not from the specific limitations of particular machines, but from the self-referential character of certain computations, a logical difficulty that raw computing power cannot overcome. Nor should we regard this result as a kind of special case: it is in fact the general situation of interacting social actors. The key characteristic of these social systems is that they are made up of linked actors each of which has a complexity level at least as great as that of a Turing machine.

The apparent nihilism of this result also needs to be put into perspective. Albin's analysis does not claim that it is always impossible to forecast the consequences of an action in finite time, only that it is impossible in general and impossible to find out except by trying. It may be that the ramifying consequences of a price increase by a particular firm are limited and can be simulated in finite time by a sufficiently sophisticated computer. But the manager of the firm cannot be sure whether this particular instance of a price increase will lead to consequences in the computable category. She has no way of finding out except to start to compute the consequences; if her program stops after a minute or a year, she will know the answer, but if the program is still running after a minute or a year, she has no way of knowing whether it will continue to run forever, or for ten thousand years, or is just one cycle away from stopping with the answer she wants. The computational complexity inherent in a social or economic system does not prevent agents from carrying out the rational decision program in some (perhaps in a large number) of contexts, but it does prevent us from accepting the logical adequacy of the rational program as a general account of what might happen.

Economists have been aware of this paradox to a certain extent since the early beginnings of the discipline. It is instructive to see how economics has attempted to cope with the challenge of complexity. The critical maneuver is to replace the computationally intractable problem of forecasting the behavior of many other complex decisionmakers and their interactions with the computationally simpler problem of calculating a posited equilibrium of the economic system. Thus the classical British political economists, Smith, Malthus, and Ricardo, employed the concept of *natural*

prices, around which day-to-day market prices of commodities fluctuate (or *gravitate*.) The implication is that the natural prices are equilibria subject to logical analysis and prediction despite the fact that the fluctuating market prices may be the result of detailed interactions too complex to predict or explain. Contemporary neoclassical economic theory rests on a similar epistemological postulate, which claims that the equilibrium of the economy is computable, even if the disequilibrium paths of the economy are not. From this epistemological viewpoint the rational economic actor need not consider the detailed reactions of other actors in all their potential complexity, since the equilibrium market prices convey enough summary information to allow her to make a rational plan. This program has had a certain amount of success, but clearly rests on a leap of faith that somehow the complexity level of the economy as a whole is lower than that of the agents who constitute it.

The study of cellular automata can throw some light on this epistemological puzzle. Complexity in cellular automata arises through the ramifications of linked local interactions. Equilibrium market prices, on the other hand, summarize information about the global state of the economic system, since they depend in principle on the states (say excess demand functions) of all the participants in the market. The traditional economist effectively assumes that this global information dominates local interactions in influencing the behavior of the agents in the system, and thus suppresses the complexity latent in the local interactions. There is surely some truth in this point of view, but it raises other serious and interesting questions that have hardly been addressed in any rigor by economic theory. The most central of these questions is where the resources come from to diffuse the information contained in market prices. The weakest aspect of economic equilibrium theory is its implicit assumption that information diffusion in the market is costless and instantaneous.

1.6.2 Organizations

The interplay of information and the complexity level of individual agents in an economy is the theme of "The complexity of social groups and social systems described by graph structures" (Chapter 7), which reveals the interplay of whimsical humor and mathematics in Albin's thought. Here Albin lays the conceptual foundations for the measurement of the complexity level of an organization, modeled as a directed graph. The nodes of the graph represent information processing units in an organization (e.g., a production unit or an accounting office) and the graph connections the communication links between them, which may be one- or two-way.

Albin proposes an extension of the concept of algorithmic complexity to measure the complexity of such directed graphs. He asks us to consider

the organization as a *rumor mill*, devoted entirely to the spreading of bits (or tidbits) of information. Each node of the organizational graph has a certain number of input and output channels along which this information must flow. The requirement that each node keep track of the rumor's path ("I heard it from Z who heard it from Y ...") for the largest possible rumor the graph can generate establishes a measure of the algorithmic complexity of each node. The sum of these algorithmic complexities over the whole organization (or graph) is a useful measure of the complexity of the organization.

Albin is able to link this measure of complexity with observable organizational phenomena, including the effectiveness and the functionality of the organization. This line of thinking raises sharply the question of whether real organizations can reasonably be supposed to reach the complexity level assumed in rational choice theory. If in most situations resources limit the complexity levels that can be sustained by organizations, they will also limit the type of prediction and projection the organization can undertake in making decisions. Under these circumstances the systems modeler may get further by identifying the particular algorithms of bounded complexity the organizations are actually using than by positing full rationality.

1.6.3 Industries and economies

Since cellular automata are a mathematically economical representation of the full range of potential system complexity, it is tempting to use them to model whole interactive economic systems, indeed, whole economies.

