Chapter 3

We the Expected

What raw day first saw life, raw itself, pregnant with the future? Four billion years, thereabouts, from the first circle of metabolic witchery to me and to thee. Raw chance? Raw improbability that ought never have occurred in billions of times the history of this universe? Raw meaninglessness that we are so very unexplained?

Is life really the unthinkable accident that follows from the calculations of Fred Hoyle and N. C. Wickramasinghe? Is time the hero of the plot, as George Wald argued? Yet we now believe there were but 300 million years or so from the cooling of the crust to clear evidence of cellular life, not the 2 billion years that Wald appealed to. Time was not there in sufficient vastness for Wald's story, and surely not for Hoyle and Wickramasinghe's tale. If we the living are wildly improbable, then we are unaccountable mysteries in the span of space and time. But if this view is wrong, if there is some reason to believe that life is probable, then we are not mysteries in the exploding cosmos, we are natural parts of it.

Most of my colleagues believe that life emerged simple and became complex. They picture nude RNA molecules replicating and replicating and eventually stumbling on and assembling all the complicated chemical machinery we find in a living cell. Most of my colleagues also believe that life is utterly dependent on the molecular logic of template replication, the A–T, G–C Watson–Crick pairing that I wrote about in Chapter 2. I hold a renegade view: life is not shackled to the magic of template replication, but based on a deeper logic. I hope to persuade you that life is a natural property of complex chemical systems, that when the number of different kinds of molecules in a chemical soup passes a certain threshold, a self-sustaining network of reactions—an autocatalytic metabolism—will suddenly appear. Life emerged, I suggest, not simple,

but complex and whole, and has remained complex and whole ever since—not because of a mysterious élan vital, but thanks to the simple, profound transformation of dead molecules into an organization by which each molecule's formation is catalyzed by some other molecule in the organization. The secret of life, the wellspring of reproduction, is not to be found in the beauty of Watson–Crick pairing, but in the achievement of collective catalytic closure. The roots are deeper than the double helix and are based in chemistry itself. So, in another sense, life—complex, whole, emergent—is simple after all, a natural outgrowth of the world in which we live.

The claim that life emerges as a natural phase transition in complex chemical systems is so radical a departure from past theories that I owe you caveats. Do we know that such a view is at least theoretically coherent? Do we know it to be physically and chemically possible? Is there evidence for such a view? Is evidence attainable? Do we know that life began as I shall suggest it did? The most that can be said at this stage is that good, careful theoretical work strongly supports the possibility I shall present. That work appears to be consistent with what we know about complex chemical systems. Scant experimental evidence supports this view as yet, but stunning developments in molecular biology now make it possible to imagine actually creating these self-reproducing molecular systems—synthesized life. I believe that this will be accomplished within a decade or two.

The Networks of Life

As noted in Chapter 2, most researchers are focusing their attention on the capacity of RNA, or RNA-like polymers, to self-reproduce by template replication. The attention is understandable. No one looking at the beautiful double helix of DNA or RNA and regarding the Watson–Crick pairing rules can avoid being struck by the beauty of nature's apparent choice. The fact that Leslie Orgel and his colleagues have not yet succeeded in getting such polymers to replicate without an enzyme does not mean that the efforts will always fail. Orgel has been at it for perhaps 25 years; nature took something like 100 million years. Orgel is very smart, but 100 million years is long enough, measured in three-year National Institutes of Health grants, to try lots of possibilities. Let us try a different tack. Suppose that the laws of chemistry were slightly different, that nitrogen had four rather than five valence electrons, say, allowing four rather than five bonding partners. Ignore the wrench this would throw into quantum mechanics—one can sometimes get away

with being wretched to quantum mechanics when making a philosophical point. If the laws of chemistry were slightly different so that the beautiful double-helix structure of DNA and RNA were no longer possible, would life based on chemistry be impossible? I do not want to think that we were quite so lucky. I hope we can find a basis for life that lies deeper than template self-complementarity.

The secret, I believe, lies in what chemists call catalysis. Many chemical reactions proceed only with great difficulty. Given a long expanse of time, a few molecules of A might combine with molecules of B to make C. But in the presence of a catalyst, another molecule we'll call D, the reaction catches fire and proceeds very much faster. The usual metaphor is the lock and key: A and B fit into slots on D, in just such a way that they are far more likely to combine to form C. As we shall see, this is a vast oversimplification, but for now it will suffice to get the point across. While D is the catalyst that joins A and B to make C, the molecules A, B, and C might themselves act as catalysts for other reactions.

At its heart, a living organism is a system of chemicals that has the capacity to catalyze is own reproduction. Catalysts such as enzymes speed up chemical reactions that might otherwise occur, but only extremely slowly. What I call a collectively autocatalytic system is one in which the molecules speed up the very reactions by which they themselves are formed: A makes B; B makes C; C makes A again. Now imagine a whole network of these self-propelling loops (Figure 3.1). Given a supply of

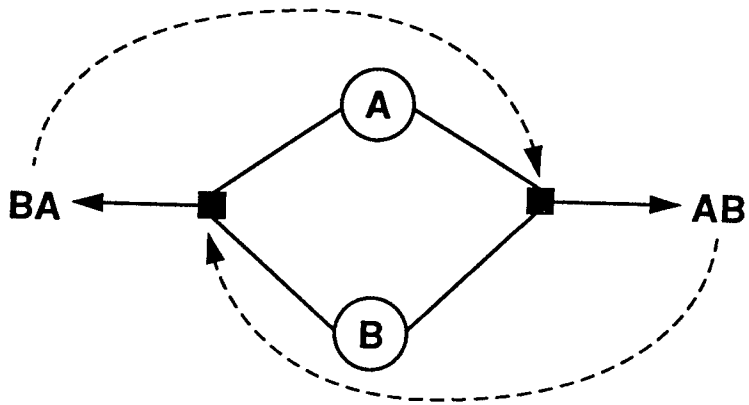

Figure 3.1 *A simple autocatalytic set. Two dimer molecules, AB and BA, are formed from two simple monomers, A and B. Since AB and BA catalyze the very reactions that join As and Bs to make the dimers, the network is autocatalytic: given a supply of "food" molecules (As and Bs), it will sustain itself.*

food molecules, the network will be able to constantly re-create itself. Like the metabolic networks that inhabit every living cell, it will be alive. What I aim to show is that if a sufficiently diverse mix of molecules accumulates somewhere, the chances that an autocatalytic system—a self-maintaining and self-reproducing metabolism—will spring forth becomes a near certainty. If so, then the emergence of life may have been much easier than we have supposed.

What I aim to show is simple, but radical. I hold that life, at its root, does not depend on the magic of Watson–Crick base pairing or any other specific template-replicating machinery. Life, at its root, lies in the property of *catalytic closure* among a collection of molecular species. Alone, each molecular species is dead. Jointly, once catalytic closure among them is achieved, the collective system of molecules is alive.

Each cell in your body, every free-living cell, is collectively autocatalytic. No DNA molecules replicate nude in free-living organisms. DNA replicates only as part of a complex, collectively autocatalytic network of reactions and enzymes in cells. No RNA molecules replicate themselves. The cell is a whole, mysterious in its origins perhaps, but not mystical. Except for "food molecules," every molecular species of which a cell is constructed is created by catalysis of reactions, and the catalysis is itself carried out by catalysts created by the cell. To understand the origin of life, I claim, we must understand the conditions that enabled the first emergence of such autocatalytic molecular systems.

Catalysis alone, however, is not sufficient for life. All living systems "eat": they take in matter and energy in order to reproduce themselves. This means that they are what is referred to in Chapter 1 as open thermodynamic systems.

In contrast, closed thermodynamic systems take in no matter or energy from their environments. A great deal is understood about the behavior of closed thermodynamic systems. The theorists of thermodynamics and statistical mechanics have studied such systems for over 100 years. In contrast, remarkably little is understood about the possible behaviors of open thermodynamic systems. Not so surprising, this ignorance. The vast flowering of all life-forms over the past 3.45 billion years is merely a hint of the possible behaviors of open thermodynamic systems. So too is cosmogenesis itself, for the evolving universe since the Big Bang has yielded the formation of galactic and supragalactic structures on enormous scales. Those stellar structures and the nuclear processes within stars, which have generated the atoms and molecules from which life itself arose, are open systems, driven by nonequilibrium processes. We have only begun to understand the awesome creative powers of nonequilibrium processes in the unfolding universe. We are

all—complex atoms, Jupiter, spiral galaxies, warthog, and frog—the logical progeny of that creative power.

Since I hope to persuade you that life is the natural accomplishment of catalysts in sufficiently complex nonequilibrium chemical systems, I had best take a moment to sketch what catalysts accomplish and how equilibrium and nonequilibrium chemical systems behave. Chemical reactions occur spontaneously, some rapidly, some slowly. Typically, chemical reactions are more or less reversible: A transforms to B, but B transforms to A. Since such reactions are reversible, it is easy to think about what would occur in a beaker that began with an initial concentration of A molecules and no B molecules and that was closed to the addition of matter or energy. The A molecules would begin to convert to B molecules, but as that occurred, the new B molecules would begin to convert back to A molecules. Starting with only A molecules, the B concentration would build up to the point at which the rate of conversion of A to B was exactly equal to the rate of conversion of B to A. This balance is called chemical equilibrium. At chemical equilibrium, the net concentrations of A and B do not change over time, but any given A molecule may convert to B and back again thousands of times per minute. Of course, the equilibrium is statistical. Minor fluctuations in A and B concentrations occur all the time.

Chemical equilibrium is not limited to a pair of molecules, A and B, but will occur in any closed thermodynamic system. If the system has hundreds of different types of molecules, it will ultimately settle down to an equilibrium in which the forward and reverse reactions between any pair of molecules balance out.

Catalysts, of which protein enzymes and ribozymes are examples, can speed up both the forward and the reverse reaction by the same amount. The equilibrium between A and B is not altered; enzymes simply hasten the rate at which this state of balance is reached. Suppose, at equilibrium, the ratio of A and B concentrations is 1, so the concentrations of the two are equal. If the chemical system starts out displaced from equilibrium—say, with a high concentration of B and almost no A—then the enzyme will vastly shorten the time it takes to reach the equilibrium ratio where the two concentrations are equal. In effect, then, the enzyme increases the rate of production of A.

How does catalysis happen? There is an intermediate state between A and B, called the transition state, in which one or more bonds among the atoms of the molecule are severely strained and distorted. The transition-state molecule is therefore rather unhappy. The measure of this unhappiness is given by the energy of the molecule. Low energy corresponds to unstrained molecules. High energy corresponds to strained

molecules. Think of a spring. At its rest length, it is happy. If stretched beyond its rest length, it has stored energy—it is unhappy—and can release that energy by snapping back to its rest length, whereupon it has low energy again.

Not surprisingly, the transition state passing from A to B is exactly the same as the transition state passing from B back to A. Enzymes are thought to work by binding to the transition state and stabilizing it. This makes it easier for both A and B molecules to jump to the transition state, increasing the rate of conversion of A to B, and of B to A. Thus an enzyme increases the rate at which the equilibrium ratio of A and B concentrations is approached.

We should be thankful that our cells are not at chemical equilibrium; for a living system, equilibrium corresponds to death. Living systems are, instead, open thermodynamic systems persistently displaced from chemical equilibrium. We eat and excrete, as did our remote ancestors. Energy and matter flow through us, building up the complex molecules that are the tokens in the game of life.

Open nonequilibrium systems obey very different rules from those of closed systems. Consider a simple case: we have a beaker into which we add A molecules continuously from some outside source at a constant rate, and we take any B molecules out of the beaker at a rate proportional to the concentration of B. A will convert to B and B will convert to A as before, but the two molecules can never reach the equilibrium balance they attained before because of the constant addition of A and the removal of B. Common sense says that the system will settle down to a steady state at which the ratio of A molecules to B molecules is higher than it was when the system was closed. In short, the ratio of A to B will be tipped from the thermodynamic equilibrium ratio. In general, this commonsense view is correct. In simple cases, such systems, open to the flux of matter and energy, settle down to a steady state different from that found in closed thermodynamic systems.

Now consider a vastly more complex open system, the living cell. The cells of your body coordinate the behaviors of about 100,000 different kinds of molecules as matter and energy cross their boundaries. Even bacteria coordinate the activities of thousands of different kinds of molecules. To think that understanding the behavior of very simple open thermodynamic chemical systems takes us far toward understanding the cell is hubris. No one understands how the complex cellular networks of chemical reactions and their catalysts behave, or what laws might govern their behavior. Indeed, this is a mystery we will begin to discuss in the next chapter. Yet simple open thermodynamic systems are at least a start and are already fascinating on their own. Even simple nonequilibrium chemical systems can form remarkably complex patterns of

chemical concentrations varying in time and space in striking ways. As noted in Chapter 1, Ilya Prigogine called these systems dissipative because they persistently dissipate matter and energy in order to maintain their structures.

Unlike the simple steady-state system in the thermodynamically open beaker, the concentrations of the chemical species in a more complex dissipative system may not fall to a steady state, unchanging in time. Instead, the concentrations can start to oscillate up and down in repeated cycles, called limit cycles, which are sustained for long periods of time. Such systems can also generate remarkable spatial patterns. For example, the famous Belosov–Zhabotinski reaction, made of some simple organic molecules, sets up two kinds of spatial patterns. In the first pattern, spreading concentric circular waves of blue propagate outward over an orange background from a central oscillating source. The blue and orange colors arise because of indicator molecules that track how acidic or basic the reaction mixture is at any point in space. In the second pattern, spiral pinwheels of blue on orange cartwheel about a center (Figure 3.2). Such patterns have been studied by a number of researchers. A fine book by my friend Arthur Winfree, *When Time Breaks Down: The Three-Dimensional Dynamics of Electrochemical Waves and Cardiac Arrhythmias,* summarizes much of the work. Among the most immediate human implications is this: the heart is an open system, and it can beat according to patterns analogous to the Belosov–Zhabotinski

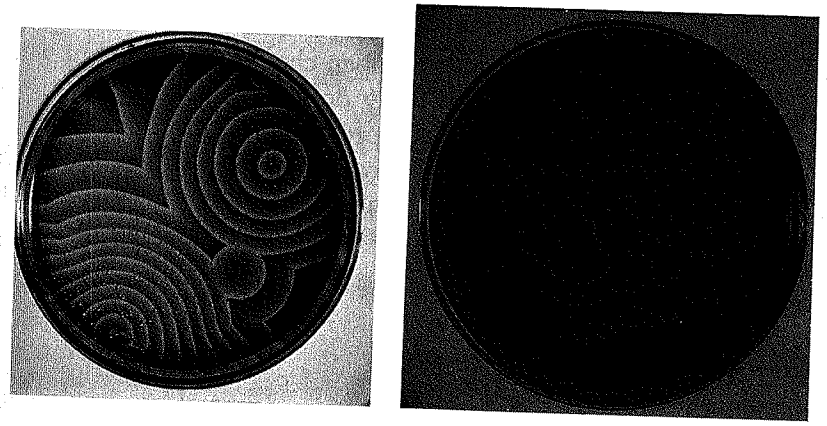

Figure 3.2 *Self-organization at work. The famous Belosov–Zhabotinski reaction showing the spontaneous emergence of order in a simple chemical system.* (a) *Concentric circular waves propagate outward.* (b) *Radially expanding pinwheels cartwheel about a center. (From Winfree, 1987.)*

reaction. Sudden death caused by cardiac arrhythmias may correspond to a switch from the analogue of the concentric-circles pattern (a steady beating) to the spiral-pinwheels pattern in your myocardium. The blue propagating wave can be thought of as corresponding to the chemical conditions in muscle cells that lead them to contract. Thus the concentric spreading pattern of the evenly spaced blue circles corresponds to ordered contraction waves. But in the spiral pattern, the blue pinwheels are very close together near the center of the spiral and are spaced farther apart the farther out on the spiral they go. This pattern corresponds to chaotic twitching of the heart muscle in the vicinity of the spiral center. Winfree has shown that simple perturbations, such as shaking the petri plate that holds the chemical reactants of the Belosov–Zhabotinski reaction, can switch the system from the concentric to the spiral pattern. Thus Winfree has suggested that simple perturbations can switch a normal heart to the spiral chaotic pattern and lead to sudden death.

The relatively simple behaviors of nonequilibrium chemical systems are well studied and may have a variety of biological implications. For example, such systems can form a standing pattern of stripes of high chemical concentrations spaced between stripes of low chemical concentrations. Many of us think that the natural patterns such systems form have a great deal to tell us about the spatial patterning that occurs in the development of plants and animals. The blue and orange stripes in the Belosov–Zhabotinski reaction may foretell the stripes of the zebra, the banding patterns on shells, and other aspects of morphology in simple and complex organisms.

However intriguing such chemical patterns may be, they are not yet living systems. The cell is not only an open chemical system, but a collectively autocatalytic system. Not only do chemical patterns arise in cells, but cells sustain themselves as reproducing entities that are capable of Darwinian evolution. By what laws, what deep principles, might autocatalytic systems have emerged on the primal earth? We seek, in short, our creation myth.

A Chemical Creation Myth

Scientists often gain insight into a more complex problem by thinking through a simpler toy problem. The toy problem I want to tell you about concerns "random graphs." A random graph is a set of dots, or nodes, connected at random by a set of lines, or edges. Figure 3.3 shows an example. To make the toy problem concrete, we can call the dots

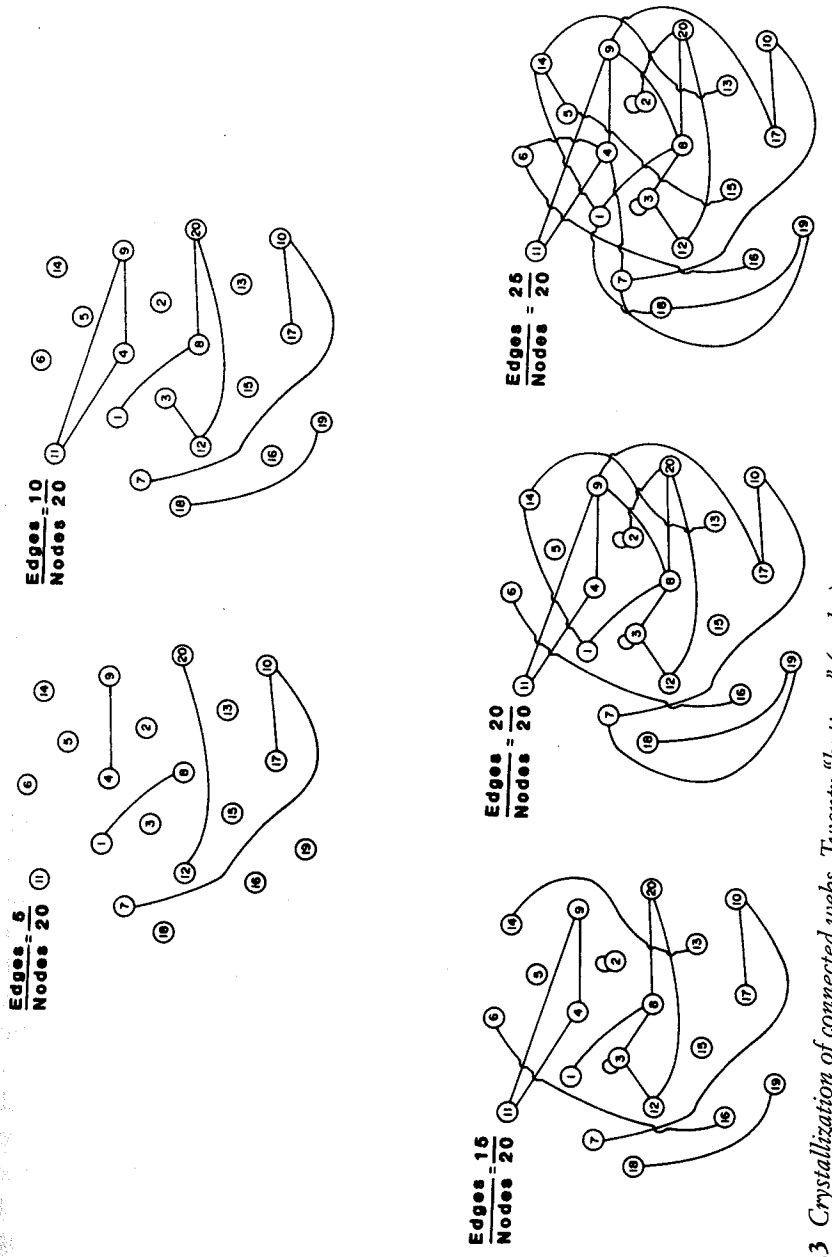

Figure 3.3 *Crystallization of connected webs. Twenty "buttons" (nodes) are connected at random by an increasing number of "threads" (edges). For large numbers of buttons, as the ratio of threads to buttons increases past a threshold of 0.5, most points become connected in one giant component. As the ratio passes 1.0, closed pathways of all lengths begin to emerge.*

"buttons" and the lines "threads." Imagine 10,000 buttons scattered on a hardwood floor. Randomly choose two buttons and connect them with a thread. Now put this pair down and randomly choose two more buttons, pick them up, and connect them with a thread. As you continue to do this, at first you will almost certainly pick up buttons that you have not picked up before. After a while, however, you are more likely to pick at random a pair of buttons and find that you have already chosen one of the pair. So when you tie a thread between the two newly chosen buttons, you will find three buttons tied together. In short, as you continue to choose random pairs of buttons to connect with a thread, after a while the buttons start becoming interconnected into larger clusters. This is shown in Figure 3.3*a*, which is limited to 20 rather than 10,000 buttons. Every now and then, lift up a button and see how many other buttons you pick up. The connected cluster is called a component in our random graph. As Figure 3.3*a* shows, some buttons may not be connected to any other buttons. Other buttons might be connected in pairs or triples or larger numbers.

The important features of random graphs show very regular statistical behavior as one tunes the ratio of threads to buttons. In particular, a *phase transition* occurs when the ratio of threads to buttons passes 0.5. At that point, a "giant cluster" suddenly forms. Figure 3.3 shows this process, using only 20 buttons. When there are very few threads compared with the number of buttons, most buttons will be unconnected (Figure 3.3*a*), but as the ratio of threads to buttons increases, small connected clusters begin to form. As the ratio of threads to buttons continues to increase, the size of these clusters of buttons tends to grow. Obviously, as clusters get larger, they begin to become cross-connected. Now the magic! As the ratio of threads to buttons passes the 0.5 mark, all of a sudden most of the clusters have become cross-connected into one giant structure. In the small system with 20 buttons in Figure 3.3, you can see this giant cluster forming when the ratio of threads to buttons is half, 10 threads to 20 buttons. If we used 10,000 buttons, the giant component would arise when there were about 5,000 threads. When the giant component forms, most of the nodes are directly or indirectly connected. If you pick up one button, the chances are high that you will pull up something like 8,000 of the 10,000 buttons. As the ratio of threads to buttons continues to increase past the halfway mark, more and more of the remaining isolated buttons and small clusters become cross-connected into the giant component. So the giant component grows larger, but its rate of growth slows as the number of remaining isolated buttons and isolated small components decreases.

The rather sudden change in the size of the largest connected cluster of buttons, as the ratio of threads to buttons passes 0.5, is a toy version of the phase transition that I believe led to the origin of life. In Figure 3.4, I show qualitatively the size of the largest cluster among 400 nodes as the ratio of edges to nodes increases. Note that the curve is S-shaped, or sigmoidal. The size of the largest cluster of nodes increases slowly at first, then rapidly, then slows again as the ratio of edge to nodes increases. The rapid increase is the signature of something like a phase transition (Figure 3.4). In the example in Figure 3.4 using 400 buttons, the sigmoidal curve rises steeply when the ratio of edges to nodes passes 0.5. The steepness of the curve at the critical 0.5 ratio depends on the number of nodes in the system. When the number of nodes is small, the steepest part of the curve is "shallow," but as the number of nodes in the toy system increases—from, say, 400 to 100 million—the steep part of the sigmoidal curve becomes more vertical. Were there an infinite number of buttons, then as the ratio of threads to buttons passed 0.5 the size of the largest component would jump discontinuously from tiny to enormous. This is a phase transition, rather like separate water molecules freezing into a block of ice.

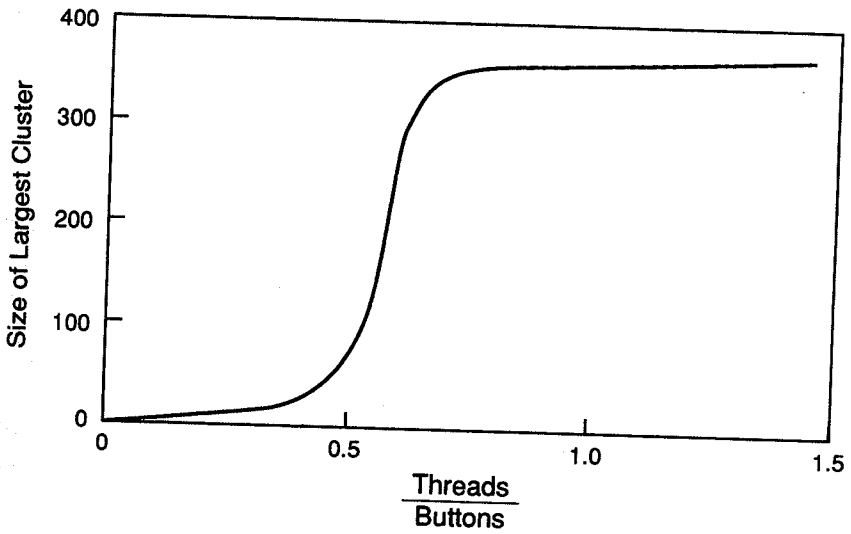

Figure 3.4 *A phase transition. As the ratio of threads (edges) to buttons (nodes) in a random graph passes 0.5, the size of the connected cluster slowly increases until it reaches a "phase transition" and a giant component crystallizes. (For this experiment, the number of threads ranges from 0 to 600, while the number of buttons is fixed at 400.)*

The intuition I want you to take away from this toy problem is simple: as the ratio of threads to buttons increases, suddenly so many buttons are connected that a vast web of buttons forms in the system. This giant component is not mysterious; its emergence is the natural, expected property of a random graph. The analogue in the origin-of-life theory will be that when a large enough number of reactions are catalyzed in a chemical reaction system, a vast web of catalyzed reactions will suddenly crystallize. Such a web, it turns out, is almost certainly autocatalytic—almost certainly self-sustaining, alive.