Albin's "Microeconomic foundations of cyclical irregularities or 'chaos'" (Chapter 3) is a pioneering effort in this direction. Here Albin seeks to map Richard Day's model of chaotic economic growth paths onto a cellular automaton. Day's aim was to show that a plausible modification of Robert Solow's model of economic growth could lead to a dynamical system with a chaotic attractor. In Solow's model the state variable of the economy is taken to be the capital stock per worker, which follows a dynamical law representing the productivity of the economy and its saving/investment behavior. Solow's original examples were carefully constructed so as to yield a system with a single equilibrium point attractor. Day showed that the addition of a plausible assumption lowering investment or output at high levels of capital per worker (which could be the reflection either of wealth effects on saving or of external pollution effects on productivity) gives rise to a dynamical system with a chaotic attractor.

Albin takes this line of thinking one step further. He disaggregates the Solow/Day analysis by considering the economy as a large number of separate firms, each with its own production function and investment decision.

The aggregation of these firms constitutes the macroeconomy; the aggregation of their investment plans corresponds to the economy-wide levels of investment in the Solow/Day setup. Albin then regards each of these microlevel firms as cells in a one-dimensional cellular automaton. In order to simplify the simulations, he restricts the investment levels for each firm to "high," "normal," and "low," leading to a three-state cellular automaton. The geometry of the line can be seen as an ordering of firms by industries, with supplier industries on the left and customer industries on the right. A cellular automaton rule in this context represents a pattern of local interaction in which each firm chooses its next period investment policy conditional on the current investment policies of its suppliers and customers. Aggregate fluctuations in investment arise from the individual fluctuations of the firms' policies. There are clearly a large number of possible rules, each corresponding to a different pattern of behavioral interactions on the part of the firms.

Albin uses this model to make two fundamental and compelling points. First, for a large class of rules, the resulting cellular automata are chaotic. In these economies there are ceaseless fluctuations of investment and growth both at the microeconomic level within industry groupings (neighborhoods), and in the economy as a whole. These chaotic fluctuations pose a serious challenge to modeling the behavior of any one firm as fully rational, that is, based on a correct prediction of the behavior of the economy or its neighbors. The problem is that a full projection of the evolution of the economy, or any piece of it, will require computational resources that grow at the same rate as the horizon of the projection. Under these circumstances the implicit assumption in rational choice models that the computational costs of projecting the consequences of behavior are negligible is unsustainable.

Second, as we have seen, one-dimensional, three-state cellular automata can produce not just type 3 chaotic patterns, but also type 4 behavior, which is conjectured to be at the same complexity level as a general-purpose computer, or an unrestricted formal language. We have no reason to believe that the interactions of the real economy exclude type 4 behavior (though by the same token we have no reason to believe that the real economic interactions lead to type 4 behavior, either). If the economy as a whole or subsystems are governed by interactions that lead to type 4 behavior, the barriers to rationality established by the paradoxes of self-reference come into play as well.

Albin's modeling of firm interaction as a cellular automaton breaks new ground in cellular automaton theory as well. Once we interpret the cells of the automaton as firms and the relation of neighbors as members of the same industry or suppliers or customers, some of the physically motivated assumptions of cellular automata theory need to be rethought. In physical models, as we have seen, it is very natural to assume that neighbor effects

are symmetrical, but if the geometry of the lattice represents asymmetric economic relations (such as supplier and customer), this assumption is no longer strongly compelling. Albin's simulation of asymmetric rules thus also extends the cellular automaton model in interesting ways.

The fluctuations that arise from the local interactions of investment policy in this model raise the question of whether some weak global information could stabilize the economy. In "Qualitative effects of monetary policy in 'rich' dynamic systems" (Chapter 4), Albin extends the model of local firm investment interaction to include a simple representation of monetary policy. The idea is to append to the set of firms a single site which is a neighbor of all of them. The state of this site represents a systemwide monetary policy ($+1$ corresponding to an expansionary policy, 0 to a neutral policy, and -1 to a contractionary policy). A procyclical monetary policy takes the value $+1$ when a majority of the firms in the economy are also in state $+1$, and therefore investing at a higher than average rate, and -1 when a majority of firms are in the -1 state and therefore investing less than average. A countercyclical monetary policy "leans against the wind" by assuming the value -1 when a majority of the firms are in state $+1$, and $+1$ when a majority are in state -1, thus offsetting the aggregate tendency of the system.

Albin uses simulation to investigate the impact of these monetary policies on the evolution of the cellular automata representing the economy of firms. His analysis is aimed not so much at measuring the effectiveness of policy in actually stabilizing the aggregate behavior of the economy as in observing the impact of the policy on the complexity level of the resulting automaton. He finds that turning on the policy interaction can result, depending on the rules governing the local interactions of the firms, in any possible change in complexity levels: type 1 point attractor systems may be transformed by the weak global interaction of the policy variable into type 2 or 3 or 4 systems, and similarly for the other complexity levels.