Reaction Networks

It is convenient to draw a metabolic reaction graph with circles representing chemicals and square representing reactions. To be concrete, we will consider four simple kinds of reactions. In the simplest, one substrate, A, converts to one product, B. Since reactions are reversible, B also converts back to A. This is a one-substrate, one-product reaction. Draw a black line leaving A and entering a small square lying between A and B, and draw a line leaving the square and ending on B (Figure 3.5). This line and the square represent the reaction between A and B. Now consider two molecules, say A and B, that are combined, or "ligated," to form a larger molecule, C. In the reverse reaction, C is "cleaved" to form A and B. We can represent these reactions with two lines leaving A and B and entering a square representing this reaction, plus a line leaving the square and entering C. Finally, we should consider reactions with two substrates and two products. Typically, this kind of reaction occurs by breaking off a small cluster of atoms from one substrate and bonding the cluster to one or more atoms on the second substrate. We can represent two-substrate, two-product reactions with pairs of lines leaving the two substrates and entering a square representing that reaction, and two more lines leaving the square and connecting to the two products. Now consider all the kinds of molecules and reactions possible in a chemical reaction system. The collection of all such lines and squares between all the chemical circles constitutes the reaction graph (Figure 3.5).

Since we want to understand the emergence of collectively autocatalytic molecular systems, the next step is to distinguish between spontaneous reactions, which are assumed to occur very slowly, and catalyzed reactions, which are assumed to occur rapidly. We want to find the conditions under which the same molecules will be catalysts for and products of the reactions creating the autocatalytic set. This depends on the

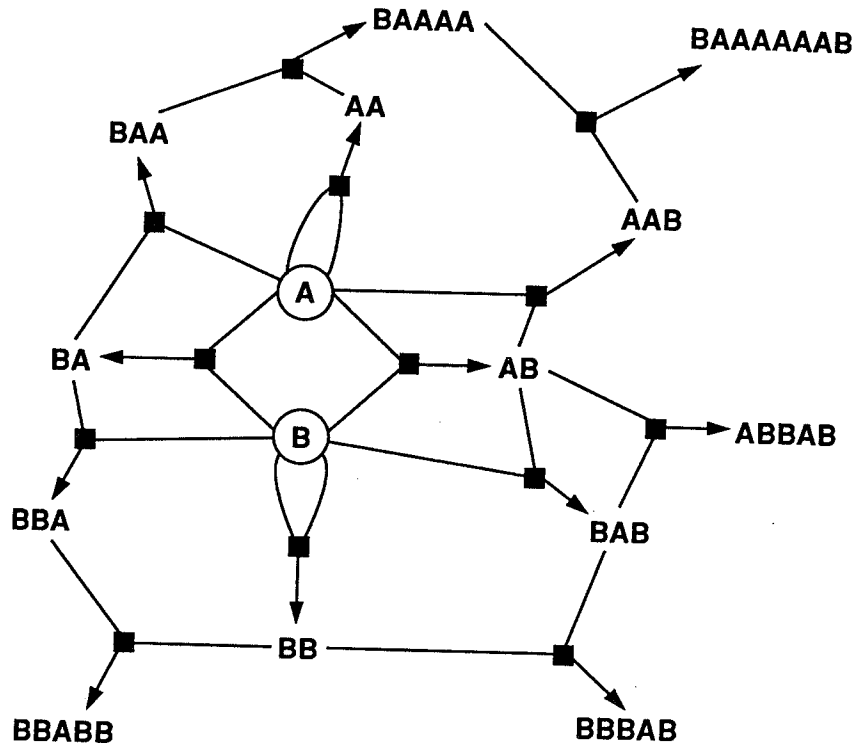

Figure 3.5 *From buttons and threads to chemicals. In this hypothetical network of chemical reactions, called a reaction graph, smaller molecules (A and B) are combined to form larger molecules (AA, AB, etc.), which are combined to form still larger molecules (BAB, BBA, BABB, etc.). Simultaneously, these longer molecules are broken down into simple substrates again. For each reaction, a line leads from the two substrates to a square denoting the reaction; an arrow leads from the reaction square to the product. (Since reactions are reversible, the use of arrows is meant to distinguish substrates from products in only one direction of the chemical flow.) Since the products of some reactions are substrates of further reactions, the result is a web of interlinked reactions.*

possibility that each molecule in the system can play a double role: it can serve as an ingredient or a product of a reaction, but it can also serve as a catalyst for another reaction. This dual role, as ingredient or catalyst, is perfectly possible, even familiar. Proteins and RNA molecules are known to play such a dual role. An enzyme called trypsin cleaves proteins you eat into smaller fragments. In fact, trypsin will cleave itself into fragments as well. And, as noted in Chapter 2, ribozymes are RNA molecules that can act as enzymes on RNA molecules. It is perfectly familiar that all kinds of organic molecules can be

substrates and products of reactions, but simultaneously act catalytically to hasten other reactions. No mystery stands in the way of a dual role for chemicals.

To proceed further, we need to know which molecules catalyze which reactions. If we knew this, we could tell whether any set of molecules might be collectively autocatalytic. Unfortunately, this knowledge is not, in general, yet available, but we can sensibly proceed by making plausible assumptions. I will consider two such simple theories, each of which allows us, in the model worlds we will consider, to assign, somewhat arbitrarily, catalysts to reactions. You should be skeptical about this maneuver. Surely, it might be thought, one must actually know which molecules catalyze which reactions to be certain that a set of molecules harbors an autocatalytic set. Such skepticism is well placed and allows me to introduce a mode of reasoning on which I am depending. One might easily object that if in the real world of chemical reactions the laws of chemistry dictated a somewhat different distribution of which molecules catalyzed which reactions, then the conclusions would not hold. My response is this: if we can show that for many alternative "hypothetical" chemistries, in which different molecules catalyze different reactions, autocatalytic sets emerge, then the particular details of the chemistry may not matter. We will be showing that the spontaneous emergence of self-sustaining webs is so natural and robust that it is even deeper than the specific chemistry that happens to exist on earth; it is rooted in mathematics itself.

Picture, as noted earlier, a reaction between a pair of molecules, A and B, as black lines or edges connecting A and B to the reaction square between them. Now picture some other molecule, C, that is able to catalyze the reaction between A and B. Represent this by drawing a blue arrow with its tail in C and its head on the reaction square between A and B (Figure 3.6). Represent the fact that the reaction between A and B is catalyzed by changing the black line between A and B to a red line. Consider each molecule in the system, and ask which reaction or reactions, if any, it can catalyze. For any such catalyst, draw a blue arrow to the corresponding reaction square, and color the corresponding reaction edges red. When you have finished this task, the red edges and the chemical nodes they connect represent all the catalyzed reactions, and collectively make up the *catalyzed reaction subgraph* of the whole reaction graph. The blue arrows and the chemical nodes from which they leave represent the molecules that carry out the catalysis (Figure 3.6).

Now consider what is required for the system to contain an autocatalytic subset: first, a set of molecules must be connected by red catalyzed reactions; second, the molecules in this set must each have its

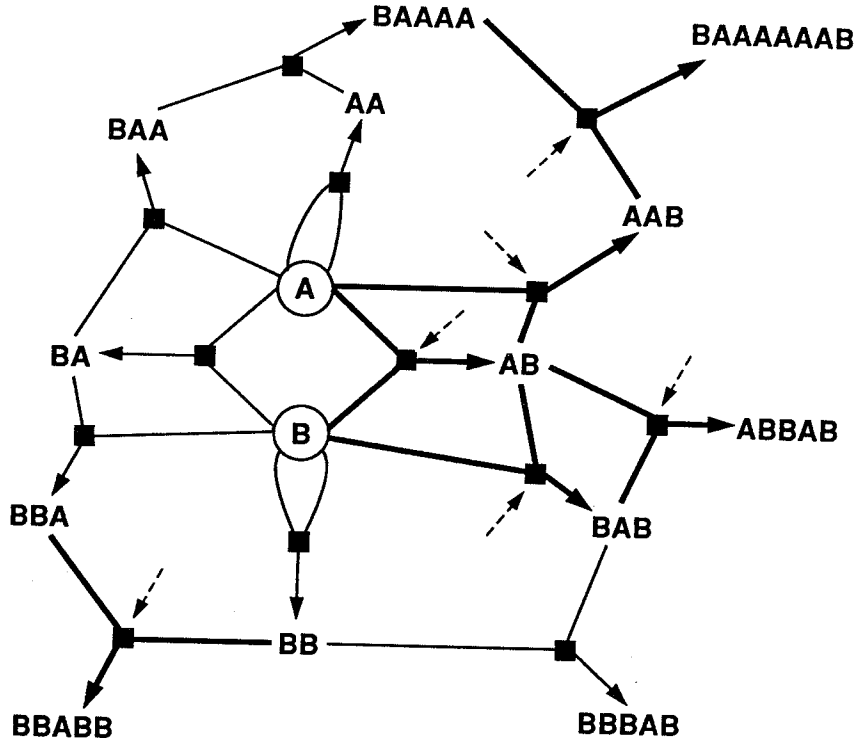

Figure 3.6 *Molecules catalyzing reactions. In Figure 3.5, all the reactions were assumed to be spontaneous. What happens when we add catalysts to speed some of the reactions? Here the reaction squares indicated by dashed-line arrows are catalyzed, and the heavy, darker lines connect substrates and products whose reactions are catalyzed. The result is a pattern of heavy lines indicating a catalyzed subgraph of the reaction graph.*

formation catalyzed by a blue arrow from some molecule in the same set or be added from outside. Call the latter molecules food molecules. If these conditions are met, we have a network of molecules that can catalyze its own formation, creating all the catalysts it needs.

The Central Idea

How likely is it that such a self-sustaining web of reactions would arise naturally? Is the emergence of collective autocatalysis easy or virtually impossible? Do we have to pick our chemicals carefully, or would just about any mixture do? The answer is heartening. The emergence of autocatalytic sets is almost inevitable.

Here, in a nutshell we will unpack later, is what happens: as the diversity of molecules in our system increases, the ratio of reactions to chemicals, or edges to nodes, becomes ever higher. In other words, the reaction graph has ever more lines connecting the chemical dots. The molecules in the system are themselves candidates to be able to catalyze the reactions by which the molecules themselves are formed. As the ratio of reactions to chemicals increases, the number of reactions that are catalyzed by the molecules in the system increases. When the number of catalyzed reactions is about equal to the number of chemical dots, a giant catalyzed reaction web forms, and a collectively autocatalytic system snaps into existence. A living metabolism crystallizes. Life emerges as a phase transition.

Now we will unpack our nutshell.

The first step is to show that as the diversity and complexity of the molecules in our system increase, the ratio of reactions to chemical dots in the reaction graph increases as well. It is easy to see why this is true. Consider a polymer consisting of four "monomers," which we can think of as atoms ABBB. Clearly, the polymer can be formed by gluing A to BBB, by gluing AB to BB, or by gluing ABB to B. So it can be formed in three ways, by three different reactions. If we increase the length of the polymer by one atom, the number of reactions per molecule will rise. ABBBA can be formed from A and BBBA, AB and BBA, ABB and BA, or ABBB and A. Since a polymer of length L has $L - 1$ internal bonds in general, a polymer of length L can be formed from smaller polymers in $L - 1$ ways. But these numbers account for only what chemists call ligation reactions, building up molecules from smaller pieces. Molecules can also be formed through cleavage. ABBB can be formed by lopping the A from the right-hand side of ABBBA. So it is rather obvious that there are more reactions by which molecules can be formed than there are molecules themselves. This means that in the reaction graph there are more lines than dots.

What happens to the ratio of reactions to molecules in the reaction graph as the diversity and complexity of those molecules increase? After some simple algebra, it is easy to show for simple linear polymers that as the length of the molecules increases, the number of kinds of molecules increases exponentially, but the number of reactions by which they convert from one to another rises even faster. This increasing ratio means that as more complex and diverse sets of molecules are considered, the reaction graph among them becomes ever denser with paths by which they can change from one into another. The ratio of reaction "lines" to dots becomes denser, a black forest of possibilities. The chemical system becomes ever more fecund with reactions by which molecules transform into other molecules.

At this point we have a flask of slow, spontaneous reactions. For the system to catch fire and generate self-sustaining autocatalytic networks, some of the molecules must act as catalysts, speeding up the reactions. The system is fecund, but not yet pregnant with life, and will not become so until we have a way to determine which molecules catalyze which reactions. Thus it is time to build some simple models. The simplest, which will do very well for a variety of purposes, is to assume that each polymer has a fixed chance, say one in a million, of being able to function as an enzyme to catalyze any given reaction. In using this simple model, we will "decide" which reactions, if any, each polymer can catalyze by flipping a biased coin that comes up heads once in a million times. Using this rule, any polymer will be randomly assigned, once and for all, the reactions it can catalyze. Using this "random catalyst" rule, we can "color" the catalyzed reactions red, draw our blue arrows from the catalysts to the reactions each catalyzes, and then ask whether our model chemical system contains a collectively autocatalytic set: a network of molecules connected by red lines and also containing the very molecules that catalyze, via the blue arrows, the reactions by which the molecules themselves are formed.

A somewhat more chemically plausible model supposes that our polymers are RNA sequences and introduces template matching. In this simplified version, Bs fit with As in a kind of Watson–Crick pairing. Thus the hexamer BBBBBB might be able to act like a ribozyme and bind two substrates, BABAAA and AAABBABA, by their two corresponding AAA trimer sites, and catalyze the ligation of the two substrates to form BABAAAAAABBABA. To make things even more chemically realistic, we might also demand that even if a candidate ribozyme has a site that matches the left and right ends of its substrates, it still has only one chance in a million to have other chemical properties that allow it to catalyze the reaction. This captures the idea that other chemical features beyond template matching may be required to achieve ribozyme catalysis. Let us call this the match catalyst rule.

Here is the crucial result: no matter which of these "catalyst" rules we use, when the set of model molecules reaches a critical diversity, a giant "red" component of catalyzed reactions crystallizes, and so collectively autocatalytic sets emerge. Now it is easy to see why this emergence is virtually inevitable. Suppose we use the random catalyst rule and assume that any polymer has a one-in-a-million chance to act as an enzyme for any given reaction. As the diversity of molecules in the model system increases, the ratio of reactions to molecules increases. When the diversity of molecules is high enough, the ratio of reactions to polymers reaches a million to one. At that diversity, on average each polymer will catalyze one reaction. A million to one, [illegible]

in a million equals one. When the ratio of catalyzed reactions to chemicals is 1.0, then with extremely high probability a "red" giant component, a web of catalyzed reactions, will form—a collectively autocatalytic set of molecules.

In this view of the origin of life, a critical diversity of molecules must be reached for the system to catch fire, for catalytic closure to be attained. A simple system with 10 polymers in it and a chance of catalysis of one in a million is just a set of dead molecules. Almost certainly, none of the 10 molecules catalyzes any of the possible reactions among the 10 molecules. Nothing happens in the inert soup save the very slow spontaneous chemical reactions. Increase the diversity and atomic complexity of the molecules, and more and more of the reactions among them become catalyzed by members of the system itself. As a threshold diversity is crossed, a giant web of catalyzed reactions crystallizes in a phase transition. The catalyzed reaction subgraph goes from having many disconnected tiny components to having a giant component and some smaller, isolated components. Your intuitions may now be tuned enough to guess that the giant component will contain a collectively autocatalytic subset able to form itself by catalyzed reactions from a supply of food molecules.

I have now related the central ideas about how I think life may have formed. These ideas are really very simple, if unfamiliar. Life crystallizes at a critical molecular diversity because catalytic closure itself crystallizes. These ideas, I hope, will become experimentally established parts of our new chemical creation story, our new view of our ancient roots, our new sense of the emergence of life as an expected property of the physical world.

In the computer-simulation movies we have made of this process, we can see this crystallization happening through an increase in either the diversity of molecules or the probability that any molecule catalyzes any reaction. We call these parameters *M* and *P*. As either *M* or *P* increases, at first nothing much happens in the dead soup; then suddenly it springs to life. The experiment has not been done with real chemicals yet, although I'll return to that later. But on the computer, a living system swarms into existence. Figure 3.7 shows what one of these model self-reproducing metabolisms actually looks like. As you can see, this model system is based on the continuous supply of several simple food molecules, the monomers A and B, and the four possible dimers: AA, AB, BA, and BB. From this, the system crystallizes a collectively autocatalytic, self-sustaining model metabolism with some 21 kinds of molecules. More complex autocatalytic sets have hundreds or thousands of molecular components.

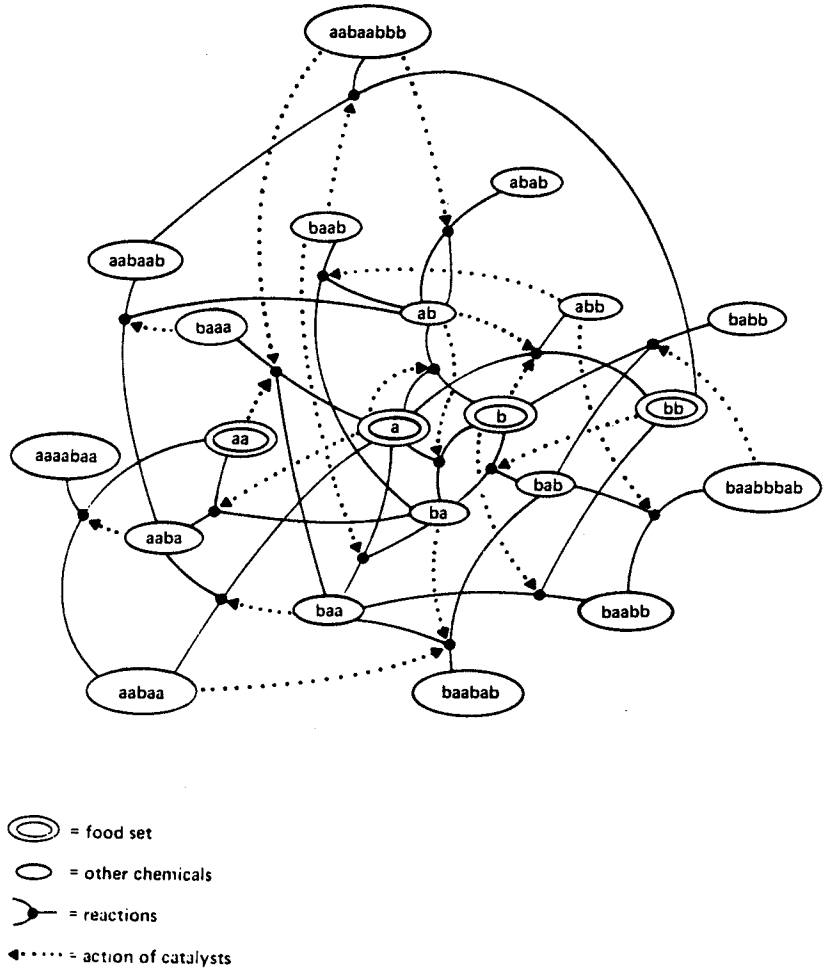

Figure 3.7 *An autocatalytic set. A typical example of a small autocatalytic set in which food molecules (a, b, aa, bb) are built up into a self-sustaining network of molecules. The reactions are represented by points connecting larger polymers to their breakdown products. Dotted lines indicate catalysis and point from the catalyst to the reaction being catalyzed.*

The same basic results are found if we use the template-matching model of catalysis. The ratio of possible reactions to polymers is so vast that eventually a giant catalyzed component and autocatalytic sets emerge. Given almost any way in which nature might determine which chemicals catalyze which reactions, a critical molecular diversity is reached at which the number of red catalyzed reactions passes a phase transition and a vast web of chemicals crystallizes in the system. This vast web is, it turns out, almost always collectively autocatalytic.

Such a system is, at minimum, self-sustaining, but such a system is very nearly self-reproducing. Suppose our collectively autocatalytic reaction system is contained within some kind of compartment. Compartmentalization must have been essential to prevent dilution of the reacting molecules. The autocatalytic system might constitute one of Alexander Oparin's coascervates, or it might create and be contained in a bilipid membrane vesicle. As the molecular constituents of the system re-create themselves, the number of copies of each kind of molecule can increase until the total has doubled. The system can then break into two coascervates, two bilipid membrane vesicles, or other compartmentalized forms. In fact, such breaking in two happens spontaneously as such systems increase in volume. Thus our autocatalytic protocell has now self-reproduced. A self-reproducing chemical system, alive by these criteria, springs into existence.

Energizing the Reactions

Now one might object that what is true for As and Bs may not be true for atoms and molecules. As Einstein said, a theory should be as simple as possible, but not too simple. One thing lacking in our model so far has been energy. As we have seen, living systems are open, nonequilibrium thermodynamic systems, sustained by a flux of matter and energy through them. As with the vastly simpler Belosov–Zhabotinski reaction, living systems maintain structures by dissipating matter and energy—in short, by eating and excreting.

The problem is this: it takes energy to create large polymers, for thermodynamics favors their breakdown into smaller constituents. A chemically realistic autocatalytic set has to obtain energy to create and sustain large molecules that may be its catalysts.

To be concrete, consider a protein with 100 amino acids linked together, or even a smaller sequence of amino acids, called a peptide. The linking of any two amino acids by a peptide bond requires energy. An easy way to see this is that the bond confines the motion of the two amino acids relative to each other. It would require some tugging to pull the amino acids apart. The tugging required is a measure of the energy of the bond. I noted earlier that almost all reactions are spontaneously reversible. This is true of a peptide bond. During its formation, a water molecule is pulled out of the reacting pair of amino acids. Thus water itself is a product of the reaction. Conversely, when a peptide bond is cleaved, a water molecule is used up. If peptides are dissolved in water, the water molecules will tend to break peptide bonds.

In a normal aqueous environment, the equilibrium ratio of cleaved amino acids to amino acid pairs (dipeptides) is about 10 to 1. But the same calculation holds for a dipeptide plus a single amino acid coming together to form a tripeptide. In an aqueous environment, the ratio of the dipeptide and amino acid to the tripeptide will be about 10 to 1 at chemical equilibrium. Note the consequence: at equilibrium, the ratio of two amino acids to the dipeptide they form is 10 to 1, and the ratio of the dipeptide plus a single amino acid to the tripeptide is also 10 to 1. Thus the ratio of single amino acids to tripeptides is not 10 to 1, but roughly 100 to 1. Similarly, at equilibrium, the ratio of amino acids to tetrapeptides is about 1,000 to 1. As the bigger polymer increases in length, its equilibrium concentration relative to the amino acids falls by a factor of about 10-fold for each increase of one amino acid in length.

The implication of the previous simple calculation is this: in an equilibrium mixture of single amino acids and various peptides up to length, say 25, the average ratio of the amino acid concentrations to that of any specific peptide of 25 amino acids would be about 1 to 10^{-25}. To be concrete, if amino acids were dissolved to the highest concentration in water that can be attained, then at equilibrium the number of copies of any specific sequence of amino acids 25 residues long would be less than one molecule in a liter of water! By contrast, the number of copies of any of the single amino acids might be on the order of 10^{20} to 10^{23}. Autocatalytic sets may use large polymers. How can high concentrations of such molecules be achieved in the face of this thermodynamic difficulty?

There are at least three fundamental ways that this vast obstacle might have been overcome. Each is remarkably simple. First, reactions can be confined to a surface rather than occurring in a volume. The reason this helps form larger polymers is simple. The rate at which a chemical reaction occurs depends on how rapidly the reaction partners collide with one another. If an enzyme is involved, the enzyme must be encountered as well. If the reaction is occurring in a volume, such as a beaker, then each molecule must diffuse in three dimensions and bump into its reaction partners. It is rather easy for molecules wandering in three dimensions to keep missing one another (recall the cartoon I described in Chapter 2). By contrast, if the molecules are confined to a very thin surface layer, such as clay or a bilipid membrane, then the search occurs in only two dimensions. It is a lot harder for the molecules to miss one another. To tune your intuition, imagine the molecules diffusing in a one-dimensional tube with a tiny diameter. Then they are bound to run into one another. In short, confining reactions to occur on surfaces strongly increases the chances of substrates hitting one another, hence enhancing the rate of formation of longer polymers.

A second simple mechanism to enhance the formation of longer polymers is to dehydrate the system. Dehydration removes water molecules, hence slowing down the cleavage of peptide bonds. In computer simulations with my colleagues Doyne Farmer, Norman Packard, and, later, Richard Bagley, we found strong evidence that even simple dehydration ought to suffice to allow real autocatalytic systems of polymers to reproduce. Our model fits the laws of chemistry and physics without straining.

Dehydration is not a cheat; it actually works. A famous reaction, called the plastein reaction, was well studied beginning almost 60 years ago. The enzyme trypsin in the stomach helps digest the proteins we eat. If trypsin is mixed with large proteins in an aqueous medium, it cleaves the proteins into smaller peptides. But if the reaction system is dehydrated, lowering the concentration of water relative to the peptides, the equilibrium shifts in favor of the synthesis of larger polymers from the small peptide fragments. Trypsin obliges by catalyzing these ligation reactions, yielding larger polymers. If these larger polymers are removed and the system is again dehydrated, trypsin obliges by synthesizing still more large polymers.

Reactions on surfaces and dehydration can be used to favor the formation of large polymers. But contemporary cells also use a more flexible and sophisticated mechanism. As cells form bonds, they obtain the needed energy by simultaneously breaking down the high-energy bonds in ubiquitous helper molecules. Adenosine triphosphate (ATP) is the most common of these. Reactions that require energy are called endergonic; those that release energy are called exergonic. Cells drive endergonic reactions by linking them to exergonic reactions.

A number of plausible candidates have been suggested for high-energy bonds that may have powered early self-reproducing metabolisms. For example, pyrophosphate, two phosphates linked together, is abundant and releases substantial energy upon cleavage. Pyrophosphate may have been a useful source of free energy to drive synthesis in early living systems. Farmer and Bagley have used computer simulations to show that model systems powered by these bonds meet plausible thermodynamic criteria and can reproduce.

What is required to link exergonic and endergonic reactions? Does some new mystery confront us beyond the achievement of catalytic closure? I think not. A problem is here, but hardly a mystery. All that is required, after all, is that the autocatalytic set include catalysts that link exergonic and endergonic reactions, so that one powers the other. The endergonic synthesis of large molecules must be coupled with the degradation of high-energy bonds supplied by food molecules or, ultimately, sunlight. But this does not seem an overwhelming obstacle.