In his investigations of the metalogic of economic forecasts, Albin already raised the possibility that the impact of policy on economic outcomes and hence on social welfare might be unpredictable. The monetary policy model complements this point by posing an abstract but not implausible model of the economy in which policy intervention may itself move the system from one level of computational complexity to another.

1.6.4 Markets

In "Decentralized, dispersed exchange without an auctioneer" (Chapter 5), Albin and Foley begin to attack the problem of representing canonical economic exchange as a cellular automaton. The basic economic setting is

chosen to be as simple as possible: 100 agents arrayed in a circle trade endowments of two goods. The agents have identical preferences, represented by the utility function x_1x_2, so that the marginal rate of substitution, (or willingness to pay) of good 2 for good 1 for an agent currently owning $\{x_1, x_2\}$ is just x_2/x_1. Trade is motivated by differences in initial endowments of the agents. The total endowment of the two goods in the economy is arranged to be equal (though the agents do not know this), so that the Walrasian market equilibrium relative price will be unity. The agents' endowments are diversified subject to the constraint that their wealth at the Walrasian equilibrium price is the same. Thus agents start with the same total $e_1 + e_2$ of the two goods, but in different proportions. As a result, the initial offer prices differ, so that agents typically have a motive to make a mutually advantageous exchange.

The model puts stringent restrictions on the information available to the agents and their economic strategies. Each agent can communicate and trade only with her neighbors within a given radius, which is a parameter of the model. When the model starts, she does not know the total endowment of the two goods, or the Walrasian equilibrium price, so that she does not know whether her endowment is typical of the whole economy or skewed in one direction or another. As a result she does not initially know whether she would be a net buyer or seller of good 1 in the Walrasian equilibrium. Each agent must inform herself about her trading possibilities by actually trying to trade with her neighbors. To do this she must pay a cost to advertise her willingness to buy or sell good 1. This advertising puts her in contact with those neighbors who advertise on the other side of the market (buyers with sellers and vice-versa). Each such pair meets and reveals its current willingness to pay. If the buyer's willingness to pay is higher than the sellers, they exchange a small amount of the goods at a price that is the geometrical average of the two marginal rates of substitution, thus accomplishing a mutually advantageous exchange. Each agent remembers the most recent willingness to pay for each of her neighbors with whom she has actually bargained, and uses this to construct a simple model of the distribution of offer prices in her neighborhood, which she uses to decide whether it is worth her while to incur the cost of further advertising.

From an economic point of view this is a model of bounded rationality. In principle, each exchange an agent makes will influence the whole future course of exchanges in the whole market. From a game-theoretic point of view, for example, we could try to work out the best strategy (in terms of advertising and revelation of prices) for an agent, supposing that all the other agents continued to follow the rules set out in their programs. It is unlikely that the program we have proposed would be the best response to itself, though it is not clear how one would even establish this proposition, since the consequences of a given change in the behavior of one agent involve

a complicated ramifying sequence of events which depend critically on facts, such as the exact or even the average endowments of the other agents, which the agent does not know. One Nash equilibrium of this economy is easy to spot, but not very interesting, in which all the agents refuse to advertise or trade at all. (If an agent knows that none of her neighbors will advertise, there is no point in her incurring the costs of doing so.) Despite the fact that the agents are not trying to be anything like fully rational, their behavior is a reasonable way to pursue the end of trading to mutual advantage. Simulations of the model reveal that it works rather well to achieve an approximation to a Pareto-efficient allocation of the goods (in which no further mutually advantageous trades exist) at a relatively low advertising cost. The main difficulty is a halting problem: occasionally two agents in a neighborhood who do not know each other's willingness to pay get caught in an endless loop, in which they both alternate between advertising as buyers and sellers in the hope of advantageous trade with each other.

A further striking feature of the simulations is that while the market, as it is supposed to, efficiently redistributes the initial endowments around the circle and evens out the relative proportions in which agents hold the two goods, it also systematically introduces inequalities of wealth into the economy. This is a manifestation of a point long acknowledged but little investigated by economic theory: when agents trade at disequilibrium prices they effectively redistribute wealth at the same time that they narrow the differences in willingness to pay. In this particular model it is not hard to see that agents who start with endowment proportions far different from 1:1, and therefore with very high or very low willingness to pay good 2 for good 1, will make most of their transactions at unfavorable prices, effectively transferring their wealth to the lucky agents whose endowment proportions are close to the average.

This model is a cellular automaton, but the states each cell (or agent) can occupy are more general than those typically studied in the cellular automaton literature. The full state of an agent at any round of trading includes her current holdings of the two goods (effectively a continuous state variable), and the record she keeps of the willingness to pay of those neighbors she has met in attempts to trade. The neighborhood structure, and locality of interactions, on the other hand, keep strictly to the cellular automaton format.