Catalysis of such coupled reactions is not fundamentally different from other reactions: an enzyme able to bind the transition state is needed. All that is required is a sufficient diversity of molecules.

An Unrepentant Holism

This theory of life's origins is rooted in an unrepentant holism, born not of mysticism, but of mathematical necessity. A critical diversity of molecular species is necessary for life to crystallize. Simpler systems simply do not achieve catalytic closure. Life emerged whole, not piecemeal, and has remained so. Thus unlike the dominant nude RNA view of the origin of life, with its evolutionary just-so stories, we have a hope of explaining why living creatures seem to have a minimal complexity, why nothing simpler than the pleuromona can be alive.

If this view is right, we should be able to prove it. We should be able to create life anew in the fabled test tube, as though it were held by some scientist driven by Faustian dreams. Can we hope to make a new life-form? Can we brazen the face of God? Yes, I think so. And God, in his grace and simplicity, should welcome our struggles to find his laws. The ways of science are genuinely mysterious. As we shall see in Chapter 7, the hope to create collectively autocatalytic sets of molecules is linked to what may become the second era of biotechnology, promising new drugs, vaccines, and medical miracles. And the concept of catalytic closure in collectively autocatalytic sets of molecules will begin to appear as a deep feature of the laws of complexity, reemerging in our understanding of ecosystems, economic systems, and cultural systems.

Immanuel Kant, writing more than two centuries ago, saw organisms as wholes. The whole existed by means of the parts; the parts existed both because of and in order to sustain the whole. This holism has been stripped of a natural role in biology, replaced with the image of the genome as the central directing agency that commands the molecular dance. Yet an autocatalytic set of molecules is perhaps the simplest image one can have of Kant's holism. Catalytic closure ensures that the whole exists by means of the parts, and they are present both because of and in order to sustain the whole. Autocatalytic sets exhibit the emergent property of holism. If life began with collectively autocatalytic sets, they deserve awed respect, for the flowering of the biosphere rests on the creative power they unleashed on the globe—awed respect and wonder, but not mysticism.

Most important of all, if this is true, life is vastly more probable than we have supposed. Not only are we at home in the universe, but we are far more likely to share it with as yet unkn

Chapter 4

Order for Free

The living world is graced with a bounty of order. Each bacterium orchestrates the synthesis and distribution of thousands of proteins and other molecules. Each cell in your body coordinates the activities of about 100,000 genes and the enzymes and other proteins they produce. Each fertilized egg unfolds through a sequence of steps into a well-formed whole called, appropriately enough, an organism. If the sole source of this order is what Jacques Monod called "chance caught on the wing," the fruit of one fortuitous accident after another and selection sifting, then we are indeed improbable. Our lapse from paradise—Copernicus to Newton in celestial mechanics, to Darwin in biology, and to Carnot and the second law of thermodynamics—leaves us spinning around an average star at the edge of a humdrum galaxy, lucky beyond reckoning to have emerged as living forms.

How different is humanity's stance, if it proves true that life crystallizes almost inevitably in sufficiently complex mixtures of molecules, that life may be an expected emergent property of matter and energy. We start to find hints of a natural home for ourselves in the cosmos.

But we have only begun to tell the story of emergent order. For spontaneous order, I hope to show you, has been as potent as natural selection in the creation of the living world. We are the children of twin sources of order, not a singular source. So far we have showed how autocatalytic sets might spring up naturally in a variegated chemical soup. We have seen that the origin of collective autocatalysis, the origin of life itself, comes because of what I call "order for free"—self-organization that arises naturally. But I believe that this order for free, which has undergirded the origin of life itself, has also undergirded the order in organisms as they have evolved and has even undergirded the very capacity to evolve itself.

If life emerged as collectively autocatalytic systems swirling in some soup, then our history only starts there. It had best not end abruptly for lack of the ability to evolve. The central motor of evolution, Darwin taught us, requires self-reproduction and heritable variation. Once these occur, natural selection will cull the fitter from the less fit. Most biologists hold that DNA or RNA as a stable store of genetic information is essential to adaptive evolution. Yet if life began with collective autocatalysis and later learned to incorporate DNA and the genetic code, we are faced with explaining how such autocatalytic sets could undergo heritable variation and natural selection without yet harboring a genome. If we required the magic of template replication and the further magic of genetic coding for proteins, the chicken-and-egg problem becomes too horrendous to contemplate. Evolution cannot proceed without these mechanisms, and we cannot have these mechanisms without evolution to tinker them together. In continuing our search for a theory of we the expected, we are led to ask this question: Is there a way that an autocatalytic set could evolve without all the complications of a genome?

My colleagues Richard Bagley and Doyne Farmer have hinted at how this might happen. We have already seen in Chapter 3 that once an autocatalytic set is enclosed in a spatial compartment of some sort—say, a coascervate or a bilipid membrane vesicle—the self-sustaining metabolic processes can actually increase the number of copies of each type of molecule in the system. In principle, when the total has doubled, the compartmentalized system can "divide" into two daughters. Self-reproduction can occur. As I noted, in experiments such compartmental systems do tend to divide spontaneously into two daughters as their volumes increase. But if daughter "cells" were always identical to the parent "cell," no heritable variation could occur.

Richard and Doyne found a natural way that variation and evolution in such systems can occur. (Richard did this work as part of his doctoral dissertation at the University of California, San Diego, with Stanley Miller as one of his examiners.) They proposed that a random, uncatalyzed reaction will occasionally occur as an autocatalytic net goes about its business. These spontaneous fluctuations will tend to give rise to molecules that are not members of the set. Such novel molecules can be thought of as a kind of penumbra of molecular species, a chemical haze surrounding the autocatalytic set. By absorbing some of these new molecular species into itself, the set would become altered. If one of these new molecules helped catalyze its own formation, it would become a full-fledged member of the network. A new loop would be added to the metabolism. Or if the molecular interloper inhibited a previously occurring reaction, then an old loop might be eliminated from the set. Either way, heritable variation was evidently possible. If the result were a more efficient network—one better able to sustain itself amid a harsh environment—then these mutations would be rewarded, the altered web crowding out its weaker competitors.

In short, there is reason to believe that autocatalytic sets can evolve without a genome. This is not the kind of evolution we are accustomed to thinking about. There is no separate DNA-like structure carrying genetic information. Biologists divide cells and organisms into the genotype (the genetic information) and the phenotype (the enzymes and other proteins, as well as the organs and morphology, that make up the body). With autocatalytic sets, there is no separation between genotype and phenotype. The system serves as its own genome. Nevertheless, the capacity to incorporate novel molecular species, and perhaps eliminate older molecular forms, promises to generate a population of self-reproducing chemical networks with different characteristics. Darwin tells us that such systems will evolve by natural selection.

In fact, such self-reproducing, compartmentalized protocells and their daughters will inevitably form a complex ecosystem. Each protocell reproduces with heritable variations; in addition, each will tend to absorb and excrete molecular species selectively in its environment, as do contemporary bacteria. In short, a molecule created in one protocell can be transported to other protocells. That molecule may promote or poison reactions in the second protocell. Not only does metabolic life begin whole and complex; but all the panoply of mutualism and competition that we think of as an ecosystem springs forth from the very beginning. The story of such ecosystems at all scales is the story not merely of evolution, but of coevolution. We have all made our worlds together for almost 4 billion years. The story of order for free continues in this molecular and organismic coevolution, as will be shown in later chapters.

But evolution requires more than simply the ability to change, to undergo heritable variation. To engage in the Darwinian saga, a living system must first be able to strike an *internal* compromise between malleability and stability. To survive in a variable environment, it must be stable, to be sure, but not so stable that it remains forever static. Nor can it be so unstable that the slightest internal chemical fluctuation causes the whole teetering structure to collapse. We have only to consider again the now familiar concepts of deterministic chaos to appreciate the problem. Recall the famous butterfly in Rio, whose energetic wing flapping, or even languid stirring, can alter the weather in Chicago. In chaotic systems, tiny changes in initial conditions can lead

to profound disturbances. From what we have said so far, there is no reason to believe that our autocatalytic sets would not be hypersensitive, chaotic, doomed from the start. A tiny change in the concentrations of the internal metabolism because some molecule from a neighboring cell is absorbed might be amplified so mightily that the network would fly apart. The autocatalytic sets I am proposing would have had to coordinate the behaviors of some thousands of molecules. The chaos that could potentially flourish in systems of this complexity boggles the mind.

The potential for chaos is not merely theoretical. Other molecules can bind to the enzymes in our own cells, inhibiting or increasing their activity. Enzymes can be "turned on" or "turned off" by other molecules in the reaction network. It is now well known that in most cells, such molecular feedback can give rise to complex chemical oscillations in time and space. The potential for chaos is real.

If we are to believe that life began when molecules spontaneously joined to form autocatalytic metabolisms, we will have to find a source of molecular order, a source of the fundamental internal homeostasis that buffers cells against perturbations, a compromise that would allow the protocell networks to undergo slight fluctuations without collapsing. How, without a genome, would such order arise? It must somehow emerge from the collective dynamics of the network, the coordinated behavior of the coupled molecules. It must be another case of order for free. As we are about to see, astonishingly simple rules, or constraints, suffice to ensure that unexpected and profound dynamical order emerges spontaneously.

The Wellsprings of Homeostasis

Allow me a simple, highly useful, idealization. Let us imagine that each enzyme has only two states of activity—on or off, and can switch between them. So at each moment, each enzyme is either active or inactive. This idealization, like all idealizations, is literally false. In reality, enzymes show graded catalytic activities. Most simply, the rate of a reaction depends on enzyme and substrate concentrations. Nevertheless, inhibition or activation of enzymes by molecules binding to sites on the enzyme or changing the enzyme in other ways is common and is often associated with a sharp change in enzyme activity. In addition, allow me to think of the substrates or products of reactions as either present or absent. This, too, is literally false. But often the concentrations of substrates and products in complex reaction networks can change very

swiftly from high to low. The "on–off" "present–absent" idealization is very useful, for we are going to consider networks with thousands of model enzymes, substrates, and products.

The point in using idealizations in science is that they help capture the main issues. Later one must show that the issues so captured are not altered by removing the idealizations. Thus in physics, analysis of the gas laws was based on models of gas molecules as hard elastic spheres. The idealization captured the main features necessary to create statistical mechanics. In Chapter 3, we presented molecules and their reactions as buttons and threads. Now let us change metaphors and think of a metabolic network of enzymes, substrates, and products as a network of lightbulbs connected by wires, an electrical circuit. A molecule catalyzing the formation of another molecule can be thought of as one bulb turning on another. But molecules can also inhibit each other's formation. Think of this as one bulb turning another bulb off.

One way to get such a network to behave in an orderly manner would be to design it with great care and craft. But we have proposed that autocatalytic metabolisms arose in the primal waters spontaneously, built from a random conglomeration of whatever happened to be around. One would think that such a haphazard concoction of thousands of molecular species would most likely behave in a manner that was disorderly and unstable. In fact, the opposite is true: order arises spontaneously, order for free. To return to our metaphor, although we wire our bulbs together at random, they do not necessarily blink on and off randomly like the twinkling lights of a vast forest of berserk Christmas trees. Given the right conditions, they settle into coherent, repeating patterns.

To see why order emerges spontaneously, I have to introduce some of the concepts mathematicians use to think about dynamical systems. If we think of our autocatalytic set as an electrical network, then it can assume a vast number of possible states. All the bulbs might be off, all might be on, and in between these two extremes can be myriad combinations. Imagine a network that consists of 100 nodes, each of which can be in one of two possible states, either on or off; the number of possible configurations is 2^{100}. For our autocatalytic metabolism, with perhaps 1,000 kinds of molecules, the number of possibilities is even vaster: $2^{1,000}$. This range of possible behaviors is called a state space. We can think of it as the mathematical universe in which the system is free to roam.

To make these notions concrete, consider a simple network consisting of just three light bulbs—1, 2, and 3—each of which receives "inputs" from the other two. (Figure 4.1*a*). The arrows show which way

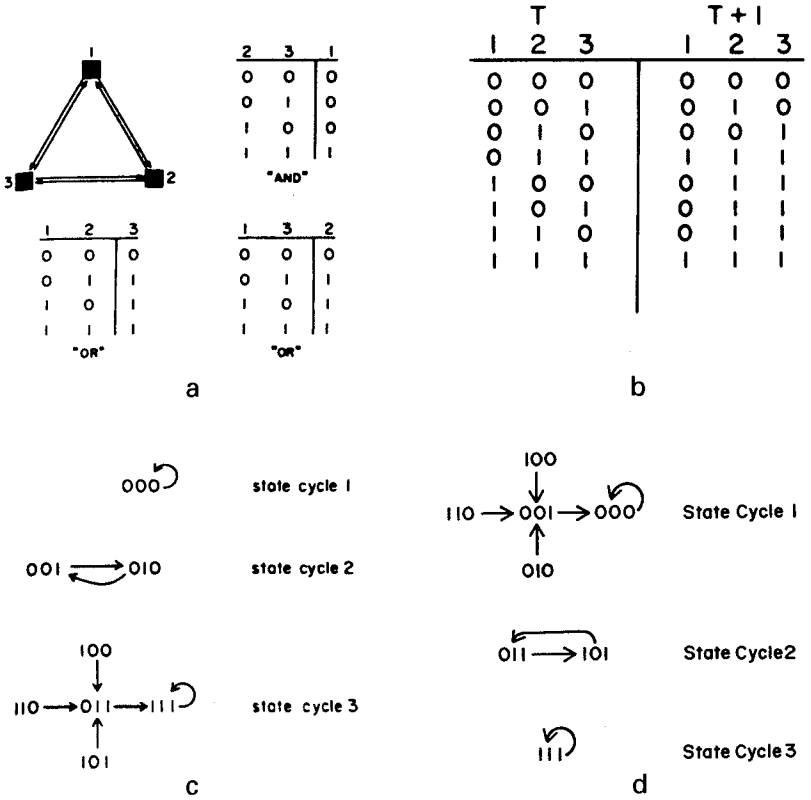

Figure 4.1 *A Boolean net.* (a) *The wiring diagram in a Boolean network with three binary elements, each an input to the other two.* (b) *The Boolean rules of* (a) *rewritten to show for all* $(2^3) = 8$ *states at time* T, *the activity assumed by each element at the next moment,* T + 1. *Read from left to right, this figure shows the successor state for each state.* (c) *The state transition graph, or "behavior field," of the autonomous Boolean network of* (a) *and* (b), *obtained by showing state transitions to successor states connected by arrows.* (d) *Effects of mutating the rule of element 2 from* OR *to* AND.

the signals flow; thus arrows point to bulb 1 from bulbs 2 and 3, signifying that bulb 1 receives inputs from bulbs 2 and 3.

In addition to denoting the wiring diagram, we need to know how each lightbulb responds to the signals it receives. Since each bulb can have only two values, on and off, which we can represent as 1 and 0, then it is easy to see that there are four possible input patterns it can receive from its two neighbors. Both inputs can be off (00), one or the other input can be on (01 or 10), or both inputs can be on (11). Using this information, we can construct a rule table specifying whether each bulb will be active (1) or inactive (0) for each of these four possible signals. For example, bulb 1 might be active only if both of its inputs were active the moment before. In the language of Boolean algebra (named in honor of George Boole, the inventor of mathematical logic in the nineteenth century), bulb 1 is an AND gate: bulbs 2 *and* 3 must be active before it will light. Or we could choose instead for the bulb to be governed by the Boolean OR function: bulb 1 will be active the next moment if bulb 2 or bulb 3 or both were active the moment before.

To complete the specification of what I will now call a Boolean network, I will assign to each lightbulb one of the possible Boolean functions. Say I assign the AND function to bulb 1 and the OR function to bulbs 2 and 3 (Figure 4.1*a*). At each tick of the clock, each bulb examines the activities of its two inputs and adopts the state 1 or 0 specified by its Boolean function. The result is a kaleidoscopic blinking as pattern after pattern unfolds.

Figure 4.1*b* shows the eight possible states that the network can assume, from (000) to (111). Read along vertical columns, the right half of Figure 4.1*b* specifies the Boolean rule governing each lightbulb. But read from left to right, Figure 4.1*b* shows, for each current state at time *T*, the next state of the entire network one moment later, at $T + 1$, when all lightbulbs simultaneously adopt their new activities, 1 or 0.

Now we are in a position to begin to understand the behavior of this little network. As we can see, the system can be in a finite number of states, here eight. If started in one state, over time the system will flow through some sequence of states. This sequence is called a trajectory (Figure 4.1*c*). Since there is a finite number of states, the system must eventually hit a state it has previously encountered. Then the trajectory will repeat. Since the system is deterministic, it will cycle forever around a recurrent loop of states called a state cycle.

Depending on the initial state in which we start our network—the pattern of on and off bulbs—it will follow various trajectories, falling at some point into an ever repeating state cycle (Figure 4.1*c*). The simplest possible behavior would occur if the network fell immediately into a state cycle consisting of a single pattern of 1s and 0s. A system started in such a state never changes; it is said to be stuck in a cycle of length 1. Alternatively, the length of the state cycle could conceivably be the total number of states in state space. A system caught in such a cycle will repeat, one after another, every pattern it is capable of displaying. For our three-bulb system, this would result in a steady twinkling as the system passed through its eight possible states. Since the number of states is so small, we could soon detect the pattern of its blinking. Now imagine a larger network, with 1,000 bulbs and thus $2^{1,000}$ possible states. If the network were on a state cycle passing through every one of this hyperas-

tronomical number of states, and if it took a mere *trillionth of a second* per state transition, we could never in the lifetime of the universe see the system complete its orbit.

So the first thing to appreciate about Boolean networks is this: any such network will settle down to a state cycle, but the number of states on such a recurrent pattern might be tiny—as small as a single steady state—or so hyperastronomical that numbers are meaningless. If a system falls into a small state cycle, it will behave in an orderly manner. But if the state cycle is too vast, the system will behave in a manner that is essentially unpredictable. The state spaces through which molecular networks with only a few thousand kinds of molecules can roam are beyond our common reckoning. For our autocatalytic networks to be orderly, they must avoid veering off on seemingly endless tangents and must settle down into small state cycles—a repertoire of stable behaviors.

To gain insight into how likely it is that autocatalytic sets would be stable enough to endure, we must ask these questions: How does one make orderly networks with short state cycles? Is the creation of tiny state cycles difficult, meaning that it is something of a miracle that stable autocatalytic metabolisms emerged? Or does it happen naturally? Is it part of order for free?

To answer these questions we need to understand the concept of an attractor. More than one trajectory can flow into the same state cycle. Start a network with any of these different initial patterns and, after churning through a sequence of states, it will settle into the same state cycle, the same pattern of blinking. In the language of dynamical systems, the state cycle is an attractor and the collection of trajectories that flow into it is called the basin of attraction. We can roughly think of an attractor as a lake, and the basin of attraction as the water drainage flowing into that lake.

Just as a mountainous region may harbor many lakes, a Boolean network may harbor many state cycles, each draining its own basin of attraction. The little network in Figures 4.1*a–c* has three state cycles. The first state cycle has the single steady state (000), which drains no basin of trajectories. It is an isolated steady state. It can be reached only if we start the network there. The second state cycle has two states, (001) and (010). The network oscillates between these two. No other states drain into this attractor. Launch the network with one of these two patterns and it will remain in the cycle, blinking back and forth between the two states. The third state cycle consists of the steady state (111). This attractor lies in a basin of attraction draining four other states. Start the network with any one of these patterns and it will quickly flow to the steady state and freeze up, displaying three lighted bulbs.

Under the right conditions, these attractors can be the source of order in large dynamical systems. Since the system follows trajectories that inevitably flow into attractors, tiny attractors will "trap" the system into tiny subregions of its state space. Among the vast range of possible behaviors, the system settles into an orderly few. The attractors, if small, create order. Indeed, tiny attractors are a prerequisite for the order for free that we are seeking.

But tiny attractors are not enough. For a dynamical system, such as an autocatalytic net, to be orderly, it must exhibit homeostasis; that is, it must be resistant to small perturbations. Attractors are the ultimate source of homeostasis as well, ensuring that a system is stable. In large networks, any state cycle typically drains an enormous basin; many states flow into the attractor. Moreover, the states within that basin can be very similar to the states on the state cycle to which they drain. Why is this important? Suppose we arbitrarily choose a single lightbulb and flip it to the opposite state. All or most such perturbations leave the system in the same basin of attraction. So the system will return to the same state cycle from which it was perturbed! That is the essence of homeostatic stability. State cycle 3 in Figure 4.1*c* is stable in this way; if the network is in this basin, flipping the activity of any single lightbulb will have no long-term impact on its behavior, for the system will return to the same state cycle.

But homeostatic stability does not always arise. State cycle 1, by contrast, is an isolated steady state and is unstable to the slightest perturbation. After any such flip, the system is shoved into a different basin of attraction. It can't come home again. If the network had the property that all attractors were unstable in this way, we can imagine that slight perturbations (the flapping of the butterfly's wings) would persistently bump the system out of attractors and send it veering off on an endless, never repeating journey through state space. The system would be chaotic.

If we are to believe that life began with the spontaneous generation of autocatalytic nets, then we had better hope that they were homeostatic. Is it natural that certain kinds of large networks will exhibit homeostasis? Is homeostasis hard to create, making the emergence of stable networks vastly unlikely? Or can it, too, be part of order for free?

What we need are laws describing which kinds of networks are likely to be orderly and which are likely to succumb to chaos. Any Boolean network has attractors, each draining some basin of attraction, but the state spaces of networks with thousands of kinds of molecular species are hyperastronomical. If we change metaphors and think of each state cycle as a galaxy in space, how many attractor galaxies are there spread

gadzillions of states in state space, one might have gadzillions of attractors. If there are vast numbers of attractors, and the system might be located on any one of them, that does not sound like order.

Collectively autocatalytic sets presumably evolved, and contemporary organisms do evolve, by mutations that permanently change the functional connections among the molecular species in the system. Will such permanent mutational changes cause an autocatalytic system to collapse into chaotic twinkling through its space of molecules, poisoning its own capacity to catalyze its own reproduction? Will minor mutational variations typically cause catastrophic changes? In the language of Boolean networks, another way to perturb a network is to permanently "mutate" its wiring diagram, changing the inputs or the Boolean function governing when a bulb is on or off. In Figure 4.1*d*, I show the result of changing the rule governing lightbulb 2 from OR to AND. As you can see, this causes the network to assume a new dynamical form. Some state cycles remain, but others are changed. New basins of attraction will steer the network into different patterns.

Darwin supposed that living systems evolve by mutations that cause small modifications in the properties of the organism. Is this graceful property of minor changes hard to achieve? Or is it, too, part of order for free? A pure Darwinist might argue that this kind of graceful stability could arise only after a series of evolutionary experiments, yet this begs the question. We are trying to explain the origin of the very ability to evolve! However life started, with nude replicating RNA molecules or with collectively autocatalytic sets, this stability cannot be imposed from outside by natural selection. It must arise from within as a condition of evolution itself.

All these properties we need, I believe, all the order we require, arise spontaneously. We next must show how order for free supplies the small ordered attractors we need, the homeostasis we need, and the graceful stability we need. Order for free, utterly natural, if previously mostly unknown, will change our view of life.

The Requirements for Order

We have seen that Boolean networks can exhibit profound order, but Boolean networks can also exhibit profound chaos. Consequently, we seek the conditions under which orderly dynamics can emerge in such systems. I will now present the results of about 30 years of work.

The main results are simple to summarize: two features of the way networks are constructed can control whether they are in an ordered regime, a chaotic regime, or a phase transition regime between these,

"on the edge of chaos." One feature is simply how many "inputs" control any lightbulb. If each bulb is controlled by only one or two other lightbulbs, if the network is "sparsely connected," then the system exhibits stunning order. If each bulb is controlled by many other lightbulbs, then the network is chaotic. So "tuning" the connectivity of a network tunes whether one finds order or chaos. The second feature that controls the emergence of order or chaos is simple biases in the control rules themselves. Some control rules, the AND and OR Boolean functions we talked about, tend to create orderly dynamics. Other control rules create chaos.

The way I and others have done this work is pretty straightforward. One way to ask what kinds of lightbulb networks exhibit order of chaos is to construct very specific networks and study them. But this would leave us with a vast number of very specific networks to study—another of our hyperastronomical numbers, big beyond meaning. The approach I have taken asks whether networks of certain *general kinds* exhibit order or chaos. To answer this question, the natural approach is to carefully define the "kind" of networks in question, and then use computers to simulate large numbers of networks drawn at random from the pool. Then, like a pollster, we can build up a portrait of the typical, or generic, behaviors of members of the class.

We might, for example, study the pool of networks with 1,000 bulbs (we'll call this variable N) and 20 inputs per bulb (the variable K). Given $N = 1000$ and $K = 20$, a vast ensemble of networks can be built. We sample this ensemble by randomly assigning to each of the 1,000 bulbs 20 inputs and, again at random, one of the possible Boolean functions. Then we can study the network's behavior, counting the number of attractors, the lengths of attractors, the stability of attractors to perturbations and mutations, and so forth. Throwing the dice again, we can randomly wire another network with the same general characteristics and study its behavior. Sample by sample, we build up a portrait of a family of Boolean nets, and then we change the values of N and K and build up another portrait.

After years of such experiments, networks with various parameters become as familiar as old friends. Consider networks in which each lightbulb receives input from only one other. In these $K = 1$ networks, nothing very interesting happens. They quickly fall into very short state cycles, so short that they often consist of but a single state, a single pattern of illumination. Launch such a $K = 1$ network and it freezes up, saying the same thing over and over for all time.

At the other end of the scale, consider networks in which $K = N$, meaning that each lightbulb receives an input from all lightbulbs, including itself. One quickly discovers that the [illegible]

state cycles is the square root of the number of states. Consider the implications. For a network with only 200 binary variables—bulbs that can be on or off—there are 2^{200} or 10^{60} possible states. The length of the state cycles is thus on the order of 10^{30} states. Start the network with some arbitrary pattern of on-bulbs and off-bulbs, 1s and 0s, and it will be pulled by an attractor into a repeating cycle, but a cycle so long as to be all but fathomless. Suppose the network took a millionth of a second to pass from state to state. Then the little network would require 10^{30} millionths of a second to traverse its state cycle. This is equal to billions of times the 15-billion-year history of the universe! So we could never actually observe the fact that the system had "settled" onto its state cycle attractor! We could never tell from the twinkling patterns of the lightbulbs that the network was not just wandering randomly around in its entire state space!