The emergence of a close-to-Pareto-efficient allocation in this model is an example of the self-organization of a complex system based solely on local interactions undertaken with no information about the global state. Because no agent can ever make a trade that leaves her worse off, the opportunities for mutually advantageous exchange are gradually exploited in the evolution of the system. The market effectively computes a Pareto-

efficient allocation (though not the one that is generated by Walrasian *tâtonnement*, in which each agent makes all her transactions at the same, market-clearing, price) and a corresponding almost-uniform willingness to pay. The simultaneous emergence of a definite wealth distribution as the consequence of market transactions at nonequilibrium prices is a further example of such self-organization.

1.6.5 The local interaction multiperson Prisoners' Dilemma

In "Approximations of cooperative equilibria in multiperson Prisoners' Dilemma played by cellular automata" (Chapter 6), Albin proposes an elegant generalization of the two-person repeated Prisoners' Dilemma game that social scientists have come to take as the paradigmatic representation of the paradoxes of rationality and social interaction. The society is represented as a cellular automaton in two dimensions with $k = 2$ states, and the "Moore" neighborhood, consisting of the eight adjacent sites in the lattice. Each cell represents an agent. The kernel of the model is a one-shot game in which each agent chooses one of two strategies, cooperation or defection. The payoff to each agent depends on her strategy and the total number of cooperators in her neighborhood. The payoffs, as in the two-person Prisoners' Dilemma, are structured so that defection dominates cooperation: no matter what the neighbors do, the payoff is higher to a defector than to a cooperator. But also, as in the two-person Prisoners' Dilemma, the payoff to a neighborhood where everyone cooperates is higher than to a neighborhood where everyone defects. From a social point of view everyone will be better off if everyone cooperates (the Pareto-efficient solution), but everyone has an incentive to increase her individual payoff by defecting.

Many interpretations could be given to this model. The quality of life in city neighborhoods, for example, depends to a considerable degree on the willingness of residents to incur real costs to maintain their property. There are considerable joint benefits to be had from uniformly high levels of property maintenance, and considerable joint costs imposed if everyone lets their property go unmaintained. Local industrial pollution poses a similar scenario, in which all the producers may enjoy lower costs if they all control their pollution, but any individual producer is tempted to pollute.

Although this model is a minimal and simple extension of the Prisoners' Dilemma, and, given the level of abstraction, plausibly represents a wide range of important social interactions, its structure poses a severe challenge to standard noncooperative game theory. The relevant strategy set would allow each agent to condition her strategy in any one play of the game on the whole past history of her and the other participants' actions. But she

is linked indirectly to all of the agents in the game by the neighborhood structure. Is it plausible to assume that the agent can even observe the actions of agents far away from her geographically? If she cannot, must she undertake to infer their behavior from the behavior of her neighbors? Even if we assume the agent can observe the whole history of the global game, the problem of calculating best responses is formidable. Of course, some game theoretic results are fairly immediate. Uniform defection, for example, is a Nash equilibrium, but not a plausible or interesting one. A "trigger" strategy that cooperates until some agent anywhere in the game defects, and then shifts to uniform defection might sustain uniform cooperation and be a best response, but most people would not view it as particularly practical in the contexts where some level of defection is almost certain, but it still pays collectively to maintain high average levels of cooperation.

Albin approaches these difficulties by restricting attention to a subset of possible strategies, those conditioned simply on the last action of the immediate neighbors. He adopts, in other words, a bounded rationality perspective on the grounds that computational complexity problems that arise in this limited setting cannot be presumed to get any better when the strategy space is expanded and the model thereby made even more complicated. This limitation also has the appealing feature of creating a strict equivalence between strategies and the rules governing the evolution of a two-dimensional cellular automaton with the Moore neighborhood. The current state of each cell is the decision of the agent to cooperate or defect, and the next state, given the strategy of the agent, will depend only on her current strategy and the current strategies of her neighbors. Thus holding the strategies of agents constant, the game evolves as a cellular automaton, about which we have considerable knowledge from earlier research.

In the context of this model Albin raises a subtle but extremely interesting question. If there were a Nash equilibrium in this subset of strategies different from universal defection, in other words, a Nash equilibrium that could sustain social cooperation, what would the complexity type of the resulting cellular automaton have to be? Consider a society in which all the agents have adopted a uniform strategy that leads to a cellular automaton evolution of types 1, 2, or 3. Suppose that a particular agent now considers changing her strategy. To work out the exact consequences of such a change assuming that the other agents continue with their existing strategies is in general computationally infeasible. Albin supposes that the agent, in the spirit, if not the letter, of Nash equilibrium reasoning, assumes that the complexity type of the society will not change when she changes her strategy. If she confronts a predictable type 1 or type 2 cellular automaton, she will do better shifting to uniform defection. Even if she confronts a type 3 situation, in which the behavior of her neighbors appears to be chaotic, it is plausible to suppose that she will view her environment as stochastic,

and realize that uniform defection is also her dominating policy against any randomized strategy of her neighbors. Thus it is implausible to imagine that even boundedly rational agents would maintain a strategy that could sustain cooperation if this strategy resulted in a social environment of complexity level lower than type 4.

An agent who lives in a society corresponding to type 4 complexity, however, might not choose to shift to uniform defection. She might perceive her neighbors as at least potentially reacting thoughtfully and predictably to defection, and punishing it selectively in such a way that uniform defection would lead to a lower average payoff. Despite the difficulty of proving the stability of a type 4 strategy, Albin conjectures that strategies which implement the game of "Life" may have this property. "Life" is known to support type 4 complexity, and is known to be capable of emulating a generalized Turing machine. In support of his conjecture, Albin presents simulations showing the ability of "Life" to generate systematic and selective punishment of uniform defectors.

Whether or not "Life" turns out to be a stable equilibrium strategy for the local interaction multiperson Prisoners' Dilemma, Albin's investigation underlines the philosophical theme that runs through his work. We live in a social environment created by human beings, who are, considered as systems, highly complex. Albin does not believe that there is any way to reduce this essential complexity to produce a simple theory of social interaction that is also robust. Despite occasional apparent successes of economic theory in using the assumption of competition to simplify the rational choice problem of individual agents, Albin argues that there remains an irreducible kernel of latent complexity in even the most stylized and abstract models of society, such as the local interaction multiperson Prisoners' Dilemma. This is probably not news to human beings who are living their way through social interactions, but it poses some fundamental challenges to social and economic theory.

In this perspective, the questions Albin poses about the complexity of social institutions, despite the difficulty or perhaps the impossibility of answering them with currently available conceptual tools, involve fundamental issues of method and metatheory. They suggest that the kind of knowledge we can hope to have about complex social systems is different in kind from the kind of knowledge we have gained about physical systems. The program of reducing complex physical interactions to computable models, which has had such striking success in the physical sciences, has much narrower conceptual limits in the context of social interactions, precisely because the constituent subsystems of societies, human beings, are themselves emergent, complex adaptive systems.

These implications of Albin's work need not lead us to nihilism about the possibility of some important and powerful kinds of knowledge of social

interactions. The ability to recognize, name, and classify the complexity levels of particular social interactions, as Albin argues, is valuable in itself. Furthermore, a rigorous ability to recognize the limits of what is knowable, which is what the complexity analysis of social systems offers, is the starting point for scientific advance.

1.7 Complexity, rationality, and social interaction

Peter Albin's investigations establish a strong presumption that the bounds to rationality constituted by computation costs in chaotic environments and the barriers to rationality raised by the self-referential paradoxes of complex environments are inherent in our conception of social systems as interactions of rational individuals. The systematic analysis of complexity reveals an internal conceptual contradiction in the program that seeks to reduce social phenomena to the interactions of self-interested individuals. This challenge to the dogmas of rationality takes a new and unexpected form. The rational choice/social interaction paradigm has often been criticized on various grounds: that human beings are in fact not motivated by individual self-interest; that individual personality is a social construct which cannot be posited independent of the social context in which it develops; that social interaction leads to the emergence of specifically social interests and forms that cannot be reduced to individual actions. These external critiques question the premises, relevance, and explanatory power of the rational choice paradigm.

The issues raised by the consideration of computational costs and complexity, on the other hand, are fundamentally internal to the rational choice paradigm. The complexity analysis accepts the premises of rationality, and pushes them rigorously to their limits. This process first of all revealed what might appear to be a gap or oversight in the rational choice theory. In assuming that agents could map their actions onto consequences, rationality implicitly abstracts from computation costs. This abstraction, like many others, would do no harm if it were appropriate, in the sense that ignoring computation costs made little practical difference to the predictions and explanations of the theory. In many familiar human situations, after all, computation costs seem negligible. We are automatically provided with a powerful general-purpose pattern recognition and computation device in the form of our brains, which typically have plenty of spare capacity to deal with day-to-day decision making, like finding the nut-bearing trees or the boar hiding in the forest, with enough left over to write epic poetry, speculate on the nature of the universe, and devise elaborate insults for

each other. It is easy to see how theorists slipped into the universal presumption that the human brain could solve the problems predicting the consequences of actions. Rational choice theory inherits from theology a preoccupation with issues of will, morality, and efficacy: what is the right decision in a given context.

One response to the criticisms of rationality based on complexity and computation costs might be to see them as requiring simply an extension of the field of rational choice. The rational decisionmaker must decide not just what action to take, but how much scarce computational resources to devote to working out the consequences of actions, presumably balancing the costs and benefits of future computation at the margin. As we have argued, this easy fix is untenable, because it leads to an infinite regress: the problem of estimating the costs and benefits of computation at the margin is itself just as complex as the original problem, and just as intractable.