I hope this gives you pause. We are searching for laws that suffice to yield orderly dynamics. Our Boolean networks are nonequilibrium, open thermodynamic systems. Since a little network with only 200 lightbulbs can twinkle away for an eternity without repeating a pattern, order is in no way automatic in nonequilibrium, open thermodynamic systems.

Such $K = N$ networks do show signs of order, however. The number of attractors in a network, the number of lakes, is only N/e, where e is the basis of the natural logarithms, 2.71828. So a $K = N$ network with 100,000 binary variables would harbor about 37,000 of these attractors. Of course, 37,000 is a big number, but very very much smaller than $2^{100,000}$, the size of its state space.

Suppose, then, that we perturb the network, flipping a bulb from off to on, or vice versa. In $N = K$ networks, we get an extreme version of the butterfly effect. Flip a bit, and the system almost certainly falls under the sway of another attractor. But since there are 37,000 attractors with lengths up to $10^{15,000}$ states, the tiny fluctuation will utterly change the future evolution of the system. $K = N$ networks are massively chaotic. No order for free in this family.

Even worse, try evolving such a network by randomly swapping the Boolean rule of some lightbulb. You will alter half the state transitions in the network and scatter all the old basins of attraction and state cycles to the dustbin of network history. Small changes here cause massive changes in behavior. There are no graceful minor heritable variations for selection to act on in this family.

Most Boolean networks are chaotic, and they are graceless with respect to minor mutations. Even networks in which K is much less than N, $K = 4$ or $K = 5$, exhibit unpredictable, chaotic behavior similar to that seen for $K = N$ networks.

Whence cometh the order? The order arises, sudden and stunning, in $K = 2$ networks. For these well-behaved networks, the length of state cycles is not the square root of the number of states, but, roughly, the square root of the number of binary variables. Let's pause to translate this as clearly as we can. Think of a randomly constructed Boolean network with $N = 100,000$ lightbulbs, each receiving $K = 2$ inputs. The "wiring diagram" would look like a madhatterly scrambled jumble, an impenetrable jungle. Each lightbulb has also been assigned at random a Boolean function. The logic is, therefore, a similar mad scramble, haphazardly assembled, mere junk. The system has $2^{100,000}$ or $10^{30,000}$ states—megaparsecs of possibilities—and what happens? The massive network quickly and meekly settles down and cycles among the square root of 100,000 states, a mere 317.

I hope this blows your socks off. Mine have never recovered since I discovered this almost three decades ago. Here is, forgive me, stunning order. At a millionth of a second per state transition, a network, randomly assembled, unguided by any intelligence, would cycle through its attractor in 317-millionths of a second. This is a lot less than billions of times the history of the universe. Three hundred seventeen states? To see what this means in another way, one can ask how tiny a fraction of the entire state space the network squeezes itself into. A mere 317 states compared with the entire state space is an extremely tiny fraction of that state space, about 1 divided by $10^{29,998}$!

We seek order without careful crafting. Recall our discussion in Chapter 1 of closed thermodynamic systems in which the gas molecules diffuse from improbable configurations—clumped in one corner or spread parallel to one face of a box—toward homogeneous configurations. The improbable configurations constituted order. Here, in this class of open thermodynamic systems, the spontaneous dynamics drive the system into an infinitesimal corner of its state space and hold it there, quivering for an eternity. Order for free.

Order expresses itself in these networks in diverse ways. Nearby states converge in state space. In other words, two similar initial patterns will likely lie in the same basin of attraction, hence driving the system to the same attractor. Thus such systems do not show sensitivity to initial conditions; they are not chaotic. The consequence is the homeostasis we seek. Once such a network is on an attractor, it will return to the same attractor with very high probability if it is perturbed. Homeostasis is free in this neck of the network woods.

For the same reason, these networks can undergo a mutation that alters wiring or logic without veering into randomness. Most small mutations cause our hoped-for small, graceful alteration in the behavior of the network. Basins and attractors change only slightly. Such systems

evolve readily. So selection does not have to struggle to achieve evolvability.

Finally, these networks are not *too* orderly. Unlike the $N = 1$ network, they are not frozen like a rock, but are capable of complex behaviors.

Our intuitions about the requirements for order have, I contend, been wrong for millennia. We do not need careful construction; we do not require crafting. We require only that extremely complex webs of interacting elements are sparsely coupled.

As I show in my book *The Origins of Order: Self-Organization and Selection in Evolution,* there are ways to tune networks in which K is greater than 2 so that they are also orderly, not chaotic. My colleagues Bernard Derrida and Gerard Weisbuch, both solid-state physicists at the Ecole Normale Supérieure in Paris, have shown that a variable called P can be tweaked to make a chaotic network become orderly.

The P parameter is very simple. Figure 4.2 shows three Boolean functions, each with four inputs. In each, the response of the regulated lightbulb must be specified for each of the 16 possible states of the four

A	B	C	D	E
0	0	0	0	0
0	0	0	1	1
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0
1	0	0	0	1
1	0	0	1	0
1	0	1	0	0
1	0	1	1	1
1	1	0	0	0
1	1	0	1	0
1	1	1	0	1
1	1	1	1	1

a

A	B	C	D	E
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	0
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	0
1	0	0	0	1
1	0	0	1	0
1	0	1	0	0
1	0	1	1	0
1	1	0	0	0
1	1	0	1	0
1	1	1	0	0
1	1	1	1	0

b

A	B	C	D	E
0	0	0	0	1
0	0	0	1	1
0	0	1	0	1
0	0	1	1	0
0	1	0	0	1
0	1	0	1	1
0	1	1	0	1
0	1	1	1	1
1	0	0	0	1
1	0	0	1	1
1	0	1	0	1
1	0	1	1	1
1	1	0	0	1
1	1	0	1	1
1	1	1	0	1
1	1	1	1	1

c

Figure 4.2 *Tinkering with the* P *parameter.* (a) *A Boolean function of four inputs, in which eight of the 16 input configurations yield a 0 response, while eight of the 16 yield a 1 response.* P = *8/16 = .50.* (b) *The response is 0 for 15 of the 16 possible input configurations.* P = *15/16 = .9375.* (c) *Fifteen of the 16 possible input configurations yield a 1 response.* P = *15/16 = .9375.*

input lightbulbs, from (0000) to (1111). For the Boolean function shown in Figure 4.2*a*, half the responses of the regulated lightbulb are 1, and the other half are 0. For the Boolean function shown in Figure 4.2*b*, 15 of the responses are 0, and only a single input pattern gets a 1 response from the regulated bulb. The Boolean function in Figure 4.2*c* is similar to that in Figure 4.2*b*, except that the preferred output response is 1, not 0. Fifteen of the 16 input patterns lead to a 1 response. P is just a parameter that measures the bias away from half 1 and half 0 responses in a Boolean function. So P for the Boolean function in Figure 4.2*a* is 0.5, while P for the Boolean function in Figure 4.2*b* is 15/16, or 0.9375, and P for the Boolean function in Figure 4.2*c* is also 15/16, or 0.9375.

What Bernard and Gerard showed is, after the fact, pretty intuitive. If different networks are built with increasing P biases, starting from the no-bias value of 0.5 to the maximum value of 1.0, networks with $P = 0.5$ or only slightly greater than 0.5 are chaotic and networks with P near 1.0 are orderly. This can be easily seen in the limit when their P parameter is 1.0. Then the bulbs in the network are of only two types. One type responds with a 0 to any input pattern; the other responds with a 1 to any input pattern. So if you start the network in any state at all, the 0-type bulbs respond with 0, the 1-type bulbs respond with 1, and the network freezes into the corresponding pattern of 0 and 1 values and remains at that steady state forever. So when the P parameter is maximum, networks are in an ordered regime. When the P parameter is 0.5, networks with many inputs per lightbulb are in a chaotic regime, twinkling away for an eternity. And, for any network, Bernard and Gerard showed that there is a critical value of P where the network will switch from chaotic to ordered. This is the edge of chaos, to which we will return in a moment.

The summary is this: two parameters suffice to govern whether random Boolean lightbulb networks are ordered or chaotic. Sparsely connected networks exhibit internal order; densely connected ones veer into chaos; and networks with a single connection per element freeze into mindlessly dull behavior. But density is not the only factor. If networks have dense connections, tuning the P bias parameter drives networks from the chaotic regime to the ordered regime.

These rules apply to networks of all sorts. In Chapter 5, I will show that the genome itself can be thought of as a network in the ordered regime. Thus some of the orderliness of the cell, long attributed to the honing of Darwinian evolution, seems likely instead to arise from the dynamics of the genomic network—another example of order for free. Again, I hope to persuade you that selection is not the sole source of

order in the living world. The powerful spontaneous order we are discussing now is likely to have played a role not only in the emergence of stable autocatalytic sets, but in the later evolution of life.

The Edge of Chaos

Living systems, from the collectively autocatalytic protocells we discussed in Chapter 3 to cells in your body to whole organisms, surely must have networks that behave stably, that exhibit homeostasis and graceful minor modifications when mutated. But cells and organisms must not be too rigid in their behavior if they are to cope with a complex environment. The protocell had best be able to respond to novel molecules floating its way. The *E. coli* in your intestine copes with an enormous variety of molecules by sending internal molecular signals cascading among its enzymes and genes, triggering a variety of changes in enzyme and gene activities bent on protecting the cell from toxins, metabolizing food, or, occasionally, exchanging DNA with other cells.

How do cell networks achieve both stability and flexibility? The new and very interesting hypothesis is that networks may accomplish this by achieving a kind of poised state balanced on the edge of chaos.

We have already seen hints of an axis running from orderly behavior to chaotic behavior in our lightbulb models. Sparsely connected networks, with $K = 1$ or $K = 2$, spontaneously exhibit powerful order. Networks with higher numbers of inputs per lightbulb, $K = 4$ or more, show chaotic behavior. So tuning the number of inputs per lightbulb—hence the density of the web of connections among the bulbs—from low to high tunes networks from orderly to chaotic behavior. In addition, we saw that adjusting the P bias parameter from 0.5 to 1.0 also tunes whether networks are in a chaotic or an ordered regime.

We should not be too surprised if some kind of sharp change in behavior, some kind of phase transition from order to chaos, occurred along this axis. In fact, in Chapter 3 we saw such a sharp change in behavior in our toy model of the origin of life. Recall that we were connecting buttons with threads and found that the size of the largest connected cluster suddenly jumped from small to huge when the ratio of threads to buttons passed the magic value of 0.5. Below that value, only small connected clusters of buttons existed. Above that value, a giant component composed of most of the buttons emerged. This is a phase transition.

A very similar kind of phase transition occurs in our lightbulb network models. Once again, a giant cluster of connected elements will appear. But the connected cluster will not be buttons; it will be a giant cluster of lightbulbs, each of which is frozen into a fixed activity, 1 or 0. If this giant frozen component forms, the network of bulbs is in the ordered regime. If it does not form, the network is in the chaotic regime. Just between, just near this phase transition, just at the edge of chaos, the most complex behaviors can occur—orderly enough to ensure stability, yet full of flexibility and surprise. Indeed, this is what we mean by complexity.

One way to visualize what is happening in random lightbulb networks is to make a mental movie. Picture starting the network in some initial state. As the network passes along its trajectory toward, then around, its state cycle, two kinds of behavior might be seen at any lightbulb. That lightbulb might twinkle on and off in some more or less complex pattern, or the lightbulb might settle to a fixed activity, always on or always off. Let's envision two colors to distinguish these two behaviors: color lightbulbs that are twinkling on and off green, and color those that are fixed on or fixed off red.

Now consider a network in the chaotic regime, say an $N = 1{,}000$, $K = 20$ network. Almost all the lightbulbs are twinkling on and off; hence they are colored green. Perhaps a few bulbs or small clusters of bulbs are fixed on or fixed off; hence these are colored red. In short, only tiny clusters of frozen red bulbs exist in a vast sea of twinkling green bulbs. So a network in the chaotic regime has a vast sea of twinkling green lightbulbs and may have a few islands of frozen red bulbs.

Alternatively, suppose we simulate a lightbulb network in the ordered regime, say $N = 100{,}000$ and $K = 2$, a vast tangle of a network with a complexity equal to your genome or to a very large autocatalytic set. Start the network in an initial state and follow it along its trajectory toward a state cycle, then around the state cycle. At first, most of the lightbulbs are twinkling on and off, and are colored green. But as the network converges onto its state cycle, then orbits the cycle, more and more of the lightbulbs settle into fixed states of activity, frozen on or frozen off. So most of the lightbulbs are now colored red.

And now the magic. If you think of all the red lightbulbs, and ask whether they are connected to one another, just as we asked if the buttons were connected to one another by threads, you will find that the frozen red lightbulbs form a vast giant cluster of interconnected lightbulbs! A giant frozen component of lightbulbs, each frozen into either the on or into the off state, exists in Boolean networks in the ordered regime.