A second response would be to argue that there are at least some contexts where the complexity levels of the environment are low, so that individual agents can plausibly be assumed to bear the costs of complete computation. For example, some economists believe that competitive economic systems are well represented by Walrasian equilibrium systems that have a relatively simple dynamic behavior, for example, unique point attractors. One of the major thrusts of Albin's work is to show how quickly even very simple models of local interaction in fact develop highly complex dynamic behavior.

1.7.1 How complex are social systems?

This line of thinking raises an important research issue, which is to find methods to measure the actual complexity of real social systems. Albin's proposed complexity measures for social organizations are a first step in this direction.

As Albin argues, human social systems are composed of agents, human beings, each of whom has a complexity level at least as high as a general purpose computer, a Turing machine. As a result it is impossible to rule out a priori, similarly complex behavior on the part of the social system as a whole. Social theory and economic theory in particular, however, have developed without explicitly addressing the level of complexity of the social system. Some of the most apparently successful social theories, such as economic equilibrium theory, picture the social environment as having limited complexity.

Two possibilities suggest themselves to reconcile these points of view. First, something may mitigate the potential complexity of social interaction, leading to a social system that has lower complexity than its component subsystems. In economic equilibrium theory we see an embryonic

argument along these lines, resting on the effects of competition and weak global interactions to suppress potential complexity in the behavior of the economy as a whole. We need to understand these claims better, and to analyze them carefully with the tools of complexity theory. Albin's model of monetary policy, in which the central bank is represented as a neighbor of all the firms in the economy, shows one way to approach this issue. The results of his simulations are ambiguous, since he finds that the addition of the weak global interaction represented by the central bank can, depending on the complexity level of the economy without intervention, lead to any possible shift of complexity level, from type 4 to types 2 and 1, to be sure, but also in the other direction. Thus it seems unlikely that there is a general principle that any kind of weak global interaction will simplify complex interactions.

One difficulty in pursuing these investigations in economic theory is the absence in received economic equilibrium theory of a generally accepted account of how market prices emerge from the decentralized interactions of competitive economic agents. Economic theory has relied on parables like Walras's auctioneer, a centralized agent who is imagined to "cry out" trial price systems in order to discover the equilibrium, to cover over this lacuna.

A second possibility is that we as human beings have difficulty in perceiving the complexity level of real social and economic interactions. The world of social interactions might seem less complex than it is, or indeed than we ourselves are. Albin's reflections on the metalogic of economic prediction suggests why this might be so. The only way to grasp a system of type 4 complexity is to simulate it on another system of equal complexity. If the interactions of a large number of complex subsystems like human beings leads to a system of even higher complexity than ourselves, our own complexity level may not be high enough to represent the interacting system as a whole. Just as two-dimensional creatures would experience a third dimension only as an abstract conception, we may be capable of direct contact with the complexity of our own social existence only through abstractions like complexity theory itself.

In this kind of situation, where a subsystem of given complexity, a human being, confronts a system of higher complexity, the society, how will things work out? The individual agent will have to find some way of representing the social system with a less complex surrogate. There may, in fact, be many such surrogates, many forms of imperfect representation of the complex social system.

Social theory based on the rational choice paradigm seeks to explain agents' representations of their social world as more or less accurate reflections of its true nature. Agents in this perspective believe the society to work in a certain way because it actually does work that way. The rational

economic actor believes that market prices reach an equilibrium because they actually do, that democratic political regimes somehow make policy choices by aggregating the preferences of the citizens, and so forth. But if complexity intervenes to prevent agents in principle from representing the system to themselves faithfully, there is another level of social explanation that must be developed. We need an explicit theory of the ways in which agents achieve simplified representations of complex social interactions. This is the program of which studies of bounded rationality form a part; Albin's work contributes to its first tentative stages of development.

1.7.2 How smart do agents need to be?

To carry out the rational choice program, agents have to be smart enough to figure out the consequences of their actions in the relevant context. As Albin's work shows, in a wide variety of plausible social and economic contexts this requirement is far too demanding to be plausible. The complexity of the social environment reflects the complexity level of the agents who interact in it, creating an unresolvable self-referential paradox.

There is, however, reason to think that the rational choice program demands far too much of agents. The rational choice program commits an agent to expending arbitrarily large computational resources to achieve a better outcome, no matter how small the potential gain may be. Ordinary experience suggests that human beings, at least, sensibly reject this counsel of perfection. In many situations algorithms with relatively low computation cost achieve quite good, if not optimal results. An example is the trading algorithm in "Decentralized, dispersed exchange without an auctioneer" (Chapter 5), which requires quite limited computational resources, does not commit agents to pursuing elaborate chains of inferential reasoning, but achieves approximately efficient allocation at a moderate resource cost.