Of course, not all the lightbulbs in our $N = 100{,}000$, $K = 2$ network need be frozen; typically, small and large clusters of connected light-

bulbs continue to twinkle on and off. These twinkling clusters are colored green. It is just the twinkling patterns of the clusters of connected green lightbulbs that constitute the cycling behavior of Boolean networks in the ordered regime. The lightbulbs in the giant frozen cluster of red lightbulbs do not twinkle at all.

If we looked into a typical network with $N = 100{,}000$ and $K = 2$, we would see a further important detail. The clusters of twinkling green lightbulbs are not themselves all interconnected. Instead, they form independent twinkling clusters, like twinkling green islands in a vast sea of frozen red lightbulbs.

So a Boolean network in the chaotic regime, as presented earlier, has a sea of ever-changing green lightbulbs twinkling on and off, with perhaps a few clusters of red lightbulbs that are frozen on or off. In contrast, a Boolean network in the ordered regime has a vast giant cluster of red lightbulbs that are frozen either on or off, a giant red cluster, with isolated islands of twinkling green bulbs. Your antennae should quiver. The phase transition from order to chaos, as parameters such as the number of inputs per lightbulb, K, or the bias parameter, P, are tuned, occurs when the giant frozen red cluster forms, leaving isolated twinkling green islands behind.

A particularly easy way to see this is to make a very simple Boolean network model on a square lattice. Here each lightbulb is connected to its four neighbors: north, south, east, and west. Each lightbulb is controlled by a Boolean function that tells it how to turn on or off depending on the current activities of its four inputs. Figure 4.3 shows such a lattice network, studied by Derrida and Weisbuch. They tuned the P bias parameter close enough to 1.0 so that the network is in the ordered regime, let the network settle into its state cycle, and then recorded the cycling period of each lightbulb. A lightbulb with a cycling period of 1 is therefore either frozen on or frozen off. In our mental picture, any such lightbulb should be colored red. Other lightbulbs are twinkling; hence these should be colored green. As Figure 4.3 shows, the period-1 frozen lightbulbs form a giant connected component that spreads across the entire lattice, leaving behind a few small and large twinkling clusters.

With Figure 4.3 in view, it is easy to explain the sensitivity to changes in initial conditions in chaotic networks and the lack of sensitivity to such perturbations in ordered networks. If a single lightbulb is flipped, one can follow the cascading changes radiating from that perturbation. In the ordered regime, such as in Figure 4.3, those rippling changes cannot penetrate the period-1 frozen red component. The giant frozen component is rather like a gigantic wall of constancy blocking off the

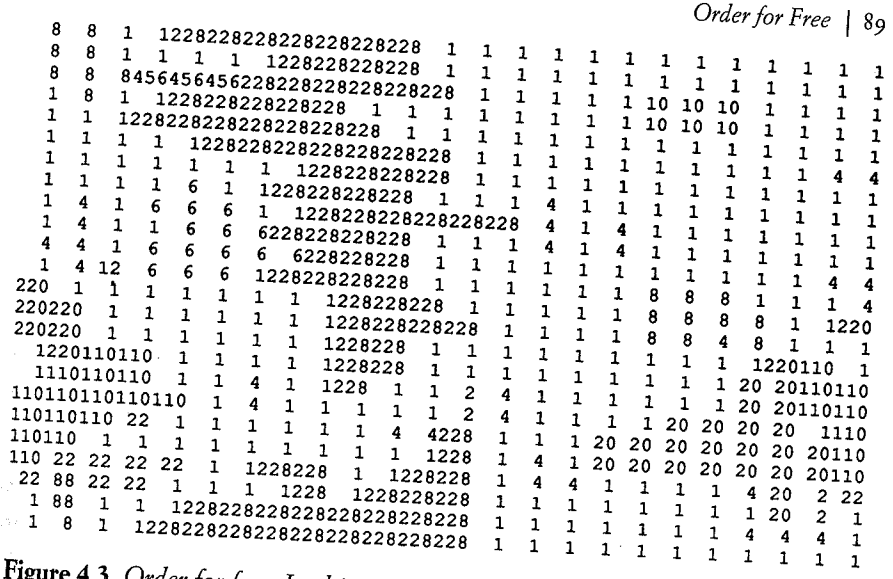

Figure 4.3 *Order for free. In this two-dimensional lattice, each site (lightbulb) is coupled to its four neighbors and governed by a Boolean function. When* P, *the bias in favor of a 1 or 0 response by any single variable, is increased above a critical value,* Pc, *percolation of a frozen component of lightbulbs, each fixed at 1 or 0, spans across the lattice and leaves isolated islands of twinking lightbulbs free to vary between 1 and 0 values. The number at each point represents the cycling period of each lightbulb. Thus sites with 1 correspond to red lightbulbs frozen in either the on or the off state. Sites with numbers greater than 1 are green, twinkling on and off and forming isolated "unfrozen" islands in the sea of frozen sites. (The two-dimensional lattice is bent into a donut, or torus, by "gluing" the top edge to the bottom edge, and the left edge to the right edge. Therefore, all lightbulbs have four neighbors.)*

twinkling islands from one another. Perturbations can cascade within each twinkling island, but rarely propagate any further. That is fundamentally why our lightbulb networks in the ordered regime exhibit homeostasis.

But in the chaotic regime, a vast sea of twinkling lightbulbs extends across the entire network. If any such lightbulb is flipped, the consequences cascade throughout that unfrozen sea, creating massive changes in the activity patterns of the lightbulbs. So chaotic systems show massive sensitivity to small perturbations. Here, in our Boolean networks in the chaotic regime, is the butterfly effect. Flap your wings, oh butterfly, moth, or starling, briskly or languidly, and you will change the behavior of lightbulbs from Alaska to Florida.

Protocells and your cells, early life and all life, must be capable of orderly yet flexible behavior. What kinds of networks of interacting mol-

cules, or interacting anything, are naturally capable of such ordered yet flexible behavior? Is such behavior hard to achieve? Or might it, too, be part of order for free? Now that we begin to understand order and chaos in networks coupling hundreds of thousands of lightbulbs, an answer, crisp and lovely, perhaps even true, suggests itself: perhaps networks just at the phase transition, just poised between order and chaos, are best able to carry out ordered yet flexible behaviors.

Here is a beautiful working hypothesis. Chris Langton at the Santa Fe Institute has stressed this important possibility more than any other scientist, and we can see intuitively that the edge of chaos might be an attractive regime to coordinate complex behavior. Suppose one wished to have a lattice of lightbulbs that coordinated the activities of two widely separated lightbulb sites on the lattice; suppose the lattice were in the chaotic regime, with an unfrozen sea. Then minor perturbations of the activities of one lightbulb would unleash cascades of alterations in activities, which would propagate throughout the lattice and dramatically undo any hoped-for coordination. Chaotic systems are too chaotic to coordinate behavior between distant sites. The system cannot send a reliable signal across the lattice.

Conversely, suppose the lattice is deep in the ordered regime. A frozen red sea is spread across the lattice, leaving twinkling tiny green islands. Suppose we wish to coordinate a series of actions by distant sites. Alas, no signal can propagate across the frozen sea. The twinkling unfrozen islands are functionally isolated from one another. No complex coordination can occur.

But at the edge of chaos, the twinkling unfrozen islands are in tendrils of contact. Flipping any single lightbulb may send signals in small or large cascades of changes across the system to distant sites, so the behaviors in time and across the webbed network might become coordinated. Yet since the system is at the edge of chaos, but not actually chaotic, the system will not veer into uncoordinated twitchings. Perhaps, just perhaps, such systems might be able to coordinate the kinds of complex behavior we associate with life.

To complete this part of the story, I will present evidence for an idea that I will more fully develop in the next chapter: *the reason complex systems exist on, or in the ordered regime near, the edge of chaos is because evolution takes them there.* While autocatalytic networks arise spontaneously and naturally because of the laws of complexity, perhaps natural selection then tunes their parameters, tweaking the dials for K and P, until they are in the ordered regime near this edge—the transitional region between order and chaos where complex behavior thrives. After all, systems capable of complex behavior have a decided survival advantage, and thus natural selection finds its role as the molder and

shaper of the spontaneous order for free. In order to test this hypothesis, Bill Macready, a postdoctoral fellow, Emily Dickinson, a computer scientist, and I have been using computer simulations to "evolve" Boolean networks to play simple and hard games with one another. In these games, each network must respond with a "proper" pattern of lightbulb activities to the prior pattern of lightbulb activities by the network it is playing. Our evolving networks are free to mutate connections between lightbulbs in each network and the Boolean rules turning lightbulbs on and off in each network. Thus our networks can change the different parameters that tune their positions on the order–chaos axis. In order to test the locations of our networks on the order–chaos axis, Bill, Emily, and I make use of a simple feature that distinguishes the ordered regime from the chaotic regime. In the chaotic regime, similar initial states tend to become progressively more dissimilar, and hence to *diverge* farther and farther apart in state space, as each passes along its trajectory. This is just the butterfly effect and sensitivity to initial conditions. Small perturbations amplify. Conversely, in the ordered regime, similar initial states tend to become more similar, hence *converging* closer together as they flow along their trajectories. This is just another expression of homeostasis. Perturbations to nearby states "damp out." We measure average convergence or divergence along the trajectories of a network to determine its location on the order–chaos axis. In fact, in this measure, networks at the phase transition have the property that nearby states neither diverge nor converge.

What are the results? As the networks play their games with one another, trying to match one another's lightbulb patterns, the computer simulation selects fitter mutant variants—that is, the networks that play better. What we have found for the modestly complex behaviors we are requesting is that the networks do adapt and improve and that they evolve, not to the very edge of chaos, but to the ordered regime, not too far from the edge of chaos. It is as though a position in the ordered regime near the transition to chaos affords the best mixture of stability and flexibility.

It is far too early to assess the working hypothesis that complex adaptive systems evolve to the edge of chaos. Should it prove true, it will be beautiful. But it will be equally wonderful if it proves true that complex adaptive systems evolve to a position somewhere in the ordered regime near the edge of chaos. Perhaps such a location on the axis, ordered and stable, but still flexible, will emerge as a kind of universal feature of complex adaptive systems in biology and beyond.

We turn to these beautiful possibilities in more detail in the following chapters, for the hypothesis that complex systems may evolve to the edge of chaos or to the order-

account for a very large number of features of ontogeny, that magnificent, ordered dance of development from fertilized egg to bird, fern, bracken, flea, and tree. But caveats again, for at this stage a potential universal law is best held as a fascinating working hypothesis.

In the meantime, we may begin to suspect, the exquisite power of self-organization, which we begin to understand in our simple models of enormous Boolean networks, may be the ultimate wellspring of dynamical order. The order in these open nonequilibrium thermodynamic systems derives from the ordered regime; in turn, the order of the ordered regime derives from the fact that nearby states tend to converge. The system therefore "squeezes" itself onto tiny attractors. Ultimately, it is this self-squeezing into infinitesimal volumes of state space that constitutes the order. And while I have called it order for free, meaning that such order is natural and spontaneous, it is not "for free" thermodynamically. Rather, in these open systems, the self-squeezing of the system into tiny regions of state space is "paid for" thermodynamically by exporting heat to the environment. No laws of thermodynamics are violated or even contested. What is new is that vast open thermodynamic systems can spontaneously lie in the ordered regime. Such systems may be the natural source of the order required for stable self-reproduction, homeostasis, and graceful heritable variation.

If we, and past eons of scholars, have not begun to understand the power of self-organization as a source of order, neither did Darwin. The order that emerges in enormous, randomly assembled, interlinked networks of binary variables is almost certainly merely the harbinger of similar emergent order in whole varieties of complex systems. We may be finding new foundations for the order that graces the living world. If so, what a change in our view of life and our place must await us. Selection is not the sole source of order after all. Order vast, order ordained, order for free. We may be at home in the universe in ways we have hardly begun to comprehend.

Chapter 5

The Mystery of Ontogeny

At least since the Cambrian explosion, 550 million years ago, and probably for the past 700 million years, multicelled organisms have mastered a mystery no human mind yet comprehends: ontogeny. Through some mysterious evolutionary creativity, the new creatures of the Cambrian—and *Homo sapiens* much more recently—began life as a single cell, the zygote, the fruit of parental union. Somehow that single cell knew to give rise to a complete structure, an organized whole, an organism. If the swarm of stars in a spiral galaxy, clustered swirling in the high blackness of space, astonishes us with the wonder of the order generated by mutually gravitating masses, think with equal wonder at our own ontogeny. How in the world can a single cell, merely some tens of thousands of kinds of molecules locked in one another's embrace, know how to create the intricacies of a human infant? No one knows. If *Homo habilis* wondered, if Cro-Magnon wondered how they came to be, so too must we.

Begin, then, with the zygote. After fertilization of egg by sperm, the human zygote undergoes rapid cleavage—cell divisions that create a small mass of cells. These cells migrate down the fallopian tube and enter the uterus. While migrating, the mass of cells hollows out, forming a ball. A small number of cells, called the inner cell mass, migrates inward from one pole of the hollow ball and lodges nestled against the remaining outer layer. All mammals derive from the inner cell mass. The outer layer of cells in humans has specialized to burrow into the uterine lining and form the extraembryonic membranes, placenta and otherwise, that support us before birth.

Already, even at this most primitive stage, we witness the two fundamental processes of ontogeny, or development: the first is cell differentiation; the second is morphogenesis. The zygote is both a single cell and