Furthermore, in many human social contexts optimization of an objective function may be less important than avoidance of disaster. There may be computationally simple strategies that lead to high rates of survival even in complex environments. In human contexts where imitation and learning play a major role in diffusing strategies, such low-cost, high-survival probability behavior may have a considerable evolutionary advantage.

Strict rationality, on the other hand, may be evolutionarily disadvantageous in some contexts. Certainly rational strategies that commit an agent to incurring extremely high computation costs impose a resource overhead burden that may offset the small gains in performance they secure. One way out of the paradoxes of self-reference created by the presupposition of rational behavior is to generalize Simon's concept of bounded rationality to

consider behavioral strategies primarily in terms of their survival and evolutionary effectiveness in relation to their computational cost, rather than in terms of their ability to approximate abstract standards of optimality.

The methodological criticism most often leveled at the program of bounded rationality is that of inherent indeterminism. The set of boundedly rational behaviors is huge, and there seems to be no systematic way to explore it, or to discover those parts of it most relevant to the explanation and prediction of particular social interactions. Rational behavior, on the other hand, appears to be generically unique, since there must be one best way to solve any given problem. Albin's work strongly challenges this methodological presumption by suggesting that in many relevant contexts the set of strictly rational behaviors is either empty or effectively beyond our knowledge because of the paradoxes of self-reference.

1.7.2.1 We are what we compute

In positive terms, the contemplation of the paradoxes of self-reference and unbounded computation cost suggests a reconceptualization of human beings as algorithms rather than preferences. Axel Leijonhufvud (1993) proposes to substitute *algorithmic man* [sic] as the protagonist of twenty-first century social theory in place of the eighteenth century's *homo economicus*.

What is characteristic of human beings as actors, in this perspective, is the computational resources and methods they deploy in decision making, rather than their pursuit of particular goals by rational means. The regularities to be discovered by social science, then, are reflections of these computational capacities rather than reflections of the rational action of the individuals. The First Welfare Theorem of economic theory, for example, shows that Walras's concept of market equilibrium logically entails Pareto's concept of efficient allocation (a situation in which it is impossible to improve any agent's state without harming some other agent). This is a paradigmatic example of a global social property derived from the presumed rationality of individual agents. From a computational point of view the analogous insights are those of Albin's local interaction multiperson Prisoners' Dilemma, in which the complexity level of the resulting social interaction reflects the algorithmic constitution of the individual agents.

1.8 Toward a robust theory of action and society

From a computational cost point of view, the rational choice paradigm's most serious weakness appears to be the lack of robustness of its predictions. Small changes in the environment or in the informational situation of

the rational actor tend to lead to large changes in rational actions, or radical alterations of basic patterns of behavior. In some situations even the formulation of a rational strategy becomes impossibly difficult in response to what appear to be relatively small changes in the posited environment.

Take, for example, the fundamental economic problem of an economic agent confronting a market. Traditional economic theory assumes that a competitive market appears to the agent as a set of established market prices at which she can exchange arbitrarily large amounts of commodities. The market prices establish a well-defined budget set of alternative combinations of commodities the agent can afford, and rational choice theory recommends that she choose that combination of commodities she most prefers among those in her budget set. Consider now the slightly altered representation of the market in Chapter 5, "Decentralized, dispersed exchange without an auctioneer." Agents still have well-defined preferences over commodity bundles, and still face the problem of organizing social exchanges when there are strong motives to trade due to differences in endowments. The agents, however, lack the convenience of a central information source, like Walras's auctioneer, to communicate uniform market prices to the market as a whole. The individual agents have to elicit this information by advertising and probing the willingness to trade in their local neighborhood.

One might suppose that this relatively small change in informational assumptions should not lead to drastic changes in the theory of rational agent behavior in the market. A robust theory would adapt smoothly to the reduction of centralized information and the rise in costs of discovering real trading opportunities. But it turns out to be very difficult even to conceptualize the fully rational behavior of an agent in the decentralized, dispersed market setting. Her final consumption plan is the result of a large number of individual trades strung out over time, each one made with slightly different information. Her own actions in advertising and announcing willingness to pay have potentially global consequences which it is impossible for her to calculate accurately. As a result it is very difficult to map her well-defined preferences for the commodities into a ranking of her available trading strategies. What appears to be a clear-cut, unambiguous application of the rational choice principle to the problem of market demand in the Walrasian setting dissolves into a maze of competing modeling strategies when we eliminate the central information.

The welfare economist or policy maker confronting this kind of decentralized market must consider possible interventions in the informational infrastructure defining the market interactions in addition to the classical economic tools of taxes and price control. The introduction of institutions that alter the flow of information in the market, and the provision of new technologies to implement this flow, can have powerful impacts on the

equity and efficiency of the final outcome.

A similar methodological problem arises in the repeated Prisoners' Dilemma. The two person one-shot Prisoners' Dilemma seems to pose no serious difficulty to the rational choice paradigm, since there is a unique Nash equilibrium in which both players choose dominant strategies. Even the elementary generalization of this model to the case of repeated plays poses serious problems for the rational choice approach, though it is possible to make some headway in defining Nash equilibria. When we take the further step, which appears modest and natural in commonsense terms, to Albin's local interaction multiperson Prisoners' Dilemma played on the lattice, however, the fundamental tasks suggested by rational choice theory, such as calculating the best response of an agent when the strategies of her neighbors are given, become computationally infeasible.

Historically rational choice theory developed in tandem with a particular method of analyzing economic behavior on markets. The tendency to represent markets as given systems of price that establish computationally trivial budget sets strongly reinforced the tendency to argue that human beings will act so as to maximize their utility given their constraints. In the abstract market setting the rational choice prescription is both feasible and plausible, and the combination of the two abstractions reinforce each other powerfully. In the other social sciences, including politics, sociology, and anthropology, rational choice methods have played a much less dominant role because the paradigmatic institutional settings—for example, elections and diplomatic negotiations, the establishment of social networks, and the evolution of human cultures—are inherently more complex. The rational choice paradigm can be imported into the other social sciences only at the cost of drastically simplifying the abstract representation of their characteristic institutional problems. Albin's application of complexity methods to economic scenarios has the important consequence of revealing the fragile interdependence of method and institutional abstraction on which traditional economic theory is based.

We are far from the point where we can propose anything like a complete computationally robust alternative to the rational choice program for economics or other social sciences. But Albin's reflections on the interplay of individual action and aggregate outcomes in complex economic interactions offers us some vital clues to the nature of such alternatives.

A more robust conception of human action in complex social contexts must explicitly address the computational resources available to and deployed by the agents. Recent advances in abstract modeling of language and computation provide the tools for at least a qualitative categorization of models in terms of their complexity levels. We ought to be able to make some progress in understanding social interactions where the complexity level of individual strategies is limited, say to the equivalent of regular lan-

guages or pocket calculators. These investigations will naturally lead to a systematic exploration of models with higher assumed complexity levels.

A robust alternative to the rational choice program must also be more explicitly context-dependent. The rational choice paradigm itself, of course, is already context-dependent, in that the conception of self-interest that operates changes from context to context: a single human is imagined to maximize quite different functions in the roles of corporate executive and mother, for example. But the rational choice program posits the pattern of explanation of action in terms of maximization subject to constraints to be universal. It seems likely, however, that in reality humans deploy very different modes of action in different contexts, and that these modes of behavior differ in their computational complexity. There is good reason to believe that the stock market, for example, is in effect a system of type 4 complexity, capable of universal computation (though it is not at all obvious what function it might be working out). But many participants in the market appear to model it at lower levels of complexity: as a periodic type 2 system in some cases, or as a chaotic type 3 system subject to statistical regularities in others.

We can also see the centrality of concepts of imitation, learning, adaptation, innovation and evolution to robust approaches to explaining human action. The rational choice program is rooted in seventeenth- and eighteenth-century conceptions of natural law, property rights, and individual autonomy. The notion of individual sovereignty is deeply bound up in its structure with the postulate that preferences are given, exogenous parameters in social theory. During the last half of the nineteenth century, and the first half of the twentieth century, while other life and human sciences were revolutionized to incorporate evolutionary ideas, economics turned increasingly to an axiomatization of rationality that is indifferent if not hostile to evolutionary ideas. But one promising path to discovering a knowable orderliness in human behavior that is context-dependent and computationally bounded is to ask whether and when the social environment favors some behaviors over others and why. Evolutionary models, in which the distribution of strategies in a population change in response to their success, and in which mutation and innovation lead to new behaviors, are a logical way to attack these issues.

Peter Albin's foray into the intersection of economics and complexity theory reveals puzzling and perhaps even insurmountable problems for the rational choice paradigm. But the rational choice paradigm is not the only theoretical path to a scientific understanding of human behavior. Albin's work also provides crucial hints to the construction of fertile alternative explanatory methods.

The role and significance of complexity theory in the physical and biological sciences are at the moment highly controversial subjects. Critics

question whether the methods and insights founded on simple and highly abstract models like cellular automata can be applied successfully to real-world systems and problems. The same questions and doubts arise legitimately in considering the application of complexity methods to social and economic interactions. There is a long step between exhibiting the possibility of interesting regularities or categorization of behaviors in abstract models and the explanation or prediction of real phenomena. But the essays of Peter Albin in this book serve economic and social theory discourse unambiguously at another level as well. In clarifying the role of implicit computational complexity assumptions in received social theory, this work poses fundamental and inescapable methodological questions to theories based on rational choice axioms.